

# **Methoden und Werkzeuge für eine datengetriebene Entwicklung digitaler Gesundheitsanwendungen**

Zur Erlangung des akademischen Grades eines

**Doktors der Ingenieurwissenschaften (Dr.-Ing.)**

von der KIT-Fakultät für  
Wirtschaftswissenschaften  
des Karlsruher Instituts für Technologie (KIT)

genehmigte

**DISSERTATION**

von

**M. Sc. Markus Schinle**

geb. in Schramberg

Tag der mündlichen Prüfung:

31.08.2022

Hauptreferent:

Prof. Dr. rer. nat. Stefan Nickel

Korreferent:

Prof. Dr. rer. nat. Wilhelm Stork



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung -  
Keine Bearbeitungen 4.0 International Lizenz (CC BY-ND 4.0):  
<https://creativecommons.org/licenses/by-nd/4.0/deed.de>

# Kurzfassung

Dem Paradigma der Präzisionsmedizin folgend schaffen digitale Gesundheitsanwendungen die Grundlage für eine personalisierte Versorgung, um damit die Effizienz und Effektivität von Gesundheitssystemen zu erhöhen. Im Kontext weltweit entstehender digitaler Gesundheitsökosysteme stehen dabei Daten als treibender Faktor im Mittelpunkt des Entwicklungsprozesses. Welche Methoden und Werkzeuge benötigt werden, um das dadurch mögliche Zusammenspiel zwischen einer datengetriebenen und einer wissensbasierten Entwicklung von digitalen Gesundheitsanwendungen zu unterstützen, wird in dieser Arbeit untersucht und anhand eines Rahmenwerks beschrieben. Durch Anwendung der Design Science Research Methode werden diesbezügliche Artefakte einem problemorientierten Ansatz folgend entworfen, implementiert und durch quantitative sowie qualitative Methoden evaluiert. Dafür wird zunächst ein Vorgehensmodell abgeleitet, welches die zu beantwortenden Fragen in den Phasen der Digitalisierung, Automatisierung und Optimierung bis hin zur Translation in die medizinische Versorgung adressiert. Unter Beachtung entsprechender Normen findet eine Verknüpfung von interdisziplinären Methoden, Anforderungen sowie technologischen Ansätzen zu einer Wissensbasis statt, womit die Grundlage für zu entwickelnde Werkzeuge gelegt wird. Diese werden im Anwendungskontext dementieller Syndrome eruiert und pro Artefakt demonstriert sowie im Detail mit  $n$  Probanden multiperspektivisch validiert. In Kooperation mit einer gerontopsychiatrischen Klinik werden diesbezüglich domänenspezifische Anforderungen an digitale Gesundheitsanwendungen bestimmt. Hierfür findet exemplarisch die explorative Entwicklung eines ambulanten Systems zur Messung kognitiver Leistungsparameter statt. Eine im Kontext dieser Zusammenarbeit durchgeführte Feldstudie ( $n = 55$ ) mit kognitiv eingeschränkten Personen zeigt Potentiale und Herausforderungen, welche durch die digitale Erfassung, Vernetzung und Auswertung von neuropsychologischen Daten entstehen. Dabei werden ebenfalls Anforderungen bezüglich der zielgruppenspezifischen Gestaltung einer gebrauchstauglichen Nutzerschnittstelle ( $n = 91$ ) gesammelt, welche in einem Leitfaden zusammenfließen und in einer grafischen Benutzeroberfläche iterativ implementiert werden. Aus der Perspektive von Datensubjekten ( $n = 238$ ) wird zusätzlich untersucht, welchen Stellenwert ein selbstbestimmter Umgang mit dieser Art von personenbezogenen Daten hat und für welche Zwecke diese aus deren Sicht eingesetzt werden sollten. Im Zuge dieses Entwicklungsprozesses sind ebenfalls Ansätze zur Automatisierung und Optimierung der Datenauswertung für die Ableitung des Gesundheitszustandes notwendig. Diese Schritte liefern als Artefakte, neben den Ergebnissen zum Vergleich verschiedener Algorithmen aus dem Bereich des maschinellen Lernens, die Identifikation von dafür geeigneten Leistungs- und Optimierungsmaßen sowie Merkmalsselektionsverfahren. Im Vergleich mit schwellwertbasierten Verfahren zur Operationalisierung von Bewertungsmetriken (maximaler Cohen's Kappa  $\kappa = 0,67$ ) erreicht die durch maschinelles Lernen gestützte Softwareanwendung eine höhere durchschnittliche Sensitivität von 83% bei einer 93%igen Spezifität (maximaler Cohen's Kappa  $\kappa = 0,79$ ) für die Erkennung von kognitiven Einschränkungen. Die automatisierte Erfassung hierfür notwendiger Merkmale erfolgt durch neu entwickelte Ansätze und zeigt zukünftige Forschungsaktivitäten auf, welche die damit verbundenen Herausforderungen adressieren. Dabei werden Indikatoren identifiziert, wodurch sich die Potentiale in computergestützten Modellen aufzeigen. Diese liefern zusätzliche Erkenntnisse über das Spannungsfeld zwischen einer zuverlässigen Erfüllung klinischer Leitlinien sowie regulato-

rischer Implikationen insbesondere hinsichtlich der Erklärbarkeit datengetriebener Optimierungs- und Automatisierungsansätze. Eine Untersuchung der Transferpotentiale in die deutsche Regelversorgung aus der Perspektive unterschiedlicher Interessenvertreter unterstreicht diese Punkte. Hierfür konzipierte Werkzeuge und Methoden ermöglichen einerseits die empirische Untersuchung der Adhärenz solcher digitaler Lösungen bezüglich der Nutzungsbereitschaft ( $n = 29$ ) sowie deren zeitliche Entwicklung ( $n = 18$ ). Andererseits werden damit die Akzeptanzkriterien der kassenärztlich organisierten Leistungserbringer im deutschen Gesundheitswesen ( $n = 301$ ) erhoben und dargestellt, welchen Einfluss diese auf Markteintrittsstrategien haben. Darauf aufbauend werden Wege definiert, um einen Beitrag zur Entlastung des Gesundheitssystems zu leisten. Die gesammelten Erkenntnisse werden hierfür in einem ganzheitlichen Plattformkonzept zur Entwicklung personalisierter Präventions- und Behandlungsprogramme gebündelt.

# Abstract

Following the paradigm of precision medicine, digital health applications create the basis for personalized care, thereby increasing the efficiency and effectiveness of healthcare systems. In the context of globally emerging digital health ecosystems, data is at the center of the development process as a driving factor. The methods and tools needed to support the resulting interaction between data-driven and knowledge-based development of digital health applications are explored in this work, as well as described using a framework. By applying the Design Science Research Method, related artifacts are designed following a problem-initiated approach, implemented, and evaluated by quantitative as well as qualitative methods. For this purpose, a process model is first derived, which addresses the questions to be answered in the phases of digitization, automation, and optimization up to translation into medical care. Taking into account the relevant standards, interdisciplinary methods, requirements, and technological approaches are linked to form a knowledge base, thus laying the foundation for tools to be developed. These will be elicited in the application context of dementia syndromes and demonstrated as well as validated per artifact in detail with  $n$  subjects multiperspectively. In cooperation with a gerontological psychiatric clinic, domain-specific requirements for digital health applications will be determined. For this purpose, the exploratory development of an ambulatory system for the measurement of cognitive performance parameters is exemplary. A study ( $n = 55$ ) with cognitive impaired persons conducted in the context of this collaboration shows the potential and challenges that arise from the digital recording, interconnection, and analysis of neuropsychological data. Requirements regarding the target group-specific design of a usable user interface ( $n = 91$ ) are also collected, which are summarised in a guideline and iteratively implemented in a graphical user interface. From the perspective of data subjects ( $n = 238$ ), the importance of self-determined handling of this type of personal data and the purposes for which it should be used from their point of view will also be investigated. In the course of this development process, approaches to automate and optimize data evaluation for the derivation of health status are also necessary. These steps provide artifacts, in addition to the results for comparing different algorithms from the field of machine learning, the identification of performance and optimization measures suitable for this purpose, as well as feature selection procedures. Compared with threshold-based methods for operationalizing assessment metrics (maximum Cohen's Kappa  $\kappa = 0,67$ ), the machine learning-based software application achieves a higher average sensitivity of 83% with a 93% specificity (maximum Cohens Kappa  $\kappa = 0,79$ ) for detecting cognitive impairment. Regarding the automated acquisition of necessary features for this purpose, new approaches are developed as well as future research activities that address the associated challenges are highlighted. Indicators are identified that show the potential in computerized models. These provide insights into the tension between a reliable fulfillment of clinical guidelines as well as regulatory framework conditions, especially with regard to the explainability of data-driven optimization and automation approaches. A consideration of the transfer to German standard care from the perspective of different stakeholders supports these points. Tools and methods designed for this purpose enable, on the one hand, the empirical investigation of the adherence of such digital solutions with regard to the willingness to use ( $n = 29$ ) as well as their development over time ( $n = 18$ ). On the other hand, acceptance criteria of health care providers in the German healthcare system ( $n = 301$ )

are surveyed and their influence is shown. Based on that, ways are defined to contribute to relieving the burden on the health care system. Therefore, the insights gathered in this process are combined into a holistic platform concept for the development of personalized prevention and treatment programs.

# Danksagung

Die vorliegende Dissertationsschrift entstand im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter am FZI Forschungszentrum Informatik und dem Karlsruher Institut für Technologie. Mein besonderer Dank gilt Prof. Dr. rer. nat. Wilhelm Stork für die Promotionsmöglichkeit. Die gelebte Interdisziplinarität an seinem Lehrstuhl und die zahlreichen Gespräche über innovative Ansätze sowie das entgegengebrachte Vertrauen in meine wissenschaftlichen Fähigkeiten und Ideen haben sowohl diese Dissertation als auch meine persönliche Entwicklung nachhaltig geprägt. Prof. Dr. rer. nat. Stefan Nickel danke ich für die Übernahme des Referats und das damit verbundene Engagement. Für die erfolgreiche Zusammenarbeit und die Unterstützung bei der Problemsuche und Lösungsfindung danke ich insbesondere Dr.-Ing. Johannes Schneider, Dr.-Ing. Timon Blöcher und Christina Erler. Des Weiteren möchte ich mich bei meinen Kollegen und Freunden Dr.-Ing. Johannes Bach, Christoph Zimmermann, Dr.-Ing. Jennifer Zeilfelder, Dr.-Ing. Simon Stock, Marius Gerdes, Kai Zhou und Lukas Kohout sowie den von mir betreuten Studenten für das kollegiale und inspirierende Arbeitsumfeld bedanken. Außerdem möchte ich allen Projektpartnern für die Ausgestaltung des Forschungsumfeldes danken, insbesondere Frau Anne Röhm von der PP.rt in Reutlingen, die durch ihre Unterstützung maßgeblich zum Gelingen dieser Arbeit beigetragen hat. Ganz herzlich danke ich auch meiner Partnerin, Familie und Freunden, die abseits der Arbeitswelt einen Rückzugsort für mich bieten. Besten Dank.

Karlsruhe, im August 2022

*Markus Schinle*



*Nachfolgend sind weibliche, diverse oder männliche Geschlechteridentitäten an jeder Stelle gleichermaßen angesprochen. Des Weiteren wird aufgrund von integrierten Animationen und verlinkten Referenzen empfohlen das Dokument in seiner digitalen Version zu lesen.*



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Motivation und adressierte Herausforderungen	2
1.2	Vorgehensweise und Aufbau der Arbeit	3
1.3	Zielsetzung und Forschungsfragen	4
1.4	Beiträge und wissenschaftliches Umfeld	5
<b>2</b>	<b>Grundlagen und domänenspezifische Rahmenbedingungen</b>	<b>9</b>
2.1	Verwendete Methoden, Metriken und Prozesse	9
2.1.1	Übergeordnetes wissenschaftliches Vorgehensmodell	9
2.1.2	Methode zur Prüfung der Gebrauchstauglichkeit	10
2.1.3	Vorgehensmodell zur Wissensentdeckung	10
2.1.4	Statistische Verfahren und Metriken	12
2.2	Gesundheitswesen und medizinisches Hintergrundwissen	18
2.2.1	Das deutsche Gesundheitswesen	18
2.2.2	Standardisierung von Gesundheits- und Behandlungszuständen	21
2.2.3	Demenz	21
2.2.4	Diagnose und Behandlung von Alzheimer Demenz	25
2.2.5	Psychometrie	29
2.3	Juristische Grundlagen und regulatorische Randbedingungen	38
2.3.1	Regulatorische Einflüsse auf das Gesundheitswesen und die Medizintechnik	38
2.3.2	Datenschutz	45
2.4	Technische Begriffserklärungen	49
2.4.1	Gesundheitsdaten und deren Verwaltung	49
2.4.2	Digitalisierung im Gesundheitswesen	50
2.4.3	Verfahren zur Automatisierung und Optimierung	57
<b>3</b>	<b>Einordnung in den Stand der Wissenschaft und Technik</b>	<b>67</b>
3.1	Digitale Transformation im Gesundheitswesen	67
3.1.1	Standardisierung und Interoperabilität	68
3.1.2	Gesundheitsdatenmobilität und -infrastrukturen	69
3.2	Datengetriebene Ansätze in der medizinischen Versorgung	72
3.2.1	Datenaustausch für personalisierte Medizin	72
3.2.2	Datengrundlage zur gemeinsamen Entscheidungsfindung	74
3.3	Demenz-spezifische Potentiale und Lösungen	75
3.3.1	Abwägung der Mechanismen zur Früherkennung und Behandlung in der Primärversorgung	75
3.3.2	Digitalisierung	76
3.3.3	Computergestützte Automatisierung und Optimierung	79
3.3.4	Vertrauenswürdigkeit computergestützter Modelle	80
3.3.5	Translation in die Versorgung	86

3.4	Kritik . . . . .	86
<b>4</b>	<b>Anforderungsanalyse und Konzeption des Rahmenwerks für eine datengetriebene Entwicklung . . . . .</b>	<b>89</b>
4.1	Anforderungen an den Umgang mit Gesundheitsdaten im Kontext entstehender Gesundheitsdatenökosysteme . . . . .	89
4.1.1	Der Datenkreislauf als Rahmenwerk . . . . .	91
4.1.2	Abgeleitetes Vorgehen . . . . .	93
4.2	Zieldefinition einer digitalen Gesundheitsanwendung . . . . .	96
4.2.1	Identifikation und Kombination relevanter Parameter . . . . .	97
4.2.2	Grundsätzliche Anforderungen an eine digitale Gesundheitsanwendung . . . . .	102
<b>5</b>	<b>Nutzerzentrierte Digitalisierung am Beispiel psychometrischer Tests . . . . .</b>	<b>105</b>
5.1	Koordination und Modellierung der Erfassungsmöglichkeiten . . . . .	105
5.1.1	Koordination der Parametererfassung . . . . .	106
5.1.2	Modellierung einzelner Testabläufe . . . . .	106
5.1.3	Systemarchitektur einer digitalen Gesundheitsanwendung . . . . .	109
5.1.4	Technologieauswahl . . . . .	110
5.2	Gebrauchstauglichkeit der Benutzerschnittstelle . . . . .	111
5.2.1	Anforderungen an eine grafische Benutzerschnittstelle . . . . .	111
5.2.2	Entwicklung einer grafischen Benutzerschnittstelle . . . . .	115
5.3	Selbstbestimmte Datenverwaltung . . . . .	119
5.3.1	Stufenweise Realisierung der Datenverwaltung . . . . .	120
5.3.2	Selbstbestimmtheit durch Einwilligung . . . . .	121
5.3.3	Notwendige Interoperabilitätseigenschaften . . . . .	127
5.4	Validierung und Einflüsse der Digitalisierung . . . . .	129
5.4.1	Datengrundlage und Merkmale . . . . .	130
5.4.2	Vergleich zwischen analoger und digitaler Lösung . . . . .	134
5.4.3	Untersuchung weiterer Einflussfaktoren . . . . .	136
5.4.4	Diskriminationsfähigkeit . . . . .	138
<b>6</b>	<b>Digitale Automatisierungs- und Optimierungsansätze zur Testauswertung . . . . .</b>	<b>141</b>
6.1	Automatisierte Extraktion einzelner Merkmale am Beispiel von Test J . . . . .	141
6.1.1	Vorgehensweise . . . . .	142
6.1.2	Kantendetektion . . . . .	143
6.1.3	Eckendetektion . . . . .	145
6.1.4	Operationalisierung der Bewertungsmetrik anhand der Eckendetektion . . . . .	148
6.1.5	Modelloptimierung und -validierung . . . . .	150
6.2	Entwurf einer Bewertungsmetrik . . . . .	153
6.2.1	Merkmalsselektion . . . . .	154
6.2.2	Merkmalspezifische Schwellwerte . . . . .	155
6.2.3	Auf kognitiven Domänen basierte Schwellwerte . . . . .	155
6.2.4	Gegenüberstellung der Performanz . . . . .	157
6.3	Optimierung der Klassifikationsergebnisse mittels Modellen des Maschinellen Lernens . . . . .	158
6.3.1	Vorgehen und Merkmalsselektion . . . . .	159
6.3.2	Modellvalidierung . . . . .	160
6.4	Plausibilisierung, Risikominimierung und Wissensgenerierung durch erklärbares Maschinelles Lernen . . . . .	163
6.4.1	Auswahl und Anwendung der Verfahren . . . . .	163

6.4.2	Evaluation ausgewählter Verfahren . . . . .	166
<b>7</b>	<b>Translation in die Regelversorgung . . . . .</b>	<b>171</b>
7.1	Perspektiven unterschiedlicher Akteure . . . . .	171
7.1.1	Adhärenz bei Leistungsempfängenden . . . . .	171
7.1.2	Akzeptanz bei Leistungserbringenden . . . . .	173
7.2	Der datengetriebene Weg zu personalisierten Präventions- und Behandlungsprogrammen	177
7.2.1	Doctor-in-the-Loop . . . . .	178
7.2.2	Patient-in-the-Loop . . . . .	179
7.3	Markteinführung . . . . .	181
7.3.1	Rahmenbedingungen für Markteintrittsstrategien . . . . .	181
7.3.2	Organisatorisch-technische Implikationen . . . . .	183
<b>8</b>	<b>Diskussion der Ergebnisse . . . . .</b>	<b>185</b>
8.1	Digitalisierungsphase . . . . .	185
8.2	Automatisierungs- und Optimierungsphase . . . . .	187
8.3	Herausforderungen und Potentiale in der Translation . . . . .	188
<b>9</b>	<b>Abschließende Betrachtung . . . . .</b>	<b>191</b>
9.1	Zusammenfassung . . . . .	191
9.2	Fazit . . . . .	191
9.3	Ausblick . . . . .	193
<b>A</b>	<b>Anhang . . . . .</b>	<b>195</b>
A.1	Anforderungen an den Umgang mit Gesundheitsdaten . . . . .	195
A.2	Beispiel für Usage Policy . . . . .	200
A.3	Fragebogen zur Umfrage Datennutzung . . . . .	200
A.4	Demografische Daten zu Datennutzung . . . . .	207
A.5	Umfrageergebnisse zu Datennutzung . . . . .	208
A.6	Beispielhafte FHIR-Ressource . . . . .	209
A.7	Fragebogen zur Technikerfahrung . . . . .	210
A.8	Studienteilnehmer der Beobachtungsstudie . . . . .	214
A.9	DemPredict nach Iteration II . . . . .	215
A.10	Mögliche digitale Parameter . . . . .	216
A.11	Normalverteilungsannahme für mit DemPredict erhobene Parameter . . . . .	219
A.12	Temporale Effekte bezüglich der Testzeitpunkte . . . . .	221
A.13	Aggregierter Vergleich zwischen CERAD und DemPredict . . . . .	223
A.14	Individueller Vergleich zwischen CERAD und DemPredict . . . . .	224
A.15	Korrelation zwischen Technologie-Erfahrung und den domänenspezifischen Parametern	225
A.16	Diskriminationsfähigkeit zwischen Diagnose und den domänenspezifischen Parametern	226
A.17	Mögliche Schwellwerte auf Merkmalsebene . . . . .	229
A.18	Algorithmus für Explainable k-Nearest Neighbors . . . . .	231
A.19	Evaluationsdaten der XAI-Ansätze . . . . .	231
A.20	Demografische Daten zur Umfrage der Adhärenz . . . . .	235
A.21	Umfrage zur Akzeptanz bei Leistungserbringern . . . . .	235
A.22	Demografische Daten zur Umfrage bei Leistungserbringern . . . . .	250
A.23	Umfrageergebnisse zur Akzeptanz bei Leistungserbringern . . . . .	251
A.24	Fragebogen zur ISO NORM 9241-110 . . . . .	261
A.26	DemPredict nach Iteration IV . . . . .	266
A.25	Umfrageergebnisse zur Usability . . . . .	268

<b>Literaturverzeichnis</b> . . . . .	<b>269</b>
<b>Abkürzungsverzeichnis</b> . . . . .	<b>305</b>
<b>Abbildungsverzeichnis</b> . . . . .	<b>313</b>
<b>Tabellenverzeichnis</b> . . . . .	<b>316</b>
<b>Eigene Veröffentlichungen</b> . . . . .	<b>317</b>
Journalartikel . . . . .	317
Konferenzbeiträge . . . . .	317

# 1 Einleitung

NACH jahrelangen Verzögerungen sichert Deutschland seinen gesetzlich versicherten Bürgern seit dem Jahr 2021 die freiwillige Nutzung einer elektronischen Patientenakte zu. Hiermit soll ab dem Jahr 2023 auch die Möglichkeit geboten werden, Daten für Forschungszwecke zu spenden, um die Potentiale digitaler Gesundheitsdaten besser nutzen zu können. Diese Entwicklung trifft mit Quantified Self auf eine gesellschaftliche Bewegung, welche die kontinuierliche Vermessung des eigenen Ichs in den Vordergrund stellt [284]. Mit Hilfe von sogenannten Self-Tracking-Werkzeugen werden unterschiedliche Parameter erfasst, welche Auskunft über die Person selbst und deren Umfeld zulassen. Der Markt von hierfür geeigneten digitalen Werkzeugen wie Wearables, Smartphones und entsprechenden Applikationen ist in den letzten Jahren kontinuierlich gewachsen [284]. Im Gesundheitswesen stehen dabei den Vorteilen, wie zum Beispiel aus erhobenen Daten abgeleitetes neues medizinisches Wissen, auch Risiken gegenüber, welche in Überwachungs- und Diskriminierungspotentialen durch Missbrauch sowie mangelnder Qualität der Geräte und Daten gesehen werden [189, 284]. Dennoch spricht die Verbreitung dieser Technologien für eine gestiegene Bereitschaft in der Bevölkerung personenbezogene Daten, insbesondere Gesundheitsdaten sowohl für sich selbst zu nutzen als auch zum Wohle der Allgemeinheit zur Verfügung zu stellen. Als jüngstes Beispiel lässt sich die, während der COVID-19-Pandemie im Jahr 2020 durch das Robert-Koch-Institut entwickelte, Applikation zur Corona-Datenspende nennen [350]. Bereits wenige Wochen nach dem offiziellen Beginn der Pandemiebekämpfung stellten über eine halbe Million Menschen in Deutschland pseudonymisierte Vitaldaten zur Verfügung, um Symptome einer Coronavirus-Infektion zu erkennen und deren Ausbreitung in der Bevölkerung nachvollziehen zu können [110]. Übertragen auf andere Krankheitsbilder können diesem Vorbild folgend wissenschaftliche Erkenntnisse zu digitalen Diagnose- und Behandlungsmöglichkeiten durch weitreichende Datenerhebungs- und -analysestudien gewonnen werden. Daneben zielt, seit deren Einführung im Jahr 2020, die Digitale Gesundheitsanwendung (DiGA) konzeptionell darauf ab, neue medizinische Erkenntnisse möglichst zeitnah in Form von Angeboten digitaler Unterstützungsleistungen für den Umgang mit Erkrankungen in der Regelversorgung zu etablieren (vgl. Unterabschnitt 2.3.1). Als ein Anwendungsfeld mit ausgeprägtem gesundheitsökonomischem Potential hierfür lassen sich psychische Erkrankungen hervorheben. Nach Angaben des Statistischen Bundesamtes wurden 13,1% (44,3 Milliarden €) der jährlichen deutschen Gesundheitsausgaben im Jahr 2015 für psychische Leiden ausgegeben [391]. Mit 15,1 Milliarden € verursachten alleine dementielle Erkrankungen und deren Folgen gut ein Drittel dieser Ausgaben [391]. Auch deshalb wird, bei einer zu erwartenden Überalterung der Gesellschaften in Nationen der Ersten Welt, ein signifikanter Einfluss von durch altersbedingte Effekte begünstigten psychischen Leiden auf zukünftige Gesundheitssysteme prognostiziert (vgl. Unterunterabschnitt 2.2.3). Diesen Umstand berücksichtigend, beschloss die Bundesregierung am 01. Juli 2020 eine Nationale Demenzstrategie, welche Ziele und Maßnahmen definiert, damit Menschen mit Demenz und ihre Angehörigen besser unterstützt werden [86]. Die Entwicklung digitaler Unterstützungslösungen, um den damit verbundenen Herausforderungen entgegenzutreten, setzt geeignete Methoden und Werkzeuge voraus, deren Erarbeitung deshalb als grundlegende Motivation dieser Arbeit gilt.

## 1.1 Motivation und adressierte Herausforderungen

Die zuvor beschriebenen Entwicklungen können die Qualität und Verfügbarkeit der Gesundheitsversorgung verbessern, indem, im weitesten Sinne dem Paradigma der Präzisionsmedizin folgend, individuelle Diagnose- und Behandlungsmaßnahmen durch digitale Unterstützungslösungen so früh und präzise wie möglich durchgeführt werden. Einer dafür notwendigen Kombination von gesundheitsbezogenen Daten und darauf anzuwendender Algorithmen stehen zum Zeitpunkt der Anfertigung dieser Arbeit in der deutschen Versorgungslandschaft neben regulatorischen Anforderungen, der Wunsch nach einem selbstbestimmten Umgang mit persönlichen Daten sowie ein Mangel an Digitalisierung und Vernetzung des Gesundheitswesens gegenüber. Durch die Umsetzung nationaler Digital-Strategien und pandemiebedingten Verhaltensänderungen besteht die Hoffnung auf eine wachsende gesellschaftliche Akzeptanz von digitalen Gesundheitsanwendungen. Gerade im Bereich chronifizierter Erkrankungen sieht die medizinische Forschung und Versorgung das Potential in der Anwendung digitaler Werkzeuge (vgl. Abschnitt 3.2). Dank moderner Informations- und Kommunikationstechnologie (IKT) findet bereits eine zunehmende Digitalisierung aller Lebensbereiche und damit verbunden eine enge Vernetzung unterschiedlicher Akteure statt. Weltweit reagieren Gesellschaften auf die damit verbundenen Chancen und Risiken unterschiedlich. In Europa wird beispielsweise den impliziten Risiken einer personenbezogenen Datenauswertung mit juristischen Regelungen begegnet, die vor allem den Schutz des Individuums in den Vordergrund stellen und somit die Nutzung dieser Potentiale gewissen Rahmenbedingungen unterstellt. Hierbei sei vor allem auf die seit dem Jahr 2018 geltende Datenschutz-Grundverordnung (DSGVO) verwiesen, welche diese Art von Daten unter besonderen Schutz stellt (vgl. Unterunterabschnitt 2.3.2). Nahezu zeitgleich erfreut sich im Zuge der Digitalisierung das Gesundheitswesen an der Durchsetzung einer patientenzentrierten Verwaltung von Gesundheitsdaten, die spätestens seit dem im November 2019 verabschiedeten Digitale-Versorgung-Gesetz (DVG) und dem im April 2020 verabschiedeten Patientendatenschutzgesetz (PDSG) auf juristischen Säulen steht. Als Bestandteil dieser Gesetze ist vorgesehen, dass die Behandlungsdaten aller Kassenpatienten unter Einhaltung der gesetzlichen Rahmenbedingungen gesammelt und für Forschungszwecke ausgewertet werden können (vgl. Unterunterabschnitt 2.3.1 und Unterunterabschnitt 2.3.2). Dabei steht die Verknüpfung und Analyse der in medizinischen Einrichtungen wie Arztpraxen oder Kliniken gesammelten Daten im Vordergrund. Daraus gewonnene Erkenntnisse sollen für personalisierte Gesundheitsdienstleistungen zur optimalen Unterstützung erkrankter Personen genutzt werden können (vgl. Abschnitt 3.2). Zum Zeitpunkt der Anfertigung dieser Arbeit wird diskutiert, wie die ab dem Jahr 2021 eingeführte elektronische Patientenakte unter Berücksichtigung von Forschungscompatibilität diesbezüglich gestaltet werden muss (vgl. Unterunterabschnitt 2.3.1). Als Ziel gilt es, das Individuum in die Lage zu versetzen, eine ganzheitliche Sicht auf die im Rahmen von Gesundheitsdienstleistungen erhobenen Daten zu erhalten. Eine weitere gesetzliche Neuerung sieht das DVG durch die Anpassung des Sozialgesetzbuchs vor, da somit die Bezahlung der durch den Einzelnen zum Beispiel mittels mobiler Applikationen in Anspruch genommenen ambulanten Gesundheitsdienstleistungen durch Krankenkassen übernommen werden kann. Infolgedessen machte die „App auf Rezept“ als Pressemeldung im Jahr 2019 Schlagzeilen [88]. Die Anbieter dieser digitalen Gesundheitslösungen haben nach vorläufiger Aufnahme in das Verzeichnis für Digitale Gesundheitsanwendungen (DiGAVZ) ein Jahr lang Zeit den Nachweis über die Wirksamkeit der von ihnen damit erbrachten Dienstleistung zu erbringen, um diese Leistung zu verstreifen. Dabei müssen positive Versorgungseffekte nachgewiesen sowie weitere Anforderungen erfüllt und durch das Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) geprüft werden (vgl. Unterunterabschnitt 2.3.1). Durch diese weltweit einzigartige Geschäftsmodellmöglichkeit sollten nachhaltige Anreize für privatwirtschaftlich motivierte Applikationsentwicklungen im Gesundheitswesen geschaffen

werden. Damit geht, je nach Einsatzzweck und Umsetzung, eine Aufnahme personenbezogener Daten einher, welche auf den vom jeweiligen Anbieter eingesetzten Systemen verarbeitet werden. Möglichkeiten zur Integration dieser unterschiedlichen Gesundheitsdatenquellen sind aus staatlicher Sicht durch die Einführung einer elektronischen Patientenakte gegeben (vgl. Unterabschnitt 2.4.1), weshalb bei der Entwicklung dieser Lösungen auch Interoperabilitätsanforderungen zu berücksichtigen sind (vgl. Unterabschnitt 2.4.2). Das Potential und der wirtschaftliche Wert von Gesundheitsdaten haben ebenso privatwirtschaftliche Interessen geweckt, weshalb die angesprochene Notwendigkeit von Datenintegrationsplattformen bereits in den späten 2000er Jahren beispielsweise von amerikanischen Software- und Internetkonzernen adressiert wurde. Diese konnten zumindest in Deutschland als Betreiber nur wenige Anwender davon überzeugen ihre Daten in ihren zentralisierten Cloud-Lösungen abzulegen, weshalb dieses Geschäftsfeld teilweise wieder aufgegeben wurde [389]. Gemäß der Datenschutzgrundverordnung steht es dem Individuum zu, über die Verwendung personenbezogener Daten selbst zu bestimmen. Eine ganzheitliche Sicht auf alle Gesundheitsdaten und selbstbestimmte Informationskoordination scheint daher unausweichlich. Neben einer dadurch möglichen digitalen Datenaustauschmöglichkeit mit medizinischem Fachpersonal sollen durch sogenannte Datenspenden gesammelte Daten mit selbst festgelegten Zweckbestimmungen auch Dritten zur Verfügung gestellt werden können [239, 365]. Hierdurch wird die Bereitstellung selbst erhobener Daten beispielsweise für die Teilnahme an klinischen Studien oder die Unterstützung krankheitsassoziierter Längsschnittdatenbanken ermöglicht. Insbesondere Krankheiten, deren Verläufe langwierig sind, häufig unentdeckt bleiben oder aus medizinischer Sicht noch nicht vollumfänglich erschlossen wurden, können davon profitieren. Jedoch hängt die Akzeptanz dieser Anwendungen von der Wahrnehmung unterschiedlicher Interessenvertreter ab. Chronisch erkrankte Menschen, vor allem im Bereich der mentalen Gesundheit, können zukünftig zu dieser Art von Profiteuren zählen, sofern digitale Werkzeuge bei der Erfassung, dem Teilen und der Analyse dieser Daten unterstützen [168, 362]. Vor allem die Integration von sogenanntem Patient Reported Outcome (PRO) - durch den Patienten selbstständig erhobenen Gesundheitsdaten - in den Entwicklungsprozess darauf aufbauender medizinischer Applikationen stellt eine zentrale Herausforderung dar, wofür Lösungskonzepte notwendig werden. Auf Grundlage dieser motivierenden Aspekte beleuchtet diese Arbeit am Beispiel von dementiellen Erkrankungen die Herausforderungen und Chancen der Digitalisierung im Kontext einer selbstbestimmten Datenerfassung und -bereitstellung unter den Prämissen der zuvor skizzierten Ausgangssituation.

## 1.2 Vorgehensweise und Aufbau der Arbeit

Die zur Anfertigung der vorliegenden Arbeit notwendigen Forschungsaktivitäten werden primär unter Anwendung der Design Science Research Methode nach Peffers et al. durchgeführt [323]. Diese untergliedert sich innerhalb eines Prozesses in sechs Schritte, wobei in Bezug auf die Zielsetzung notwendige Artefakte iterativ entwickelt und evaluiert werden. Indem neu gewonnene Erkenntnisse durch eine konkrete Umsetzung beispielhaft illustriert werden und deren Nutzen unmittelbar getestet wird, werden die Anforderungen der Theorie mit denen der Praxis verbunden (vgl. Unterabschnitt 2.1.1). Im ersten Prozessschritt wird neben der Problemidentifikation, die Eruierung der Relevanz und Motivation von Lösungen dafür vorgenommen. Im Kontext dieser Arbeit werden hierzu reale Herausforderungen aufgegriffen und sowohl der Mehrwert von Lösungen für Individuen, Organisationen als auch die Gesellschaft erörtert (vgl. Abschnitt 1.1). Daraus ergibt sich im zweiten Schritt die Definition der Ziele, welche mittels Forschungsfragen den Lösungsraum hinsichtlich der adressierten Herausforderungen abstecken (vgl. Abschnitt 1.3). In der dritten Stufe des Prozesses findet das Design und die Entwicklung

der Artefakte statt. In den Kapiteln 4 bis 7 werden Lösungsinstrumente sowohl in Form von domänen-spezifischen Anforderungskatalogen als auch Modellen, Algorithmen und Softwareartefakten erarbeitet. Im vierten Schritt werden die Lösungsartefakte demonstriert und deren Einsatz diskutiert. Hierbei wird ebenfalls auf die Chancen und Risiken eingegangen, welche der Verwendung dieser Artefakte aktuell gegenüberstehen. In der anschließenden fünften Phase des Prozesses werden die Lösungsartefakte gemäß den definierten Zielen und Evaluationskriterien bewertet (vgl. Kapitel 8). Anhand von Feldstudien wird hierfür untersucht, inwieweit die Artefakte den realweltlichen Anforderungen genügen und welchen Teil diese zur Problemlösung beitragen. Hierzu werden die Ergebnisse mit den zuvor festgelegten Zielen verglichen und entschieden, bis zu welchem Grad die Ziele dadurch erfüllt werden können beziehungsweise ob weitere Schritte zur Erreichung dieser unternommen werden müssen. Als abschließende Phase steht die Kommunikation der Erkenntnisse und wissenschaftlichen Beiträge im Vordergrund. Diese erfolgt neben den in dieser Arbeit dokumentierten Ergebnissen in Form von Publikationen auf nationalen und internationalen Fachkonferenzen (vgl. Abschnitt 1.4).

## 1.3 Zielsetzung und Forschungsfragen

In der vorliegenden Arbeit wird erforscht, wie im Gesundheitswesen das Zusammenspiel von einer wissenschaftsbasierten und einer datengetriebenen Entwicklung digitaler Lösungen unter Berücksichtigung der zuvor skizzierten Rahmenbedingungen unterstützt werden kann. Ziel der zu entwickelnden Methoden und Werkzeuge ist eine schnellere Überführung dieser Lösungen in die Regelversorgung. Daher steht zum Erhalt der Forschungstransferpotentiale eine holistische Betrachtung der Lösungsentwicklung im Fokus. Neben konkreten technischen Herausforderungen aus den Themenfeldern Digitalisierung, Automatisierung und Vernetzung werden übergeordnete interdisziplinäre Fragestellungen aufgegriffen und multiperspektivisch betrachtet. Die folgende zentrale Forschungsfrage bildet den Kern dieser Arbeit:

*Welche Methoden und Werkzeuge werden für eine datengetriebene Entwicklung von digitalen Gesundheitsanwendungen benötigt?*

Die zur Beantwortung dieser Leitfrage notwendigen Untersuchungen finden primär im Kontext von Anwendungen für die Versorgung dementieller Erkrankungen statt und lassen sich in folgende Teilfragen untergliedern:

- FF1** Welche Anforderungen bestehen an digitale Gesundheitsanwendungen und deren datengetriebene Entwicklung?
- FF2** Welche Konzepte werden hierfür benötigt?
- FF3** Wie lässt sich deren Umsetzung unterstützen?
- FF4** Unter welchen Rahmenbedingungen lassen sich diese in die Versorgungslandschaft transferieren?

Hierfür werden Artefakte entwickelt, welche in ein übergeordnetes Rahmenwerk münden, das anhand der Entwicklung digitaler Lösungen für das Diagnose- und Behandlungsmanagement im Kontext dementieller Erkrankungen dargestellt wird. Einem diesbezüglich problem-initiierten Ansatz folgend werden einerseits übergeordnete Herausforderungen, die bei der Entwicklung digitaler Gesundheitsanwendungen adressiert werden müssen, verfolgt. Andererseits werden für die Anwendungsdomäne spezifische Problemlösungen entworfen und entwickelt. Als primäres Ziel dieser Arbeit wird eine nachhaltige Verwertung der Lösungsinstrumente unter realen Bedingungen angestrebt. Dies soll den Transfer über die Grenzen dieser

wissenschaftlichen Arbeit hinaus beinhalten. Dieses Merkmal dient als zentrales Abgrenzungskriterium und als Richtschnur für die Evaluation der Ergebnisse.

## 1.4 Beiträge und wissenschaftliches Umfeld

Die in dieser Arbeit erschlossenen Artefakte (A) stehen wie nachfolgend dargestellt in Bezug zu den in Abschnitt 1.3 definierten Forschungsfragen (FF):

- A1** Anforderungen an eine datengetriebene Entwicklung digitaler Gesundheitsanwendungen im Kontext weltweit entstehender Gesundheitsdateninfrastrukturen (**FF1**).
- A2** Datengetriebenes Vorgehensmodell unter Beachtung entsprechender Normen und Leitlinien (**FF3**).
- A3** Domänenspezifische Anforderungen an digitale Gesundheitsanwendungen (**FF1 & FF4**).
- A4** Methoden und Algorithmen zur Verarbeitung psychometrischer Daten (**FF3**).
- A5** Konzepte zur Integration und Personalisierung von ambulanten Gesundheitsanwendungen im Bereich komplexer Psychosyndrome (**FF2**).
- A6** Softwareartefakte für das Erfassen, Teilen und Analysieren von neuropsychologischen Parametern (**FF2**).
- A7** Werkzeuge zur Erfassung und Evaluation von Anforderungen an die Translation digitaler Gesundheitsanwendungen aus der Perspektive unterschiedlicher Interessenvertreter (**FF4**).

Diese Beiträge entstanden zu großen Teilen in durch das Bundesministerium für Bildung und Forschung (BMBF), das Ministerium für Soziales und Integration Baden-Württemberg (MSIBW) sowie durch die Helmholtz-Gemeinschaft Deutscher Forschungszentren e.V. (HGDF) geförderten Forschungs- und Forschungstransferprojekten. Hierbei konnten sowohl auf nationalen als auch auf internationalen Fachkonferenzen nachfolgende Veröffentlichungen getätigt werden. Das dabei generierte Wissen ist in diese Arbeit mit eingeflossen, wurde aber um zahlreiche neue Aspekte erweitert, sodass diese für sich als eigenständige wissenschaftliche Veröffentlichung steht. Eine Übersicht über alle Publikationen findet sich in Anhang A.26.

### **PflegeCoDe – Pflegecoaching für die optimale Unterstützung von Menschen mit Demenz (BMBF, Laufzeit: September 2015 - September 2018)**

- Markus Schinle, Daniel Wyszka, Frank Schwärzler, Kirsten Volz, Marie-Anne Ruby, Emir Sejdinovic, Wilhelm Stork. An Approach to digitalize Psychological Tests to support Diagnosis of Alzheimer's Disease in Ambulatory Care. In *2018 IEEE International Symposium on Medical Measurements and Applications*, pages 1–6, Piscataway, NJ, 2018. IEEE. doi : 10.1109/MeMeA.2018.8438761
- Markus Schinle, Ioannis Papantonis, Wilhelm Stork. Personalization of Monitoring System Parameters to support Ambulatory Care for Dementia Patients. In *2018 IEEE Sensors Applications Symposium*, pages 1–6, Piscataway, NJ, 2018. IEEE. doi : 10.1109/SAS.2018.8336724
- Markus Schinle, Johannes Schneider, Timon Blöcher, Jochen Zimmermann, Sebastian Chiriac, Wilhelm Stork. A Modular Approach for Smart Home System Architectures Based on Android

Applications. In *MobileCloud 2017*, pages 153–156, Piscataway, NJ, 2017. IEEE. doi : 10.1109/MobileCloud.2017.20

- Marianne Behrends, Mathias Witte, Ralf Eckert, Markus Schinle, Matthias Gietzelt. Die Wohnung als Standort der Gesundheitsversorgung: Informatik in der guten Stube. In *Health — Exploring Complexity*. URL: <https://www.egms.de/static/en/meetings/gmds2016/16gmds126.shtml> [zuletzt abgerufen am 09.12.2021], doi : 10.3205/16GMDs126

### **Assistent – Assistierte Kommunikation in ambulanten Betreuungsformen (MSIBW, Laufzeit: April 2016 - März 2019)**

- Markus Schinle, Thomas Födisch, Wilhelm Stork. Assistierte Kommunikation in ambulanten Betreuungsformen: Teilhabe mit AAL. In *AAL Kongress*. URL: [https://hci.w-hs.de/wp-content/uploads/2018/11/pub\\_AAL\\_2018\\_Tagungsband\\_web.pdf](https://hci.w-hs.de/wp-content/uploads/2018/11/pub_AAL_2018_Tagungsband_web.pdf) [zuletzt abgerufen am 09.12.2021]

### **ODin – Organisationsgrenzen übergreifender Informationsaustausch mittels Distributed Ledger Technology (BMBF, Laufzeit: Januar 2019 - Juni 2020)**

- Markus Schinle, Christina Erler, Philip Nicolai Andris, Wilhelm Stork. Integration, Execution and Monitoring of Business Processes with Chaincode. In *2020 2nd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS)*, pages 63–70. IEEE, 9/28/2020 - 9/30/2020. doi : 10.1109/BRAINS49436.2020.9223283
- Markus Schinle, Christina Erler, Adrian Rudolf Vetter, Wilhelm Stork. How to disclose selective Information from permissioned DLT-based Traceability Systems? In *2020 IEEE International Conference on Decentralized Applications and Infrastructures (DAPPS)*, pages 153–158. IEEE, 8/3/2020 - 8/6/2020. doi : 10.1109/DAPPS49028.2020.00020
- Markus Schinle, Christina Erler, Wilhelm Stork. Distributed Ledger Technology for the systematic Investigation and Reduction of Information Asymmetry in Collaborative Networks. In Tung Bui, editor, *Proceedings of the 53rd Hawaii International Conference on System Sciences*, Proceedings of the Annual Hawaii International Conference on System Sciences. Hawaii International Conference on System Sciences, 2020. doi : 10.24251/HICSS.2020.651

### **BloG<sup>3</sup> – Blockchain-basiertes Gesundheitsdatenmanagement für gesamtheitliche Gesundheitsprofile (BMBF, Laufzeit: März 2020 - März 2023)**

- Christina Erler, Markus Schinle, Michael Dietrich, Wilhelm Stork. Decision model to design a blockchain-based system for storing sensitive health data. In *An Online AIS Conference (Hg.) 2022 – In ECIS 2022 Proceedings*. URL: [https://aisel.aisnet.org/ecis2022\\_rp/151](https://aisel.aisnet.org/ecis2022_rp/151)
- Markus Schinle, Christina Erler, Wilhelm Stork. Data Sovereignty in Data Donation Cycles: Requirements and Enabling Technologies for the Data-driven Development of Health Applications. In Tung Bui, editor, *Proceedings of the 54th Hawaii International Conference on System Sciences*, Proceedings of the Annual Hawaii International Conference on System Sciences. Hawaii International Conference on System Sciences, 2021. URL: <http://hdl.handle.net/10125/71098> [zuletzt abgerufen am 09.12.2021]

- Markus Lücking, Raphael Manke, Markus Schinle, Lukas Kohout, Stefan Nickel, Wilhelm Stork. Decentralized patient-centric data management for sharing IoT data streams. In *2020 International Conference on Omni-layer Intelligent Systems (COINS)*, pages 1–6. IEEE, 8/31/2020 - 9/2/2020. doi : 10.1109/COINS49042.2020.9191653

#### **Metis Neurotec – Mit KI gegen neurologische Störungen (HGDF, Laufzeit: März 2021 - April 2022)**

- Markus Schinle, Christina Erler, Simon Leenstra, Simon Stock, Marius Gerdes, Wilhelm Stork. A Decision Process Model for De-Identification Methods on the Example of Psychometric Data. In *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pages 1–6. IEEE, 10/7/2021 - 10/8/2021. doi : 10.1109/ICECCME52200.2021.9591139
- Simon Stock, Florian Mazura, Fernando Gomez de La Torre, Marius Gerdes, Markus Schinle, Wilhelm Stork. MMAI - Mobile Moods AI; Electroencephalography Artifact Detection; Towards Objective Assessment of Mental States. In *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pages 01–06. IEEE, 10/7/2021 - 10/8/2021. doi : 10.1109/ICECCME52200.2021.9590972
- Markus Schinle, Christina Erler, Timon Schneider, Joana Plewnia, Wilhelm Stork. Data-driven Development of Digital Health Applications on the Example of Dementia Screening. In *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6. IEEE, 6/23/2021 - 6/25/2021. doi : 10.1109/MeMeA52024.2021.9478676

#### **ALFREDO – KI-basierte Plattform für die ambulante Früherkennung und individualisierte Behandlung von Demenz (MSIBW, Laufzeit: September 2021 - Dezember 2023)**

- Markus Schinle, Christina Erler, Maximilian Hess, Wilhelm Stork. Explainable Artificial Intelligence in Ambulatory Digital Dementia Screenings. *Studies in health technology and informatics*, 294:123–124, 2022. doi : 10.3233/SHTI220411
- Markus Schinle, Christina Erler, Mayumi Kaliciak, Christopher Milde, Simon Stock, Marius Gerdes, Wilhelm Stork. Digital Health Apps in the Context of Dementia: Questionnaire Study to Assess the Likelihood of Use among Physicians. *JMIR Formative Research*, 2022. doi : 10.2196/35961



# 2 Grundlagen und domänenspezifische Rahmenbedingungen

IN diesem Kapitel wird eine Wissensbasis geschaffen, die das Verständnis für die Inhalte und Motivation dieser Arbeit fördern soll. Aufgrund der Interdisziplinarität der bearbeiteten Problemstellung werden deshalb die aus Sicht des Autors beim Leser nicht vorauszusetzenden aber notwendigen Begrifflichkeiten definiert sowie wo nötig detailliert erklärt und voneinander abgegrenzt.

## 2.1 Verwendete Methoden, Metriken und Prozesse

Die Beschreibung der eingesetzten wissenschaftlichen Werkzeuge und Definitionen dient zur besseren Nachvollziehbarkeit der Vorgehensweise. Durch Definitionen wird ebenfalls ein Nachschlagewerk für angeführte mathematische Formeln, Begrifflichkeiten sowie verwendete Technologien und Algorithmen zur Verfügung gestellt.

### 2.1.1 Übergeordnetes wissenschaftliches Vorgehensmodell

Als übergeordnetes wissenschaftliches Vorgehensmodell wird in dieser Arbeit die Design Science Research Methodology (DSRM) verwendet. Diese ist dazu angedacht, Probleme an der Schnittstelle zwischen Informationstechnologie (IT) und Organisationen zu lösen, indem sogenannte Artefakte entwickelt werden, die zur Lösung wissenschaftlicher Probleme beitragen [188]. Eine Ausprägung dieses Ansatzes stellt der problem-initiierte empirische Forschungsprozess nach Peffers et al. dar, der für Forschungskategorien im Bereich von Informationssystemen verbreitet ist, in welche diese Arbeit ebenfalls einzuordnen ist (vgl. Abbildung 2.1). Dieser Prozess wird jeweils iterativ für die identifizierten Teilprobleme durchlaufen. Durch die Integration neu gewonnenen Wissens in die agile Entwicklung von IT-Artefakten zur Beantwortung der Forschungsfragen, werden diese den Anforderungen der Realität immer weiter angenähert. Die Vorgehensweise lässt sich grundsätzlich in sechs Prozessschritte unterteilen [323]:

*Prozessschritt 1* Problemidentifikation und Motivation,

*Prozessschritt 2* Definition von Zielen zur Lösung,

*Prozessschritt 3* Design und Entwicklung des Lösungsartefakts,

*Prozessschritt 4* Demonstration des Lösungsartefakts,

*Prozessschritt 5* Bewertung der Wirksamkeit und Effizienz,

Prozessschritt 6 Kommunikation.

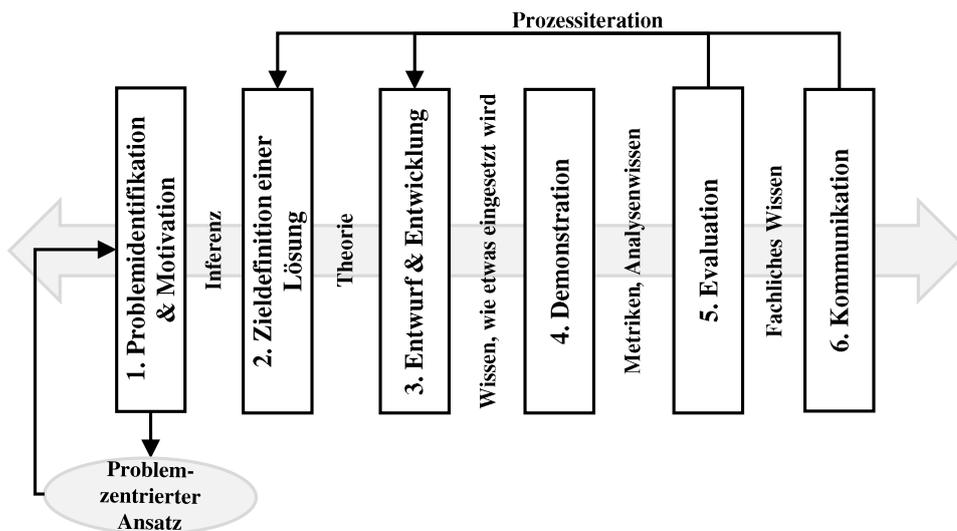


Abbildung 2.1: DSRM nach Peffers et. al. [323].

## 2.1.2 Methode zur Prüfung der Gebrauchstauglichkeit

Unter „Usability Engineering“ wird die gebrauchstaugliche Gestaltung von Systemen verstanden [302]. Um diese auf Gebrauchstauglichkeit zu testen existieren neben Normen (vgl. Unterunterabschnitt 2.3.1) in der Forschung erprobte Methoden, welche bei der Erfassung und Bewertung der Anforderungen an ein gebrauchstaugliches Produkt unterstützen (vgl. Absatz 2.4.2). Zur Untersuchung der Interaktion zwischen Menschen und Technik, insbesondere auch mit digitaler Technik, kann die Thinking Aloud Methode (TAM) verwendet werden, welche von Nielsen als „die wertvollste Usability-Engineering-Methode“ beschrieben wird [301, 302]. Hierbei werden Personen gebeten laut nachzudenken während sie ein Problem lösen [411]. Die dabei verbalisierten Gedanken werden protokolliert und analysiert [411]. Da sich diese Methode für anwendungsnahe Forschung sowohl zu kognitiven Prozessen als auch für den Wissenserwerb bezüglich des Aufbaus wissensbasierter IT-Systeme eignet, wird auch in dieser Arbeit darauf zurückgegriffen [411].

## 2.1.3 Vorgehensmodell zur Wissensentdeckung

Unter „Knowledge Discovery“ oder auch Wissensentdeckung wird der Identifikationsprozess von Mustern in Datenbeständen durch Analysen verstanden, die neu, gültig, verständlich und nützlich sind [408, S. 40 f.][379, S. 52]. Hierfür entstanden in den letzten Jahren verschiedene strukturierte Vorgehensmodelle. Je nach Anwendungsfall werden darin Fachrichtungen wie das Wissensmanagement, Maschinelles Lernen (ML), Künstliche Intelligenz (KI), Statistik, Datenbank- und Informationssysteme im Zuge von Data Mining (DM) verknüpft sowie entsprechend interdisziplinär genutzt, um Wissen aus Daten zu extrahieren (vgl. Unterabschnitt 2.4.3) [379, S. 52]. Während sich die akademische Welt an Modellen wie das Knowledge Discovery in Databases (KDD) orientiert, werden im industrienahem Umfeld die Modelle Sample, Explore, Modify, Model, and Assess (SEMMA) oder Cross-industry Standard Process for Data Mining (CRISP-DM) eingesetzt [276, 339, 408]. CRISP-DM stellt im Gegensatz zu KDD die betriebswirtschaftliche Problemstellung und ein zyklisches Vorgehen in den Vordergrund, was sich jedoch im

KDD-Modell auch in einer iterativen Vorgehensweise im wissenschaftlichen Ziel der Wissensfindung widerspiegelt [35]. Unterschiede zwischen den Modellen liegen somit in deren Betrachtungsweisen und Zielen, letztlich dokumentieren aber beide die gleichen Prozessabläufe [35, S. 185]. Da die Motivation der Arbeit auch einem ökonomischen Ziel folgt und der Transfer von Wissen in ein digitales Produkt in Form einer Gesundheitsanwendungen stattfinden soll, orientiert sich diese Arbeit an der nachfolgend detaillierter ausgeführten CRISP-DM. Die CRISP-DM wird aufgrund einer intuitiven und einfachen Beschreibung als Standard-Prozess des Data Minings gesehen [380]. Trotz des Fokus auf ein industrienahes Umfeld, findet dieser Prozess auch Anwendung in der Forschung [244]. Er lässt sich in die folgenden sechs Schritte unterteilen:

- *Business Understanding*: Initial geht es darum, die Ziele des Data Minings abzustecken und zu verstehen [380]. Hierbei spielen sowohl der Verwendungszweck, das -gebiet, dafür benötigte Ressourcen (wie Hard- und Software) als auch Anforderungen an die Performanz, die Robustheit und die Definition der Verständlichkeit von Modellen für Menschen eine Rolle [213].
- *Data Understanding*: Das Verständnis der Daten muss im nächsten Schritt erlangt werden, indem eine hinreichende Exploration einer initialen Datenmenge stattfindet [213]. Dies wird als vollständig angenommen, sobald ersichtlich wird, woher die Daten stammen, eine Beschreibung dieser Daten vorliegt, die Art, Qualität und Umfang sowie Merkmale, welche die zu untersuchende Hypothese stützen und deren Datentypen beschrieben sind [213, 380]. Da das Verständnis für die Daten und die Domäne iterativ entsteht, wird diese Phase häufig im Wechsel mit der Phase des Business Understanding durchlaufen.
- *Data Preparation*: Diese Phase umfasst Vorverarbeitungsschritte der Daten, wozu beispielsweise die Merkmalsextraktion, die Standardisierung der Daten oder die Ausreißerentfernung gehören [380]. Eine Aufteilung der Daten in Merkmale (Features) und Referenzmarker (Labels) ist ebenfalls Teil dieser Phase [380].
- *Modeling*: Dieser Schritt umfasst das Erstellen eines gewünschten Modells und ist daher eng verknüpft mit der Data Preparation, da zur Modellierung benötigte Daten durch diese Phase zur Verfügung gestellt werden. Die Modellierung stellt einen iterativen Prozess dar, in dem das Modell schrittweise verbessert wird [380].
- *Evaluation*: Mit der Evaluationsphase erfolgt die Qualitätsbeurteilung der erstellten Modelle. Dafür muss auf Wissen der Phase des Business Understanding zurückgegriffen werden, um die erzielten Ergebnisse richtig einordnen zu können [380].
- *Deployment*: Das Modell, welches am besten geeignet ist, wird schließlich innerhalb des Anwendungsfalls realisiert [380].

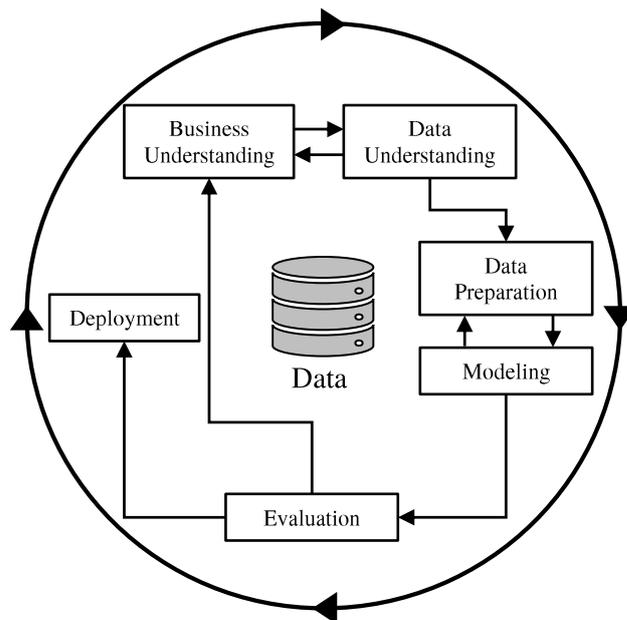


Abbildung 2.2: CRISP-DM nach Shearer [380].

## 2.1.4 Statistische Verfahren und Metriken

Nachfolgend werden die in dieser Arbeit verwendeten mathematischen Definitionen und Verfahren dargestellt sowie deren Einsatzzweck kurz erläutert. Hierbei spielen sowohl die Teilgebiete der deskriptiven als auch der induktiven Statistik eine Rolle. Die Inhalte der verwendeten statistischen Werkzeuge und Konzepte stammen primär von Sachs und Hedderich [358]. Grundlegend unterscheiden diese zwischen konfirmatorischen und explorativen Ansätzen [358]. Während die explorativen Ansätze dazu dienen Daten zu beschreiben, darzustellen und Auffälligkeiten zu identifizieren, um Hypothesen und passende Modelle aufzuspüren, werden die konfirmatorischen Ansätze dazu genutzt, diese induktiv zu überprüfen und zu beurteilen [358]. Die Kenntnis grundsätzlicher statistischer Maße und Darstellungsarten insbesondere bezüglich der deskriptiven Statistik in den explorativen Ansätzen wird vorausgesetzt.

### Parametrische und nicht-parametrische Verfahren

Grundsätzlich werden die konfirmatorischen Ansätze in parametrische oder nicht-parametrische Verfahren unterteilt [358]. Der Entscheidung welches Verfahren eingesetzt wird liegt eine Verteilungsannahme der Zufallsvariable zugrunde [358]. Während für parametrische Verfahren eine bestimmte Verteilungsannahme wie zum Beispiel eine Normalverteilung getroffen werden muss, sind nicht-parametrische Verfahren unabhängig von der Verteilung [358]. Als Vorteile werden den parametrischen Verfahren eine höhere Genauigkeit und dadurch mehr statistische Power zugesprochen, nicht-parametrische Verfahren gelten als konservativer, da sie gegebenenfalls länger die Nullhypothese bestätigen als notwendig [358].

## Transformation einer Zufallsvariablen

Als Berechnungsvorschrift zur Standardisierung der Werte einer Zufallsvariablen wird die z-Transformation angewandt. Hierbei werden alle Werte auf den Mittelwert  $\bar{x} = 0$  und die Varianz  $\sigma^2 = 1$  standardisiert [65]:

$$z = \frac{X - \bar{x}}{\sigma} \quad (2.1)$$

## Korrelations- und Regressionsanalyse

Zur Analyse der Korrelation zwischen zwei Variablen werden Korrelationskoeffizienten als Maß eingesetzt [358]. Deren Wertebereich ist auf -1 bis 1 normiert und sagt aus, dass eine perfekte negative Korrelation bei -1 vorliegt sowie bei +1 gegenteilig eine perfekte positive Korrelation [358]. Der Wert 0 zeigt an, dass keine Korrelation festgestellt werden kann [358]. Rückschlüsse auf einen linearen Zusammenhang in der Grundgesamtheit lassen sich mit dem Korrelationskoeffizient nach Pearson ziehen, der von einer Normalverteilungsannahme ausgeht [358]. Sollte diese Annahme verworfen werden müssen, bietet eine weitere Möglichkeit der Rang-Korrelationskoeffizient nach Spearman [358]. Dieser beobachtet ausschließlich die Abhängigkeit der Ränge von Variablenausprägungen und prüft diese auf eine allgemeinere monotone Abhängigkeit [358]. Auf Grundlage von Korrelationen verfolgt die Regressionsanalysen das Ziel Zusammenhänge zwischen Variablen zu nutzen, um bestmögliche Vorhersagen für Variablen zu treffen.

## Statistische Hypothesentests

Durch statistische Hypothesentests können Hypothesen ( $H_1$ ) über Eigenschaften von einer oder mehreren Zufallsvariablen aufgestellt und überprüft werden [358]. Hierfür werden Nullhypothesen ( $H_0$ ) aufgestellt, die als Gegenhypothese der zu überprüfenden Hypothese gegenübergestellt werden [358]. Durch das empirische Widerlegen der Nullhypothese wird die Hypothese angenommen [358]. Als Prüfgröße hierfür existiert der sogenannte p-Wert, der angibt welches Signifikanzniveau akzeptiert wird, damit die Nullhypothese gerade noch abgelehnt werden kann [358]. Dadurch wird die Wahrscheinlichkeit einen Fehler erster Art zu begehen, nämlich fälschlicherweise die Nullhypothese abzulehnen, auf das gewählte Signifikanzniveau  $\alpha$  reduziert [358]. In der Literatur findet sich üblicherweise ein  $\alpha$  von 5% [358]. Je geringer der Umfang einer zugrundeliegenden Stichprobe ist desto stärker muss der untersuchte Effekt sein, um das Signifikanzniveau zu erreichen [97]. Ein kleiner Effekt kann in einer umfangreichen Stichprobe trotz der geringen Effektstärke dennoch signifikant sein [97].

**Shapiro-Wilk-Test:** Mit Hilfe des Shapiro-Wilk-Test (SWT) kann die Annahme der Normalverteilung einer Zufallsvariable geprüft werden, indem die Nullhypothese aufgestellt wird, dass Beobachtungen einer Stichprobe dieser Verteilung folgen. Der SWT gilt als das Verfahren zur Überprüfung einer Normalverteilungsannahme mit der höchsten statistischen Power [358].

**Kruskal-Wallis-Test:** Der Kruskal-Wallis-Test (KWT) oder auch H-Test wird der Gruppe von nicht-parametrischen Verfahren zugeordnet und ermöglicht die Prüfung auf gleiche Verteilung unabhängiger Stichproben. Als unabhängig gelten die Stichproben, wenn deren Zufallsvariablen unabhängig voneinander sind. Hierfür wird die Nullhypothese aufgestellt, dass eine Anzahl an k unabhängigen Stichproben

derselben Grundgesamtheit entstammen und damit der gleichen Verteilung folgen. Als alternatives parametrisches Verfahren dient bei bekannter Verteilung die einfaktorielle Analysis of Variance (ANOVA) [358].

**Wilcoxon-Rangsummen-Test:** Der Wilcoxon-Rangsummen-Test (WRT)

oder auch U-Test wird als nicht-parametrisches Verfahren eingesetzt, um die Ränge zwischen zwei unabhängigen Stichproben zu vergleichen. Dabei wird die Nullhypothese überprüft, dass mit gleicher Wahrscheinlichkeit eine Ausprägung von einer Stichprobe größer oder kleiner ist als eine zufällig gezogene Ausprägung einer anderen Stichprobe [358]. Diese Art von Test prüft, ob systematische Unterschiede zwischen den Werten von Stichproben bestehen und wird beispielsweise angewandt, um die Wirksamkeit von Medikamenten zwischen Interventions- und Kontrollgruppen zu vergleichen. Als parametrische Alternative steht ein sogenannter t-Test bei unabhängigen Stichproben zur Verfügung [358].

**Wilcoxon-Vorzeichen-Rang-Test:** Ähnlich wie der WRT ist der Wilcoxon-Vorzeichen-Rang-Test (WVRT) ein Hypothesentest, der zur Gruppe der nichtparametrischen Verfahren zählt. Hierbei wird jedoch für gepaarte Stichproben verglichen, ob die Differenzen zwischen den Ausprägungen der Stichproben um den Wert Null symmetrisch verteilt liegen [358]. Wie durch den WRT wird dadurch vergleichbar, ob die Werte von Stichproben systematisch größer oder kleiner sind, jedoch zwischen gepaarten und somit abhängigen Stichproben. Damit ist beispielsweise ein Vergleich der Entwicklungen von Leistungen derselben Schüler innerhalb eines Schuljahres in einer Klasse über die Testzeitpunkte hinweg möglich. Um die Wahrscheinlichkeit zufällig angenommener Signifikanz für Merkmale (Fehler erster Art) zwischen zwei Messzeitpunkten zu minimieren findet eine Korrektur der  $\alpha$ -Fehlerkumulierung statt. Neben der konservativen Bonferroni-Korrektur, lässt sich durch die Holm-Prozedur eine höhere Power erzielen, weshalb diese vorzuziehen ist [358, S. 430]. Lokale p-Werte, die den Schwellwert für die Signifikanz darstellen, werden gemäß der Gleichung 2.2 definiert.

$$p_{\text{lokal}}(i) = \frac{\alpha}{n - i + 1} \quad (2.2)$$

Wobei  $i$  für den Rang des p-Werts bei aufsteigender Sortierung steht, das zuvor definierte Signifikanzniveau  $\alpha$  gewählt wird und die Anzahl an Testdurchführungen durch  $n$  definiert ist. Der t-Test für gepaarte Stichproben stellt hierfür das parametrische Äquivalent für normalverteilte Zufallsvariablen dar [358].

## Indizes und Maße

Insbesondere im medizinischen Umfeld, beispielsweise bei der Entwicklung diagnostischer Tests, werden Maße zur Bewertung der Leistungsfähigkeit von Modellen und Schätzern benötigt [286]. In diesem Abschnitt werden Metriken vorgestellt, die in dieser Arbeit als Bewertungskriterium für die Güte solcher Modelle eingesetzt werden.

**Genauigkeit und Interrater-Reliabilität:** Als zentrales Bewertungskriterium für die Güte eines Schätzers gilt dessen Genauigkeit (engl. *accuracy*) [286]. Am Beispiel einer binären Klassifikation eines medizinischen Tests ist das der Anteil an Fällen, der den richtig klassifizierten Fällen, also den Richtig-Positiv (RP) und Richtig-Negativ (RN) zugeordnet wird. Diese stehen im Verhältnis zu allen Vorhersagen

also der Summe aus richtigen Vorhersagen und falschen Vorhersagen, die als Falsch-Positiv (FP) und Falsch-Negativ (FN) klassifiziert sind (vgl. Gleichung 2.3).

$$accuracy = \frac{RP + RN}{RP + RN + FP + FN} \quad (2.3)$$

Die Fehlerrate (error rate) entspricht demnach der in Gleichung 2.4 dargestellten Formel.

$$errorrate = 1 - accuracy \quad (2.4)$$

Um bei einer unbalancierten Datenlage der Verzerrung dieses Maßes durch einen zu großen Anteil an richtigen Zufallsübereinstimmungen für kategoriale Merkmale entgegenzuwirken, wird zur Überprüfung der Übereinstimmung zwischen zwei Ratern (Interrater-Reliabilität) der sogenannte Cohen's Kappa-Wert ( $\kappa$ ) herangezogen [358]. Hierbei werden nach Cohen folgende zwei Größen definiert [96, 358]:

- Anteil der Fälle, bei denen die Schätzer übereinstimmen ( $p_0$ ).
- Anteil der Fälle, bei denen eine zufällige Übereinstimmung der Schätzer zu erwarten ist ( $p_c$ ).

Dabei gibt es verschiedene Definitionen des kappa-Werts, wobei in dieser Arbeit die in der Literatur gängigste Definition nach Cohen verwendet wird [96, 249] (vgl. Gleichung 2.5).

$$\kappa = \frac{p_0 - p_c}{1 - p_c} \quad (2.5)$$

Diese hat den Vorteil auf den Wertebereich zwischen 0 und 1 normiert zu sein. Dabei werden unterschiedliche Wertebereiche zur Beurteilung der Übereinstimmungsstärke beider Rater angegeben (vgl. Tabelle 2.1).

Wertebereich von $\kappa$	Stärke der Übereinstimmung
< 0	keine
0,10 - 0,40	schwach
0,41 - 0,60	deutlich
0,61 - 0,80	stark
0,81 - 1,00	fast vollständig

Tabelle 2.1: Stärke der Übereinstimmung für  $\kappa$ -Werte nach Landis et al. [248].

**Zusammenhänge zwischen Genauigkeitsmaßen:** Um den zuvor angedeuteten Limitationen des Genauigkeitsmaßes (vgl. Gleichung 2.3) entgegenzutreten, wird die Entscheidungsperformanz binärer Klassifikatoren mittels zweier gepaarter Indizes bewertet [286]. Die Sensitivität (vgl. Gleichung 2.6), auch Richtig-Positiv-Rate (RPR) oder Recall genannt, sowie Spezifität (vgl. Gleichung 2.7), auch Richtig-Negativ-Rate (RNR) genannt, sind Maße, welche in Bezug auf die Erkennung einer Krankheit den Anteil der richtigerweise positiv beziehungsweise richtigerweise negativ klassifizierten Fälle in Bezug auf die wirklich positiven beziehungsweise negativen Fälle angeben [286].

$$\text{RPR} = \frac{\text{RP}}{\text{RP} + \text{FN}} \quad (2.6)$$

$$\text{RNR} = \frac{\text{RN}}{\text{RN} + \text{FP}} \quad (2.7)$$

Dem gegenüber stehen die Falsch-Negativ-Rate (FNR) und Falsch-Positiv-Rate (FPR) (vgl. Gleichung 2.8 und Gleichung 2.9) [286].

$$\text{FNR} = \frac{\text{FN}}{\text{FN} + \text{RP}} = 1 - \text{RPR} \quad (2.8)$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{RN}} = 1 - \text{RNR} \quad (2.9)$$

Mit dem Positiver Prädiktiver Wert (PPW) oder auch Präzision (precision) und Negativer Prädiktiver Wert (NPW) wird die Wahrscheinlichkeit ausgedrückt, wonach ein positives beziehungsweise negatives Ergebnis einer Testung auch tatsächlich einem positiven oder negativen Ergebnis entspricht (vgl. Gleichung 2.10 und Gleichung 2.11).

$$\text{PPW} = \frac{\text{RP}}{\text{RP} + \text{FP}} \quad (2.10)$$

$$\text{NPW} = \frac{\text{RN}}{\text{RN} + \text{FN}} \quad (2.11)$$

Zum Vergleich der Güte verschiedener Klassifikatoren bietet sich der F1-Score als sogenannter harmonischer Mittelwert zwischen der Sensitivität und der Präzision an (vgl. Gleichung 2.12) [280].

$$F1 = 2 \frac{\text{PPW} \times \text{RPR}}{\text{PPW} + \text{RPR}} \quad (2.12)$$

Da bei einer medizinischen Testung möglichst viele der nicht als gesund einzustufenden Testteilnehmer erkannt werden sollen, würde eine reine Optimierung der Sensitivität dazu führen, dass jede Testperson als krank klassifiziert wird. Dadurch würde jedoch auch die Spezifität entsprechend beeinflusst. Ziel dieser gepaarten Metriken ist es die Abwägungsentscheidung zwischen Sensitivität und Spezifität abzubilden, um als relevantes Optimierungskriterium eingesetzt werden zu können. Der zuvor definierte Cohen's Kappa-Wert (vgl. Gleichung 2.5) oder nachfolgend erläuterte Mechanismen, können diese Abwägungsentscheidung abbilden (vgl. Absatz 2.1.4).

**Receiver Operating Characteristic und Area Under the Curve:** Mit Hilfe der zuvor definierten Indizes wie RPR und RNR lässt sich die Genauigkeit von Verfahren differenziert ausdrücken [286]. Jedoch hängt deren Aussagekraft ebenso von einem gewählten Schwellwert ab, der in diesem binären Fall den Entscheidungspunkt zur Differenzierung zwischen der positiven oder negativen Klasse festlegt [286]. Die Receiver Operating Characteristic (ROC)-Kurve versucht diesen Zusammenhang empirisch zu beschreiben, indem alle Kombinationen der relativen Häufigkeiten richtiger und falscher Entscheidungen auf Basis der möglichen Schwellwerte beschrieben werden [286]. Wie die ROC-Kurve in Abbildung 2.3 veranschaulicht, wird die RPR auf der Ordinate gegen die FPR auf der Abszisse des Diagramms aufgetragen. Werden die Klassifizierungsergebnisse in Abhängigkeit von einer Reihe verschiedener Schwellenwerte wie beispielsweise  $c \in [0, 1]$  aufgetragen, erhält man eine gekrümmt aufsteigende Kurve. Diese sagt aus, dass eine perfekte Klassifikation und damit ein perfektes Modell dann erreicht ist, wenn die Schwelle so gewählt ist, dass der Punkt (0,1) erreicht wird, also 100% aller Instanzen in die Klassen RN und RP eingeordnet werden. Die Punkte (0,0) für  $c = 1$  und (1,1) für  $c = 0$  beschreiben triviale Klassifikatoren, bei

denen die Instanzen nur in die Klassen der Negativen beziehungsweise Positiven prognostiziert werden. Die erste Winkelhalbierende stellt die Basislinie (baseline) der ROC-Kurve dar, welche für den binären Klassifikator eine Trefferwahrscheinlichkeit von 50% angibt. Ein sinnvoller Schwellwert sollte also zu einem Punkt oberhalb der Basislinie führen. Der optimale Schwellwert ist durch den R-Wert mit dem größten Normalabstand zur Diagonalen also zum obersten, linken Punkt auf der ROC-Kurve definiert beziehungsweise durch eine minimale Distanz zum Punkt (0,1) [242]. Um insgesamt die Qualität eines

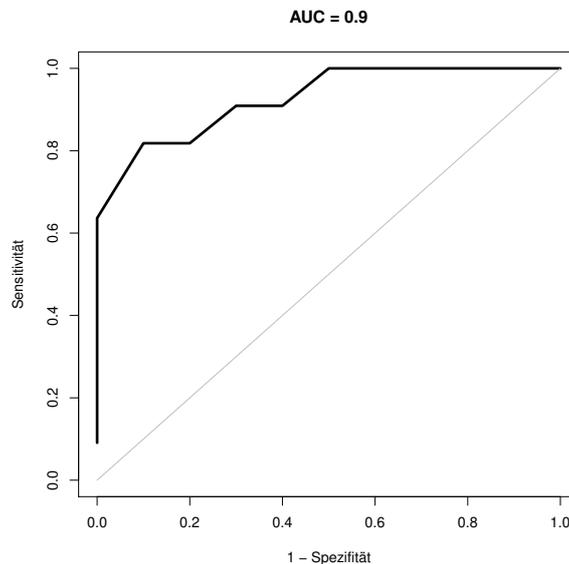


Abbildung 2.3: Beispiel einer ROC-Kurve.

Prädiktors zu bestimmen, also ein Maß zu finden, welches unabhängig vom Schwellwert eine Beurteilung der Modellqualität zulässt, wird die Fläche unter der ROC-Kurve verwendet und als Area Under the Curve (AUC) bezeichnet. Aufgrund dieser Eigenschaft ist der AUC-Wert für generelle Vergleiche zwischen unterschiedlichen Modellen und Datensätzen geeignet [398, S. 700], [211, S. 245]. Er ist nur für Analysen in binäre Zieldimensionen definiert. Der AUC-Wert weist gegenüber ungleichen Gruppengrößen eine hohe Stabilität auf [41, S. 300] und ist daher im Rahmen prädiktiver Analysen auf Daten mit ungleich verteilten Klassen geeignet [41, 142, 263, 398].

**Ähnlichkeitsmaß:** Als relatives Streuungsmaß, mit dem die Variationen der Werte unabhängig von deren Größenordnung beurteilt werden, gilt der Variationskoeffizient. Dieser lässt sich als Streuung in Prozent um den Mittelwert interpretieren (vgl. Gleichung 2.13). Zu beachten ist, dass dieser nur für Werte mit den gleichen Vorzeichen sinnvoll ist, da ansonsten ein Mittelwert von 0 auftreten kann [234].

$$v(X) = \frac{sd(X)}{E(X)} \quad (2.13)$$

## 2.2 Gesundheitswesen und medizinisches Hintergrundwissen

Da die Einordnung der Problemstellung medizinisches sowie gesundheitsökonomisches und -politisches Wissen voraussetzt wird nachfolgend dieses Themenfeld umrissen. Aufgrund des Umfangs und der Komplexität dieser Themenbereiche kann nachfolgend lediglich ein Überblick zur Verdeutlichung gegeben werden.

### 2.2.1 Das deutsche Gesundheitswesen

In diesem Abschnitt wird auf relevante Begriffe und Institutionen aus dem deutschen Gesundheitswesen eingegangen, die insbesondere für das Verständnis der juristischen Rahmenbedingungen relevant sind (vgl. Abschnitt 2.3). Da sich weltweit die Gesundheitssysteme und deren Aufbau unterscheiden, wird nachfolgend eine Übersicht über das Zusammenspiel im deutschen Gesundheitssystem gegeben. Der gesetzliche Rahmen hierzu findet sich primär im Sozialgesetzbuch (SGB) V (vgl. Abschnitt 2.3).

#### Wie funktioniert das deutsche Gesundheitssystem?

Das deutsche Gesundheitssystem lässt sich grundlegend durch die vier Eigenschaften Versicherungspflicht, Beitragsfinanzierung, Solidaritäts- und Selbstverwaltungsprinzip charakterisieren [252]. Grundsätzlich bedeutet das für deutsche Staatsbürger, dass sie dazu verpflichtet sind sich zu versichern. Hierfür stehen die Organisationskonstrukte Gesetzliche Krankenversicherung (GKV) und Private Krankenversicherung (PKV) zur Verfügung. Für circa 90% erfolgt dies im Rahmen einer GKV, die als „Vollkaskoversicherung ohne Selbstbeteiligung“ interpretiert werden kann [252]. Hierbei gilt das Solidaritätsprinzip, das jedem gesetzlich Versicherten unabhängig von seiner individuellen Leistungsfähigkeit einen äquivalenten Versorgungsanspruch garantiert und sich durch Beiträge aller Krankenversicherten sowie der Arbeitgeber finanziert [78, 252]. Von gesetzlich Versicherten (Leistungsnehmer) über den ersten Gesundheitsmarkt in Anspruch genommene medizinische Sach- oder Dienstleistungen werden mit wenigen Ausnahmen direkt zwischen den Leistungserbringer (LE) und der GKV als Kostenträger abgerechnet [78, S.24]. Eine Ausnahme hierfür bilden Bevölkerungsgruppen, die beruflich selbstständig sind, einer Anstellung als Beamte nachgehen oder deren Einkünfte über einem gewissen Sockelbetrag (Versicherungspflichtgrenze) liegen [252]. Diese organisieren sich in PKV, welche einen individuellen Tarif auf Grundlage der Präferenzen und Risiken des Versicherten berechnen. Hierbei muss der Versicherte bei Inanspruchnahme von medizinischen Versorgungsleistungen in der Regel zunächst selbst in Vorleistung treten [78, S.24]. Leistungen wie die sogenannte Individuelle Gesundheitsleistung (IGeL), müssen im gesetzlichen und privaten System durch den Leistungsnehmer selbst getragen werden und sind somit dem sogenannten zweiten Gesundheitsmarkt zuzurechnen.

Das Prinzip der Selbstverwaltung im Gesundheitswesen basiert auf der Aufteilung zwischen Zuständigkeiten in der medizinischen Versorgung. Einerseits existiert die staatliche Zuständigkeit für die Gestaltung gesetzlicher Rahmenbedingungen [252]. Andererseits die Zuständigkeit der Einrichtungen zur Selbstverwaltung für die konkrete Finanzierung, Ausgestaltung und Organisation medizinischer Leistungen [252]. Diese werden durch die nachfolgend erklärten Institutionen als Vertreter der LE, Kostenträger und Leistungsnehmer gemeinsam diskutiert und eigenverantwortlich durchgeführt.

## Welche Akteure gibt es?

Die Notwendigkeit einer prinzipiellen Unterscheidung zwischen Leistungserbringenden, Leistungsempfangenden und Kostenträger wurde zuvor erläutert. Die erste Ebene der Selbstverwaltung bilden die staatlichen Stellen (vgl. Abbildung 2.4). Das oberste Gremium auf Bundesebene im Rahmen der gesetzlichen Krankenversicherung ist der Gemeinsame Bundesausschuss (G-BA) [252]. Der G-BA legt Richtlinien fest, die bestimmen welche medizinische Leistungen durch gesetzlich Versicherte in Anspruch genommen werden können. Außerdem definiert dieses Organ Maßnahmen zur Sicherung der Qualität in Krankenhäusern und Praxen [160]. Dabei steht der G-BA unter Rechtsaufsicht des Bundesministerium für Gesundheit (BMG) und setzt sich auf Seiten der Leistungserbringer aus Vertretern der Kassenzahnärztliche Bundesvereinigung (KZBV), Kassenärztliche Bundesvereinigung (KBV) und Deutsche Krankenhausgesellschaft (DKG) sowie der Spitzenverband Bund der Krankenkassen (SGKV) auf Seiten der Kostenträger zusammen, die sich wiederum dem föderalen System folgend in kommunale und Landes-Gruppierungen unterteilen [80, 160]. Diese können Stimm-, Mitberatungs- und Antragsrechte ausüben. Die Interessensvertretungen der Leistungsnehmer setzen sich zum Zeitpunkt der Anfertigung dieser Arbeit aus der Deutsche Arbeitsgemeinschaft Selbsthilfegruppen e.V. (DAS), dem Deutscher Behindertenrat (DBR), der Bundesarbeitsgemeinschaft der PatientInnenstellen (BAGP) und der Verbraucherzentrale Bundesverband e.V. (VB) zusammen [160]. Diese besitzen lediglich Mitberatungs- und Antragsrechte. Nachfolgend wird auf die weiteren relevanten Akteure kurz genauer eingegangen. Das Leistungsspektrum der SGKV

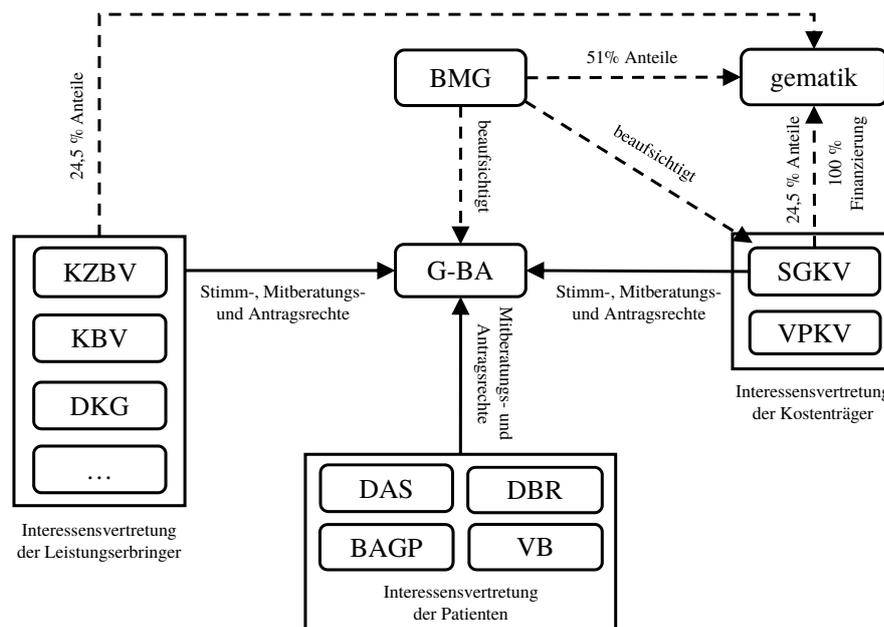


Abbildung 2.4: Akteure im selbstverwaltenden Gesundheitswesen [80, 158, 160, 252].

umfasst neben der Interessensvertretung der gesetzlich Versicherten gegenüber den Leistungserbringern und der Politik für die pflegerische und gesundheitliche Versorgung, Fragen bezüglich der Finanzierung sowie des Datenmanagements innerhalb der GKV [163]. Letzterer Punkt ist nach § 217f Absatz 2 und 7 SGB V geregelt, wonach er die Förderung der Ausgestaltung des elektronischen Datenaustausches durch die Entwicklung und Standardisierung innerhalb der GKV zuständig ist. Für dessen Umsetzung hat die SGKV im Jahre 2005 unter Beteiligungen der anderen Interessensvertreter und des BMG die Gesellschaft für Telematikanwendungen der Gesundheitskarte mbH (gematik) ins Leben gerufen. Gemäß § 291 SGB

V ist ihr Ziel die Einführung, Pflege und Weiterentwicklung der elektronischen Gesundheitskarte (eGK) und der Telematikinfrastruktur (TI) (vgl. Unterabschnitt 2.4.1), wodurch Verbesserungen in der Qualität, Transparenz und Wirtschaftlichkeit im Gesundheitswesen erreicht werden sollen [157]. Im Zuge der Einführung des Terminservice- und Versorgungsgesetz (TSVG) übernimmt das BMG im Jahr 2019 mit 51% der Gesellschafteranteile die Mehrheit an der gematik. Ziel ist die Beschleunigung der Einführung einer Elektronische Patientenakte (ePA) durch schnellere Beschlussfassungen [122] (vgl. Unterabschnitt 2.4.1). Als zusätzliche Ebene von Akteuren werden Unternehmen der Medizintechnik- und Pharma-Branche gesehen [252]. Diese stehen bei der Entwicklung und dem Inverkehrbringen von Medizinprodukten und pharmakologischen Präparaten ebenfalls unter der Aufsicht staatlicher Stellen (vgl. Abbildung 2.5). Das

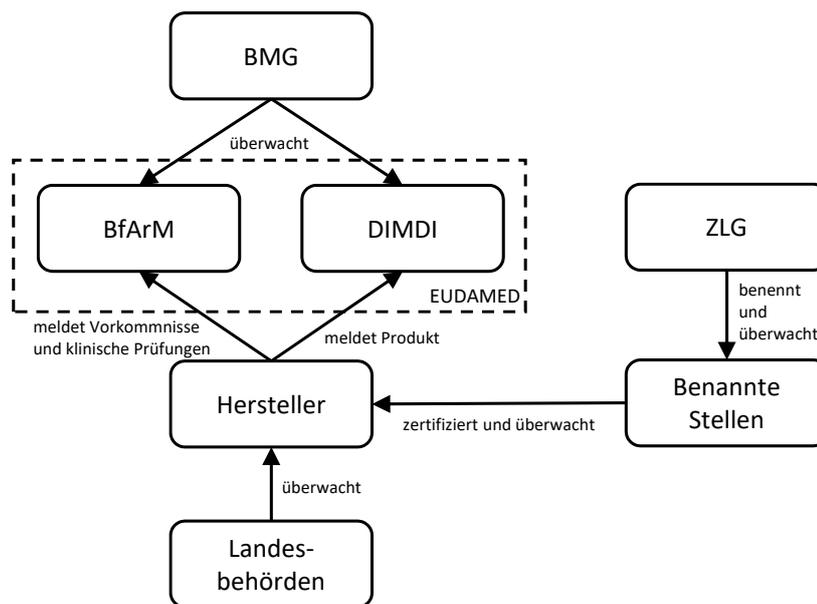


Abbildung 2.5: Akteure in der Medizinproduktherstellung nach Hastenteufel und Renaud [180].

BfArM ist für die Überwachung von Risiken sowie die Genehmigung klinischer Studien in diesem Kontext zuständig und wird durch das BMG überwacht [180]. Vorkommnisse, die während der Entwicklung und Anwendung dieser Produkte entstehen, müssen durch die Hersteller an das BfArM gemeldet werden. Das Deutsche Institut für Medizinische Dokumentation und Information (DIMDI) untersteht ebenfalls dem BMG und betreibt ein elektronisches Verzeichnis für Deutschland indem alle Hersteller ihre zugelassenen Medizinprodukte registrieren müssen [180]. Nach Inkrafttreten der Medical Device Regulation (MDR) ab Mai 2021 werden die durch das BfArM und DIMDI betriebenen Systeme auf europäischer Ebene in das gemeinschaftliche European Database on Medical Devices (EUDAMED) konsolidiert (vgl. Unterunterabschnitt 2.3.1). Um ein Medizinprodukt auf den Markt bringen zu dürfen sind für Hersteller in der Regel Zertifizierungen notwendig, insbesondere bezüglich der Einführung und Überwachung ihrer Qualitätsmanagementsysteme. Hierfür werden Benannte Stellen eingesetzt deren Auftrag neben der Konformitätsbewertung für Hersteller und ihrer Medizinprodukte, deren Überwachung ist [180]. Dafür beauftragen die zuständigen Landesbehörden Organisationen aus der Privatwirtschaft wie beispielsweise den Technischer Überwachungsverein (TÜV) [180]. Diese werden in Deutschland wiederum von der Zentralstelle der Länder für Gesundheitsschutz bei Arzneimitteln und Medizinprodukten (ZLG) benannt und überwacht [180].

Zu guter Letzt wird das Robert Koch-Institut (RKI) als deutsche Bundesbehörde für die Krankheits-

überwachung und -prävention aufgeführt [351]. In dieser Rolle steht vor allem die wissenschaftliche Untersuchung von Krankheiten im Vordergrund, die aufgrund ihrer gesundheitspolitischen Bedeutung in besonderem öffentlichen Interesse stehen.

## 2.2.2 Standardisierung von Gesundheits- und Behandlungszuständen

Um eine einheitliche und wohldefinierte Repräsentation von Informationen zu erhalten werden Terminologien und Kodierungssysteme eingesetzt, die beispielsweise Informationen über Diagnosen, Behandlungen und Prozeduren definieren [164]. Ein in der wissenschaftlichen Literatur und klinischen Praxis anzutreffender Klassifikationsstandard für Gesundheitszustände ist die International Statistical Classification of Diseases and Related Health Problems (ICD). Die ICD wird von der World Health Organization (WHO) herausgegeben und befindet sich zum Zeitpunkt der Anfertigung dieser Arbeit in ihrer elften Fassung [430]. Darin werden Gesundheitszustände definiert, klassifiziert und hierarchisch strukturiert. Sie dient somit als diagnostischer Klassifikationsstandard und Grundlage für die Ermittlung von Gesundheitstrends und -statistiken weltweit. Die damit einhergehende Codierung findet Anwendung in der medizinischen Datenerfassung und hilft bei der sprachenagnostischen Interpretation und Integration von Gesundheitsdaten. Die Abbildung von klassifizierten Erkrankungsbildern auf digitale Datenspeicherungs- und -übertragungsstandards wie Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) oder Fast Healthcare Interoperability Resources (FHIR) ist ebenfalls möglich (vgl. Unterunterabschnitt 2.4.2).

Für die Identifikation und Erfassung krankheitsassoziierter Parameter existieren Leitlinien, die medizinischem Personal evidenz- und konsensusbasierte Aussagen zu Präventions-, Diagnostik- und Behandlungsmöglichkeiten liefern. In Deutschland gibt beispielsweise die Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde (DGPPN) mit den S3-Leitlinien einen konsentierten Standard zum Umgang mit Demenz, insbesondere für die Diagnose und Behandlung heraus [17]. Auf internationaler Ebene orientiert man sich an der Diagnostic and Statistical Manual of Mental Disorders (DSM) [357].

## 2.2.3 Demenz

Das RKI zählt chronische Erkrankungen zu den bedeutendsten gesundheitsökonomischen Herausforderungen [349]. Als chronisch gelten Erkrankungen demnach sobald sie nicht vollständig heilbar sind und diesbezüglich regelmäßig Leistungen des Gesundheitssystems in Anspruch genommen werden müssen. Die Demenz lässt sich durch den unheilbaren und langwierigen Verlauf mit zunehmender Verschlechterung des Krankheitsbildes ebenfalls den chronischen Erkrankungen zurechnen. Es gibt verschiedene Ursachen für Demenz, die als solche nicht als Krankheit sondern als Syndrom gilt, welchem eine fortschreitende oder chronische neurodegenerative Erkrankung zugrunde liegt [430]. In der Regel gehen mit diesem Syndrom multifaktorielle kognitive Abbauprozesse einher, die vom zentralen Nervensystem ausgehen und höhere kortikale Funktionen sowie unter anderem Gedächtnis, Orientierung, Lernfähigkeit oder Sprache beeinträchtigen [423, 430]. Auf diese Eigenheiten des Syndroms wird nachfolgend genauer eingegangen, um als Ausgangspunkt für die in dieser Arbeit erschlossenen Erkenntnisse betrachtet werden zu können.

## Krankheitsbild und Ätiologie

Grundlegend kann zwischen primären und sekundären Demenzformen unterschieden werden [26, 423]. Lediglich 10% der Fälle lassen sich den sekundären Demenzen zuordnen, die in Folge von anderen Grunderkrankungen wie beispielsweise Depressionen und Schilddrüsenerkrankungen oder durch Wechselwirkungen von Medikamenten entstehen [26]. Bei frühzeitiger Intervention gelten diese als reversibel [26, 423]. Rund 90% aller Fälle können den primären Demenzen zugeordnet werden, welchen irreversible Erkrankungen des Gehirns zugrunde liegen [26, 423]. Hierbei sterben, aus in der medizinischen Forschung noch nicht vollumfänglich geklärten Gründen, Neuronen im Gehirn ab, wodurch Verbindungen zwischen den Zellen dauerhaft verloren gehen und betroffene Hirnregionen unterschiedliche Symptome auslösen (vgl. Abbildung 2.6). Zum Zeitpunkt der Anfertigung dieser Arbeit bestehen keine Therapieansätze mit Heilungschancen. Eine möglichst frühe Diagnose wird aber als positiv gewertet, da eine Verzögerung und Milderung des Verlaufs durch entsprechende Interventionen bewirkt werden kann [153, 262, 300]. Der medizinischen Fachliteratur folgend, gilt mit 55-80% aller diagnostizierten primären Fälle die Alzheimer Demenz (AD) als häufigste Form des Demenzsyndroms [33, 423, 431]. Die AD (ICD-11 Code 8A20) ist seit ihrer erstmaligen Beschreibung durch Alois Alzheimer im Jahr 1906 Teil des wissenschaftlichen Diskurses [30, 430]. Hierbei stellt er bereits einen Bezug zu einer mikroskopisch feststellbaren Veränderung der Hirnrinde als anatomische Abweichung zu den damals bereits bekannten Fällen von Demenz als charakteristisches Kennzeichen heraus [30]. Die damit einhergehende „fortschreitende Verblödung“ führt er auf eine zu beobachtende Verklumpung der Fibrillen zurück, was zur Zerstörung der Nervenzellen im Gehirn führt (vgl. Abbildung 2.6) [30]. Aus ätiologischer Sicht können bereits Jahre vor dem Auftreten erster Symptome, durch sogenannte Biomarker, Veränderungen im Gehirn der Betroffenen erkennbar sein [143]. Hierzu zählt vor allem die chronische Degeneration von sogenannten kortikalen Neuronen (Neuronen der Hirnrinde). Außerdem ist der Hippocampus betroffen, was vor allen Dingen mit Lernfähigkeits-, Wahrnehmungs- und Gedächtnisbeeinträchtigungen in Bezug gesetzt wird sowie mit der Schwächung von Exekutivfunktionen [91, 128]. Durch das Fortschreiten der Krankheit werden vermehrt Protein-Ablagerungen, sogenannte Amyloid-Plaques und neurofibrilläre Bündel festgestellt. Eine verbreitete Arbeitshypothese besagt, dass diese zu einer Unterbrechung der neuronalen Signalübertragungen und schließlich zum Absterben der Neuronen in den betroffenen Arealen des Gehirns führen. Dieser kausale Zusammenhang, der auch zur Rechtfertigung der bislang erfolglosen Entwicklung entsprechender Medikamente genutzt wurde, steht zum Zeitpunkt der Anfertigung dieser Arbeit noch im Diskurs innerhalb der medizinischen Forschung [294].

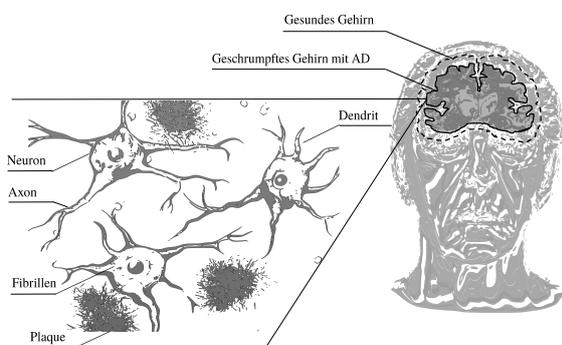


Abbildung 2.6: Anatomische Auswirkungen einer AD nach [31].

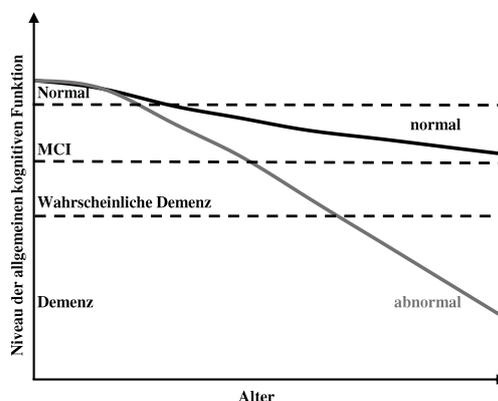


Abbildung 2.7: Vergleich kognitiver Abbauprozesse [172].

Wie zuvor skizziert, wird Demenz im Allgemeinen als klinisches Syndrom verstanden mit dem ein erworbener kontinuierlicher Rückgang des kognitiven Leistungsniveaus einhergeht, welchem verschiedene Ursachen und Erkrankungsprozesse zugrunde liegen können (vgl. Abbildung 2.7) [106, 143, 273, 357, 430]. Diese Abhängigkeit hat Einfluss auf den prognostizierten Verlauf sowie daran geknüpfte Behandlungsentscheidungen [143]. Hierbei wird, durch einen kognitiven Abbauprozess, die Alltagskompetenz negativ beeinflusst [26]. Als kognitiv werden in der Psychologie die mit dem Denken einhergehenden Prozesse bezeichnet, als neurokognitive Störungen, die Gruppe von Denkprozessstörungen, die in den meisten Fällen als quantitativ messbar angesehen werden und durch neuronale Substrate gewisse Symptome verursachen [357]. Einen Ansatz um diese Domänen in ihre kognitiven Funktionen zu untergliedern liefert das DSM, indem eine einheitliche Nomenklatur für neuropsychiatrische Störungen aufgestellt wird. In der fünften Fassung dieses Rahmenwerks werden für die Diagnose neurokognitiver Störungen relevante kognitive Domänen wie in Abbildung 2.8 dargestellt definiert und zur Abgrenzung verschiedener ätiologischer Unterarten von Störungen klassifiziert [273, 357]. Hierbei wird eine Unterteilung in sechs Domänen mit jeweiligen Subdomänen beziehungsweise Funktionen vorgenommen [273, 357]. Um den

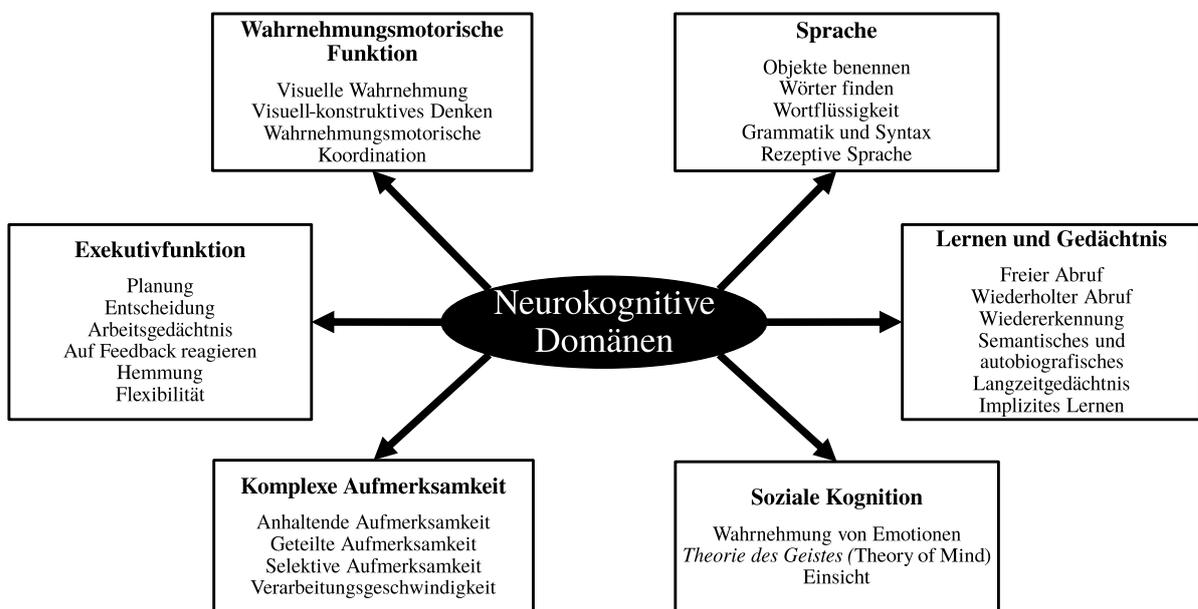


Abbildung 2.8: Neurokognitive Domänen nach der DSM-V [357].

Vorgaben der DSM-V gerecht zu werden, wird für die Diagnose einer AD das Vorliegen von Einschränkungen im Bereich Lernen und Gedächtnis und mindestens einer weiteren kognitiven Domäne, sodass die Alltagsfähigkeit eingeschränkt ist, vorausgesetzt [273, 357]. Diese lassen sich erneut in verschiedene Subdomänen unterteilen. Da an dieser Stelle nicht jede weiter ausgeführt werden kann, stellt ein Beispiel das Arbeitsgedächtnis als eine Subdomäne der Exekutivfunktion dar. Als solches stellt es ein kognitives System im System dar, welches nach dem Modell von Baddeley aus mehreren Komponenten besteht, die räumlich-visuelle für nicht-verbale und phonologische für verbale Informationen speichern [42]. Diese zwei Komponenten werden von einer zentralen Komponente gesteuert, indem Aufmerksamkeit zwischen ihnen gelenkt wird, um auf das Langzeitgedächtnis zuzugreifen [42]. Daran sind kognitive Prozesse gekoppelt, die es Menschen ermöglichen deren Umgebung zu erfassen sowie neues Wissen zu akquirieren, Ziele zu formulieren und Probleme zu lösen [42]. Sind diese gestört leidet das Verständnis für Sprache sowie Rechnen und weitere wahrnehmungsmotorische Fähigkeiten [42]. Studien von Belleville et al. legen nahe, dass durch eine AD das Arbeitsgedächtnis sehr früh betroffen ist [50].

Eine Einteilung der Schweregrade dieses Symptomkomplexes erfolgt in Gruppen. Aus Sicht der medizinischen Forschung findet bei AD eine Unterteilung in drei ineinander übergehende Krankheitsphasen statt [143]:

- *PAD*: Präklinische AD
- *AD-LKB*: AD-assoziierte leichte kognitive Beeinträchtigung
- *KAAD*: Klinisch apparentes dementielles Syndrom bei vorliegender AD

Die präklinische Form der AD wurde für Forschungszwecke definiert und bezieht sich auf Veränderungen von Biomarkern noch bevor sich erste kognitive Beeinträchtigungen beobachten lassen [143]. In der Regel unterscheidet man also lediglich zwischen der majoren (KAAD) und minoren (AD-LKB) Form der Demenz [273]. Die majore Form charakterisiert sich durch ausgeprägte kognitive Leistungsminderungen, welche in Unselbstständigkeit und Hilfebedürftigkeit der Betroffenen münden [273]. Mit der minoren Form gehen unterschwellige Leistungseinbußen einher, welche Funktionsstörungen hervorrufen, die keine Alltagsfunktionen beeinträchtigen [273]. Die letztere Form steht dem Konzept der Mild Cognitive Impairments (MCI) nahe, welches häufig als Vorstadium einer AD interpretiert wird und sich durch leichte Gedächtnisdefizite und Fehler im abstrakten Denken kennzeichnet [143, 273]. MCI entstehen somit zwischen einer als normal zu bezeichnenden altersbedingten Abnahme kognitiver Fähigkeiten und den Einschränkungen der kognitiven Fähigkeiten als Folge von einem dementiellen Syndrom (vgl. Abbildung 2.7). Die Bewältigung grundlegender Alltagsaktivitäten ist dabei noch uneingeschränkt möglich [424]. MCI lässt sich wiederum in zwei Typen einteilen. Die amnestische MCI ist im Vergleich zum altersbedingten kognitiven Abbau mit einem stärkeren Gedächtnisverlust verbunden [172, 424]. Zwischen 10-15% der Patienten mit amnestischer MCI erkranken innerhalb eines Jahres an AD [172]. Die nicht-amnestischen MCI ist mit einer Verschlechterung anderer Fähigkeiten neben der Gedächtnisleistung, beispielsweise der Rechenfähigkeit, verbunden und ein möglicher Hinweis auf eine vaskuläre Demenz [172]. Die AD selbst wird der majoren Form zugerechnet. In der ICD-11 wird dieses Syndrom weiter differenziert, indem assoziierte mentale und verhaltensbezogene Symptome gelistet und codiert werden [430]. Hierzu zählen unter anderem Verhaltensauffälligkeiten durch psychotische oder psychische Störungen, Ängste, Stimmungsschwankungen, Apathie, Aggression, Enthemmung und Unruhezustände. Des Weiteren findet eine Schweregradeinteilung der AD statt, welche sich in drei Stufen gliedert. Man spricht nach der Diagnose von einem milden (ICD-11 Code XS5W), moderaten (ICD-11 Code XS0T) und schwerwiegenden (ICD-11 Code XS25) Stadium einer AD. Eine Definition wie diese Einteilung in unterschiedliche Stufen vorzunehmen ist, wird in der ICD nicht gegeben. Die Diagnose erfolgt durch eine Differenzialdiagnostik, da die Demenzform und daran geknüpfte unterschiedlich stark ausgeprägte Symptome und Ursachen nach dem Ausschlussverfahren zugeordnet werden müssen (vgl. Unterabschnitt 2.2.4) [143]. Ein Werkzeug für die Einteilung in diese Stufen bietet beispielsweise das Clinical Dementia Rating (CDR) von Hughes et al. [194]. Dieses adressiert das mit der Diagnose einhergehende Beurteilungsproblem, indem sowohl Verhaltens- als auch psychometrische Daten bezüglich der kognitiven Domänen in die Beurteilung mit einfließen. Aufgrund der natürlichen Ursache einer AD bildet dieser nicht-invasive Ansatz jedoch keine objektiv validierbare Möglichkeit zur Unterscheidung der Stufen, stellt aber einen praxisrelevanten Ansatz dar [194].

## Epidemiologische Fakten und ökonomische Implikationen

Schätzungen zufolge litten im Jahr 2020 weltweit circa 54 Millionen Menschen an AD, in Deutschland wird für das Jahr 2018 von 1,7 Millionen betroffenen Menschen ausgegangen [109, 332]. Die in Abbildung 2.10 abgebildete global gemittelte Inzidenzrate beschreibt den Anstieg an Neuerkrankungen innerhalb eines Jahres pro Altersgruppe [109, 332]. Ebenfalls zeigt Abbildung 2.10 die geschätzte Prävalenz in Europa über die Altersgruppen hinweg [60]. Beide epidemiologischen Kennzahlen zeigen deutlich, dass die Wahrscheinlichkeit an AD zu erkranken mit zunehmenden Alter steigt, was bei einer Überalterung der Gesellschaft zu einer höheren absoluten Häufigkeit dieser Krankheit führt. Weltweit wird daher mit einer Zunahme von Menschen mit Demenz bis zum Jahr 2050 auf 131 Millionen gerechnet (vgl. Abbildung 2.9) [332]. Vor allem in den gemäß der von der Weltbank vorgenommenen Bruttonationaleinkommenkategorisierung klassifizierten Ländern mit geringen und mittleren Einkommen werden im Vergleich zu Ländern mit hohem Einkommen die Zahl der an AD erkrankten Menschen stark steigen [332, 429]. Als Gründe für diesen Trend werden demografische Effekte und das zunehmende Bevölkerungswachstum in diesen Ländern genannt [332]. Nicht nur die Betroffenen selbst, auch deren Angehörige leiden unter den Auswirkungen dieses Psychosyndroms [331]. Insgesamt werden die direkten und indirekten Kosten weltweit auf eine Billion Dollar für das Jahr 2018 geschätzt [332]. Allein in Deutschland sind, bei 300.000 neuen Demenzfällen pro Jahr und insgesamt 1,7 Millionen Erkrankten, Kosten in Höhe von 15,1 Milliarden € im Jahr 2015 entstanden [109, 392].

Studien zeigen, dass ungefähr ein Drittel aller Demenzen bei rechtzeitiger Diagnose und Behandlung entsprechender Risikofaktoren wie beispielsweise Hypertonie, Schlaf- oder Bewegungsmangel verhindert werden können [153, 300]. Insbesondere deshalb müssen kostengünstige Wege und Mittel gefunden werden, um diese Erkrankung möglichst frühzeitig zu erkennen und das Fortschreiten zu verzögern oder gar zu stoppen.

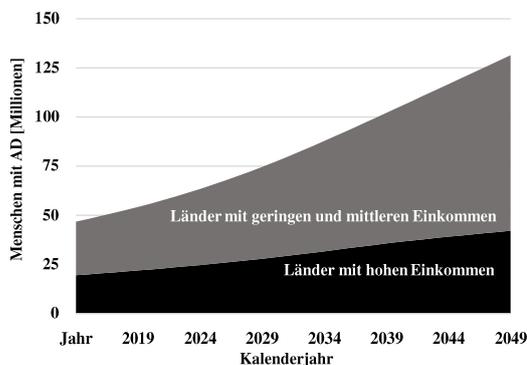


Abbildung 2.9: Anzahl von Menschen mit AD im Ländergruppenvergleich [332].

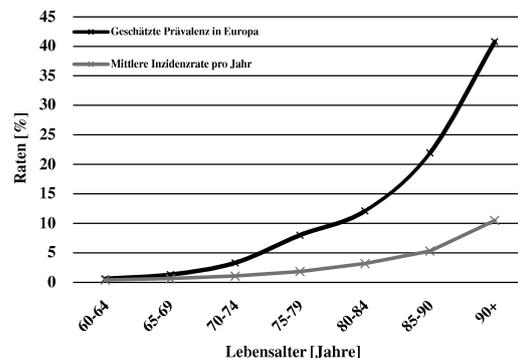


Abbildung 2.10: Geschätzte Prävalenz in Europa und Inzidenzrate [60, 109, 332].

### 2.2.4 Diagnose und Behandlung von Alzheimer Demenz

Da es sich bei der Diagnose dementieller Syndrome um eine sogenannte Ausschlussdiagnose handelt, müssen andere möglicherweise mit einer ähnlichen Symptomatik einhergehenden Erkrankungen ausgeschlossen werden. Hierfür existiert eine Reihe von Methoden und Werkzeugen, die nachfolgend zusammengefasst werden. Wie zuvor bereits angedeutet, gibt es für AD keine Therapie, die eine Heilung der erkrankten Personen herbeiführen kann [150]. Jedoch sind in gewissen Fällen präventive und

therapeutische Interventionen möglich, um eine Erkrankung zu verhindern, zu lindern oder zu verzögern [150, 153, 262, 300]. Auch hierüber wird in diesem Abschnitt ein Überblick gegeben.

### Diagnose

Wie eine medizinische Diagnose erfolgt, ist durch entsprechende Empfehlungen und Leitlinien festgelegt. So ist für dementielle Erkrankungen ein Prozess vorgesehen, der sich weltweit am Klassifikationssystem ICD sowie am Leitfaden der DSM orientiert (vgl. Unterabschnitt 2.2.2). Darauf aufbauend geben nationale Fachgesellschaften Leitlinien heraus wie zum Beispiel die S3-Leitlinien Demenzen der DGPPN [17]. Darin dokumentierte anerkannte Kriterien werden nachfolgend als notwendige Schritte und Verfahren des Diagnoseprozesses abgebildet [17, 32]. Da es sich bei einer Demenzdiagnose um eine Syndromdiagnose handelt dient der globale klinische Eindruck als Entscheidungsgrundlage [17, 143]. Hierfür werden ausführliche Untersuchungen notwendig, um den Gesundheitszustand des Betroffenen aus physiologischer, neurologischer und psychiatrischer Sicht zu differenzieren sowie andere ursächliche Erkrankungen ausschließen zu können. Dabei wird einer frühestmöglichen Erkennung aus ärztlicher Sicht zur möglichen Intervention und weiteren Lebensplanung ein Mehrwert beigemessen [155]. Grundsätzlich stehen dafür sowohl systembiologische als auch neuropsychologische Verfahren zur Verfügung. Entgegen den Erwartungen wurden in die 2013 herausgegebene DSM-V keine systembiologischen Ansätze aufgenommen. Stattdessen werden auf der Symptomatik aufbauende neuropsychologische Diagnose-Ansätze weiter präferiert. Dennoch fordern Experten, dass systembiologische Verfahren mit in den Diagnose-Prozess integriert werden sollten [258]. Bereits 2010 schlugen Eschweiler et al. einen zweistufigen Diagnose-Prozess vor [136]. Nach diesem ist zunächst das dementielle Syndrom nach den ICD-Kriterien festzustellen [431]. Diese erfordern über einen Zeitraum von mindestens sechs Monaten das Vorliegen einer Beeinträchtigung des Gedächtnisses und mindestens einer weiteren kognitiven Domäne, den Ausschluss eines Delirs (Zustand geistiger Verwirrung) sowie eine aus dem Syndrom resultierende Beeinträchtigung des alltäglichen Lebens. Im zweiten Schritt wird die Ätiologie durch weitere Untersuchungen eingegrenzt. Dazu sind medizinische, historische und klinische Daten notwendig. Diese stammen aus unterschiedlichen Quellen wie einer (Fremd-)Anamnese, den Ergebnissen einer allgemeinmedizinischen, neurologischen und psychiatrischen Untersuchung, der beobachteten Leistungsfähigkeit in neuropsychologischen Screening-Tests (Schnelltests), den Ergebnissen aus bildgebenden Verfahren zur Untersuchung des Gehirns sowie weiteren Labortests zum Nachweis relevanter Biomarker [136]. Auch Fiedler et al. beschreiben einen Diagnose-Prozess, in dem der klinische Eindruck als erste Indikation einer Demenz dient [143]. Ist dieser auffällig, sind weitere Tests zur Einschätzung der kognitiven Fähigkeiten durchzuführen. Grundsätzlich setzt eine Diagnose das Vorliegen der Leitsymptome voraus, die sich bei mehr als 80% aller Fälle auf amnestische Störungen früh in Form von Einschränkungen in der Merkfähigkeit und der Gedächtnisleistung bemerkbar machen [423]. Diese müssen über einen Zeitraum von mehr als sechs Monate vorliegen, da ohne Unterstützung diagnostischer Biomarker nicht ausgeschlossen werden kann, dass temporär vorliegende kognitive Defizite beispielsweise auf häufig auftretende Depressionen im Alter zurückzuführen und dadurch reversibel sind [423]. Die unterschiedlichen Diagnoseschritte können daher je nach Krankheitsbild, Demenzform und Fortschritt differieren. Bei AD kann die PAD Phase nach Fiedler et al. durch Auffälligkeiten in sogenannten Biomarkern charakterisiert werden [143, 233]. Es handelt sich hierbei um ein präsymptomatisches Stadium, welches nicht durch neuropsychologische Untersuchungen identifiziert werden kann [233]. Jedoch können genetische Mutationen aber auch Ablagerungen von Amyloid-Plaques im Gehirn durch bildgebende Verfahren wie Positronen-Emissions-Tomographie (PET) Aufschluss über AD-assoziierte Risikofaktoren geben [143, 233]. Mit Hilfe dieser, aus ökonomischer Sicht aufwändigeren

Verfahren, kann eine Demenz in bestimmten Fällen bereits Jahre vor dem Ausbruch identifiziert werden [143, 233].

In der Phase AD-LKB oder MCI wird eine leichte kognitive Beeinträchtigung vorausgesetzt. Diese kann mittels neuropsychologischer Verfahren in Form von standardisierten psychometrischen Tests aufgedeckt und quantifiziert werden (vgl. Unterabschnitt 2.2.5). Eine differenzierte Diagnose der Demenzart sowie eine Verlaufsprognose für die Erkrankung kann mit dieser Art von existierenden Verfahren bislang nicht ausreichend gesichert werden [233]. Weitere diagnostische Verfahren die durch Biomarker zur Früherkennung in dieser Phase beitragen können, beziehen sich auf die Erkennung von charakteristischen Stoffwechselfdefiziten sowie Volumenverlusten (80% - 90% Spezifität und Sensitivität) in früh betroffenen Hirnarealen [233]. Auch Analysen der Hirnaktivität mittels bildgebender Verfahren bietet die Möglichkeit AD-Patienten von Patienten mit anderen Arten von Demenz zu unterscheiden [233]. Letztlich lässt sich der Übergang von Phase AD-LKB zur Phase KAAD über die Konzentration von mit AD verbundenen Proteinablagerungen (Sensitivität von 95% und einer Spezifität von ca. 85%) vorhersagen [233]. In der Phase KAAD wird von einer manifestierten AD ausgegangen. Insgesamt wird durch die Kombination verschiedener Biomarker und neuropsychologischer Verfahren eine erhebliche Verbesserung bei der Erkennung und Prognose von AD erwartet [233]. Nachfolgend werden die einzelnen diagnostischen Schritte gemäß der S3-Leitlinie nochmal zusammengefasst [17].

**Fremd- und Selbstanamnese:** Die Anamnese bildet den Einstieg in die Diagnostik der AD. Bei Fragen an den Patient selbst spricht man von Eigen- oder Selbstanamnese, bei der Befragung von dem Patient nahestehenden Personen von einer Fremdanamnese. Hierbei werden durch einen Selbst- und Fremdbildabgleich relevante Informationen bezüglich der Diagnosestellung extrahiert. Ein Hinweis auf eine Demenz liefern Aussagen über die mit dem potentiellen Syndrom verbundene kognitive Einschränkungen der Betroffenen und deren Auswirkungen auf ihre Alltagsbewältigungsfähigkeiten [17, 143]. Des Weiteren werden darüber zum Ausschluss anderer Ursachen Fragen zum zeitlichen Verlauf der Krankheitsgeschichte bezüglich der Symptomatik, den Rahmenbedingungen wie familiären Erkrankungen, Medikamenteneinnahmen oder möglichem Suchtverhalten gestellt. Um die Betrachtung der Symptomatik zu objektivieren, wird ebenfalls eine Sozial- und Familienanamnese durch Befragung des persönlichen Umfelds der Betroffenen durchgeführt. Standardisierte Fragebögen und Interviewleitfäden sind hierfür verfügbar [155].

**Physiologische Untersuchung mit Labordiagnostik:** Im Anschluss an die Anamnese folgt die Untersuchung verschiedener Körperfunktionen des Patienten [31, 143]. Innerhalb dieser Untersuchungen werden sowohl Vitalparameter wie Blutdruck oder der Puls zur Abgrenzung von Herzkrankheiten, als auch die Lungen-, Seh- und Hörfunktionen überprüft, um den körperlichen Gesundheitszustand des Patienten einschätzen zu können. Mittels labordiagnostischer Ansätze, können durch die Analyse von Blut- und Urinproben weitere Parameter gesammelt und die Ursachen für das Vorliegen der Symptome eingegrenzt werden [143]. Besteht daraufhin der Verdacht auf eine vorliegende Hirnleistungsstörung, werden neurologische Untersuchungen nötig.

**Neurologische Befunderhebung durch psychometrische Testung:** Mittels psychometrischer Tests werden bei Verdacht auf Hirnfunktionsstörungen kognitive Domänen (vgl. Abbildung 2.8) getestet [31, 143, 200]. Psychometrische Tests sind standardisierte Testverfahren, um Einbußen in der Gedächtnisleistung festzustellen und zu objektivieren (vgl. Unterabschnitt 2.2.5). Zusätzlich tragen sie zu einer vollständigen Dokumentation der Erkrankung bei [200]. Die Auswahl des Testverfahrens hängt von dem Krankheitsstadium der zu untersuchenden Person ab [200]. Werden mehrere Tests kombiniert spricht man von sogenannten Testbatterien.

**Liquordiagnostik und Gentests:** Durch die Entnahme und Untersuchung von Liquor (Nervenwasser) können mit einer AD assoziierte erhöhte Konzentrationen von gewissen Proteinen bereits im Frühstadium erkannt und Therapieoptionen erwogen werden [143]. Über die Kombination von verschiedenen Messverfahren können Tau-Proteine und Amyloid- $\beta_{1-42}$  nachgewiesen werden, wodurch eine Sensitivität und Spezifität von 80-90% zur differenzialdiagnostischen Abklärung einer AD erreicht werden [62].

Bei Patienten für die bekannt ist, dass ein Familienmitglied in der Vergangenheit mit AD diagnostiziert wurde, können Gentests Aufschluss über eine mögliche genetische Veranlagung geben [31]. Zwischen 1-5% lassen sich Schätzungen zufolge der dominant vererbaren Form von AD zuordnen [31, 353].

**Bildgebende Verfahren:** Mit bildgebenden Verfahren werden physiologische Veränderungen des Gehirns untersucht, die bei der Diagnose einer AD zur Abgrenzung anderer Demenzerkrankungen eine Rolle spielen [31, 143]. Hierfür stehen verschiedene Verfahren zur Verfügung, um eine strukturelle Bildgebung wie eine Magnetresonanztomographie (MRT) oder Computertomographie (CT) durchzuführen [143]. Als Teil der klinischen Differenzialdiagnostik werden hiermit andere Demenzursachen wie beispielsweise vaskuläre Demenzformen oder Tumorerkrankungen ausgeschlossen, indem zerebrale Änderungen des Gehirns untersucht werden. Neben der Abnahme des Hirnvolumens insgesamt lassen sich hiermit auch Atrophien in früh betroffenen Regionen wie dem Hippokampus beobachten, was eine frühe Diagnose und Informationen zur Bestimmung des Übergangs in die unterschiedlichen Phasen einer AD ermöglicht [143].

Daneben steht mit der PET die Möglichkeit metabolische Veränderungen einer AD nachzuweisen wie zum Beispiel die Amyloid-PET, welche Amyloid-Plaques am lebenden Organismus darstellen kann. Diese sind oftmals früher nachzuweisen als strukturelle Änderungen und werden mit 80-90% Sensitivität und Spezifität angegeben [143]. Bildgebende Verfahren ermöglichen somit den Nachweis pathologischer Änderungen in frühen Stadien einer AD und eignen sich für die Abgrenzung anderer zugrundeliegender Erkrankungen.

Zusammenfassend stehen sowohl systembiologische als auch neuropsychologische Verfahren zur Verfügung. Der wesentliche Vorteil der systembiologischen Verfahren besteht darin, dass sie mehr Informationen bezüglich einer Differenzialdiagnose liefern [143]. Um eine Diagnose zu stellen, werden damit verschiedene Biomarker analysiert, die mit Hilfe von bildgebenden Verfahren oder neurochemischer Liquordiagnostik erfasst werden können [423]. Demgegenüber untersuchen neuropsychologische Verfahren die kognitiven Leitsymptome der Demenz. Dabei werden systematisch die unterschiedlichen kognitiven Domänen getestet (vgl. Abbildung 2.8). Dazu existieren sowohl zeitökonomisch durchführbare Screening-Tests als auch umfangreiche Testbatterien, die in der klinischen Praxis zum Einsatz kommen [423]. Hierbei ist anzumerken, dass neuropsychologische Tests die klinische Symptomatik erfassen, was eine Früherkennung erst ab dem Zeitpunkt des Auftretens von Symptomen ermöglicht. Ab dem Stadium der MCI ist deren Einsatz also möglich. Der Vorteil gegenüber systembiologischen Ansätzen ist, dass sie Aufschluss über die durch die Symptome auftretende tatsächlichen Beeinträchtigungen des alltäglichen Lebens einer betroffenen Person geben können [423]. Damit besteht zumindest zu einem gewissen Grad auch für neuropsychologische Tests die Möglichkeit einer Differenzialdiagnose.

## Prävention und Behandlung

Eine individuelle Prognose für Menschen mit Demenz hängt von der pathologischen Ursache ab, jedoch werden zwischen 9-33% aller Demenzen als potenziell reversibel oder präventiv vermeidbar eingestuft

[95, 262, 293]. Da das Fortschreiten der Krankheit eine umfangreiche Pflegebedürftigkeit erwarten lässt und die Lebenserwartung reduziert, sind möglichst frühe Interventionen sinnvoll [423]. Zahlreiche Studien, die in den letzten Jahren unterschiedliche Ansätze verfolgten, um pathophysiologische Eigenschaften der Krankheit zu adressieren, konnten keine Verbesserung erzielen [412]. Aus pharmakologischer Sicht existieren lediglich Antidementiva, die mit dem Ziel einer Stabilisierung der kognitiven Leistungsfähigkeit eingesetzt werden [150]. Um Begleiterscheinungen und Symptome wie Depressionen, Unruhezustände und Aggressivität zu lindern, werden zusätzlich Neuroleptika eingesetzt [150]. Neben der medikamentösen Behandlung werden in der wissenschaftlichen Literatur in den letzten Jahren vermehrt psychosoziale Maßnahmen zur Linderung, Verhinderung oder Verzögerung des Ausbruchs dieser Erkrankung diskutiert [153, 262, 300, 423]. Die Einschätzung und gezielte Adressierung von Risikofaktoren werden hierfür als Ansätze herangezogen. Neben dem Alter als bedeutendsten Risikofaktor werden in der Literatur verschiedenen Faktoren diskutiert [153, 262, 300, 423]. Livingston et al. identifizieren insgesamt neun Risikofaktoren, welche durch Präventionsmaßnahmen adressiert werden sollten: niedrige Bildung, Schwerhörigkeit im mittleren Lebensalter, Fettleibigkeit, Bluthochdruck, Depression im späten Lebensalter, Rauchen, körperliche Inaktivität, Diabetes und soziale Isolation [262]. Ansätze für validierte Risikoscoringverfahren und deren digitale Umsetzung liefern Kivipelto et al. und Sindi et al. [232, 383]. Je früher eine beginnende Demenz erkannt wird, desto effektiver kann der Krankheitsverlauf positiv beeinflusst werden, insbesondere bei Patienten, die noch keine kognitiven Beeinträchtigungen zu verzeichnen haben [52]. Bestenfalls sollte der Ausbruch des dementiellen Syndroms durch präventive Maßnahmen vollständig verhindert oder bis zum Lebensende hinausgezögert werden [412]. Der Nachweis für die Wirksamkeit passender Behandlungs- und Präventionsstrategien gilt als schwierig zu führen, da umfangreiche Zeiträume und Probandengruppen betrachtet werden müssen, um zu klären, welche Kombinationen von Maßnahmen in welchem Ausmaß und zu welchem Zeitpunkt bei verschiedenen Risikoprofilen wirksam sind [150]. Studien, die dieses Ziel in den letzten Jahren dennoch verfolgen, liefern Isaacson et al., Bredesen et al., Ngandu et al. und Toups et al. [72, 73, 74, 203, 300, 406].

### 2.2.5 Psychometrie

Wie aus den Abschnitten zuvor ersichtlich wird, zählen psychometrische Testverfahren zu den grundlegenden Werkzeugen der Demenzdiagnostik. Ihl und Grass-Kapanke bewerten deren Spezifität und Sensitivität zur Abgrenzung von einer Depression mit über 90% [200]. Verschiedene Verfahren können je nach Zweck und zur Verfügung stehenden Ressourcen gewählt werden, um beispielsweise ein Screening in der Primärversorgung wie einer Hausarztpraxis durchzuführen oder weiterführende Einschätzungen des Schweregrades sowie die Verlaufskontrolle nach Interventionen zu evaluieren [200]. Die Grundlagen der hierfür erforderlichen Testtheorie werden dem Fachgebiet der Psychometrie zugeordnet, welche die mathematischen Grundlagen zur Objektivierung psychologischer Befunde liefert [414]. Unter Beteiligung von Patienten wurde in den letzten Jahrzehnten die Genauigkeit und Interpretierbarkeit dieser Verfahren erhöht und ein besseres Verständnis für die verwendeten psychometrischen Items erarbeitet [414]. Hierfür kommen in der Medizin sogenannte Scoring-Modelle zum Einsatz, welche in diesem Fall die Ergebnisse psychometrischer Tests operationalisieren und Schwellwerte definieren, um diese Gesundheitszuständen zuordnen zu können und die Entscheidungsfindung zu unterstützen [94]. Zur Bestimmung dieser Schwellwerte und Beurteilung der Validität werden sowohl etablierte Methoden wie die der Regressions-, Korrelations- oder ROC-Kurvenanalyse als auch durch die zunehmende Verbreitung der Computerisierung umfangreiche probabilistische Modelle eingesetzt [414]. Die Vielzahl an ausführlichen neuropsychologischen Tests, die für eine differenzierte Diagnose eingesetzt werden sowie Kurztests für

das Screening, macht es notwendig, dass nachfolgend ein Überblick über etablierte, sowie in dieser Arbeit referenzierte Tests gegeben wird. Da die kontinuierliche Weiterentwicklung, Validierung und Normierung dieser Testverfahren und psychometrischen Variablen dem Forschungsfeld der empirischen Psychologie zuzuordnen ist, wird hier lediglich ein Ausschnitt der grundlegenden Testcharakteristiken und -abläufe dargestellt.

### **Mini-Mental State Examination**

Die quantitative Bewertung von kognitiver Leistungsfähigkeit durch klinische Tests nimmt Zeit für die Durchführung und Administration in Anspruch [147]. Diesem zeitlichen Aufwand steht die kurze Aufmerksamkeitsspanne und Kooperationsbereitschaft älterer Patienten gegenüber [147]. Folstein et al. erkannten diesen Umstand bereits im Jahre 1975 und entwickelten daraufhin einen Schnelltest für das Screening einer Demenz. Die Mini-Mental State Examination (MMSE) bietet den Vorteil, dass sie auf Basis eines einfach auszuwertenden Scoring-Systems eine Beurteilung des Schweregrads der vorliegenden Demenz ermöglicht [147].

**Ziel und Testablauf:** Mit der MMSE sollte eine möglichst zeiteffiziente Testung entstehen (5-10 Minuten), indem Studienteilnehmer anhand von elf Aufgaben Auskunft über ihre kognitive Leistungsfähigkeit geben [147]. Diese werden auf zwei Abschnitte aufgeteilt, der Reihe nach durchgeführt und direkt bewertet.

**Adressierte kognitive Domänen und relevante Maße:** Im ersten Abschnitt des MMSE werden Fragen zur Prüfung der Orientierungsfähigkeit, des Gedächtnisses und der Aufmerksamkeit gestellt, die verbal beantwortet werden müssen. Hierbei lassen sich bis zu 21 Punkte erreichen. Im zweiten Abschnitt werden die Testteilnehmer dazu aufgefordert sich an die im ersten Abschnitt gelernten Objekte zu erinnern sowie schriftliche und zeichnerische Lösungen von Aufgaben zu erfüllen, die mit maximal 9 Punkten bewertet werden. Insgesamt können bis zu 30 Punkte erreicht werden, eine zeitliche Komponente existiert für die Bewertung nicht. Zur Untersuchung der Validität des Tests unterscheiden Folstein et al. zwischen den erreichten Durchschnittswerten von Menschen mit Demenz (9,7), mit von Depressionen verursachte kognitive Einschränkungen (19,0) und depressiven Menschen mit affektiven Störungen (25,1). Als gesund klassifizierte Menschen erreichen im Durchschnitt 27,6 Punkte. Untersuchungen ergeben, dass das Alter sowie Lerneffekte durch Testwiederholungen (Reliabilität) keinen signifikanten Einfluss auf die Scores haben. Insgesamt wird sowohl auf die Störung der zeitlichen und räumlichen Orientierung, die während einer Demenz auftreten kann, als auch die zusätzlichen Domänen wie Gedächtnis, Aufmerksamkeit, Rechnen und Sprache geprüft [147, 395]. Eine im Jahr 2009 durchgeführte Meta-Analyse ergab für die Performanz der MMSE, je nach Einsatzort der zugrundeliegenden Studien, eine Sensitivität von 71,1-85,5% sowie eine Spezifität von 81,3-95,6% [291]. Insgesamt wird darin für die MMSE aufgrund der damit erreichten moderaten Genauigkeit der größte Nutzen in der Primärversorgung für den Ausschluss einer Demenzdiagnose zugesprochen [291].

### **Free and Cued Selective Reminding Test**

Die DGPPN empfiehlt den Free and Cued Selective Reminding Test (FCSRT) zur Überprüfung des episodischen Gedächtnisses als Teil des Langzeitgedächtnisses für eine vertiefte neuropsychologische Diagnostik bei Alzheimer Demenz [17, S. 34]. Der FCSRT basiert auf Arbeiten von Buschke et al. [92]

sowie Grober et al. [170] und testet in seiner Gesamtheit sowohl die kognitiven Domänen Lernen und Gedächtnis als auch Sprache (vgl. Abbildung 2.8). Klinische Studien zeigen, dass der Test eine Trennung von Menschen mit kognitiven Einschränkungen und gesunden Menschen erlaubt [170]. Ebenfalls ist er dazu geeignet eine leichte Demenz (vgl. Unterabschnitt 2.2.4) in der Primärversorgung zu erkennen [171]. Der Test wurde für neun Altersgruppen zwischen 56 und 98 Jahren normiert [204]. Das Urheberrecht liegt beim US-amerikanischen Albert Einstein College of Medicine in New York und kann nach deren Zustimmung für nicht-kommerzielle Zwecke verwendet werden.

**Ziel und Testablauf:** Als Ziel gilt in einem mehrstufigen Lern- und Erinnerungsprozess verschiedene grafisch dargestellte Objekte anhand von dazu gehörigen Überbegriffen zu identifizieren, korrekt zu benennen, einzuprägen und zu erinnern. Eine genaue Beschreibung des Ablaufs findet sich in Unterabschnitt 5.1.2.

**Adressierte kognitive Domänen und relevante Maße:** Dieser Test versucht bei den Teilnehmern durch die Wiedererkennung von bekannten Objekten und des freien und wiederholten Erinnerns der gelernten Objekte die Lernfähigkeit und Gedächtnisleistung zu testen. Durch deren Benennung wird gleichzeitig die Sprachfähigkeit abgeprüft (vgl. Abbildung 2.8). Für die Bewertung der Funktionsfähigkeit werden zwei Maße zur Erkennung von Demenz als sinnvoll erachtet [171]:

- freies Erinnern (Anzahl der richtig erinnerten Objekte in der Phase des freien Erinnerns, nach drei Durchläufen liegt dieser Wert zwischen 0-48)
- gesamtes Erinnern (die Summe der richtig erinnerten Objekte aus freiem Erinnern und Erinnern nach Hinweisreiz, nach drei Durchläufen liegt dieser Wert zwischen 0-48)

Grober et al. beschreiben, dass das freie Erinnern für die Erkennung bereits vorliegender und zur Vorhersage zukünftiger Demenzen das beste Maß liefert [171]. Das gesamte Erinnern trägt stärker zur Diagnose einer vorliegenden AD als zum Ausschluss dieser Erkrankung bei [171]. Die Schwellwerte zur Klassifikation der Ergebnisse anhand von erreichten Punktzahlen wurden von den Autoren so gewählt, dass eine Spezifität von 90% erreicht wird. Aufgrund des negativen Einflusses einer positiven Diagnose für Betroffene und deren Angehörige, soll dadurch die Zahl der fälschlicherweise positiven Testergebnisse verringert werden. Die Schwellwerte für das freie Erinnern wurden daher auf 24 von 48 gesetzt, um 78% Sensitivität bei 90% Spezifität zu erreichen, für das gesamte Erinnern wurde der Schwellwert auf 46 von 48 gesetzt, um bei 54% Sensitivität eine Spezifität von 90% zu erreichen [171]. Die Referenzdiagnose wurde nach Aussagen der Autoren durch Standardkriterien ohne Zugriff auf die Ergebnisse der Testbatterie getroffen. Für die Durchführung des FCSRT werden 12-15 Minuten angegeben [171].

## DemTect

Als Testbatterie zur Früherkennung von kognitiven Einschränkungen wurde mit dem DemTect durch Kalbe et al. ein weiteres psychometrisches Screening-Verfahren entwickelt [220]. Hierfür wurden fünf Aufgaben identifiziert, welche durch das Erlernen und Abfrage von Wörtern einer Wortliste, die Transkodierung von Zahlen, Überprüfung der Wortflüssigkeit und das Lernen von Ziffern sowie deren Abruf in umgekehrter Reihenfolge kognitive Leistungseinbußen feststellbar machen.

**Ziel und Testablauf:** Der DemTect wurde mit dem Ziel entwickelt, möglichst sensitiv bezüglich der Erkennung von MCI und Demenzen in einem frühen Stadium zu sein. Hierfür wurden fünf Untertests ausgewählt, die mit möglichst geringem Aufwand eine ausreichende Bandbreite kognitiver Fähigkeiten

abdeckt, um Demenz mit hoher Sensitivität zu diagnostizieren [220]. Die Beeinträchtigung des Gedächtnisses wird hierbei durch Erlernen sowie sofortigem und verzögertem Erinnern von Wörtern einer Wortliste getestet [220]. Der Trade-off liegt zwischen einer ausreichenden Länge der Liste, um genügend Sensitivität zu erreichen und einer genügenden Kürze, damit Bodeneffekte und nachlassende Motivation vermieden werden [220]. Studien haben gezeigt, dass zehn Items mit zwei Durchgängen hierfür besonders geeignet sind, um das verbale Gedächtnis zu testen [220]. Auch die Verständlichkeit der Anleitung spielt eine Rolle sowie die Auswahl der Items, die sich nach leicht vorstellbaren, mit vergleichbarer Silbenlänge und phonetisch unterschiedlichen Wörtern richtet [220]. Eine weitere Aufgabe umfasst das Umwandeln von Zahlen in unterschiedliche Codierungssysteme, wie beispielsweise von ausgeschriebenen Worten in arabische Ziffern, da hierbei bei Menschen mit frühen Formen einer Demenz bereits Störungen festgestellt wurden [220, 221]. Diese Transkodierung, testet direkt und indirekt die kognitiven Domänen der Sprach-, Zahlenverarbeitung sowie Exekutivfunktion [220]. Ebenfalls ist die sogenannte verbale Flüssigkeit ein frühzeitig betroffener Bereich bei dementiellen Erkrankungen [220, 227]. Dieser wird über die „Supermarktaufgabe“ getestet, indem als sogenannte semantische Kategorie Dinge genannt werden sollen, die in einem Supermarkt gekauft werden können [220, 227]. Hierbei werden für Menschen mit AD eine Sensitivität von 100% bei einer Spezifität von 90% erreicht [220, 227]. Diese hohen Werte hängen mit der Vielzahl hierfür erforderlicher kognitiver Domänen zusammen, wie Aufmerksamkeit, semantisches Gedächtnis, Arbeitsgedächtnis, Sprache und Flexibilität [227]. Als letzter der fünf Tests wird das Erlernen von Ziffern, die der Reihe nach genannt werden sowie deren Abruf in umgekehrter Reihenfolge durchgeführt. Hierdurch wird das Kurzzeit- und Arbeitsgedächtnis geprüft [220].

**Adressierte kognitive Domänen und relevante Maße:** Der DemTect dauert zwischen 8-10 Minuten [220]. Die maximale Gesamtpunktzahl von 18 Punkten wird als erreichbar in Unabhängigkeit von Bildung und Alter angegeben. Hierbei wird bei einer Punktzahl von 13-18 Punkten von altersgerechter Leistungsfähigkeit, 9-12 Punkten von MCI und 0-8 Punkten von einer Demenz ausgegangen [220]. Als Klassifikationsraten für den DemTect werden für die Gruppen MCI und AD Sensitivitäten von 80% beziehungsweise 100% angegeben, mit den entsprechenden Cut-Off-Scores werden bei 81% der Fälle mit MCI und 85% der Fälle mit AD richtig klassifiziert, insgesamt schließen Kalbe et al. auf eine Klassifikationsrate von 85,4%.

### Rowland Universal Dementia Assessment Scale

Der von J. Rowland entwickelte Rowland Universal Dementia Assessment Scale (RUDAS) ist der Versuch eine portable, leicht durchzuführende und interkulturell valide Testmethode zur Erkennung von Demenz zu finden [395]. Deren Anwendbarkeit in der Primärversorgung steht dabei im Fokus [395].

**Ziel und Testablauf:** Als Ziel gilt es, eine Methode zu entwickeln, welche kulturelle Einflüsse auf die Testergebnisse ausschließt, um international einsetzbar zu sein [395]. Insgesamt werden hierfür sechs Items identifiziert und evaluiert. Als Durchführungsdauer sind 10 Minuten angegeben [395]. Um potentielle Ängste zu minimieren werden die Tests in einer bestimmten Reihenfolge, wie nachfolgend skizziert, abgefragt [395].

Begonnen wird mit dem Erlernen von vier Wörtern, die Lebensmittel darstellen (Tee, Speiseöl, Eier und Seife), indem diese fünfmal vom Patienten genannt und dadurch auswendig gelernt werden müssen. Im Anschluss daran werden Körperteile erfragt, die durch den Patienten an seinem Körper selbst und am Körper der durchführenden Partei benannt und gezeigt werden müssen. Danach werden die Patienten gebeten eine motorische Aufgabe nachzumachen und für etwa zehn Sekunden auszuführen, bei der im

Wechsel eine Hand in vertikaler Richtung zur Faust geballt wird während die andere Hand mit der Handfläche auf dem Tisch liegt. Als vierte Aufgabe gilt es eine vorgezeichnete geometrische Darstellung eines Würfels auf ein Stück Papier abzuzeichnen. Darauf folgend muss der Patient beschreiben, wie er eine stark befahrene Straße ohne Ampel oder Fußgängerüberweg überqueren würde. Diese Beschreibung wird durch die durchführende Partei hinterfragt, falls zwei Angaben zur Bewertung der Lage bezüglich des Verkehrs und der Sicherheit durch den Patienten nicht erfolgt sind. Anschließend muss der Patient die in der ersten Aufgabe gelernten Lebensmittel abrufen. Zuletzt werden die Patienten gebeten innerhalb von einer Minute so viele Tiere zu benennen, wie es ihnen möglich ist.

**Adressierte kognitive Domänen und relevante Maße:** Basierend auf der DSM-IV werden entsprechend der oben genannten Tests durch den RUDAS die Domänen Lernen und Gedächtnis, Komplexe Aufmerksamkeit, Wahrnehmungsmotorische Funktion, Exekutivfunktion und Sprache getestet. Als relevante Maße werden für richtige Antworten Punkte verteilt beziehungsweise Punkte abgezogen, falls Hilfestellungen notwendig werden oder Aufgabenkriterien nicht vollumfänglich erfüllt sind. Für eine detaillierte Darstellung der Punkteverteilung und -gewichtung muss an dieser Stelle auf Storey et al. verwiesen werden [395]. Die maximal erreichbare Punktzahl liegt bei 30. Auf Grundlage einer ROC-Kurvenanalyse wurde als optimaler Cut-Off-Score 23 Punkte ermittelt. Bewertungen, die niedriger ausfallen werden als Demenz eingestuft und besitzen mit einer Sensitivität von 89% und einer Spezifität von 98% eine signifikante Aussagekraft. Neben der Validität gegenüber Einflussfaktoren wie Geschlecht, Sprache oder Bildungsniveau, wird auch die Reliabilität mittels standardisierter logistischer Regressionskoeffizienten für die Retest-Reliabilität sowie Interrater-Reliabilität geprüft und als sehr hoch eingestuft [395].

### Groton Maze Learning Test

Bereits zu Beginn des 20. Jahrhunderts wurde zur Untersuchung der Lernfähigkeit von Lebewesen ein Test eingesetzt, der das Erlernen und Abrufen eines Pfades in einem Labyrinth erfordert [46, 288, 306]. Zur Untersuchung von blinden Kindern wurde dieser Test, auch um taktiles Feedback zu erhalten, als eine Variante mit in einer Matrix angeordneten quadratischen Blöcken, die blockweise erkundet wird, entwickelt und evaluiert [46]. Diese Variante existiert auch als sogenannter Hidden Pathway Maze Learning Task (HPMLT) in Form einer angepassten digitalen Version und ist Bestandteil der CogState Testbatterie, welche mehrere angepasste neuropsychologische Tests enthält [326]. Weitere Studien haben gezeigt, dass sich diese Art von Test auch zur Überprüfung kognitiver Domänen bei Menschen mit Demenz im ambulanten Umfeld eignet [287, 386].

**Ziel und Testablauf:** Ziel ist es den Lernprozess zu beobachten und die damit einhergehende Lernfähigkeit einzuschätzen [46, 288, 306]. Hierfür wird eine Matrix aus Blöcken dargestellt und Probanden dazu aufgefordert den Pfad zu erkunden, der sich verdeckt innerhalb dieser Matrix befindet [326, 386]. Der Pfad verläuft von einem vordefinierten Start- zu einem Endpunkt. Ausgehend vom Start versucht der Teilnehmer einen Weg zu finden, indem er sich Block für Block vorarbeitet während er nach Versuch und Irrtum jeweils ein Signal für seine Entscheidung (richtiger oder falscher nächster Block) erhält [326, 386]. Der gesamte Weg ist nie vollständig sichtbar, da vorher berührte Blöcke wieder unsichtbar werden [326, 386]. Ein Teilnehmer darf weder diagonal gehen noch muss er bei einer falschen Auswahl auf das letzte gültige Feld des Pfades zurückkehren [326, 386]. Der gleiche Block darf nicht mehrfach berührt werden. Nachdem der Pfad erfolgreich erkundet wurde, wird dieser durch den Teilnehmer in der Matrix erneut repliziert. Nachfolgend werden die drei existierenden Fehlertypen bezüglich der vorgestellten Regeln und der relativen Position des aktuell berührten Blocks vorgestellt. Ein Durchgang endet,

nachdem das Ende des richtige Pfades erreicht ist. Für fünf erfolgreiche Lerndurchgänge auf einer 10 x 10-Matrix werden 10-15 Minuten veranschlagt [326, 386].

**Adressierte kognitive Domänen und relevante Maße:** In seiner ursprünglichen Variante werden durch den Groton Maze Learning Test (GMLT) aus der CogState-Testbatterie die kognitiven Domänen der komplexen Aufmerksamkeit (Verarbeitungsgeschwindigkeit), psychomotorischen Funktion sowie des Lernens und der Exekutivfunktion (Arbeitsgedächtnis) bei gesunden Erwachsenen getestet [279, 326]. Als Fehlermaß wird zwischen der Anzahl von drei verschiedenen Fehlerarten (ER, EL, EP), die beim Abruf des Pfades auftreten können, unterschieden [386]:

1. Rulebreak Error (*ER*): Beim Erkunden des Pfades wird kein Nachbarfeld des zuletzt besuchten Feldes ausgewählt (bspw. lückenhafter Pfad).
2. Legal Error (*EL*) oder auch Exploratory Error: Es wird ein falsches Nachbarfeld ausgewählt (bspw. in die falsche Richtung).
3. Perseverative Error (*EP*): Wiederholung eines bereits begangenen Fehlers.

Des Weiteren existieren Ausprägungen des GMLT in der CogState-Testbatterie, die zur Messung der Verarbeitungsgeschwindigkeit eingesetzt werden, indem möglichst viele richtige Blöcke in eingeschränkter Zeit gefunden werden müssen oder einem blockweise aufleuchtenden Pfad schnellstmöglich gefolgt werden muss [326]. Diese Maße werden sowohl für die Phase des Erlernen eines Pfades als auch für den Abruf des Pfades aufgenommen [326]. Papp et al. haben herausgefunden, dass Personen mit amnestischer MCI signifikant mehr Fehler beim Erlernen des Pfades machen als gesunde Personen und den Pfad weniger schnell über die Lernphasen hinweg erlernen [320]. Einschränkungen in der visuell-räumlichen Orientierung und dem Lernen lassen sich damit bereits früh offenlegen [320].

## Go/No-Go-Aufgaben

Go/No-Go-Aufgaben werden in der Neuropsychologie als Stimuli zur Herbeiführung von Entscheidungssituationen genutzt [166, 167]. Gordon et al. schlagen diese Art von Test vor, um den Zugriff auf die Sprache und die lexikalische Entscheidungsfähigkeit anhand der Abfrage von Wörtern zu untersuchen [165, 166, 167].

**Ziel und Testablauf:** Bei Go/No-Go-Aufgaben geht es um das Reagieren auf bestimmte Objekte [99]. Diese Objekte können spezifische Symbole oder Wörter sein, die in einer Menge an zufällig angeordneten Objekten nacheinander gezeigt werden und auf die der Teilnehmer mechanisch reagieren muss sobald das vorgegebene Objekt erscheint [478].

**Adressierte kognitive Domänen und relevante Maße:** Um den kognitiven Bereich der selektiven Aufmerksamkeit, Exekutivfunktion (Entscheidungen) und Sprache zu messen, werden Go/No-Go-Aufgaben eingesetzt [99]. Dabei werden die Reaktionszeit und die Fehlerrate von richtig oder falsch erkannten Objekten gemessen. Collette et al. konnten nachweisen, dass Menschen mit AD signifikant längere Reaktionszeiten im Vergleich zu gesunden Personen zeigen [99].

## Consortium to Establish a Registry for Alzheimer's Disease

Das Consortium to Establish a Registry for Alzheimer's Disease (CERAD) wurde im Jahre 1986 ins Leben gerufen, um eine Testbatterie zur Unterstützung einer AD-Diagnose zu entwickeln, die auf standardisierten, validierten und zuverlässigen neuropsychologischen Tests basieren [145, 295]. Über die Jahre hat sich die dadurch entstandene Testbatterie CERAD-NP als nützliches und frei zugängliches Werkzeug erwiesen, um sinnvolle Daten für eine Differenzialdiagnose zu liefern [364].

**Ziel und Testablauf:** Die Testbatterie verfolgt das Ziel einer differenzierten Erfassung von kognitiven Defiziten bei Menschen mit Demenz [364]. Hierfür werden Tests kombiniert, die auch in Teilen ihre Anwendung in anderen Testbatterien finden. Der CERAD-Plus ist eine erweiterte Version, der zusätzlich zur ursprünglichen Version drei weitere Tests integriert. Für die Dauer der Durchführung bei gesunden Probanden werden 20 Minuten angegeben, bei Menschen mit kognitiven Einschränkungen zwischen 30 und 45 Minuten [21]. Unter den ausgewählten Tests befindet sich auch der zuvor vorgestellte MMSE. Außerdem ist eine auf 15 Items reduzierte Version des Boston Naming Test (BNT) integriert, der ähnlich wie der FCSRT durch die visuelle Konfrontation der Patienten mit gezeichneten Objekten und deren anschließenden Benennung abläuft [127, 422]. Ebenso wird die verbale Flüssigkeit geprüft, indem durch die Teilnehmer möglichst viele Tiere innerhalb einer Minute aufgezählt werden müssen. Ebenfalls muss eine Wortliste mit zehn Wörtern in drei Durchgängen gelernt und direkt abgerufen werden. Die konstruktive Praxis wird durch das Abzeichnen von vier Figuren geprüft (Kreis, Rhombus, überlappende Vierecke, Würfel). Danach folgt ein verzögerter Abruf der zuvor gelernten Wortliste, genannte Wörter die nicht auf der Liste sind, werden ebenfalls notiert (Intrusionen). Eine weitere Aufgabe testet die Wiedererkennung der Worte von der Wortliste im Vergleich mit zehn Worten, die zuvor nicht gezeigt wurden (Distraktoren). Nach dieser Unterbrechung müssen erneut die zuvor gezeichneten Figuren ohne Vorlage gezeichnet werden. Zusätzlich wird der Trail Making Test (TMT) in der Variante A und B eingesetzt, bei dem Zahlen beziehungsweise die Kombination von Zahlen und Buchstaben abwechselnd in aufsteigender Reihenfolge verbunden werden müssen [405] (vgl. Abbildung 2.12).

**Adressierte kognitive Domänen und relevante Maße:** Mit der erweiterten Testbatterie CERAD-Plus lassen sich die kognitiven Funktionsbereiche der Orientierung, Sprache, psychomotorische Geschwindigkeit, konstruktive Praxis und (verbales, visuell episodisches und semantisches) Gedächtnis überprüfen [290, 364].

Mistridis et al. stellen anhand des CERAD-Plus in einer über zwölf Jahre hinweg laufenden Beobachtungsstudie fest, dass insbesondere die damit gemessenen Aussagen über die kognitiven Funktionsbereiche des Gedächtnisses und die Exekutivfunktionen bereits in einem präklinischen Stadium von MCI acht Jahre vorher beziehungsweise zwei Jahre vor der klinischen Diagnose von MCI erkannt werden können, was vergleichbar mit den Zeiträumen ist, die in Studien zur Früherkennung mittels Liquoruntersuchungen und Biomarkern genannt werden [290]. Eine multi-centrische Validierungsstudie für die deutsche Version der CERAD-Testbatterie hat ergeben, dass bei allen Variablen außer denen der konstruktiven Praxis (Figuren abzeichnen), des BNT und der Intrusionen, die Skalen des CERAD-NP zwischen Menschen mit AD und der gesunden Kontrollgruppe eine sehr gute Diskriminationsfähigkeit aufweisen [21]. Dabei stellt sich die Kombination der Skalen für die Tests zur Überprüfung der verbalen Flüssigkeit (Tiere benennen), des Erlernens und Abrufens der Wortliste, die Diskriminabilität von Distraktoren (Worte aus Wortliste von anderen unterscheiden) und die konstruktive Praxis (Figuren aus dem Gedächtnis zeichnen) als beste Kombination mit einer Sensitivität von 87% und einer Spezifität von 98% heraus [21].

## Story-based Empathy Task

Unter Theory of Mind (TOM) wird der Prozess verstanden, bei dem sich Menschen selbst und anderen Personen mentale Zustände zuordnen [121]. Hierfür wird primär zwischen zwei Bereichen unterschieden, nämlich zwischen dem Verständnis für die Absicht und der Emotionen anderer [121]. Um diese Attribute zu operationalisieren wurde der Story-based Empathy Task (SET) entwickelt [121].

**Ziel und Testablauf:** Ziel ist die Überprüfung von Funktionsbereichen, nämlich der Identifizierung von emotionalen Zuständen der Akteure (Emotionsattribution), von Absichten der Akteure (Intentionsattribution) sowie als Kontrollbedingung von Kausalitätsketten basierend auf dem Wissen über die physikalischen Eigenschaften von Objekten und menschlichen Körpern (Kausalitätsreaktion) [121]. Dies geschieht mit Hilfe von Bildergeschichten mit jeweils sechs Bildern, von denen drei einen bereits geschehenen Ablauf beschreiben und eines der verbleibenden drei (A, B oder C) als wahrscheinlichste Antwort für die Fortsetzung ausgewählt werden muss (Abbildung 2.11). Für jede Geschichte gibt es nur eine richtige Antwort. Pro richtiger Antwort wird ein Punkt vergeben, insgesamt werden pro Funktionsbereich sechs Punkte vergeben, was eine maximale Punktzahl von 18 ermöglicht [121]. Für die Durchführung werden 20 Minuten angegeben [121].

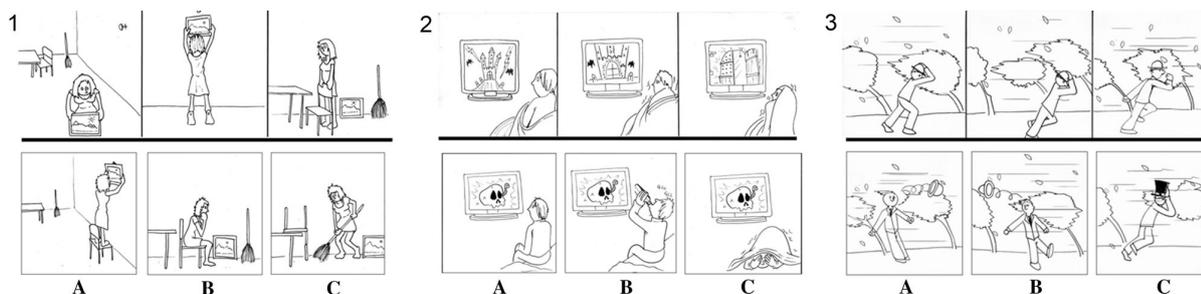


Abbildung 2.11: Beispiele für Intentions-, Emotionsattribution und Kausalitätsreaktion von links aus Dodich et al. [121].

**Adressierte kognitive Domänen und relevante Maße:** Die TOM geht mit der Fähigkeit auf die Absichten und Emotionen anderer zu schließen einher. Zur Unterstützung einer Diagnose wird auch die Messung der sozialen Kognition empfohlen, da sich ein sozialer Rückzug bereits in frühen Phasen der Demenz zeigt [57, 151]. Auch bereits in der Phase der MCI kann eine Reduktion der Intentions- und Emotionsattribution sichtbar werden, die dann stetig zunehmen kann [151]. Für den SET existiert zum Zeitpunkt der Anfertigung dieser Arbeit keine Studie, die einen direkten Bezug zur Diagnose von Demenz untersucht, jedoch konnte gezeigt werden, dass der SET als Hilfsmittel zur Quantifizierung von subtilen Defiziten der sozialen Kognition eingesetzt werden kann [121].

## Montreal Cognitive Assessment

Der von Nasreddine et al. entwickelte Montreal Cognitive Assessment (MoCA) ist ein Schnelltest, der Experten beim Screening nach MCI unterstützen soll [281, 297]. Hiermit sollen die Schwächen der MMSE adressiert werden, die bei der Erkennung von Frühformen der Demenz und MCI zu Schwierigkeiten führt [297].

**Ziel und Testablauf:** Das Ziel des MoCAs ist ein zeitökonomisches Screening, womit Mediziner innerhalb von zehn Minuten eine Aussage über die kognitiven Leistungseinbußen insbesondere hinsichtlich des Vorliegens von MCI treffen können sollen [297]. Hierfür sind auf einem Blatt Papier verschiedene

Aufgaben angeordnet, die nacheinander mit den Patienten durchgeführt werden (vgl. Abbildung 2.12). Dabei gibt es maximal fünf Punkte für das Erlernen und den verzögerten Abruf von fünf Nomen. Für

MONTREAL COGNITIVE ASSESSMENT (MOCA)

NAME : \_\_\_\_\_ Education : \_\_\_\_\_ Date of birth : \_\_\_\_\_  
 Sex : \_\_\_\_\_ DATE : \_\_\_\_\_

VISUOSPATIAL / EXECUTIVE		Copy cube		Draw CLOCK (Ten past eleven) (3 points)		POINTS	
							___/5
<b>NAMING</b>							
						___/3	
<b>MEMORY</b>							
Read list of words, subject must repeat them. Do 2 trials. Do a recall after 5 minutes.		FACE	VELVET	CHURCH	DAISY	RED	No points
		1st trial					
		2nd trial					
<b>ATTENTION</b>							
Read list of digits (1 digit/ sec).		Subject has to repeat them in the forward order		[ ] 2 1 8 5 4		___/2	
		Subject has to repeat them in the backward order		[ ] 7 4 2			
Read list of letters. The subject must tap with his hand at each letter A. No points if ≥ 2 errors		[ ] F B A C M N A A J K L B A F A K D E A A A J A M O F A A B				___/1	
Serial 7 subtraction starting at 100		[ ] 93	[ ] 86	[ ] 79	[ ] 72	[ ] 65	___/3
		4 or 5 correct subtractions: 3 pts, 2 or 3 correct: 2 pts, 1 correct: 1 pt, 0 correct: 0 pt					
<b>LANGUAGE</b>							
Repeat : I only know that John is the one to help today. [ ]		The cat always hid under the couch when dogs were in the room. [ ]				___/2	
Fluency / Name maximum number of words in one minute that begin with the letter F		[ ] _____ (N ≥ 11 words)				___/1	
<b>ABSTRACTION</b>							
Similarity between e.g. banana - orange = fruit [ ] train - bicycle [ ] watch - ruler						___/2	
<b>DELAYED RECALL</b>							
Has to recall words WITH NO CUE		FACE	VELVET	CHURCH	DAISY	RED	Points for UNCLUED recall only
		[ ]	[ ]	[ ]	[ ]	[ ]	
Optional		Category cue					
		Multiple choice cue					
<b>ORIENTATION</b>							
[ ] Date		[ ] Month	[ ] Year	[ ] Day	[ ] Place	[ ] City	___/6
© Z.Nasreddine MD Version November 7, 2004		Normal ≥ 26 / 30		TOTAL		___/30	
www.mocatest.org				Add 1 point if ≤ 12 yr edu			

Abbildung 2.12: Fragebogen zur Beurteilung des Vorliegens von MCI nach Nasreddine et al. [297].

das Konstruieren einer Analoguhr (Clock-Drawing-Test, maximal drei Punkte) und das Abzeichnen einer dreidimensionalen Darstellung eines Würfels (Cube-Drawing-Test, maximal ein Punkt) gibt es insgesamt maximal vier Punkte [297]. Einer adaptierten Version des TMT-B (maximal ein Punkt) folgen eine Aufgabe zur Prüfung der Sprachflüssigkeit (maximal ein Punkt) und zur verbalen Abstraktion (maximal zwei Punkte). Des Weiteren wird das Anzeigen eines vorgegebenen Ziels (maximal ein Punkt), das kontinuierliche Subtrahieren einer Zahl (maximal drei Punkte) und das Erlernen und Abrufen von Ziffern vorwärts wie rückwärts (maximal ein Punkt pro Richtung) getestet. Danach werden zeichnerische Abbildungen von drei Tieren (Löwe, Kamel, Nashorn) dargestellt, die durch die Patienten benannt werden müssen (ein Punkt pro richtig benanntem Tier). Gefolgt von der Wiedergabe zweier vorgegebener Sätze mit syntaktisch komplexer Struktur (maximal zwei Punkte) sowie die Angabe von der aktuellen Zeit und des aktuellen Ortes der getesteten Person bilden den Abschluss dieser Untersuchung (maximal sechs Punkte).

**Adressierte kognitive Domänen und relevante Maße:** Der MoCA deckt folgende acht kognitiven Bereiche ab: Kurzzeitgedächtnis, wahrnehmungsmotorische Funktionen, Exekutivfunktionen, Aufmerksamkeit, Konzentration und Arbeitsgedächtnis, Sprache und Orientierung. Insgesamt können maximal 30 Punkte erreicht werden. Als Grenze für die Entscheidung, ob ein MCI vorliegt oder nicht (Cut-Off Score) ergibt sich bei einer maximalen Spezifität und Sensitivität von 87% beziehungsweise 90% die mindestens zu erreichende Punktzahl 26.

## 2.3 Juristische Grundlagen und regulatorische Randbedingungen

Dieser Abschnitt erklärt die einschlägigen juristischen Begrifflichkeiten, Gesetze und Normen, die für das Verständnis der nachfolgenden Lösungskonzepte notwendig sind. Der Fokus liegt auf im Laufe der Anfertigung dieser Arbeit beschlossenen und in Kraft getretenen Gesetze und gesetzlichen Änderungen. Abbildung 2.13 stellt die Chronologie deren Einführung und inhaltlichen Ziele dar. Hierbei findet keine

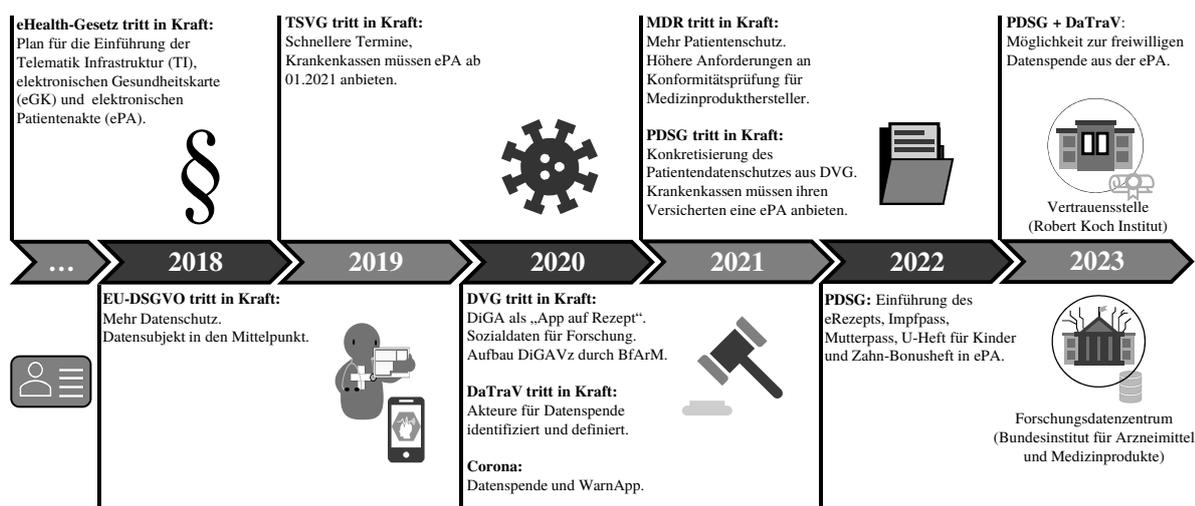


Abbildung 2.13: Chronologie relevanter Gesetzes- und Verordnungseinführungen im Zeitraum von 2015 bis 2021.

detaillierte Beschreibung der relevanten Gesetzestexte statt sondern lediglich eine Betrachtung des vorhandenen Ist-Zustandes und davon relevanter Teilaspekte. Grundlegend beziehen sich die gesetzlichen Randbedingungen und Änderungen auf Ansprüche aus dem Bürgerlichen Gesetzbuch (BGB) und SGB. Im BGB sind privatrechtliche Regelungen in insgesamt fünf Büchern definiert [112]. In dieser Arbeit tritt beispielsweise das in Buch II definierte Recht der Schuldverhältnisse im Kontext der Dokumentationspflicht bei ärztlichen Behandlungen auf. Das SGB hat die Aufgabe der Gestaltung von den zu einer sozial gerechten und sicheren Gesellschaft notwendigen Rechten (nach § 1 SGB I) und ist zum Zeitpunkt der Anfertigung dieser Arbeit in 13 Bücher unterteilt [111]. Für diese Arbeit relevante Regelungen für das deutsche Gesundheitswesen werden in Buch V geführt. Eine Kontextualisierung dieser Gesetze findet bei Notwendigkeit in den Folgekapiteln statt. Die aus dem Umgang damit resultierenden Anforderungen werden in Kapitel 4 beschrieben.

### 2.3.1 Regulatorische Einflüsse auf das Gesundheitswesen und die Medizintechnik

Seit Anfang der 2000er Jahre wird politisch versucht den Weg für ein digitales Gesundheitswesen in Deutschland zu ebnen. Im Zeitrahmen der Anfertigung dieser Arbeit (2016-2022) entstand ebenfalls eine Vielzahl an Gesetzesinitiativen und regulatorischen Anpassungen. Dieser Abschnitt soll einen Überblick darüber verschaffen und die Abhängigkeiten, die zwischen diesen gesetzlichen Änderungen bestehen, auszugsweise aufzeigen.

## **GKV-Modernisierungsgesetz**

Mit dem Inkrafttreten des Gesetzes zur Modernisierung der GKV am 1. Januar 2004 intendierte die damalige Bundesregierung das Ziel eines schlankeren und günstigeren Gesundheitssystems [90]. Durch die Änderungen an § 291 SGB V wurde der Grundstein für die TI sowie dessen Betreibergesellschaft gematik als Gesellschaft mit beschränkter Haftung (GmbH) gelegt und die Einführung der dafür notwendigen eGK als Berechtigungsnachweis für Leistungsempfänger definiert. Damit wurde erstmals der Weg zu einem digitalisierten Gesundheitssystem skizziert und die politische Weichenstellung dafür geschaffen.

## **E-Health-Gesetz**

Da die Realisierung der ursprünglich im GKV-Modernisierungsgesetz definierten Ziele, trotz zahlreicher Gesetze zur Weiterentwicklung dieser, nur schleppend voranging (vgl. Unterunterabschnitt 2.3.1), trat am 01.01.2016 das E-Health-Gesetz in Kraft, um die Potentiale der Digitalisierung endgültig für die Gesundheitsversorgung zu nutzen und weitere diesbezüglich notwendige Änderungen des SGB V zu veranlassen [114]. Mit dem Ziel einen konkreten Plan für den Aufbau von sicheren digitalen Kommunikationsmöglichkeiten und medizinischen Anwendungen zwischen Leistungserbringern, Krankenkassen und Patienten zu gewährleisten, werden darin sowohl Vorgaben und Fristen als auch Sanktionsmöglichkeiten bei Nichteinhaltung festgelegt. Darunter fällt die Anbindung der Krankenhäuser und Arztpraxen bis Mitte 2018 an die TI. Außerdem sollte ab Mitte 2016 eine Stammdatenaktualisierung durch Ärzte, Apotheken und gesetzliche Krankenkassen durchgeführt werden können sowie ab dem Jahr 2018 das Speichern von Notfalldaten auf der eGK ermöglicht werden [278]. Demnach wurde auch der Anspruch auf einen Medikationsplan in Papierform sowie ab 2018 in elektronischer Form geregelt. Ebenso die Schaffung der technischen Voraussetzungen für eine ePA bis Ende 2018. Diese sollte ab 2019 für gesetzlich Versicherte zur freiwilligen Nutzung angeboten werden, um Daten wie Arztbriefe, Notfalldaten, Medikationsplan und weitere digitalen Gesundheitsdaten verwalten zu können. Die Freigabe der Daten erfolgt gemeinsam über die Kombination der Schlüssel aus dem elektronischen Heilberufsausweis (eHA) und der eGK beim Leistungsbringer vor Ort. Einzige Ausnahmen stellen Notfalldaten und Medikationsplan dar. Neben Vergütungszuschlägen für die Erstellung elektronischer Entlassbriefe wurde auch eine finanzielle Förderung für telemedizinische Leistungen geregelt, die für das Jahr 2017 mit Befundbeurteilungen von Röntgenaufnahmen und Online-Video-Sprechstunden benannt wurden [278].

## **Terminservice- und Versorgungsgesetz**

Mit dem am 11. Mai 2019 in Kraft getretenen TSVG wird die Verbesserung der medizinischen Leistungserbringung für den Versicherten angestrebt [87]. Ein Schwerpunkt der damit bewirkten Änderung des SGB V bildete das Ziel einer schnelleren Terminfindung für medizinische Dienstleistungen, welche primär über den Ausbau koordinierender Terminservicestellen, erhöhten Mindestsprechstundenzeiten und durch die Kassenärztliche Vereinigung (KV) betriebenen Arztpraxen in unterversorgten Gebieten erreicht werden sollte. Eine zweite Säule bildete die Digitalisierung der Versorgung, indem beispielsweise Krankenkassen gesetzlich dazu verpflichtet werden, für ihre Versicherten ab dem Jahr 2021 eine ePA anzubieten (vgl. Unterabschnitt 2.4.1). Eine im Kontext dieser Arbeit erwähnenswerte Neuerung am bestehenden Konzept der ePA stellt für den Versicherten die Entkopplung der Datenzugriffsmöglichkeiten von der eGK dar (vgl. Unterabschnitt 2.4.1). Um Entscheidungsprozesse rund um diese Entwicklungen effektiver zu gestalten, hatte das BMG die mehrheitliche Beteiligung an der gematik übernommen.

## Digitale-Versorgung-Gesetz

Das im November 2019 vom Deutschen Bundestag verabschiedete DVG sorgt für die Anpassung gesetzlicher Rahmenbedingungen zur Entwicklung und Einführung digitaler Lösungen im Gesundheitswesen. Definiertes Ziel ist es, die im Zuge der digitalen Transformation zu erwartenden Innovationen schneller in die Versorgung integrieren zu können, um deren Potentiale für Deutschland besser zu nutzen [89]. Das Gesetz regelt dafür schwerpunktmäßig die Digitalisierung von Verwaltungsprozessen, Fördermöglichkeiten digitaler Innovationen sowie eine schnellere Überführung dieser in die Regelversorgung. Außerdem wird darin eine zunehmende Vernetzung von Leistungserbringern durch die TI und die Stärkung von Telemedizin angestrebt. Des Weiteren wird für gesetzlich Versicherte ein Anspruch auf digitale Gesundheitsanwendungen festgeschrieben und die Regelungen zur Datentransparenz weiterentwickelt, um die Sozialdaten der Krankenkassen für Forschungszwecke nutzen zu können. Hierfür erhält das SGB V mit §33a einen Zusatzparagraphen, der den Anspruch auf die Versorgung mit Medizinprodukten niedriger Risikoklasse festlegt (vgl. Unterunterabschnitt 2.3.1). Danach werden die Kosten für digitale Anwendungen, die vom BfArM in das DiGAVZ aufgenommen und von einem behandelnden Arzt verschrieben oder der Krankenkasse genehmigt werden, erstattungsfähig. Die Regelung dieser Vorgänge und Anforderungen, die an die Hersteller gestellt werden, erfolgt in Zusatzparagraph §139e SGB V. Demnach muss der Hersteller Nachweise erbringen, dass die Anwendung „Anforderungen an Sicherheit, Funktionstauglichkeit und Qualität des Medizinproduktes“ entspricht, den „Datenschutz und die Datensicherheit nach dem Stand der Technik gewährleistet“ ist und das Produkt „positive Versorgungseffekte aufweist“. Die Hauptfunktion muss dabei auf digitalen Technologien beruhen, dadurch erzielte positive Versorgungseffekte müssen durch ein „von einer herstellerunabhängigen Institution erstelltes wissenschaftliches Evaluationskonzept“ nachgewiesen werden [89].

Die Nutzbarmachung von Daten für Forschungszwecke wird in den §§303a bis 303f neu geregelt. Demzufolge ist eine räumliche, organisatorische und personelle Aufteilung der Aufgaben für die Datentransparenz vorgesehen, die von durch das BMG und BMBF zu bestimmende öffentliche Stellen des Bundes sowie vom SGKV übernommen werden sollen [89]. Der Ablauf und die beteiligten Institutionen sind in Abbildung 2.14 skizziert. Der SGKV soll als Datensammelstelle der Versicherten fungieren. Die Krankenkassen übermitteln hierfür die in §303b, Absatz 1 geregelten Daten. Die technische Ausgestaltung hierfür sollte bis Ende des Jahres 2021 geregelt sein. Ausgehend von dieser Datensammelstelle werden die Einzeldaten durch die Datensammelstelle zusammengeführt, geprüft und das Versichertenkennzeichen durch eine Arbeitsnummer ersetzt. Die Übermittlung der Daten an das Forschungsdatenzentrum und die Definition des Aufgabenspektrums erfolgt nach §303d. Die organisatorisch getrennte Vertrauensstelle bekommt die Zuordnung zwischen den Versichertenkennzeichen und Arbeitsnummern durch die Datensammelstelle übermittelt. Die Aufgabe der Vertrauensstelle ist es, aus den übermittelten Lieferpseudonymen und Arbeitsnummern, periodenübergreifende Pseudonyme für die jeweiligen Versicherten zu ermitteln und die Liste mit Arbeitsnummern und Pseudonymen, dem Forschungsdatenzentrum zur Verfügung zu stellen. Danach ist die Vertrauensstelle dazu verpflichtet, die beiden Ausprägungen von Pseudonymen und Arbeitsnummern zu löschen.

Die Anforderungen an die Verarbeitung von Forschungsdaten sind in §303e geregelt. Das Forschungsdatenzentrum macht den in §303e, Absatz 1 definierten berechtigten Einrichtungen die Daten zugänglich, falls diese die Voraussetzungen des §303e, Absatz 4 erfüllen. Hierin ist sowohl die Verpflichtung der verarbeitenden Personen zur Geheimhaltung nach §203 des Strafgesetzbuches als auch die Gewährleistung einer Verarbeitung, die nicht über das erforderlichen Maß hinaus geht inklusive der Sicherstellung, dass die Daten nicht kopiert werden können, festgelegt.

In §303f wird abschließend eine Gebührenordnung geregelt, die für im Rahmen der Bereitstellung von

Daten erbrachten Leistungen durch das Forschungsdatenzentrum erhoben werden dürfen. Eine weitere Ausgestaltung der gesetzlichen Anforderungen an die Datenspende erfolgt in der Datentransparenz-Verordnung (DaTraV) und dem PDSG (vgl. Unterunterabschnitt 2.3.2).

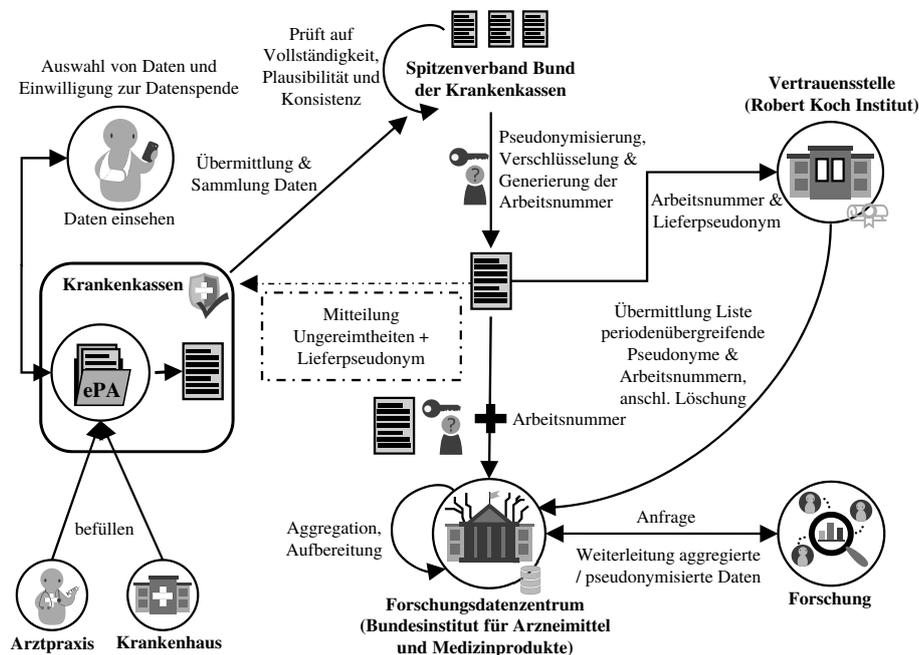


Abbildung 2.14: Datenspende ab Januar 2023 nach DVG, PDSG und DaTraV.

## Digitale-Gesundheitsanwendungen-Verordnung

Die Digitale-Gesundheitsanwendungen-Verordnung (DiGAV) regelt Verfahren und Anforderungen, welche an die Prüfung einer DiGA bezüglich der Erstattungsfähigkeit durch die GKV gestellt werden [81, 83]. Für Anbieter einer DiGA bestehen dadurch Richtlinien, welche sie erfüllen müssen, damit ihre Anwendung durch das BfArM in das DiGAVZ aufgenommen wird. Außerdem sind darin Vorschriften zum Verwaltungsverfahren, dem Aufbau des DiGAVZ, der Gebührensätze und eines Schiedsverfahrens beschrieben. Demnach muss eine DiGA, zum Zeitpunkt der Anfertigung dieser Arbeit, entweder als Medizinprodukt der Produktklasse I oder IIa zertifiziert sein. Für Hersteller besteht die Möglichkeit vorläufig in das DiGAVZ aufgenommen zu werden und im Laufe einer bis zu maximal 24 Monate dauernden Testphase den positiven Versorgungseffekt nachzuweisen, um gegebenenfalls danach dauerhaft darin verzeichnet zu sein [78, 83]. Ein vergleichbares Konzept ist 2021 durch das Digitale-Versorgungs- und -Pfleger-Modernisierungs-Gesetz (DVPMG) für eine sogenannte Digitale Pflegeanwendung (DiPA) in Kraft getreten mit dem Ziel Anwendungen zur Unterstützung der Pflege nach § 40a SGB XI erstattungsfähig zu gestalten. Die Klassifizierung als Medizinprodukt ist darin nicht verpflichtend.

## Medizinproduktegesetz und Medizinprodukteverordnung

An die Entwicklung und das Inverkehrbringen von Medizinprodukten bestehen weltweit Regularien, die unter Beteiligung unterschiedlicher Akteure umgesetzt werden. Grundlegend regelt seit Mai 2021 die MDR europaweit diesen Vorgang [138]. Für das deutsche Gesundheitswesen werden diese Regularien

durch das Medizinproduktgesetz (MPG) konkretisiert und durch die in Abbildung 2.5 identifizierten Akteure implementiert [113]. Für die Konformitätsbewertung muss zunächst die Medizinproduktklasse bestimmt werden. Diese richtet sich nach der Zweckbestimmung des Produktes gemäß § 3 Nr. 10 MPG. Ein Produkt der niedrigsten Risikoklasse I kann durch Selbstauskunft auch ohne zertifiziertes Qualitätsmanagementsystem als konform erklärt werden. Ab Risikoklasse II ist diese Vorgehensweise nicht mehr möglich. Des Weiteren wird die Einbindung einer benannten Stelle in den Zertifizierungsprozess notwendig, die den Prozess auditiert und den Herstellern im Erfolgsfall Konformität bestätigt (vgl. Unterabschnitt 2.2.1). Auch reine Softwarelösungen können den Kategorien von Medizinprodukten zugeordnet werden. Eine Differenzierung findet laut BfArM statt, indem lediglich Anwendungen, die eine Wissensbereitstellung anbieten nicht als Medizinprodukt einzustufen sind [84]. Im Gegensatz dazu werden als Orientierungshilfe Anhaltspunkte genannt, die Hinweise liefern, ob Software als Medizinprodukt einzustufen ist oder nicht. Hierdurch lassen sich Anwendungen, deren Zweckbestimmung bestimmte Funktionen wie die Diagnose, Steuerung und Überwachung von Gesundheitszuständen durch das Messen, Analysieren und Interpretieren von Daten enthalten, als Medizinprodukte oder auch Software as a Medical Device (SaMD) annehmen [84, 180]. Des Weiteren sind in Anhang VIII der MDR verschiedene Klassifizierungsregeln aufgeführt, welche die Hersteller bei der Einstufung von Medizinprodukten unterstützen sollen [180, S. 15 ff.]. Hierbei wird Software als aktives Medizinprodukt eingeordnet und insofern diese Medical Device Software (MDSW) Informationen liefert, die diagnostische oder therapeutische Entscheidungen unterstützen, mindestens der Klasse IIa zugeordnet (vgl. Tabelle 2.2).

		Bedeutung der vom MDSW zur Verfügung gestellten Informationen für eine Gesundheitssituation in Bezug auf Diagnose/Therapie		
		hoch	mittel	niedrig
		behandeln oder diagnostizieren	treibt das klinische Management an	informiert das klinische Management (alles andere)
Zustand der Gesundheitssituation oder Zustand des Patienten	kritische Situation oder Zustand des Patienten	Klasse III	Klasse IIb	Klasse IIa
	ernste Situation oder Zustand des Patienten	Klasse IIb	Klasse IIa	Klasse IIa
	nicht ernste Situation oder Patientenzustand (alles andere)	Klasse IIa	Klasse IIa	Klasse IIa

Tabelle 2.2: Klassifikation nach der International Medical Device Regulators Forum (IMDRF) für Software als Medizinprodukt [283].

### Relevante Normen

Auf europäischer Ebene existieren sogenannte harmonisierte Normen, die durch alle nationalen Normungsgremien übernommen werden und somit europaweite Gültigkeit besitzen. Für harmonisierte Normen gilt das Vermutungsprinzip, was bedeutet, dass bei Einhaltung einer harmonisierten Norm vermutet wird, dass die von dieser abgedeckten Aspekte erfüllt sind [180]. Für die Entwicklung von Medizinprodukten bestehen ebenfalls Empfehlungen zur Einhaltung von Normen, um bei Konformität die Wahrscheinlichkeit einer späteren Zulassung zu erhöhen (vgl. Abbildung 2.15). Nachfolgend wird der Schwerpunkt auf Normen zur Erfüllung der Anforderungen an die Gebrauchstauglichkeit gelegt. Die harmonisierte Norm IEC 62366-1 beschreibt die Anwendung der Gebrauchstauglichkeit auf Medizinprodukte anhand eines gebrauchstauglichkeitsorientierten Entwicklungsprozesses [116]. Hierdurch sollen Risiken identifiziert und minimiert werden, die durch den unsachgemäßen Gebrauch bestehen könnten beziehungsweise

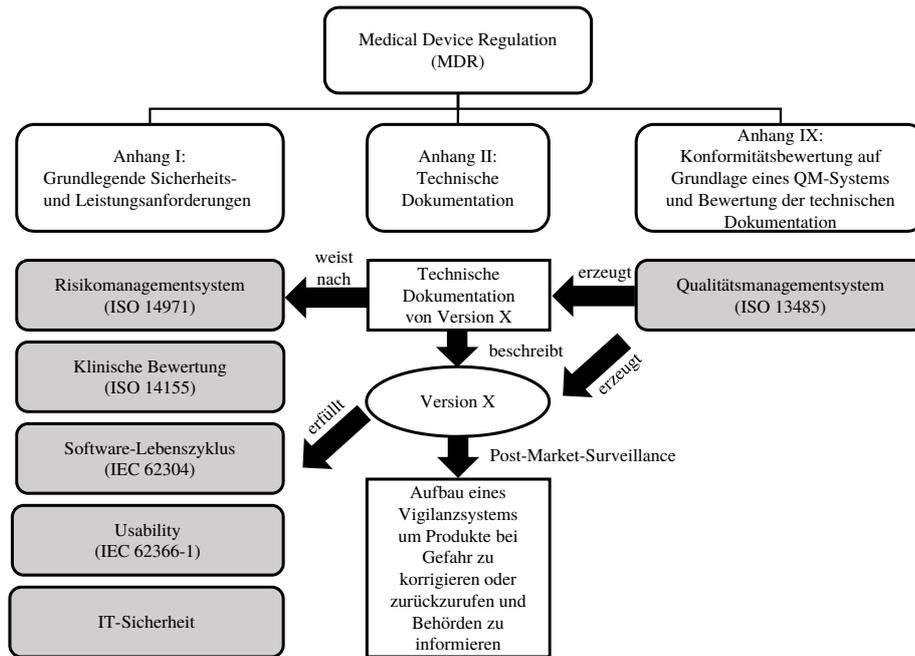


Abbildung 2.15: Zusammenhänge der Anforderungen und harmonisierten Normen der MDR erweitert nach Hastenteufel und Renaud [180, S. 39].

durch die bestimmungsgemäße Verwendung beispielsweise durch Benutzungsfehler entstehen. Der in Abbildung 2.16 dargestellte Zusammenhang zwischen den Anforderungen aus der ISO 14971 [120], welche den generellen Prozess für das Risikomanagement bei Medizinprodukten spezifiziert, sowie den Anforderungen der IEC 62366-1, zeigt den gebrauchstauglichkeitsorientierten Entwicklungsprozess als Teil des Risikomanagements. Die normenkonforme Durchführung des gesamten Prozesses umfasst neben einer formativen auch eine summative Evaluation. Da die Norm Deutsches Institut für Normung (DIN) Europäische Norm (EN) ISO 9241-110 „Grundsätze der Dialoggestaltung“ durch „Interaktionsprinzipien“ als erneuerte Version abgelöst wird, dient diese, zum Zeitpunkt der Anfertigung dieser Arbeit noch im Entwurf-Status vorliegende Norm, als Grundlage [116, 119]. Für die Informationsdarstellung existiert zusätzlich die Norm DIN EN ISO 9241-125 mit konkreten Anforderungen an die Gestaltung des User Interface (UI) [118]. Wie in Abbildung 2.17 ersichtlich wird, enthalten die Normen unterschiedliche Richtlinien, die zur Gestaltung einer benutzerfreundlichen Benutzerschnittstelle beitragen sollen. Die darin definierten einzelnen Grundsätze und Empfehlungen sind sowohl abhängig von der Benutzergruppe, dem Anwendungszweck als auch der eingesetzten Technik und diesbezüglichen Interaktionsmöglichkeiten zu gewichten.

Zur Unterstützung der Umsetzung einer geeigneten Nutzerschnittstelle bestehen generelle Anforderungen an die Informationsdarstellung nach der DIN EN ISO 9241-112, die eingehalten werden sollten. Ziel ist es für die Nutzer eine schnellere Aufgabenerfüllung zu erreichen und deren mentalen Aufwand zu reduzieren. Hierbei wird die „Entdeckbarkeit“ von Informationen und Elementen hervorgehoben, um in die Wahrnehmung der Nutzer zu gelangen. Die Geschwindigkeit der Informationsdarstellung sollte ebenfalls angepasst werden können. Dabei ist auch „Ablenkungsfreiheit“ anzustreben, indem möglichst wenige Störungen durch andere Ereignisse zugelassen werden. Die „Unterscheidbarkeit“ zwischen Informationselementen ist dabei ebenso durch die Gestaltung zu ermöglichen wie deren „Eindeutige Interpretierbarkeit“ durch den Einsatz von Vokabular mit dem die Zielgruppe vertraut ist. Der Grundsatz der „Kompaktheit“ stellt heraus, dass die Komplexität der Darstellung zu minimieren ist, um die Aufmerksamkeit der Nutzer nur auf die wichtigsten Informationen und Elemente zu lenken. Die „Konsistenz“ bei der Verwendung

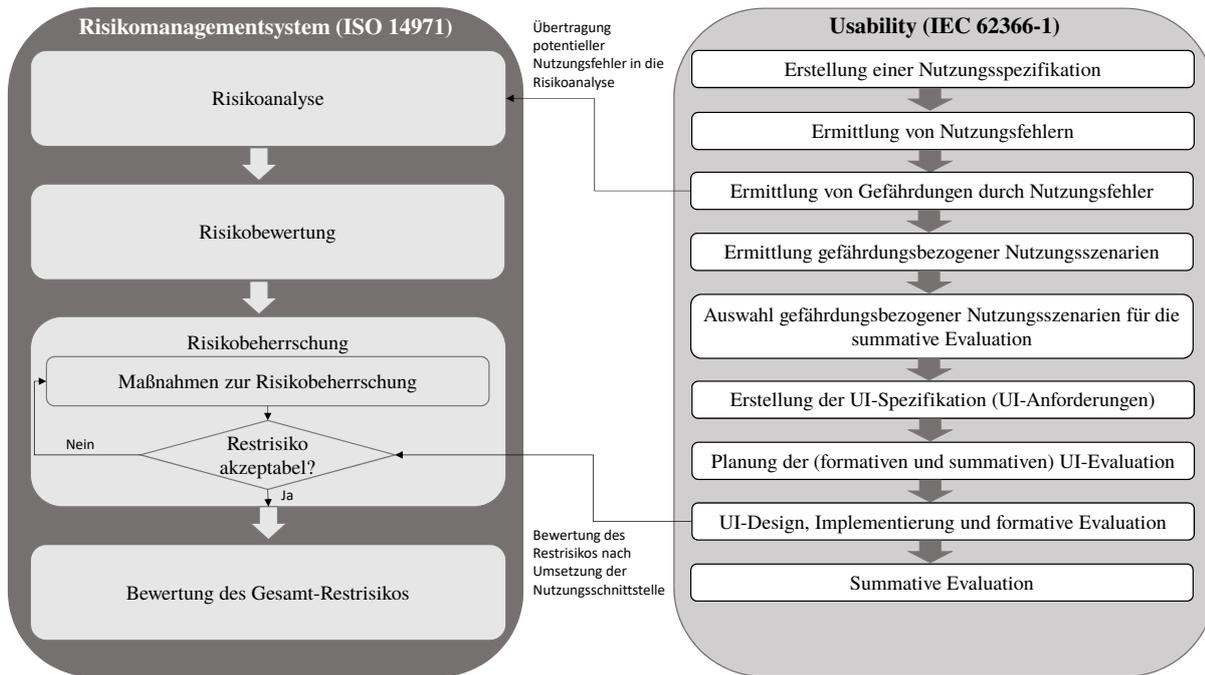


Abbildung 2.16: Gebrauchstauglichkeitsorientierter Entwicklungsprozess nach IEC 62366-1 und Zusammenhang mit dem Risikomanagement nach International Organization for Standardization (ISO) 14971 aus Hastenteufel und Renaud [180].

von Informationselementen sollte bei vergleichbarem Zweck sowohl innerhalb als auch außerhalb der Anwendung gegeben sein.

Darauf aufbauend definiert die Norm DIN EN ISO 9241-110 die nachfolgenden sieben Grundsätze der Interaktionsgestaltung, um eine gebrauchstaugliche und konsistente Benutzerschnittstelle zu ermöglichen. Demnach schreibt diese eine „Angemessenheit der Benutzeraufgabe“ vor, wonach dem Nutzer lediglich Bedienelement und Informationen angezeigt werden sollten, die für eine effiziente Bewältigung der aktuellen Aufgabe in der Anwendung notwendig sind. Des Weiteren wird die „Selbstbeschreibungsfähigkeit“ der Anwendung gefordert, wodurch sich der Nutzer zu jedem Zeitpunkt über den Fortschritt seiner Handlungen bewusst ist und versteht welche weiteren Aktionen wie ausgeführt werden können. Unter „Erwartungskonformität“ wird verstanden, dass das Systemverhalten für den Nutzer vorhersehbar gestaltet ist. Diese kann beispielsweise durch die konsistente Verwendung von Designstandards erreicht werden. Der Grundsatz der „Erlernbarkeit“ gilt, damit der Umgang mit einem System möglichst schnell angeeignet werden kann, indem ein Nutzer die dafür notwendigen Informationen und Funktionen leicht aufnehmen und ohne negative Konsequenzen ausprobieren kann. Hierfür spielt ebenfalls die „Steuerbarkeit“ eine Rolle, welche das Maß der Kontrolle über die Interaktionen mit dem System beschreibt. Sowohl die Geschwindigkeit und Reihenfolge der Aktionen als auch die Möglichkeit Interaktionen zu pausieren werden beispielsweise als Eigenschaften benannt, die hierfür ein System erfüllen sollte. Um auch gegenüber Benutzerfehlern ein tolerantes Verhalten zu zeigen, gilt es die „Robustheit“ des Systems zu erhöhen, indem Eingaben vereinfacht und Bestätigungsmöglichkeiten eingeführt werden, um versehentliche Falscheingaben abzusichern und Korrekturen zu ermöglichen. Zuletzt sieht diese Norm auch die Einführung von Mechanismen zur „Benutzerbindung“ vor, welche die Aufmerksamkeit des Nutzers

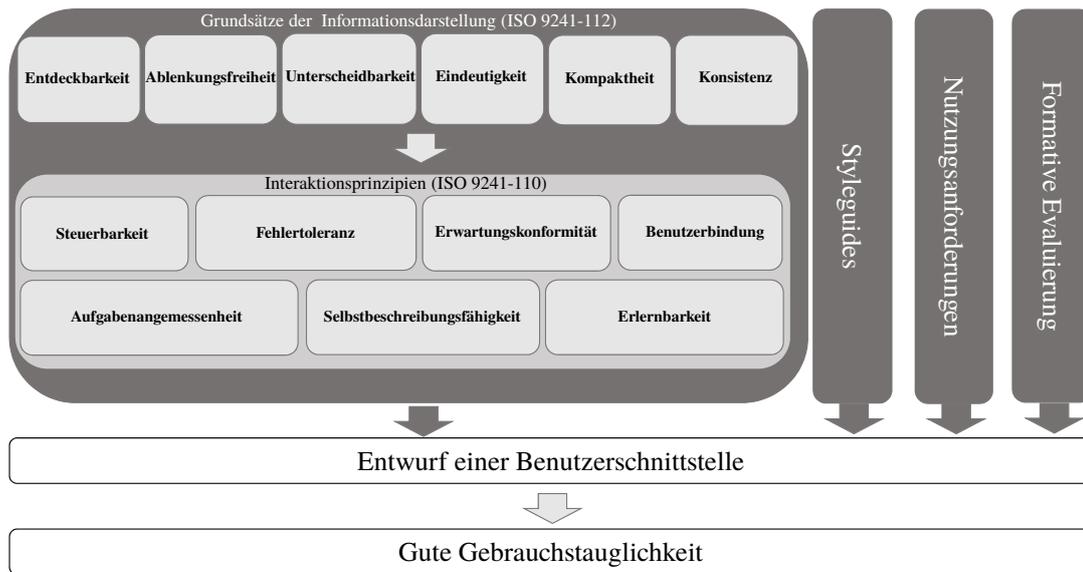


Abbildung 2.17: Anforderungen an den Entwurf einer gebrauchstauglichen Benutzerschnittstelle erweitert nach Hastenteufel und Renaud [180].

weckt und ihn motivieren soll, diese in Zukunft weiterhin zu verwenden. Als Voraussetzung für diesen Aspekt gilt der Aufbau von Vertrauen und die Einhaltung der zuvor genannten Grundsätze.

## 2.3.2 Datenschutz

Bei der Verarbeitung von personenbezogenen Daten wie Gesundheitsdaten sind die Anforderungen an den Datenschutz und die -sicherheit zu erfüllen. Innerhalb der Europäischen Union (EU) steht die unerlaubte Verarbeitung, Weitergabe oder Erhebung sogar unter Strafe. Nachfolgend werden die für diese Arbeit relevanten Verordnungen und daraus resultierende Anforderungen kurz skizziert.

### Europäische Datenschutz-Grundverordnung

Die EU-DSGVO regelt seit dem 24.05.2016 den Datenschutz in der EU [139]. Hierin wird der Rechtsrahmen für die Verarbeitung personenbezogener Daten festgelegt, der nach einer Umsetzungsfrist von zwei Jahren seit dem 25.05.2018 in jedem Mitgliedstaat unmittelbar Anwendung findet (Art. 99 Abs. 2 DSGVO). Durch den Erlass dieser Verordnung gelten die darin definierten Regeln verbindlich für alle EU-Mitgliedstaaten. Öffnungsklauseln erlauben dabei die Auslegung gewisser Regelungen zur Anwendung in nationalem Recht solange die Grundsätze der DSGVO dadurch nicht verletzt werden. Es gilt das Verbotprinzip, welches die Verarbeitung personenbezogener Daten untersagt, sollte diese nicht den Regeln gemäß dieser Verordnung folgen. Darunter fallen alle Informationen, welche natürliche Personen identifizierbar machen, das heißt ein direkter oder indirekter Bezug zu einer Identität möglich ist (Art. 4 Abs. 1 DSGVO). Unter Verarbeitung werden sämtliche manuelle oder automatisierte Vorgänge verstanden, die den Umgang mit Daten, wie zum Beispiel Erhebung, Speicherung oder Verwendung beschreiben (Art. 4 Abs. 2 DSGVO). Die von der Datenverarbeitung betroffene Person muss demnach über den Zweck der Verarbeitungsschritte informiert werden und diesem explizit durch freiwillige „Einwilligung“ zustimmen (Art. 4 Abs. 11 DSGVO). Da Gesundheitsdaten Auskunft über den Gesundheitszustand einer

natürlichen Person geben (Art. 4 Abs. 15 DSGVO), hat die Verarbeitung den Grundsätzen nach Art. 5 DSGVO folge zu leisten, die nachfolgend gekürzt wiedergegeben werden.

**Transparenz:** Unter dem Grundsatz der Transparenz wird in Art. 5 Abs. 1 lit. a DSGVO festgelegt, dass die Verarbeitung der Daten für die davon betroffene Person auf eine nachvollziehbare Weise geschieht (Erwägungsgrund 39 DSGVO). Daraus entsteht der Anspruch auf unmissverständliche Auskunft über die Art und den Umfang sämtlicher durchgeführter und geplanter Verarbeitungsschritte (Art. 12 - 15 DSGVO).

**Zweckbindung:** Als Zweckbindung wird nach Art. 5 Abs. 1 lit. b DSGVO die Festlegung eindeutiger und legitimer Zwecke verstanden, für welche personenbezogene Daten verarbeitet werden. Abweichende Verarbeitungsschritte sind nur zulässig, wenn sie unter Art. 89 DSGVO fallen. Hierin ist definiert, dass Verarbeitungszwecke denen berechnigte öffentliche Interessen zugrunde liegen unter Einhaltung bestimmter Garantien grundsätzlich möglich sind. Hierzu zählt der Gesetzgeber wissenschaftliche und historische Forschungszwecke sowie Statistik- und Archivzwecke. Eine Öffnungsklausel sieht vor, dass die EU-Mitgliedstaaten geeignete Garantien wie zum Beispiel mittels Pseudonymisierung hierfür selbst festlegen (Erwägungsgrund 156 DSGVO).

**Datenminimierung:** Nach Art. 5 Abs. 1 lit. c DSGVO hat die Zweckbindung ebenso Auswirkungen auf den Grundsatz der Datenminimierung. Demnach dürfen personenbezogene Daten nur in einem dem Zweck der Verarbeitung angemessenen und darauf reduzierten Ausmaß verarbeitet werden. Die in Art. 89 Abs. 1 geforderten Garantien sind als Maßnahmen für die Gewährleistung des Datenminimierungsgrundsatzes zu sehen.

**Richtigkeit:** Maßnahmen zur Sicherstellung der Richtigkeit von personenbezogenen Daten sind nach Art. 5 Abs. 1 lit. d DSGVO ebenfalls erforderlich, um Auswertungen vornehmen zu dürfen. Neben sachlicher Richtigkeit wird die Berücksichtigung von Aktualität vorausgesetzt. Gegebenenfalls muss vor der Verarbeitung eine Berichtigung oder Löschung der Daten stattfinden.

**Speicherbegrenzung:** Auch die Dauer der Speicherung ist nach Art. 5 Abs. 1 lit. e DSGVO an den Zweck der Datenerhebung gebunden. Die Identifikationsmöglichkeit von betroffenen Personen für deren Daten darf demnach nur so lange bestehen, wie es für den Zweck der Verarbeitung notwendig ist. Daten, deren Verwendung nach Art. 89 Abs. 1 DSGVO von öffentlichem Interesse ist, dürfen unter Einhaltung der Garantien durch Maßnahmen zum Schutz von Freiheiten und Rechten der davon betroffenen Personen länger gespeichert werden.

**Integrität und Vertraulichkeit:** Nach Art. 5 Abs. 1 lit. f DSGVO müssen ebenfalls Maßnahmen ergriffen werden, um personenbezogene Daten vor unbefugter, unrechtmäßiger oder unbeabsichtigter Verarbeitung, Verlust, Zerstörung oder Schädigung zu schützen.

**Rechenschaftspflicht:** Der Art. 5 Abs. 2 DSGVO macht eine für die Verarbeitung verantwortliche Stelle für die Einhaltung der in Art. 5 Abs. 1 lit. a-f DSGVO definierten Grundsätzen verantwortlich. Die Nachweispflicht liegt demnach ebenfalls beim Verantwortlichen.

**Einwilligung zur wissenschaftlichen Forschung:** Bei der Erhebung personenbezogener Daten für die wissenschaftliche Forschung lassen sich die Zwecke der Datenverarbeitung zum Zeitpunkt der Erhebung nicht immer vollumfänglich angeben. Nach Erwägungsgrund 33 der DSGVO wird erlaubt, dass betroffene Personen eine Einwilligung für Teilbereiche und Projekte der wissenschaftlichen Forschung erteilen, solange ethische Standards eingehalten werden.

**Personenbezogene Automatisierung von Entscheidungen:** In Art. 22 ist geregelt inwiefern eine automatisierte Entscheidungsfindung bezüglich einer Person auf Grundlage ihrer Daten stattfinden darf. Das heißt es besteht das Recht nicht ausschließlich durch computerisierte Entscheidungsfindungsalgorithmen und Profilbildung (Profiling) bewertet zu werden, falls daraus rechtliche Schritte oder Beeinträchtigungen folgen können. Zur Erfüllung vertraglicher Vereinbarungen, explizite Einwilligung oder übergeordneten Rechtsvorschriften kann dieser Anspruch aufgehoben werden.

**Recht auf Vergessenwerden:** In Art. 17 ist das Recht auf Löschung personenbezogener Daten verbrieft. Die betroffene Person kann verlangen, dass die über sie erhobenen Daten unverzüglich gelöscht werden, solange die Verarbeitung nicht erforderlich ist. In Absatz 3 sind Ausnahmen benannt, die dieses Recht einschränken wie beispielsweise sobald öffentliches Interesse wie im Bereich der Gesundheitsversorgung vorliegt oder falls die Daten zu wissenschaftlichen Zwecken benötigt werden.

### Weitere Datenschutzgesetze in Deutschland

Das Bundesdatenschutzgesetz (BDSG) flankiert die EU-DSGVO als ergänzendes Gesetz, indem es dessen Öffnungsklauseln durch für die Bundesrepublik Deutschland geltenden, spezifischen Klauseln präzisiert und erweitert [85]. Es gliedert sich in mehrere Teile. In dem Teil zu den gemeinsamen Bestimmungen werden neben der Definition von zuständigen Stellen und deren Befugnissen zur Durchsetzung des Gesetzes, darin auch die Verarbeitung besonderer Kategorien personenbezogener Daten, deren Weiterverarbeitung zu anderen Zwecken und Datenübermittlungen durch öffentliche Stellen angegeben [85]. In Bezug auf Gesundheitsdaten spricht man nach Artikel 9, Absatz 1 der DSGVO und § 22, Absatz 1 BDSG von einer dieser besonderen Kategorien personenbezogener Daten, die der höchsten Datenschutzstufe unterliegen, um die Wahrung der Interessen von betroffenen Personen zu gewährleisten. Grundsätzlich ist deren Verarbeitung untersagt, solange für diese nicht ein Erlaubnistatbestand erfüllt ist. Als Beispiel hierfür lässt sich die Verarbeitung der Daten durch medizinisches Personal nennen, welche zum Zwecke der Diagnose und Behandlung auf diese Daten im Sinne des Patientenwohls zugreifen dürfen [243].

Zusätzlich zum BDSG gestalten die Datenschutzgesetze der Länder und Kirchen für ihren spezifischen Anwendungsbereich die Zuständigkeiten und gesetzliche Anforderungen weiter aus. Je nachdem, welche Institutionen an der Datenerhebung und -verarbeitung beteiligt sind und in welcher Region sich diese befinden, können also unterschiedliche Anforderungen an den Datenschutz in Deutschland bestehen, grundsätzlich richten sich diese jedoch an den übergeordneten Gesetzgebung der EU und des Bundes aus.

### Patientendaten-Schutzgesetz und Datentransparenzverordnung

Seit dem Jahr 2020 konkretisiert das PDSG die Anforderungen aus dem DVG aus datenschutzrechtlicher Perspektive (vgl. Unterunterabschnitt 2.3.1). Hierbei wird aber auch der Patientendatenschutz für den Zulassungsweg für DiGAs verkürzt. Außerdem wird die Verarbeitung der Daten nach § 363 aus der elektronischen Patientenakte im Zuge einer sogenannten „Datenspende“ zu Forschungszwecken freiwillig und pseudonymisiert ab dem Jahr 2023 geregelt. Die Zugriffsrechte auf die gespendeten Daten legt das Datenforschungszentrum fest, welches unter der Leitung des BMG steht. Das Gesetz definiert hierfür legitime Zwecke der Verarbeitung, den Patienten wird jedoch freigestellt, in welchem Umfang diese ihre Daten freigeben. Des Weiteren wird darin der Anspruch auf die freiwillige Nutzung der ePA von Versicherten und durch medizinische Einrichtungen auf den 01.01.2021 festgelegt. Über die ePA werden demnach Schritt für Schritt ärztliche Befunde, Behandlungsberichte, Bilder, der Impfpass, Medikationspläne sowie weitere Dokumente digital verwaltet. Ziel dabei ist die selbstbestimmte Verwaltung der

Gesundheitsdaten. Die DaTraV listet den Umfang der Datenübertragung durch den SGKV auf. Die Aufgabe der geplanten Vertrauensstelle ist an das RKI übertragen worden und das Forschungsdatenzentrum wird vom BfArM betrieben (vgl. Abbildung 2.14).

## 2.4 Technische Begriffserklärungen

Die hierin gegebenen Definitionen und Begriffserklärungen bilden die Grundlage für die technisch-wissenschaftliche Einordnung der Arbeit. Es werden sowohl Themenfelder bezüglich der eingesetzten Technologien adressiert als auch domänenspezifische Teilaspekte abgedeckt.

### 2.4.1 Gesundheitsdaten und deren Verwaltung

Um den Ausbau digitaler Infrastrukturen im Gesundheitswesen zu steuern und das Diskussionspotential damit einhergehender Konflikte aufzulösen, hat der Gesetzgeber in den letzten Jahren die in Abschnitt 2.3 skizzierten Gesetze erlassen. Das BGB und SGB legen die juristische Grundlage für folgende verschiedenen Aktenkonzepte, welche ursprünglich eine Datenhaltung für das deutsche Gesundheitswesen regeln sollten (vgl. Tabelle 2.3):

- elektronische Arztakte (eAA),
- elektronische Gesundheitsakte (eGA),
- einrichtungsübergreifende elektronische Patientenakte (eEPA),
- elektronische Fallakte (eFA).

Die eAA ist durch § 630f BGB definiert und schreibt dem LE vor, dass zur Dokumentation von Behandlungsleistungen eine Patientenakte in Papierform oder elektronisch geführt werden muss, wobei inhaltliche Änderungen aufgezeigt werden müssen, indem dokumentiert wird, wann und durch wen diese vorgenommen wurden. Diese Art von Aktenmodell wird pro Patient von jedem LE und somit dezentral erstellt, verwaltet und begutachtet [313]. Sich darin befindende Daten bilden die Behandlungshistorie des Patienten für die jeweilige Einrichtung in deren Primärsystem ab. Der LE hat somit alleinigen Lese- und Schreibzugriff (read/wright). International wird diese Art von Aktenkonzept auch als Electronic Medical Record (EMR) bezeichnet.

Eine Erweiterung dieses Konzepts stellt die eEPA oder international auch Electronic Health Record (EHR) dar [313]. Hiermit soll ein Transfer der Daten von der eAA aus den Primärsystemen der Gesundheitsorganisationen heraus in andere Aktensysteme veranlasst werden können und somit einrichtungsübergreifend nutz- und erweiterbar gemacht werden. Die Übertragung der darin gesammelten Daten soll in Deutschland gemeinsam mit dem Patienten über die TI erfolgen, welche in Kombination mit der eGK und dem eHA als Datenübertragungsinfrastruktur zwischen LE im Gesundheitswesen und gesetzlichen Krankenversicherern dient. Die eGK selbst muss nach Abs. 3, § 291a SGB V sowohl für die lokale Speicherung von Notfalldaten selbst als auch den Zugang zu Anwendungen der TI, wie zum Beispiel die eEPA unterstützen. Dabei erhalten lediglich die in Abs. 4, § 291a SGB V beschriebenen Angehörigen von Heilberufen mit einem eHA in Kombination mit der eGK des Versicherten in den jeweiligen Heilanstalten vor Ort Zugriff auf die Anwendungen der TI über einen TI-Konnektor. Lediglich auf Notfalldaten können Versicherte selbst oder Angehörige anderer Heilberufe lesend zugreifen.

Die eFA stellt einen Sonderfall dar, der vor allem in Deutschland diskutiert wird [313]. Hierbei handelt es sich um ein Aktenkonzept, womit krankheitsfallspezifische Daten durch LE gesammelt und für den Austausch zwischen mehreren LE, die beispielsweise für eine fallspezifische Nachsorge in Form von Behandlungs- und Rehabilitationsprozessen zusammenarbeiten, angedacht ist.

Die eGA, international auch als Personal Health Record (PHR) bekannt, ist ein durch den Patienten

moderiertes Modell einer Gesundheitsakte zur elektronischen Übermittlung und Speicherung von Gesundheitsdaten. Das klassische Modell beinhaltet vom Patienten selbst erhobene Daten. Diese Dienstleistungen können durch den Versicherten von Dritten wahrgenommen und nach § 68 SGB V durch die gesetzlichen Krankenkassen finanziell unterstützt werden. Neben privatwirtschaftlichen Anbietern, bieten auch die IT-Dienstleister der deutschen Krankenkassen solche Dienste für ihre Versicherten an. Je nach Anbieter findet eine zentrale oder dezentrale Datenhaltung und davon abhängig ein Daten- oder Zugriffsberechtigungstransfer statt. Als Erweiterung wird hierfür die Persönliche einrichtungsübergreifende Elektronische Patientenakte (PEPA) definiert, welche die Elemente aus der eGA und eEPA verbindet und somit eine patienten-moderierte, sektorenübergreifende und interoperable Aktenlösung bietet. Aus dem, im Jahre 2019 verabschiedeten, TSVG geht hervor, dass ab dem Jahr 2021 gesetzliche Krankenkassen in Deutschland den Versicherungsnehmern eine Patientenakte zur freiwilligen Nutzung anbieten müssen (vgl. Unterunterabschnitt 2.3.1). Dieses Konzept wird, den zuvor beschriebenen gesetzlichen Neuerungen folgend, zusammengefasst als ePA nach § 341 SGB V bezeichnet. In dieser ePA sollen Befunde, Diagnosen, Therapiemaßnahmen, Behandlungsberichte, Notfalldatensatz, elektronische Medikationspläne sowie elektronische Arztbriefe und Impfungen fall- und einrichtungsübergreifend dokumentiert werden, was die Grundlage für den Übergang in das zuvor beschriebene Konzept der PEPA schafft, jedoch parallel zur eigentlichen eEPA für die Kommunikation zwischen LE betrieben werden soll.

<b>Akte</b>	<b>eAA</b>	<b>eEPA</b>	<b>eGA</b>	<b>eFA</b>
<b>Grundlage</b>	§ 630f BGB	§ 291a SGB V	§ 68 SGB V	§ 67 SGB V
<b>Moderator</b>	LE	LE	Patient	LE
<b>Interoperabilität</b>	einrichtungs- bezogen	einrichtungs- übergreifend	einrichtungs- übergreifend	einrichtungs- übergreifend
<b>Dauer</b>	10 Jahre nach Abschluss der Behandlung	lebenslang	lebenslang	fallspezifisch
<b>Datenhaltung</b>	dezentral	zentral	zentral/dezentral	dezentral
<b>Transfer</b>	Daten	Berechtigung	Daten/ Berechtigung	Daten
<b>Zugriff</b>	LE (r/w)	LE (r/w) / Patient (r)	Patient (r/w) / LE (r/w)	LE (r/w)

Tabelle 2.3: Übersicht über deutsche Aktenlösungen für eine elektronische Gesundheitsdatenverwaltung in Anlehnung an [159, 191, 313].

## 2.4.2 Digitalisierung im Gesundheitswesen

Als grundlegender Aspekt bei der Entwicklung einer digitalen Gesundheitsanwendung gilt die Phase der Digitalisierung. Wolf und Strohschen verstehen unter Digitalisierung, „wenn analoge Leistungserbringung durch Leistungserbringung in einem digitalen, computerhandhabbaren Modell ganz oder teilweise ersetzt wird“ [427]. Als Teil des Prozesses der Digitalisierung bezeichnen Bengler und Schmauder „sowohl die Überführung von Informationen von einer analogen in eine digitale Speicherung als auch den Prozess, der durch die Einführung digitaler Technologien beziehungsweise der darauf aufbauenden Anwendungssysteme hervorgerufenen Veränderungen. Die Umwandlung von analogen Daten (Text, Bild, Ton) in digitale (diskrete/abgestufte) Daten kann dabei als die Digitalisierung im engeren Sinne gesehen

werden“ [51, S.1]. Diesen Definitionen folgend, sind in diesem Abschnitt technische Begrifflichkeiten und Konzepte beschrieben, die für das Verständnis der Inhalte dieser Arbeit im Kontext der Digitalisierung von Gesundheitsanwendungen vorausgesetzt werden.

### **Technologische Konzepte für eine digitale und vernetzte medizinische Versorgung**

Seit Mitte Oktober 2020 können in Deutschland Ärzte sogenannte DiGAs auf Rezept verschreiben und über die GKV abrechnen [78, S. 4] (vgl. Unterunterabschnitt 2.3.1). Eine DiGA basiert auf Technologien, die nachfolgend skizziert und in das Gesamtkonzept des digitalen Gesundheitswesens eingeordnet werden.

**E-Health und M-Health:** Die Digitalisierung ist in allen Lebensbereichen anzutreffen, im Gesundheitssektor spricht man von Electronic Health (eHealth) oder E-Health, wenn elektronische Geräte und Applikationen zum Einsatz kommen. Diese befassen sich einerseits mit unterschiedlichen Bereichen aus der Primär-, Sekundär- und Tertiärversorgung als Bestandteil des täglichen Lebens von Patienten, tangieren aber andererseits auch Bereiche des Sozialstaats, Versicherungs- und Gesundheitswesens [222, S. 361][161, S. 6]. Damit umfasst eHealth den „gesamten Einsatz elektronischer Geräte bei der medizinischen Versorgung sowie bei gesundheitsnahen Dienstleistungen“ [25, S. 51]. Für Mediziner soll deren Praxisalltag erleichtert werden, indem gewisse Vorgänge und Verwaltungsprozesse schneller abgewickelt werden, um somit die gesamte Versorgungskette unter dem Einsatz von technischen Hilfsmitteln zu verbessern [140, S. 1][161, S. 1][307, S. 4]. M-Health oder Mobile Health (mHealth) lässt sich als Teil des eHealth betrachten, da hierbei lediglich mobile Endgeräte zum Einsatz kommen [161, S. 1]. Diese Geräte sind Bestandteil des Alltags einer überwiegenden Mehrheit geworden [24, S. 417], wodurch der Zugang zu Gesundheitsdienstleistungen niederschwellig und insbesondere bei Personen mit chronischen Erkrankungen, die lange Zeit medizinisch versorgt werden, einen Beitrag zu einer verbesserten Versorgung leisten können [226, S. 1]. Hierbei wird häufig nicht auf neuartige Versorgungskonzepte gesetzt, sondern bestehende auf digitale Medien übertragen [241, S. 101].

**Mobile, Web und Cloud Computing:** Spätestens seit der Vorstellung des *iPhones* im Jahre 2007 ist die Verbreitung von mobilen Endgeräten und somit dem Konzept der mobilen Applikation (App) fortgeschritten. Grundlegend sind diese miniaturisierte Personal Computer (PC)s mit berührungsempfindlichen Bildschirmen, die als Eingabegeräte für den Anwender dienen. Für die von den Herstellern abhängigen Betriebssysteme und Plattformen existieren Rahmenwerke mit denen Anwendungen für das entsprechende Ökosystem nativ entwickelt werden können. Neben diesen nativen Softwareanwendungen, die spezifisch für verschiedene Geräte und deren Betriebssysteme entwickelt werden, existieren auch auf Webstandards basierende Anwendungen. Diese, durch das World Wide Web (WWW) hervorgebrachten und darüber erreichbaren Webanwendungen, funktionieren durch den Ressourcen-Abruf und die -Darstellung in Browsern, welche deren Ausführungsumgebung darstellen. Mittels Uniform Resource Identifier (URI) werden Zeichenfolgen definiert, die diese Ressourcen eindeutig identifizieren und demnach einer vordefinierten Menge von Syntaxregeln folgen [54]. Das *Semantic Web* baut auf dem Konzept des WWW auf, indem das Ziel verfolgt wird, die Beschreibung von Daten als Ressourcen besonders im Hinblick auf die Interoperabilität und Integration von Informationen und Diensten zu gestalten [190]. Diese Beschreibung erfolgt durch gerichtete Graphen, welche eine Zusammensetzung dezentral gespeicherter Informationen ermöglichen [190, S. 36 ff.]. Dies fördert die semantische Interoperabilität, indem Beziehungen zwischen zuvor

nicht zusammenhängenden und unstrukturiert vorliegenden Datenelementen etabliert werden [190, 376]. Damit lassen sich gemeinsame Schemata definieren, die formell das Konzept hinter bestimmten Anwendungsdomänen beschreiben, welche als Ontologien bezeichnet werden und diese Definitionen enthalten sowie untereinander durch übergeordnete Ontologien verknüpft werden können [190]. Diese Beziehungen werden durch sogenannte Tripel, die sich aus Subjekt, Prädikat und Objekt zusammensetzen, definiert. Im Kontext der Graphen-basierten Darstellung ist das Subjekt der ausgehende Knoten, das Prädikat die Kante und das Objekt der Endknoten oder ein Literalwert [376]. Durch diese sogenannten Tripel werden die vorliegenden Daten zusammenhängend darstellbar und mittels wohldefinierten sowie verbreiteten Definitionen für Prädikate und Objekte um ein universales Verständnis für die Beziehungen zwischen Knoten erweitert. Hiermit lassen sich beispielsweise Webseiten durch maschinenlesbare Sprachen wie Resource Description Framework (RDF) oder Web Ontology Language (OWL) um inhaltliche Zusatzinformationen durch Metadaten erweitern [56]. Hierfür wird durch RDF die Struktur des Webs erweitert indem es URIs verwendet, um die Beziehung zwischen den Objekten sowie die beiden Enden der Verknüpfung zu benennen [190, S. 36 ff.]. RDF gilt als Standardformat für die Beschreibung von Metadaten im Web, es existieren Werkzeuge, die das Modellieren und Zusammenführen von Daten und Schemata damit erleichtern [190, S. 36 ff.][251]. Die Möglichkeit einer Extensible Markup Language (XML)-basierten Schreibweise und verschiedenen syntaktischen Weiterentwicklungen wie Turtle, die Notation 3 (N3)-Notation oder N-Triples, ermöglichen eine menschenlesbare Beschreibung der RDF-Dokumente [190, S. 40 ff.]. Mittels Resource Description Framework Schema (RDFS) kann der Entwurf von in RDF kodierten Vokabularen unterstützt werden, insbesondere durch die Verwendung vordefinierter Ressourcen, die zur Modellierung von Konzepten, Eigenschaften und Rollen genutzt werden können [77]. Dennoch lässt sich damit Wissen nicht hinreichend genau modellieren, da neben der Ausdrucksstärke einer Sprache auch eine effiziente Umsetzung logischen Schlussfolgerns und damit die Ableitung impliziten Wissens als wichtig erachtet wird [190, S. 125 ff.]. Mit OWL existiert eine auf der Prädikatenlogik basierende Ontologiesprache, die diesen Umstand berücksichtigt. OWL2 ist eine Erweiterung der OWL [415]. Diese bietet die Möglichkeit über Klassen, Eigenschaften, Personen und Datenwerte Ontologien zu erstellen, die kompatibel mit RDF-kodierten Informationen sind und sich gemeinsam verwenden lassen. Um einen Mehrwert aus dieser Art der Darstellung und Verknüpfung von Daten zu ziehen geben die Linked-Data-Prinzipien vor, wie Daten mit RDF(S) zu veröffentlichen sind. Diese definiert Tim Berners-Lee wie folgt [55]:

1. Benutzung von URIs um Konzepte (Dinge, Personen, Orte) zu benennen.
2. Benutzung von Hypertext Transfer Protocol (HTTP) URIs, damit Menschen und Maschinen Dinge nachschlagen können.
3. Nach Aufruf dieser URIs, relevante Daten unter Benutzung standardisierter Technologien wie RDF oder SPARQL Protocol and RDF Query Language (SPARQL) bereitstellen.
4. Hinzufügen von Links zu diesen Daten, um auf andere URIs zugreifen zu können, sodass man darüber weitere Objekte (Personen, Orte) entdecken kann.

Unter Einhaltung dieser Grundprinzipien lassen sich Daten aus unterschiedlichen Quellen in Beziehung setzen und integrieren. Diese sind über das Web verfü- und verknüpfbar, was eine Ausführung von Diensten notwendig macht, welche skalierbar und verfügbar sind. Dieser Bedarf ist die Motivation hinter der Entwicklung von Cloud Computing Technologien [208]. Dabei werden physische Infrastrukturen, welche, in Form von Rechenzentren, zur Speicherung und Vernetzung der Anwendungen mit diesbezüglichen Verwaltungsprozessen gekoppelt sind, in virtuelle Teile zerlegt und deren Nutzung kommerziell

angeboten [208]. Diesem Prinzip liegt das Konzept der Virtualisierung zugrunde, welches sowohl auf Betriebssystemebene als auch auf Applikationsebene erfolgen kann. Für die erste Ebene werden sogenannte Images als Abbilder von Betriebssystemen in isoliert voneinander betriebenen virtuellen Maschinen ausgeführt, welche wiederum verfügbare physische Ressourcen zugeteilt bekommen, um diese möglichst effizient auf Betriebssystemebene zu nutzen [368]. Die Bereitstellung von isolierten Anwendungen kann auch durch sogenannte Containervirtualisierung stattfinden [285, 368]. Hierbei kommt das Konzept der Kapselung zum Einsatz, welches Container als eine Softwareeinheit beschreibt, die eine ausführbare Anwendung beinhaltet [285, 368]. Dazu werden sowohl der Programmcode einer Anwendung als auch dessen Abhängigkeiten zu weiteren Programmbibliotheken gebündelt, um eine Softwareeinheit funktionsfähig auf ein anderes System portieren zu können [285, 368]. Die darin gebündelten Informationen stellen ein Abbild der Anwendung dar, welches weiterentwickelt und verteilt werden kann [285, 368]. Damit kann eine Anwendung als Container in einem Zielsystem durch den Einsatz einer sogenannten Engine ausgeführt werden und zeigt, unabhängig vom zugrundeliegende System, dasselbe Verhalten [285, 368]. Diese Fähigkeit besteht aufgrund der Isolation aller für die Anwendung benötigter Informationen vom umgebendem System. Mehrere Container können gleichzeitig ausgeführt werden und den Systemkern teilen [285, 368].

**Distributed Ledger Technology:** Distributed Ledger Technology (DLT) ist der Überbegriff für technische Kommunikationsprotokolle, welche eine konsistente und manipulationssichere, dezentral verwaltete sowie über mehrere Teilnehmer verteilte Transaktionsdatenhaltung ermöglichen [129]. Wie der Name schon sagt, besteht der Kern der DLT aus einem Hauptbuch (Ledger), das über alle Teilnehmer (Nodes) eines Peer-to-Peer-Netztes verteilt ist [129]. Ein Ledger ist eine Datenstruktur, in der Transaktionen gespeichert werden können. DLTs ermöglichen eine dezentrale Verwaltung des Ledgers, das heißt durch alle Knoten des Netzwerks und nicht zentral durch eine möglicherweise vertrauensunwürdige Instanz. Daher verfügt jeder Knoten über eine identische Kopie des Ledgers. Wird eine neue Transaktion in das Ledger eingetragen, prüfen die Knoten zunächst die Echtheit der Transaktion, um mit Hilfe eines Konsensmechanismus einen Konsens über die wahre Version zu erzielen. Sobald ein Konsens erreicht ist, übernimmt jeder Teilnehmer die neue Transaktion in seine Kopie des Hauptbuchs. Zur Identifizierung eines Knotens im Netz und zur Verknüpfung mit Transaktionen besitzt jeder Teilnehmer einen öffentlichen und einen privaten kryptografischen Schlüssel (asymmetrische Verschlüsselung). Mit dem privaten Schlüssel kann ein Teilnehmer eine Signatur erzeugen, die ihn direkt identifiziert. Die Signatur kann dann von anderen Teilnehmern mit Hilfe des öffentlichen Schlüssels überprüft werden. Der bekannteste und am weitesten verbreitete Vertreter von DLT ist die sogenannte Blockchain [363]. Blockketten verknüpfen die Transaktionshistorien (Blöcke) in einer verketteten Liste, deren Elemente einen Fingerabdruck (Hash) ihres Vorgängers enthalten, um ein unveränderliches Ledger zu erstellen (vgl. Abbildung 2.18). In den letzten

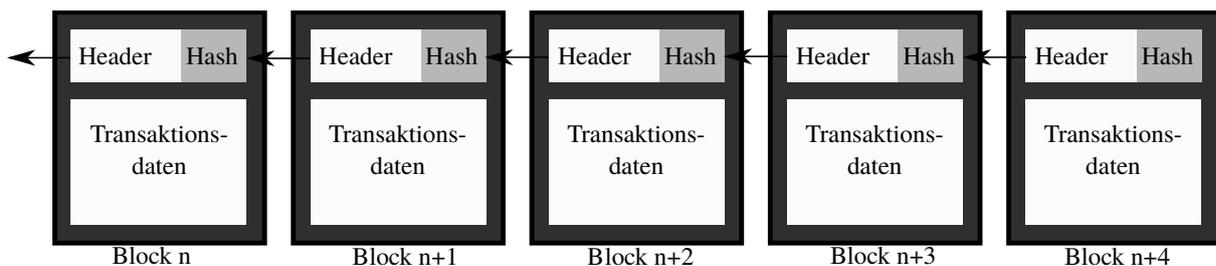


Abbildung 2.18: Konzeptioneller Aufbau einer Blockchain.

Jahren haben sich verschiedene Arten von DLT entwickelt [129]. Neben der Existenz öffentlicher ver-

teilter Ledger („public Ledger“) werden in zugangsbeschränkten Netzwerken, sogenannte „permissioned Ledger“ verwendet, die eine Authentifizierung und Autorisierung erfordern [129]. Beispielsweise bietet *Hyperledger Fabric*<sup>1</sup> (HLF) ein hochgradig modulares und konfigurierbares permissioned Blockchain Framework [196]. Es bietet auch die Möglichkeit, Smart Contracts (*chaincode*) auf der Blockchain auszuführen, was Zugriffs- und Manipulationsmöglichkeiten bezüglich der Ledger-Daten bei der Ausführung von Transaktionen erlaubt. Es ist für den Einsatz in Unternehmen konzipiert. Sein Konsensmechanismus kann auch so konfiguriert werden, dass er sich an geschäftsbezogene Anwendungsfälle anpasst [196].

Derzeit werden in der Forschung auch Kombinationen verschiedener DLT-Typen in Softwarearchitekturen diskutiert [202]. Sogenannte Sidechains ermöglichen beispielsweise eine modulare Architektur und bestehen aus einer gemeinsamen öffentlichen oder privaten Blockchain und einer Reihe privater Teilnetzwerke mit permissioned Blockchains [27, 129]. Globale Zugriffsanfragen und Integritätsprüfungen werden über die gemeinsame Blockchain verwaltet. Daher wird der Informationsaustausch zwischen Gruppen von Partnern durch die Nutzung dieser Teilnetze für lokale Transaktionen erreicht und erfordert nur einen Konsens zwischen den mit den Partnern verbundenen Knoten. Diese Konstrukte können dazu beitragen ein gewisses Maß an Privatsphäre zu wahren indem sensible Daten vor Konkurrenten verborgen werden, aber dennoch Daten mit einem ausgewählten Kreis in einem gewünschten Umfang zu teilen [129].

Hinsichtlich der Gestaltung von Blockchain-basierten Anwendungen haben Xu et al. weitere Muster identifiziert [435]. Ein Muster wird als Orakel bezeichnet, da es externe Informationen (Off-Chain-Daten) für die Blockchain bereitstellt, indem es diese mithilfe von Transaktionen in On-Chain-Daten umwandelt. Um eine Interaktion von Off-Chain-Komponenten mit den auf der Blockchain gespeicherten Daten zu ermöglichen, wird das Reverse-Orakel-Muster verwendet. Damit kann von außerhalb des Blockchain-Netzwerks anhand unveränderlicher On-Chain-Daten geprüft werden, ob erforderliche Bedingungen erfüllt sind. Als weiteres Muster wird die Verschlüsselung von On-Chain-Daten verwendet, um sensible Daten in der Blockchain für bestimmte Teilnehmer durch Verschlüsselung zu verbergen.

### **Datenaggregation, -integration und Interoperabilität**

Ab dem Jahr 2021 soll mit der ePA durch die Aggregation der Gesundheitsdaten aus unterschiedlichen Quellen schrittweise die Erstellung eines gesamtheitlichen Gesundheitsprofils ermöglicht werden (vgl. Unterabschnitt 2.4.1). Hierfür ist ein vernetztes Gesundheitswesen unabdingbar, welches auf Konzepten zur Aggregation, Integration und Interoperabilität zwischen Datenquellen aufbaut. Die darin entstehenden Daten stammen aus unterschiedlichen Quellen, die von verschiedenen Akteuren wie beispielsweise Patienten, Ärzten, Klinik-, Pflegepersonal oder automatisiert durch medizinische Geräte erhoben und verwaltet werden. Dafür sind in den letzten Jahren, je nach Art der Daten und deren Anwendungsdomäne, verschiedene Formate und Standards entstanden (vgl. Unterunterabschnitt 2.4.2). Unter dem Begriff der Datenaggregation lässt sich die Ermöglichung eines zentralen Zugriffspunkts auf eine Sammlung von Daten verstehen [324]. Die Integration von Daten impliziert den Schritt der Aggregation, erweitert durch einen zusätzlichen Schritt mittels Abbildung der Daten in ein heterogenes Format zur Sicherstellung der Wiederverwendbarkeit dieser [324]. Ein Zielsystem kann diese integrierten Daten so verwenden, als würden sie aus einem einzelnen Ausgangssystem stammen [324]. Als Grundlage dieser Schritte wird die Interoperabilität definiert, welche über Systemgrenzen hinweg besteht [185, 324]. Dabei wird zwischen verschiedenen Ebenen der Interoperabilität unterschieden, um die Fähigkeit zum Austausch, der Integration sowie der Interpretation und koordinierten Nutzung von Daten zwischen verschiedenen Systemen zu differenzieren [185, 324]. Nach der Healthcare Information and Management Systems Society (HIMSS)

---

<sup>1</sup> <https://www.hyperledger.org/projects/fabric>, zuletzt zugegriffen am 21.02.2022.

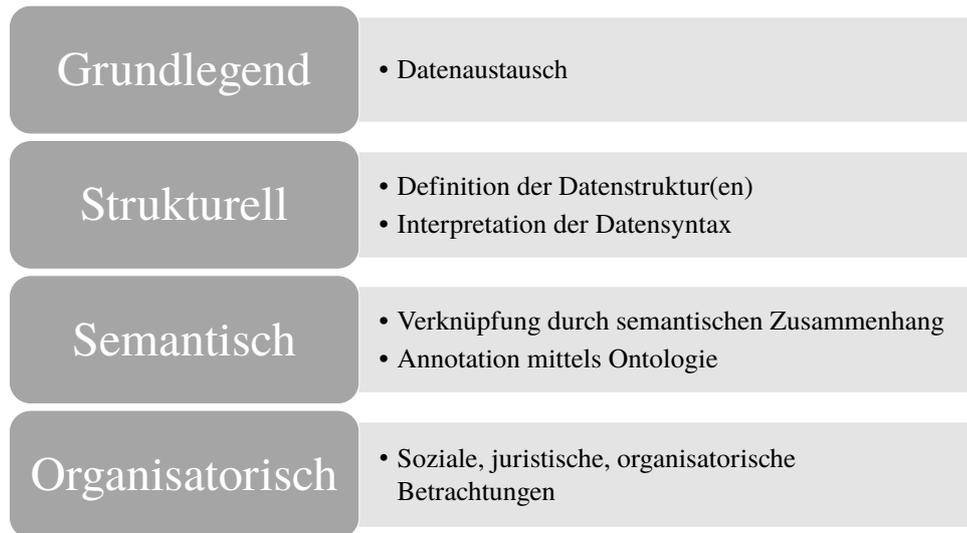


Abbildung 2.19: Ebenen der Interoperabilität nach der HIMSS [185, 324].

lassen sich nachfolgende vier Ebenen der Interoperabilität unterscheiden (vgl. Abbildung 2.19). Als unterste Ebene wird die grundlegende Interoperabilität definiert, welche durch Vernetzung zwischen Systemen erreicht ist, sobald Daten sicher von einem Ausgangssystem in ein Zielsystem übertragen, empfangen und gespeichert werden können. Auf dieser Ebene ist nicht sichergestellt, dass das Zielsystem die empfangenen Daten verarbeiten kann [69, 185]. Daher folgt auf der zweiten Ebene, die Gewährleistung struktureller Interoperabilität, welche durch die Definition von einheitlichen Datenstrukturen und Datensyntax die Interpretation der ausgetauschten Daten ermöglicht [185]. Um auf semantischer Ebene Interoperabilität zu erreichen, wird die Nutzung von standardisierten Datenkodierungssystemen und Datenmodellen zur Definition einer einheitlichen Bedeutung von Daten gesehen [185]. Damit kann ein Zielsystem die Daten von Ausgangssystemen so verarbeiten als wären sie im Zielsystem selbst erhoben worden [69]. Auf organisatorischer Ebene findet Interoperabilität durch legale, soziale und organisatorische Betrachtungen wie Richtlinien statt, wobei das Ziel einer nahtlosen, sicheren und schnellen Kommunikation zwischen den Systemen im Vordergrund steht, um für den Anwender eine vertrauenswürdige Basis für das Teilen von Daten zu ermöglichen [185].

Demnach wird für eine Datenaggregation das Erreichen der grundlegenden Ebene von Interoperabilität vorausgesetzt, die Datenintegration setzt mindestens die Ebene der strukturellen Interoperabilität voraus.

### Datenstandards in der Medizin

Die zuvor beschriebenen Ebenen der Interoperabilität zwischen Systemen, Organisationen und Geräten werden medizinische Datenstandards benötigt, die den Inhalt und Austausch von medizinischen Daten ermöglichen. Diese werden von Organisationen entwickelt, wie beispielsweise die Health Level Seven (HL7), WHO oder SNOMED CT (vgl. Unterabschnitt 2.2.2) und durch die Integrating the Healthcare Enterprise (IHE) in sogenannten Profilen gebündelt, um konkrete Anwendungsfälle zu lösen [324, 346]. Je nach Art der Daten und deren Anwendungsgebiet existieren verschiedene Standards [164]. Als grundlegender Rahmen für das Speichern patientenbezogener Gesundheitsdaten wird für das deutsche Gesundheitswesen eine angepasste Version des etablierten Industriestandards IHE Cross-Enterprise Document Sharing (XDS) verwendet [156]. Dabei schreibt die Version XDS.b nicht den Inhalt der Doku-

mente, sondern lediglich deren Speicherung innerhalb eines auf XML basierenden Archivs vor [123]. Für die Kollaboration zwischen Gesundheitsdienstleistern können über ein gemeinsames Archiv Dokumente ausgetauscht werden, indem diese sich zu einer „klinischen Affinitäts-Domäne“ zusammenschließen [123]. Um dieses Archiv durchsuchen zu können, einigen sich die beteiligten Einrichtungen auf einheitliche Metadaten in Form von Datenannotationen und Identifikatoren für Patienten [123]. Im deutschen Gesundheitssystem werden Standards der XML Data Transfer (xDT)-Gruppe für den Datenaustausch zwischen Informationssystemen bei niedergelassenen Ärzten verwendet. xDT definiert hierfür den Austausch von ärztlichen Dokumenten ebenfalls im XML-Format [224]. Für den Austausch klinischer Bilddaten wird Digital Imaging and Communications in Medicine (DICOM) als ein weit verbreiteter Standard gesehen [115, 164]. Mit FHIR<sup>2</sup> entstand zusätzlich ein, von HL7 entwickelter, webbasierter Standard für die Interoperabilität zwischen Systemen und den Austausch medizinischer Daten. Hierbei wird auf ein modulares Konzept gesetzt, welches aus Komponenten, den sogenannten Ressourcen („Resources“) aufgebaut ist, womit Aspekte des Gesundheitssystems abgebildet werden. Als Schlüsselkomponenten lassen sich Patienten (FHIR Patient), Beobachtungen (FHIR Observation) und klinische Zustände (FHIR Condition) identifizieren. Diese Ressourcen können in den Formaten JavaScript Object Notation (JSON), XML oder RDF dargestellt werden [69, 184]. Generell sind Ressourcen neben deren Inhalt, durch ihren Typ und eine eindeutige Identifikation gekennzeichnet. Gemäß der Spezifikation lassen sich Felder innerhalb der Ressource durch einzelne Einträge oder Listen von Einträgen definieren. Für einige Einträge sind sogenannte „ValueSets“ vorgeschrieben, welche eine Menge an erlaubten Werten für den Eintrag vorgeben und welches Format beispielsweise für Zeit und Datum innerhalb der Ressourcen zu verwenden ist. Die Nutzung von ValueSets ermöglicht durch die eindeutige Interpretation der Daten bereits einen beschränkten Grad der zuvor beschriebenen semantischen Interoperabilität [69, 184]. Erweitert werden kann diese Art der Interoperabilität durch die Nutzung von Kodierungssystemen innerhalb der FHIR-Ressourcen als textuelle Repräsentation zum Beispiel von Logical Observation Identifiers Names and Codes (LOINC), SNOMED CT oder ICD [69] (vgl. Unterabschnitt 2.2.2). Durch Referenzen können die FHIR Komponenten untereinander verknüpft werden, wodurch im Gegensatz zu bisher verwendeten klinischen Dokumenten, Dokumente nicht mehr aufwändig nach dem gewünschten Inhalt durchsucht werden müssen, sondern nur die jeweilige Ressource angefordert werden muss [69]. Der Abruf von FHIR-Ressourcen kann über eine webbasierte Representational State Transfer (REST) Application Programming Interface (API) erfolgen.

### **Mensch-Technik-Interaktion**

Um die digitalen Medien im Kontext medizinischer Anwendungen nutzenstiftend einzusetzen, finden Interaktionen zwischen Menschen und Maschinen statt, welche zum bidirektionalen Austausch von Daten notwendig sind. Nachfolgend werden die für diese Arbeit grundlegenden Konzepte und Begrifflichkeiten dargestellt.

**Gebrauchstauglichkeit und Visualisierung:** Für eine erfolgreiche Produktentwicklung bestehen Anforderungen an einen gebrauchstauglichkeitsorientierten Entwicklungsprozess, welche in Normen als Orientierungsgrundlage beschrieben sind (vgl. Unterunterabschnitt 2.3.1). Unter dem Begriff der Gebrauchstauglichkeit versteht die Normenfamilie ISO 9241 „das Ausmaß in dem ein Produkt, System oder Dienst durch bestimmte Benutzer in einem bestimmten Anwendungskontext genutzt werden kann, um

---

<sup>2</sup> <http://hl7.org/fhir/directory.html>, zuletzt zugegriffen am 15.04.2022.

bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen“ [116, 118, 119]. Bei Medizinprodukten wird unter Gebrauchstauglichkeit beziehungsweise Usability keine direkte Qualitätseigenschaft verstanden, sondern deren Abhängigkeit zu Charakteristiken der Nutzer, der Nutzungsumgebung und den definierten Zielen [180]. Deshalb empfiehlt es sich konkrete Gestaltungsrichtlinien für den Entwurf von Benutzerschnittstellen zu definieren. Diese beziehen sich bei einer SaMD auf die grafische Visualisierung der Interaktionselemente. Mit Hilfe von sogenannten Mockups können Ideen und Konzepte visualisiert, getestet und evaluiert werden. Ein Mockup ist ein digital gestalteter Entwurf einer Website oder einer mobilen Applikation ohne Funktionalität und wird oftmals auf Grundlage von sogenannten Wireframes erstellt [348]. Diese repräsentieren wiederum einen ersten abstrahierten Entwurf einer Anwendung und konzentrieren sich insbesondere auf den Aufbau, die Zusammenhänge zwischen einzelnen Elementen und die Benutzerführung. Damit lässt sich eine grafische Benutzerschnittstelle (GUI) unter Einbindung des Feedbacks potenzieller Nutzergruppen deren Bedürfnissen an das Design der UI und User Experience (UX) entsprechend iterativ weiterentwickeln.

**Gamification:** Durch das Konzept der Gamification wird durch spielerische Elemente versucht positive Muster bei Anwendern während der Nutzung von Diensten oder Produkten zu erzeugen [175]. Die spielerischen Erfahrungen sollen motivierend wirken, um beispielsweise Benutzeraktivität in verschiedenen Kontexten zu fördern [175]. Durch Nutzung dieses Paradigmas werden unter dem Stichwort „Serious Games“ Spiele verstanden, welche den ernsten Charakter von Problemen und des Denkens mit dem freiheitlich-experimentellen Charakter des Spielens verbinden [20]. In Bezug auf ältere Menschen beschreiben Gupta et al. das Potential und einen weitgehend unerforschten Markt, der sich an hochgradig beeinträchtigte ältere Menschen richtet, um diese Art von Spielen für Therapien oder Diagnosen zu nutzen [172].

### 2.4.3 Verfahren zur Automatisierung und Optimierung

Um Vorgänge zu automatisieren und optimieren werden Modelle benötigt, welche diese abbilden. Hierfür ist Wissen notwendig, das auf Grundlage von Datenanalysen generiert und in die Modellierung integriert wird. Durch welchen Prozess dieses Wissen generiert wird ist in Unterabschnitt 2.1.3 beschrieben. In diesem Fall beschränkt sich dieser auf Mustererkennung in Datenbeständen (Data Mining), wodurch die zuvor für Analysen entsprechend aufbereiteten Daten durch Anwendung deskriptiver und prädiktiver Modelle unter Berücksichtigung praktischer Berechnungs- und Effizienzgrenzen analysiert werden, um verschiedene Modelle und Muster aus den Daten hervorzubringen [408]. Neben der Analyse von Daten mittels statistischer Verfahren (vgl. Unterabschnitt 2.1.4) spielen Algorithmen aus dem Bereich des MLs auch eine zunehmend wichtige Rolle im Gesundheitswesen [209]. Da sich diesbezügliches Wissen größtenteils aus englischsprachiger Literatur ableitet werden nachfolgend die eingedeutschten englischen Begriffe teilweise synonym verwendet. Grundsätzlich wird zwischen überwachtem und unüberwachtem Lernen (Supervised und Unsupervised Learning) unterschieden [249]. Wenn Algorithmen dazu genutzt werden, prädiktive Modelle zu erstellen, handelt es sich um überwachtes Lernen [249]. Hierbei geht es darum, Beziehungen in Daten (Samples) zwischen einem Zielmerkmal (Target Feature) und den restlichen Merkmalen so zu modellieren, dass die Datenlage aufgrund der Ausprägungen dieser Merkmale generalisiert jedoch nicht über- oder unterrepräsentiert (Over-/Underfitting) wird [249]. Unüberwachte Ansätze verfolgen das Ziel, bestehende Daten möglichst aussagekräftig zu beschreiben, wodurch Zusammenhänge in den Daten erkennbar werden sollen, die bisher nicht offensichtlich sind [249].

## Unüberwachtes Lernen

Das Training eines deskriptiven Modells wird auch als „Unüberwachtes Lernen“ oder „Unsupervised Learning“ bezeichnet [250]. Ein Vertreter dieser Verfahren ist das sogenannte Clustering, welches Daten ohne Label anhand von Ähnlichkeitsmaßen in Gruppen unterteilt [250]. Dabei sind die Intraklassendistanzen zu minimieren und die Interklassendistanzen zu maximieren [177, 426]. Wie diese Gruppen erstellt werden, kann algorithmisch unterschiedlich abgebildet werden, die Interpretation der Ergebnisse obliegt jedoch dem menschlichen Entscheider [250].

**k-means:** Ein Beispiel bildet der k-means-Algorithmus, womit durch  $k$  die Anzahl der zu findenden Cluster definiert wird [250, 426]. Danach werden die dadurch entstehenden Clusterzentren initial im Merkmalsraum frei verteilt [250, 426]. Die Datenpunkte werden dem nächstgelegenen Zentrum zugeordnet [250, 426]. Danach werden die Clusterzentren iterativ als Mittelwert der Datenpunkte so lange neu gesetzt, bis danach kein Datenpunkt mehr einem anderen Cluster zugeordnet wird [250, 426]. Die Abhängigkeit zum initial gewählten Clusterzentrum kann dazu führen, dass kein optimales Clustering gefunden wird, mit dem k-means++ existiert allerdings eine mögliche Lösung für dieses Problem [426].

**Frequent Pattern Mining (FPM):** Ein weiteres Verfahren um häufige Muster aus Datenbeständen zu extrahieren stellt das sogenannte FPM dar (vgl. Unterabschnitt 2.1.3). Hierbei besteht das Ziel besonders häufig vorkommende Kombinationen von sogenannten Items als Itemsets zu identifizieren [183]. Demnach ist ein identifiziertes Frequent Pattern (fp) ein Itemset mit  $k > 1$  Items. Die Häufigkeit dieses Itemsets für einen Datensatz wird wie in Gleichung 2.14 als Support Count (SC) bezeichnet [183]. Als Support wird, wie in Gleichung 2.15 dargestellt, der Anteil der Instanzen bezeichnet, der die fp enthält [183]. Der gewählte Schwellenwert *minsup* legt fest, welcher Wert überschritten werden muss, um ein

$$SC = \sigma(fp) \quad (2.14) \quad support(fp) = \frac{\sigma(fp)}{N} \quad (2.15)$$

Muster als häufig interpretieren zu können [183]. Zur Beschreibung der Beziehungen zwischen einzelnen Items innerhalb eines Itemsets über den Datensatz hinweg lassen sich Assoziationsregeln ableiten. Ein Itemset mit den Items  $A, B$  kann beispielsweise eine Assoziation  $A \Rightarrow B$  enthalten,  $A$  gilt hierbei als Antezedens und  $B$  als Konsequenz. Das Konfidenz- oder Präzisionsniveau lässt sich wie in Gleichung 2.16 beschreiben [183].

$$P(B|A) = \frac{support(A, B)}{support(A)} \quad (2.16)$$

Der Support der Antezedens wird auch als Coverage bezeichnet [183].

## Überwachtes Lernen

Das Anwendungsgebiet überwachter Verfahren lässt sich in Regressions- und Klassifikationsprobleme aufteilen [249]. Die Unterschiede ergeben sich aus den Ausprägungen der Zielmerkmale, bei Klassifikationen nehmen diese kategoriale Werte (levels) an, regressive Modelle gehen von stetigen numerischen Werten aus, die sie vorhersagen [249]. Letztere sollen Eingabe- und Zielmerkmale exakt quantifizieren [249]. Für das Training solcher Modelle werden Referenzdaten als Labels benötigt, welche eine unabhängige Bewertung der Ausprägung des Zielmerkmals beinhalten, die als Wahrheit angenommen werden können. Nachfolgend werden die Eigenschaften der in dieser Arbeit zur Anwendung gebrachten Klassifikationsmodelle kurz skizziert (vgl. Abbildung 2.20).

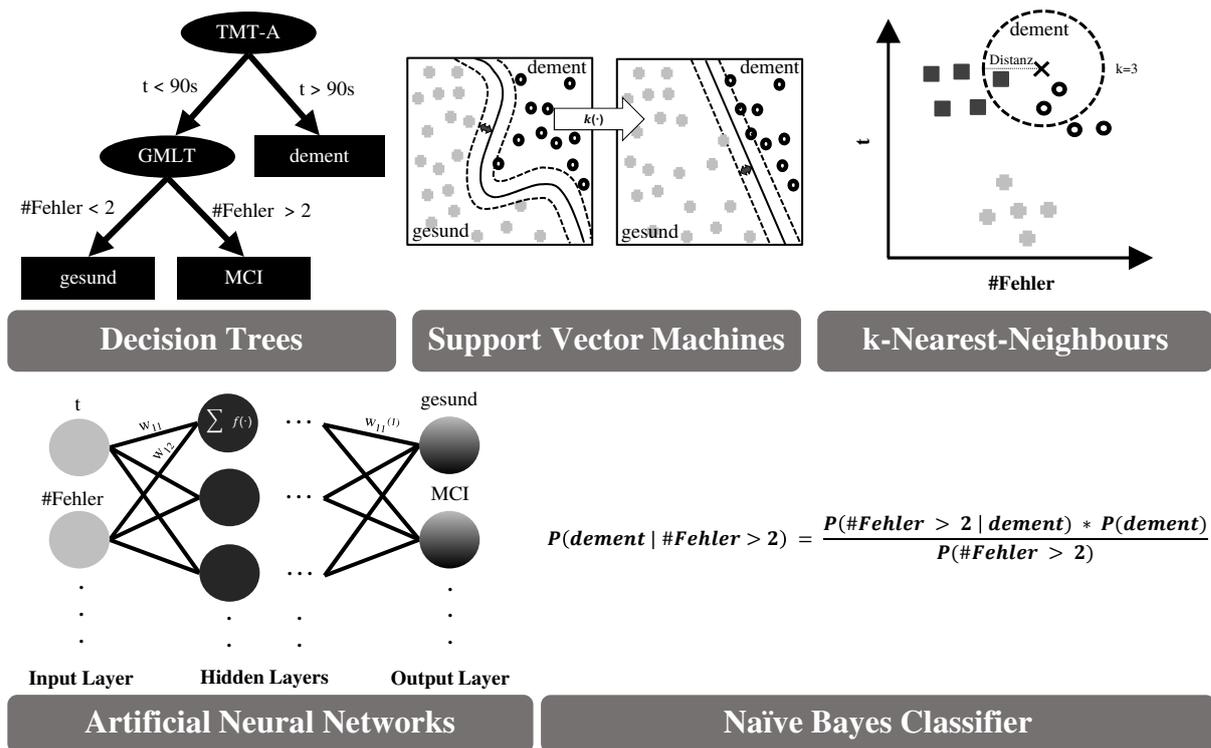


Abbildung 2.20: Schematische Darstellung der eingesetzten Klassifikationsalgorithmen.

**Decision Trees:** Entscheidungsbäume oder auch *Decision Trees* folgen der schematischen Erstellung einer hierarchischen Anordnung von Entscheidungsregeln [250]. Dabei wird ein Datensatz systematisch in mehrere Datensätze unterteilt und für jeden eine Klassifikationsentscheidung getroffen [250]. Am häufigsten in der Realität vorzufinden sind der von Breiman et al. eingeführte Classification And Regression Tree (CART) sowie der von Quinlan entwickelte C5.0-Algorithmus, welcher als Industriestandard für Entscheidungsbäume gesehen wird [75, 242, 250, 337]. In dieser Arbeit werden CARTs angewandt, die als binäre Entscheidungsbäume mittels Greedy-Algorithmus erstellt werden [296]. Der Algorithmus zur Konstruktion eines solchen Decision Trees beginnt bei der Wurzel und ermittelt mit einer Kostenfunktion, welches Attribut durch welchen Schwellwert die beste Wahl für diese darstellt [296]. Darauf aufbauend werden die zwei folgenden Entscheidungsknoten äquivalent dazu bestimmt [296]. Inwiefern ein solcher Knoten als Blatt definiert wird oder weiter aufgespalten werden muss, wird heuristisch bestimmt [296]. Kostenfunktionen sind in Abhängigkeit des Verwendungszwecks definiert, für Klassifikationsprobleme kommt, wie in Gleichung 2.17 formuliert, der Gini-Index zum Einsatz [296, S.548]. Hierin ist  $p$  die Klassenwahrscheinlichkeit und  $C$  die Anzahl der Klassen.

$$G(x) = \sum_{c=1}^C p_c(1 - p_c) = 1 - \sum_{c=1}^C p_c^2 \quad (2.17)$$

Durch die Baumstruktur entsteht auch für den menschlichen Entscheider ein nachvollziehbarer Entscheidungsprozess und aufgrund der Struktur geringe Trainings- und Klassifikationszeiten [250, 296]. Hierbei können sowohl nominale als auch numerische Merkmale mit in das Modell einbezogen werden, geringe oder große Datenmengen sowie fehlende Datenpunkte schließen ebenfalls eine mögliche Anwendung nicht aus [250]. Des Weiteren impliziert das Verfahren eine sogenannte Feature Selection, welche in Unterunterabschnitt 2.4.3 spezifiziert wird [296]. Zu den Nachteilen der CARTs werden verhältnismäßig niedrige Trefferquoten der prognostizierten Ausgangswerte gezählt, welche mit dem heuristischen

Greedy Algorithmus begründet werden [296]. Außerdem ist der Algorithmus anfällig für Instabilität, da Änderungen in den Trainingsdaten strukturelle Auswirkungen auf den Entscheidungsbaum haben [296]. Um diesen Effekt zu verringern, wird nachfolgend die Methode der Random Forest eingeführt.

**Random Forest:** Die Kombination mehrerer maschineller Lernverfahren wird unter dem Überbegriff Ensemble Learning zusammengefasst [250]. Der Random Forest (RF) ist ein Vertreter dieser Verfahren und setzt sich aus mehreren Modellen von Entscheidungsbäumen zusammen [76, 250]. Diese Prädiktoren wiederum werden jeweils mit einer zufällig ausgewählten Teilmenge der verfügbaren Merkmale trainiert und entscheiden unabhängig voneinander [76]. Über ein Abstimmungsmechanismus werden die Einzelentscheidungen zu einer übergeordneten Lösung des Entscheidungsproblems zusammengeführt [76]. Als vorteilhaft gilt dieses Verfahren im Umgang mit großen Datensätzen und einer Vielzahl an Merkmalen sowie deren gute Performanz trotz verrauschter oder fehlerhafter Daten [250]. Die Interpretierbarkeit durch menschliche Entscheider leidet aber darunter [250].

**Naïve Bayes:** Als Methode des probabilistischen Lernens gilt der Naïve Bayes Klassifikator [250]. Dieser basiert, dem Bayes-Theorem folgend (vgl. Gleichung 2.18), auf bedingten Wahrscheinlichkeiten.

$$P(Y|X) = \frac{P(Y \cap X)}{P(X)} = \frac{P(X|Y) \cdot P(Y)}{P(X)} \quad (2.18)$$

$Y$  und  $X$  sind Ereignisse.  $P(Y|X)$  ist die Posteriorwahrscheinlichkeit, dass das Ereignis  $Y$  unter der Prämisse eintritt, dass  $X$  eingetreten ist. Während  $P(X)$  die Wahrscheinlichkeit darstellt, dass  $X$  eintritt.  $P(Y)$  ist die Wahrscheinlichkeit, dass das Ereignis  $Y$  auftritt und  $P(X|Y)$  ist die Wahrscheinlichkeit, dass  $X$  eintritt, wenn das Ereignis  $Y$  eingetreten ist. Ereignisse lassen sich auch in Klassen- und Merkmalsausprägungen unterteilen, wodurch sich aus dem Bayes-Theorem der in Gleichung 2.19 dargestellte Naïve Bayes Klassifikator ableiten lässt.

$$\hat{y} = \operatorname{argmax}_y P(y) \prod_{i=1}^n P(x_i|y) \quad (2.19)$$

Hierbei stellt  $y$  eine mögliche Klasse dar sowie  $x_i$  eines der Merkmale von Eingabedaten. Für die Klassifizierung wird eine Funktion benötigt, welche das Argument des Maximums, die in diesem Fall die Klasse mit der höchsten Wahrscheinlichkeit, sowie die Wahrscheinlichkeiten aller Klassen und die bedingten Wahrscheinlichkeiten  $P(x_i|y)$  zurückgibt. Um diese zu berechnen existieren unterschiedliche Ansätze. Bei Vorliegen nominaler Attribute können diese über die relativen Häufigkeiten bestimmt werden (vgl. Gleichung 2.20).

$$P(x_i|c_j) = \frac{n_{ij}}{n} \quad (2.20)$$

Hierbei ist  $x_i$  ein Merkmal,  $c_j$  eine mögliche Klasse,  $n_{ij}$  die Anzahl der Trainingsbeispiele, in welchen die Klasse  $c_j$  und das Merkmal  $x_i$  vorhanden ist sowie  $n$  die Anzahl der Trainingsbeispiele, welche die Klasse  $c_j$  und beliebige Merkmale aufweisen. Liegen keine nominalen Werte vor, so können die Werte durch Binning in Klassen zusammengefasst werden. Ein Bin enthält dabei einen Teil des gesamten Wertebereichs, welcher in mehrere Bins so aufgeteilt wird, dass jeder vorkommende Wert einem Bin zugeordnet werden kann. Eine andere Möglichkeit für ordinale Attribute ist die Annahme einer Verteilungsfunktion über deren Dichtefunktion eine Bestimmung der Wahrscheinlichkeiten möglich wird. Für numerische Attribute kann eine Annahme über die Verteilung der Daten getroffen werden. Üblicherweise wird hier eine Normalverteilung angewandt [242]. Unter dieser Annahme wird die bedingte Wahrscheinlichkeit für eine Klasse unter Berücksichtigung der Eingabedaten durch Verwendung der (Gleichung 2.19 - Glei-

chung 2.21) bestimmt [409]. Unter der Normalverteilungsannahme lässt sich das wie folgt definieren (vgl. Gleichung 2.21).

$$P(X = x|C = c) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x-\mu^2}{2\sigma^2}} \quad (2.21)$$

Dieses (naïve) Vorgehen setzt des Weiteren die Annahmen voraus, dass alle Merkmale unabhängig voneinander und von gleicher Relevanz sind [250]. In der Realität ist diese Prämisse nur selten zu halten, dennoch liefert der Naïve Bayes Algorithmus beispielsweise bei Textverarbeitung häufig sehr gute Ergebnisse [250].

**k-Nearest-Neighbours:** Durch den k-nearest-neighbours-Algorithmus (k-NN) wird ein unbekannter Datenpunkt klassifiziert, indem die Umgebung dieses Datenpunktes betrachtet wird [250, 280]. Das  $k$  steht für die Anzahl der zu betrachtenden benachbarten Datenpunkte [250]. Die prädizierte Klassenzugehörigkeit richtet sich nach der häufigsten, in dieser Umgebung vorkommenden, Klasse (vgl. Abbildung 2.20). Für die Wahl des  $k$  werden heuristische Verfahren eingesetzt, auch für die verwendete Distanz-Funktion existieren verschiedene Optionen [250, 280], wie beispielsweise die rechenaufwändige euklidische Distanz oder ressourcenschonende Manhattan-Distanz. Damit zählen k-NNs zu den sogenannten „Lazy Learner“-Algorithmen, welche keine generalisierte Modellbildung durchführen, sondern direkt auf den zuvor gespeicherten Trainingsdaten arbeiten. Im Gegensatz zu anderen Verfahren entfallen dabei statistische Verteilungsannahmen und Trainingszeiten. Das hat einerseits den Nachteil, dass die Klassifikationszeit von der Abstandsberechnung und Anzahl von Datenpunkten abhängt, andererseits entfällt dadurch auch das Verständnis für Beziehungen zwischen den Merkmalen und Klassen [250, 280].

**Support Vector Machines:** Eine Support Vector Machine (SVM) basiert auf dem Prinzip der linearen Trennung von Daten durch sogenannte Hyperebenen [101]. Objekte werden als Vektoren in  $\mathbb{R}^d$  dargestellt und gehören einer von zwei Klassen an. Hierbei wird die optimale Hyperebene bestimmt, indem der Abstand zwischen der Ebene und den nächstgelegenen Datenpunkten (Support Vectors) maximiert wird, die Wahrscheinlichkeit, dass die Ebene beim Einfügen neuer Datenpunkte verschoben werden muss, minimiert wird sowie am besten die Trainingsdaten generalisiert. Eine Unterscheidung zwischen dem „hard margin“- und dem „soft margin“-Ansatz erlaubt bei letzterem einen toleranteren Umgang mit Ausreißern, um Daten besser abstrahieren zu können [210]. Um eine lineare Trennbarkeit zwischen den Datenpunkten zu erreichen, werden diese mittels Kernel-Funktionen  $k(\cdot)$  transformiert (vgl. Abbildung 2.20). Diese existieren in unterschiedlichen Ausprägungen wie lineare, sigmoid, polynomielle, Radial Basis Funktion (RBF) oder Gauß'sche Kernels deren Eignung vom Kontext abhängig ist und keinen generellen Entscheidungsregeln unterliegt [250]. Als Vorteile dieses Verfahrens werden in der Literatur unter anderem, der tolerante Umgang mit verrauschten Daten und gegenüber Overfitting genannt [250]. Als Nachteil wird gesehen, dass diese Verfahren für den menschlichen Entscheider intransparente Entscheidungen treffen und damit als Black-Box-Prozesse bezeichnet werden [250]. Auch die Optimierung der Parameter erfolgt durch mit Rechenaufwand verbundenes Durchprobieren verschiedener Kombinationen [250]. Um mehr als zwei Klassen voneinander zu separieren, lassen sich hybride Strategien durch Kombinationen von SVMs einsetzen [434].

**Artificial Neural Networks:** Um das Konzept der biologischen neuronalen Netzstruktur des Gehirns zu imitieren, werden bei Artificial Neural Networks (ANN) mathematische Modelle künstlicher Neuronen (Knoten) in Schichten angeordnet und über künstliche Dendriten als Eingabe- und Axone als Ausgabeparameter (Kanten) miteinander verbunden (vgl. Abbildung 2.6 und Abbildung 2.20). In die Modellbildung einzubeziehende Merkmale werden als jeweils eigenes Neuron in der sogenannten „Input Layer“ auf der linken Seite in einer eigenen Schicht abgebildet (vgl. Abbildung 2.20). Auf der rechten Seite wer-

den die Dimensionen der Zielmerkmale in der „Output Layer“ jeweils auf ein Neuron abgebildet (vgl. Abbildung 2.20). In Abhängigkeit zur Komplexität des Problems ergeben sich zur Erreichung einer optimalen Lösung zwischen „Input Layer“ und „Output Layer“ beliebig viele Dimensionen in Form von Schichten mit unterschiedlichen Anzahlen an Neuronen, die als sogenannte „Hidden Layers“ bezeichnet werden [250]. Die grundlegende Architektur von ANNs lässt sich über die Eigenschaften der Aktivierungsfunktion, Netzwerktopologie und des Trainingsalgorithmus definieren. Einzelne Neuronen werden wie in Abbildung 2.20 dargestellt aus einer Übertragungsfunktion beispielsweise  $\Sigma$  und Aktivierungsfunktion  $f(\cdot)$  zusammengesetzt, welche die Eingabewerte über die Kantengewichte  $w_{ij}$  als Netzeingabe miteinander verrechnet und auf einen Ausgabewert  $w_{ij}^{(i)}$  als Kantengewicht an die nächste Schicht weitergibt. Die Aktivierungsfunktion entscheidet anhand der zuvor bestimmten Netzeingabe, ob das Signal im Netzwerk weitergeleitet wird oder nicht [250]. Aktivierungsfunktionen wie zum Beispiel lineare, Gauß- oder Sigmoid-Funktionen gehen mit unterschiedlichen Ausgangssignalbereichen und verschiedenen Vor- und Nachteilen einher, die in Abhängigkeit zum eingesetzten Trainingsprozess stehen [250]. Für einen Trainingsprozess mit sogenannter Back-Propagation, das heißt mit der iterativen Anpassung der gelernten Kantengewichte aufgrund der schichtweisen Rückverfolgung von Fehlern, muss diese Funktion differenzierbar sein, um eine lokal optimale Lösung finden zu können [250]. Details hierzu beschreiben Rumelhart et al. [355]. Ob dieses Verfahren zum Einsatz kommen kann, hängt mit der Topologie des Netzes zusammen, welche durch die Anzahl der Schichten, der sich darin befindenden Neuronen sowie die Richtung in welche Information durch das Netzwerk fließen darf, definiert ist [250]. Letzteres ermöglicht auch die Gestaltung sogenannter Recurrent Neural Networks (RNN) bei denen die Ausgangssignale in Schleifen, verzögert oder unter Einführung sogenannter Kurzzeitgedächtnisse erneut als Eingangssignale für dieselben Neuronen genutzt werden können, um zeitliche Zusammenhänge besser abbilden zu können [250]. Je komplexer die Netzwerkstruktur ist, desto mehr Muster und Entscheidungsgrenzen werden damit abbildbar [250]. Netzwerktopologien mit mehreren Schichten werden als Deep Neural Network (DNN) bezeichnet, welche in jüngster Zeit bei Bild- und Textverarbeitungsaufgaben mit Menschen vergleichbare Performanz erreichen [250]. Dieser Vorteil resultiert aus der Möglichkeit komplexe Zusammenhänge damit modellieren zu können ohne Annahmen über diese Daten treffen zu müssen [250]. Nachteilig ist der Rechenaufwand, der für das Training aufgewendet werden muss sowie die Tendenz zum Overfitting und durch menschliche Entscheider nicht nachzuvollziehende Entscheidungen aufgrund der Struktur des Modells [250].

## Leistungsevaluation der Modelle

Um die Leistungsfähigkeit verwendeter Modelle vergleichen zu können, werden geeignete Bewertungsschemata benötigt (vgl. Abbildung 2.21) [250]. Die für das Modell noch unbekanntes Testdaten dienen dazu, die Vorhersagen des trainierten Modells gegenüber der Labels zu vergleichen und auf Grundlage eines Genauigkeitsmaßes die Leistungsfähigkeit zu bestimmen (vgl. Absatz 2.1.4) [250]. Grundsätzlich findet hierzu eine strikte Aufteilung der vorhandenen Daten in disjunkte Teilmengen statt, welche einerseits zum Training und andererseits für das Testen der Modelle verwendet werden [250]. Für die sogenannte „holdout validation“ werden zwei disjunkte Teilmengen gebildet, beispielsweise im Verhältnis 80 zu 20, wobei der größere Anteil der Daten für das Training des Modells genutzt wird und der kleinere Anteil als Testdaten für die Bestimmung der Genauigkeit [250]. Als „k-fold cross-validation“ oder k-fache Kreuzvalidierung wird das Schema bezeichnet, bei dem durch eine zufällige Auswahl  $k$  disjunkte Teilmengen (folds) gebildet werden [250]. Üblicherweise liegen die Werte für  $k$  zwischen 5 und 10, hierbei dient jede Teilmenge einmal als Testdatensatz, welcher ein Modell evaluiert, das auf

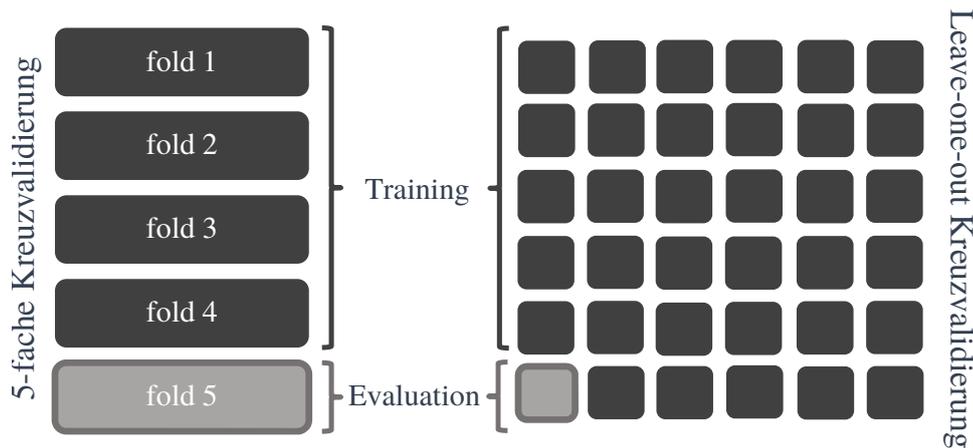


Abbildung 2.21: Beispiele für Bewertungsschemata.

Grundlage der restlichen Teilmengen trainiert wird [250]. Es werden insgesamt  $k$  Modelle trainiert, die erreichte Leistungsfähigkeit wird durch den Mittelwert der Leistungsfähigkeit aller Modelle angegeben [250]. Als Vorteil dieser Vorgehensweise wird gesehen, dass alle Daten zumindest einmalig Teil des Trainingsprozesses sind, was insbesondere bei geringen Datenmengen zu verlässlichen Ergebnissen beitragen soll [250]. Eine Erweiterung dieses Schemas stellt die „repeated  $k$ -fold cross-validation“ dar, welche eine mehrmalige Wiederholung dieses Prozesses durchführt, um die Abhängigkeit zum Zufallsprozess bei der Erstellung der folds zu reduzieren und damit robustere Modelle zu erhalten [250]. Eine weitere Form dieser Art von Validierung stellt die „leave-one-out cross-validation“ dar [250]. Hierbei wird  $k$  entsprechend der Anzahl an verfügbaren Datenpunkten gewählt, sodass für jeden Datenpunkt auf Basis von den restlichen  $n-1$  Datenpunkten, eigene Modelle trainiert werden können, die auf dem  $n$ -ten Datenpunkt evaluiert werden [250].

### Optimierung der Leistungsfähigkeit

Um die Leistungsfähigkeit der Modelle zu optimieren, existieren zwei Ansatzpunkte. Einerseits kann die Parametrierung der Modelle durch Methoden des sogenannten Parameter-Tuning adressiert, andererseits kann durch die Merkmalsauswahl (Feature Selection) die Leistungsfähigkeit beeinflusst werden [242]. Beim Parameter-Tuning geht es darum, die Kombination an Parametern für ein Modell zu finden, welches den zu modellierenden Sachverhalt bestmöglich abbildet [242]. Dieser Vorgang geschieht innerhalb des Trainingsprozesses. Je nach zu trainierendem Modell ist die Zahl an zu wählenden Parametern entsprechend umfangreich, mittels der vorgestellten Bewertungsschemata kann die Wahl optimiert werden (vgl. Unterunterabschnitt 2.4.3 und Abbildung 2.22). Je höher die Dimensionalität des verfügbaren Datensatzes, desto eher stellt sich die Frage, welche Merkmale mit in das Modell integriert werden sollen [242]. Anhand des Informationsgehalts einzelner Merkmale und der Art des Modells wird entschieden, ob das Merkmal einen Beitrag zur Leistung des Modells leistet [242]. Falls Modelle, wie beispielsweise die *Decision Trees*, implizit Merkmale entsprechend ihres Einflusses gewichten, spricht man von Embedded Feature Selection (EFS) [242]. Eine weitere Möglichkeit, um die entsprechende Merkmalsauswahl zu tätigen stellt die Filter Feature Selection (FFS) dar, welche unabhängig vom gewählten Modell anhand einer zu bestimmenden Relevanz pro Merkmal, beispielsweise durch statistische Kennzahlen, bewertet werden [242]. Mittels der Wrapper Feature Selection (WFS) werden Teilmengen von Merkmalen gewählt und deren Einfluss auf die Leistungsfähigkeit des entsprechenden Modells evaluiert [242]. Die

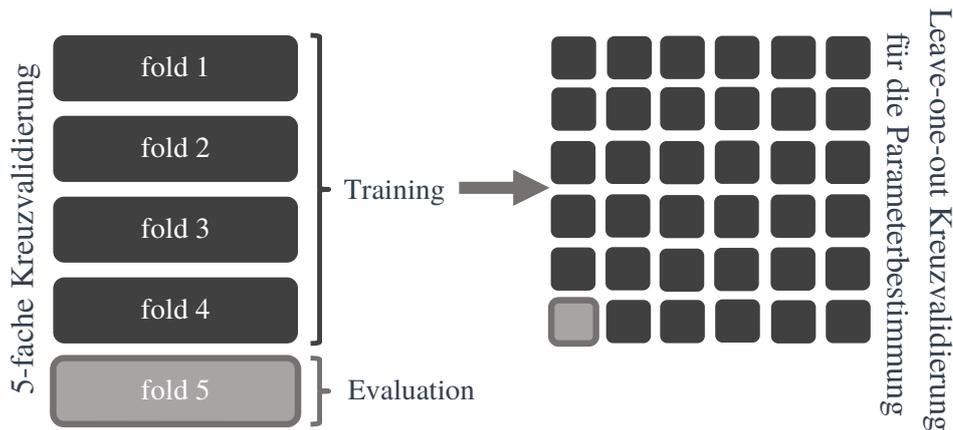


Abbildung 2.22: Optimierung der Modellleistung anhand von Parameter-Tuning.

Merkmalsgruppierung, welche die besten Leistungsdaten liefert, wird schlussendlich ausgewählt [242]. Da die Dauer des Suchprozesses von der Anzahl an Features abhängt, wird dieser durch Heuristiken abgekürzt, die nicht zwingend die optimale Teilmenge liefern, jedoch den Suchaufwand reduzieren [242]. Eine dieser Heuristiken ist die greedy-search Suchstrategie [242]. Hierbei werden zu bestehenden Teilmengen der Merkmale iterativ weitere Merkmale hinzugefügt und evaluiert, ob eine Leistungssteigerung des Modells festzustellen ist [242]. Die Merkmalsmenge, welche die beste Modell-Performanz ermöglicht wird gewählt [242]. Dieser Prozess wird solange wiederholt, bis keine Steigerung der Performanz mehr festzustellen ist [242]. Nachteilig hierbei ist, dass für jedes potentiell hinzuzufügende Merkmal der Trainingsprozess durchlaufen werden muss, weshalb Guyon et al. den sogenannten „backward selection“ Algorithmus entwickelten [173]. Dieser ist Ausgangspunkt für die sogenannte Recursive Feature Elimination (RFE), worin als erstes ein Modell auf Grundlage aller zur Verfügung stehender Merkmale gelernt wird [242]. Anschließend wird die Relevanz einzelner Merkmale pro Modell bestimmt und die am wenigsten wichtigen entfernt bevor in der nächsten Iteration erneut ein Modell mit den übrigen Merkmalen trainiert und entsprechend des zuvor beschriebenen Prozesses evaluiert wird [242].

## Interpretierbarkeit von Modellen

Unter Interpretierbarkeit wird im Kontext von Modellen die Fähigkeit verstanden, „einem Menschen etwas zu erklären oder in verständlicher Form zu präsentieren“ [124]. Interpretable Machine Learning (IML) oder Explainable Artificial Intelligence (XAI) beschreiben Methoden und Modelle, um die mittels ML beziehungsweise KI generierten Ergebnisse für menschliche Nutzer nachvollziehbar darzustellen, um das Vertrauen in solche Systeme insbesondere bezüglich kritischer Entscheidungen zu stärken [292]. Dafür spielt neben der Genauigkeit beispielsweise auch die Transparenz, Kausalität, Privatsphäre und Fairness des ML- oder KI-gestützten Entscheidungsfindungsprozesses eine Rolle [124, 292]. Um die Eigenschaften von XAI-Verfahren zu untersuchen, wird zwischen verschiedenen Kriterien differenziert nach denen sich diese Ansätze unterscheiden lassen. Grundsätzlich lässt sich unterteilen, ob eine Erklärungsmethode das Verhalten eines ML-Algorithmus (global) oder eine einzelne damit prognostizierte Instanz erklärt (lokal) [292]. Ein weiteres Kriterium stellt die Entscheidung dar, ob ein Ansatz intrinsisch durch die Begrenzung der Komplexität des ML-Modells oder post-hoc durch die Analyse der Interpretierbarkeit nach dem Modelltraining funktioniert [292].

**Intrinsische Interpretierbarkeit von Modellen:** Wie bereits zuvor erörtert ist die Transparenz bei gewissen Modellansätzen inhärent. So liefern beispielsweise Klassifikationsmodelle wie *Decision Trees* (vgl. Absatz 2.4.3) hierarchische oder Regressionsmodelle lineare Zusammenhänge, welche strukturell nachvollziehbar sind, da das Verhältnis zwischen den Merkmalen und damit auch zwischen den Eingabewerten und prädizierten Ausgabewerten sichtbar wird [292]. Je stärker in den Modellen die Interaktionen zwischen den Merkmalen berücksichtigt werden, desto stärker kann dies die Interpretierbarkeit einschränken [292]. In Tabelle 2.4 sind die Eigenschaften ausgewählter interpretierbarer Algorithmen aufgelistet. Dabei wird zwischen den möglichen Aufgaben für die diese eingesetzt werden können unterschieden [292]. Auch ob lineare oder monotone Zusammenhänge zwischen dem Merkmal und den Ausgangswerten modelliert werden sind Kriterien, die einen Einfluss auf die Interpretierbarkeit haben [292]. So lassen sich Verhältnisse zwischen Merkmalsausprägungen und deren Ausgangswerten besser annehmen, falls diese Zusammenhänge bestehen [292]. Durch die Modellierung von Interaktionen zwischen den Features wird hingegen die Interpretierbarkeit erschwert [292].

Algorithmus	Aufgabe	Monoton	Linear	Interaktion
<b>k-NN</b>	Klassifikation & Regression	Nein	Nein	Nein
<b>Decision Tree</b>	Klassifikation & Regression	Teilweise	Nein	Ja
<b>Naïve Bayes</b>	Klassifikation	Ja	Nein	Nein
<b>Lineare Regression</b>	Regression	Ja	Ja	Nein

Tabelle 2.4: Ausgewählte interpretierbare Modelle und deren Eigenschaften nach Molnar [292].

**Modell-agnostische Methoden:** Je umfangreicher die zuvor beschriebenen Interaktionen zwischen Merkmalen abgebildet werden, desto mehr bewegt man sich hin zu sogenannten „Black Box“-Modellen, deren Interpretierbarkeit für menschliche Entscheider unmöglich wird [292]. Für solche Modelle setzen modell-agnostische Methoden auf dem Konzept der Manipulation von Eingabemerkmalen und der anschließenden Beobachtung korrespondierender Prädiktionen auf. Dadurch entsteht eine Unabhängigkeit bezüglich der Modellart und -komplexität sowie der Art der Erklärung wie beispielsweise die Angabe der Wichtigkeit einzelner Merkmale oder Berechnungsvorschriften [345]. Dabei findet wie einleitend beschrieben, eine weitere Unterscheidung in lokale und globale Verfahren statt [292]. Beispiele hierfür werden in Unterabschnitt 3.3.4 genauer ausgeführt.

**Beispielbasierte Erklärung:** Die Erklärung anhand von Beispielen erfolgt über die Wahl bestimmter Instanzen des Datensatzes, welche das Verhalten des ML-Algorithmus oder die zugrundeliegende Verteilung der Daten anhand ähnlicher anderer Instanzen erklären [292]. Dem k-NN (vgl. Absatz 2.4.3) ist dieses Prinzip inhärent [292]. Diese Art von Erklärung setzt voraus, dass die Dateninstanzen genügend Kontext enthalten, um für Menschen verständlich darstellbar zu sein, wie es beispielsweise bei Texten oder Bildern der Fall ist [292]. Bei Daten mit einer hohen Anzahl an Features, die tabellarisch dargestellt werden, ist die Verständlichkeit für menschliche Entscheider weniger gegeben, weshalb Instanzen zusammengefasst werden sollten, um mögliche Beispiele oder Erklärungen einfach zu halten [292].

**Evaluation dieser Methoden:** In der wissenschaftlichen Literatur existiert zum Zeitpunkt der Anfertigung dieser Arbeit noch kein Konsens darüber, was Interpretierbarkeit im Kontext des maschinellen Lernens bedeutet [292]. Demnach sind auch keine verbindlichen Maße definiert, anhand derer die Güte

einer Erklärung bestimmt werden kann [292]. Doshi-Velez et al. leiten allgemeine Ansätze ab, um die verschiedenen XAI-Verfahren zu evaluieren [124]. Dabei unterscheiden sie zwischen drei Ebenen der Evaluation.

Als **Bewertung auf Anwendungsebene** wird die Methode bezeichnet, bei der die Erklärung direkt in das Produkt integriert ist und von Anwendern evaluiert wird. Diese Art von Bewertung erfolgt mittels Experten [124]. Daher werden als Tester Anwender benötigt, die entsprechendes Domänenwissen besitzen, um die Erklärung verstehen und anschließend evaluieren zu können inwiefern die Darstellung der Erklärung auf den tatsächlichen Sachverhalt zutrifft [292]. Nachteilig für dieses Verfahren ist daher die Verfügbarkeit von Experten [292].

Die **Bewertung auf menschlicher Ebene** vereinfacht das zuvor beschriebene Konzept indem die Erklärung nicht mit Experten sondern Probanden ohne Fachwissen evaluiert wird. Hierdurch entfällt die Abhängigkeit zu Experten, woraus eine größere potentielle Testerbasis resultiert [124]. Die Anwender wählen die für sie verständlichste Erklärung unabhängig von deren inhaltlichen Wahrheitsgehalt aus [292]. Die Verständlichkeit steht somit im Vordergrund [124].

Unabhängig von menschlichen Testern funktioniert die **Bewertung auf Funktionsebene**. Hierfür müssen vorab individuelle Kriterien wie beispielsweise Kennzahlen definiert werden, anhand derer evaluiert wird [124]. Als ein Beispiel kann die Tiefe von Entscheidungsbäumen dienen, wobei gilt, je weniger Tiefe ein Entscheidungsbaum aufweist, desto besser die Erklärbarkeit [292].

Für die in dieser Arbeit verwendeten verschiedenen Arten von Modellen finden unterschiedliche Verfahren ihre Anwendung. Daraus resultierende Ergebnisse lassen sich ebenfalls in folgende Kategorien unterteilen:

- **Feature Summary Statistic:** Eine Menge an Interpretationsmethoden liefert für jedes Feature zusammenfassende Statistiken beispielsweise Kennzahlen über die Wichtigkeit des Features [292].
- **Feature Summary Visualization:** Die Visualisierung der zusammenfassenden Statistiken wird hierbei anhand verschiedener Grafiken vorgenommen [292].
- **Model Internals:** Bezüglich intrinsisch interpretierbarer Modelle werden hierbei die gelernten Parameter und internen Strukturen des Modells externalisiert [292].
- **Data Point:** Hierin werden Methoden wie beispielbasierte Erklärungen, welche Datenpunkte als Erklärung anführen, zusammengefasst [292].
- **Intrinsically Interpretable Model:** Dieser Ergebniskategorie subsumiert Verfahren, die mit Hilfe von intrinsisch interpretierbaren Modellen sogenannte Black-Box-Modelle approximieren und dementsprechend darstellen [292].

# 3 Einordnung in den Stand der Wissenschaft und Technik

**I**N diesem Kapitel wird der zum Zeitpunkt der Anfertigung dieser Arbeit vorzufindende Stand der Wissenschaft und Technik erörtert. Die Interdisziplinarität der Problemstellung macht die Betrachtung der aktuellen Lage aus verschiedenen Perspektiven notwendig. Die Erkenntnisse in diesem Kapitel sind das Kondensat mehrerer strukturierter Literaturrecherchen, deren Vorgehensweisen und Ergebnisse teilweise bereits in den vermerkten Publikationen in Abschnitt 1.4 beschrieben sind.

## 3.1 Digitale Transformation im Gesundheitswesen

In Organisationen ermöglicht die Nutzung von IKT Effektivitäts- und Effizienzgewinne, welche sowohl durch die Reorganisation von Arbeitsabläufen und -strukturen im Zuge der Entwicklung und Einführung dafür vorgesehener Lösungen als auch durch die Digitalisierung von diesen Vorgängen selbst entstehen [238, S. 2]. Während unter Daten lediglich eine regelbasierte Aneinanderreihung von Zeichen verstanden wird, kann durch Kontextualisierung dieser, Information gewonnen und schließlich durch Vernetzung mit anderen Informationen, Wissen abgeleitet werden [238, S. 12]. Dadurch können neben digitalen Verwaltungsdaten zur Abrechnung von Dienstleistungen auch in digital erfassten und vernetzten Patientendaten weiterführende Informationen und Wissen stecken. Dieser Logik folgend, kann aus digitalen Gesundheitsdaten vertieftes Wissen über Behandlungsprozesse oder Krankheiten abgeleitet werden. Die Nutzung dieser Potentiale sollte auch das Ziel eines patientenorientierten Gesundheitssystems sein. Die fortschreitende digitale Transformation betrifft somit neben verschiedenen gesellschaftlichen Bereichen zunehmend auch den medizinischen Sektor [282]. Unter der Bezeichnung KI rücken hierin Entwicklungen in den Vordergrund, die insbesondere das verborgene Potential digitaler Daten durch tiefgreifende Analysemethoden heben sollen. Als Grundlage hierfür wird jedoch eine Kombination von Daten in ausreichender Menge und Qualität benötigt (vgl. Unterabschnitt 2.1.3). Aktuell findet die Erhebung und Auswertung dieser Daten durch unterschiedliche Akteure statt. Die Einbeziehung von Menschen als Datenlieferant ist einerseits dem privatwirtschaftlichen Anbieter der jeweiligen Dienstleistung oder des Produkts überlassen, andererseits staatlich gelenkten Institutionen im Gesundheitswesen. Jüngste gesetzliche Änderungen stellen neben dieser isolierten Betrachtung Integrationsansätze und Patientensouveränität in Aussicht mit dem Ziel einer selbstbestimmten Verwertung der eigenen Gesundheitsdaten (vgl. Unterabschnitt 2.4.1 und Unterabschnitt 3.1.2). Im deutschen Gesundheitswesen kommt die Verfolgung des Ziels eines patientenorientierten Gesundheitssystems seit Jahren nur zögerlich voran [40, S. 107-116]. Als Gründe für die Schwierigkeiten bei der Umsetzung hierfür erforderlicher kompatibler IKT-Lösungen werden primär Zielkonflikte aufgrund gegensätzlicher Interessen der sich durch Selbstverwaltung (vgl. Unterabschnitt 2.2.1) organisierten Beteiligten gesehen [40, S. 107-116]. Als ein Beispiel hierfür wird durch die digitale Erfassung von Behandlungsdaten, die aus Sicht der unterschiedlichen Gesundheitsprofessionen

neu gewonnene Transparenz bezüglich Entscheidungsprozessen und Handlungen angeführt, wodurch paternalistische Strukturen ersichtlich und gegebenenfalls angreifbar werden [40, S. 114].

### 3.1.1 Standardisierung und Interoperabilität

Um möglichst vollständige Gesundheitsprofile zu erhalten, müssen von den jeweiligen Interessenvertretern erhobene gesundheitsbezogene Datenpunkte zu diesen zusammengesetzt werden. Grundlegende Voraussetzung für eine nachhaltige Nutzung von Gesundheitsdaten stellt die Interoperabilität über Organisations-, System- und Gerätegrenzen hinweg sowie die Möglichkeit zur Integration unterschiedlicher Datenquellen dar (vgl. Unterunterabschnitt 2.4.2). Hierfür existieren technische Formate und Protokolle, welche den Austausch und die Integration von gesundheitsbezogenen Daten in Kombination mit einheitlich definierten Gesundheits- und Behandlungszuständen standardisieren. Die grundlegende Unterteilung der Standards lässt sich in zwei Kategorien vornehmen. Einerseits werden Kodierungssysteme und Terminologien benötigt, welche ein gemeinsames Verständnis von medizinischen Sachverhalten unterstützen (vgl. Unterabschnitt 2.2.2). Andererseits werden auf technischer Ebene Standards für den Austausch und die inhaltliche Beschreibung gesundheitsrelevanter Daten definiert (vgl. Unterunterabschnitt 2.4.2). Die zweite Kategorie brachte in den vergangenen Jahren Standards hervor, die sich nach Art und Anwendungsgebiet der Daten unterscheiden [298]. Basierend auf einer angepassten Version der XDS.b ist die Struktur der seit 2021 verfügbaren ePA in Deutschland definiert [159]. Als ein Format für die Beschreibung und den Austausch von Gesundheitsdaten mittels webbasierter Standards wurde FHIR entwickelt (vgl. Unterunterabschnitt 2.4.2). FHIR ist modular aufgebaut und lässt sich aus unterschiedlichen Komponenten, den sogenannten Ressourcen zusammensetzen, um die Aspekte des Gesundheitssystems abzubilden [68]. Als Schnittstellen implementiert FHIR das Konzept der API unter Verwendung des REST-Protokolls. Damit sind Abfragen von Ressourcen mittels Uniform Resource Locator (URL) über HTTP bei FHIR-Servern möglich (vgl. Absatz 2.4.2). Zusammengehörende Mengen an Ressourcen werden in sogenannten *Bundles* zusammengefasst, deren Inhalte wiederum über ein *ValueSet* definiert sind. Hierin sind Suchergebnisse, klinische Dokumente oder zu einer Transaktion zusammengefasste Anfragen an einen Server bündelbar und einheitlich kodiert. Somit lassen sich beispielsweise Patienten, deren Gesundheitszustände und diesen zugrundeliegenden Diagnosen verknüpfen und in Formaten wie JSON, XML oder RDF darstellen. Für FHIR bietet „Substitutable Medical Applications, Reusable Technologies (SMART)“ eine Plattform, welche durch quelloffene Bibliotheken bei der Integration und Entwicklung von Applikationen in Kombination mit EHR-Systemen unterstützt. Unter anderen verwendet die Health App von Apple<sup>1</sup> SMART und damit FHIR, um die Aggregation von Gesundheitsdaten aus verschiedenen Gesundheitsdatenquellen zu ermöglichen [70]. Ein weiteres Anwendungsgebiet liegt in der Integration von Systemen in ein Klinisches Informationssystem (KIS). Die darin gesammelten Daten können durch zusätzliche Software ausgewertet und zur klinischen Entscheidungsunterstützung dienen [71], wie zum Beispiel zur Erkennung von Kontraindikation bei der Erstellung von Medikationsplänen. Zur Beschreibung von sogenannten Medizinische Informationsobjekte (MIO), welche in übergeordnete Konzepte wie dem Impf- oder Mutterpass in der deutschen ePA zusammengefasst werden, soll ebenfalls FHIR eingesetzt werden [223]. Gesetzlich festgelegt ist die standardisierte und offene Gestaltung von Archiv- und Wechselschnittstellen in informationstechnischen Systemen für die vertragsärztliche Versorgung nach § 291d Abs. 1 SGB V und § 371 Abs. 1 SGB V.

---

<sup>1</sup> <https://www.apple.com/healthcare/health-records>, zuletzt zugegriffen am 15.04.2022.

### 3.1.2 Gesundheitsdatenmobilität und -infrastrukturen

Unter der Begrifflichkeit Datenmobilität wird in diesem Abschnitt die Austauschmöglichkeit von Datenmengen mit Zweckbindung verstanden. Hierbei handelt es sich nicht nur um den reinen Datentransfer sondern auch um Mechanismen, die eine gezielte Freigabe, Steuerung und Überwachung des Datengebrauchs durch den Datenanbieter ermöglichen. Bezüglich der zuvor erörterten Potentiale, die in gesundheitsbezogenen Daten stecken, könnte das zur Verfügung stellen dieser zur Personalisierung von medizinischen Dienstleistungen und dadurch zur Steigerung der Effizienz und Effektivität beitragen (vgl. Unterabschnitt 3.2.1). In diesem Abschnitt wird die Erhebung, Verwertung und Verbreitung von gesundheitsbezogenen Daten mittels Datensouveränitätstechnologien aus unterschiedlichen Perspektiven dargestellt. Die Quellen und Inhalte basieren auf einer vom Autor publizierten strukturierten Literaturrecherche und dienen als Grundlage für die in Kapitel 4 abgeleiteten Anforderungen [473].

Die Erfassung und Nutzung digitaler Gesundheitsdaten nimmt aufgrund der technischen Entwicklungen rund um die Themen der persönlichen oder elektronischen Gesundheitsakten (vgl. Unterabschnitt 2.4.1), Wearables, Internet of Things (IoT)-Sensoren, Edge- und Cloud-Computing, 5G als Kommunikationsstandard sowie KI zu [269, 396]. Die Forschung bezüglich des Austauschs medizinischer Daten ist seit Jahren aktiv. Bereits 2010 stellen Rao et al. ein Software-Werkzeug für den Austausch und die Abfrage klinischer Informationen unter Verwendung des HL7-Standards vor [341, 342]. Dennoch mangelt es derzeit noch an einheitlichen und interoperablen Infrastrukturen zur Erfassung medizinischer Daten für die Gesundheitsforschung, die produktiv genutzt und der Öffentlichkeit zur Verfügung gestellt werden, da sowohl technische als auch strukturelle, rechtliche und ethische Fragen noch immer ungelöst sind [396]. Der Wert von Gesundheitsdaten für Individuen und Gesellschaften steht dabei im Vordergrund. Diesbezüglich besteht eine Abwägung zwischen einer selbstbestimmten Nutzung von Gesundheitsdaten zur Optimierung des individuellen Nutzens und eine fremdbestimmte Zuführung der Daten im Dienste des Gemeinwohls. Eine Studie von Pickard greift diese Abwägung auf und zeigt, wie sich finanzielle Anreizmechanismen auf die Bereitschaft zum Austausch von Gesundheitsdaten auswirken [218]. Zwei Drittel der Teilnehmer sind durch eine finanzielle Belohnung stärker motiviert, diese Informationen zu teilen. Neben einem finanziellen Mehrwert könnte auch die Möglichkeit, mehr Einblick in den aktuellen Gesundheitszustand eines Individuums und mehr Selbstbestimmung sowie Möglichkeiten zur Teilnahme am Behandlungsprozess zu erhalten, eine Motivation sein, insbesondere für Menschen mit chronischen Krankheiten [313]. Im Jahr 2020 aktuelle staatliche und kommerzielle Bemühungen Gesundheitsdaten als digitales Gut zur Verfügung zu stellen, werden diesbezüglich nachfolgend kurz zusammengefasst.

Mit *MyHealthRecord* verfolgt das nationale Gesundheitsdatensystem Australiens das Ziel, digitale Patientendaten sicher zu speichern und den Bürgern über eine Lösung zentral zur Verfügung zu stellen [39]. Es ermöglicht den Patienten, Gesundheitsdaten mit Ärzten, Krankenhäusern und anderen Leistungserbringern des Gesundheitswesens auszutauschen. Das System implementiert eine Widerspruchslösung, das heißt, die Teilnahme ist standardmäßig für jeden Australier aktiviert. Die Daten aus *MyHealthRecord* stehen für Forschungszwecke zum Zeitpunkt der Anfertigung dieser Arbeit noch nicht zur Verfügung, deren forschungsbezogene Nutzung ist aber geplant. Grund für die bisherige Zurückhaltung stellt die noch ausstehende Etablierung robuster Prozesse und Regelungen zu den Themen Verwaltung, Sicherheit, Datenschutz und technischer Vorkehrungen zu deren Umsetzung dar [396]. *FINDATA* ist die nationale Behörde Finnlands, welche für die Umsetzung einer zentralen Verwaltung und Freigabe von Gesundheits- und Sozialdaten der finnischen Bevölkerung zuständig ist [146]. Dieses Vorhaben trat via Gesetz im Mai 2019 in Kraft. Zu diesem Zweck werden Forscher über die Datenverfügbarkeit informiert, Genehmigungen für die sekundäre Nutzung erteilt und die pseudonymisierte Datenverarbeitung in der von ihnen bereitgestellten sicheren Umgebung ermöglicht. Der Zugriff auf die Gesundheits- und

Sozialdaten wird auch über deren nationale eHealth-Infrastruktur *Kanta* ermöglicht. *23andMe* ist ein US-amerikanisches Biotechnologie-Unternehmen, das Gen-Tests auf der Grundlage von Speichelproben direkt an den Verbraucher verkauft, um die Abstammung und genetische Veranlagung für bestimmte Krankheiten und Leiden zu identifizieren [394]. Das Geschäftsmodell beinhaltet auch die Möglichkeit, freiwillig die Testergebnisse in Verbindung mit Informationen über das eigene Surfverhalten im Web und einer Selbstauskunft an Dritte für Forschungszwecke oder kommerzielle Anwendungen weiterzugeben. Dafür erhält *23andMe* eine finanzielle Gegenleistung. Aufgrund dieses Geschäftsmodells und der ethischen Implikationen geriet das Unternehmen in den Fokus von Medien, Wissenschaft und der Food and Drug Administration (FDA) [394, 396]. Neben staatlichen und kommerziellen Ansätzen hinsichtlich der Erfassung und des Austauschs von Gesundheitsdaten, scheint es auch in der Wissenschaft einen Trend zu mehr Transparenz zu geben, um die Erschließung von weiterem Wissen und dessen Reproduzierbarkeit zu gewährleisten. Dies erfordert die aktive Beteiligung von finanzierenden und veröffentlichenden Organisationen, die derzeit Forschungsergebnisse sammeln [393]. In diesem Zusammenhang identifiziert Chan et al. Probleme mit der Kommunikation von Forschungs- und Studiendaten [34]. Sie stellen fest, dass oft unvollständige Daten vorliegen oder die Wiederverwendung dieser Daten erschwert wird, was zu finanziellen Auswirkungen, Verzerrungen, weniger Nutzen für die Forschung und damit für die Versorgung der Patienten führt. Um diese Probleme zu vermeiden, geben Chan et al. drei Empfehlungen [34]. Erstens sollten akademische Institutionen und Geldgeber für die vollständige Weitergabe ihrer Forschungsdaten belohnt werden. Zweitens sollten Gesetzgeber, Ethikkommissionen, Geldgeber und veröffentlichende Organisationen einheitliche Richtlinien für Studienregistrierungen und die Verfügbarkeit vollständiger Forschungsdaten durchsetzen. Darüber hinaus müssen Standards für den Inhalt von Protokollen und vollständigen Studienberichten sowie für die Praktiken des Datenaustauschs entwickelt werden. Da die semantische Komplexität der Medizin als hoch angesehen wird, stellen Dugas et al. einen Community-basierten Ansatz in Form einer mehrsprachigen Plattform für den Austausch und die Diskussion von Datenmodellen in der medizinischen Forschung und im Gesundheitswesen vor. Ziel ist es den Entwurf medizinischer Datenmodelle durch den Austausch von erfolgreichen Konzepten und stärker standardisierten Datenmodellen mit semantischer Annotation sowohl zu verbessern als auch zu beschleunigen [126]. Um den zuvor beschriebenen Problemen in Deutschland auf nationaler Ebene zu begegnen, verfolgt die Medizininformatik-Initiative (MII)<sup>2</sup> das Ziel, eine Infrastruktur für die Integration klinischer Daten aus der Patientenversorgung und medizinischen Forschung zu schaffen [374]. Diese Initiative befindet sich zum Zeitpunkt der Anfertigung dieser Arbeit in der Entwicklungs- und Vernetzungsphase, in welcher jedes der insgesamt vier Konsortien in Zusammenarbeit mit den beteiligten Universitätskliniken und Partnerinstitutionen über vier Jahre hinweg ein Datenintegrationszentrum aufbaut, um IT-Lösungen für spezifische Anwendungsfälle in Neurologie, Kardiologie, Onkologie, Pneumologie und Infektionskontrolle zu entwickeln [174, 330, 333, 425]. Diese Datenintegrationszentren sollen zunächst demonstrieren, wie Daten, Informationen und Wissen aus der Patientenversorgung sowie der klinischen und biomedizinischen Forschung über die Grenzen verschiedener Standorte hinweg verknüpft werden können. Diese konzentrieren sich auf universitäre medizinische Einrichtungen und befassen sich in dieser Phase nicht mit Arztpraxen, Regionalkrankenhäusern oder der Patientenversorgung im stationären und ambulanten Bereich, sodass sie sich noch in einem frühen Stadium der Suche nach einer umfassenden Lösung befinden. Ein diesbezügliches Forschungsprojekt auf europäischer Ebene stellt das Projekt MyHealthMyData (MHMD)<sup>3</sup> dar. Ziel des Projekts ist es, das erste offene biomedizinische Informationsnetzwerk zu schaffen, das sich auf die Verbindung zwischen Organisationen und Einzelpersonen konzentriert. Es möchte Krankenhäuser ermutigen, damit zu beginnen, anonymisierte Daten

---

<sup>2</sup> <https://www.medizininformatik-initiative.de>, zuletzt abgerufen am 20.07.2020.

<sup>3</sup> <http://www.myhealthmydata.eu/>, zuletzt abgerufen am 20.07.2020.

für die Forschung offen zur Verfügung zu stellen und Bürger als Eigentümer dazu befähigen, das Beste aus den verfügbaren Gesundheitsdaten zu machen, indem ein selbstbestimmter Umgang damit ermöglicht werden soll. Hierin wird DLT (vgl. Absatz 2.4.2), insbesondere die Blockchain-Technologie, als Schlüsselement untersucht, um die Akteure im Gesundheitswesen miteinander zu verbinden und den Prozess der gemeinsamen Datennutzung proaktiv, sicher, offen und dezentralisiert zu steuern [47, 235]. Daher wird ein föderiertes Blockchain-Netzwerk verwendet, in dem die autorisierten Teilnehmer die Zugangskontrollrichtlinien im Sinne der Blockchain definieren. Für die Speicherung von Gesundheitsdaten wird ein Off-Chain-Ansatz (vgl. Absatz 2.4.2) verwendet, um die DSGVO-Konformität hinsichtlich Datenlöschung zu gewährleisten, da in der Blockchain nur Referenzen auf die Daten über Hash-Werte gespeichert werden und die sensiblen Gesundheitsdaten ihre Speicherorte nicht verlassen sollen [47, 235]. Dadurch ist es möglich, die Daten bei Bedarf zu löschen, aber auch eine Historie von Datenänderungen nachvollziehen zu können. Dieser Off-Chain-Ansatz wird im Kontext des Datenmanagements im Gesundheitswesen von weiteren Publikationen diskutiert und verfolgt [217, 269, 421, 433]. Welche Entscheidungspunkte bei der Entwicklung eines Blockchain-basierten Systems zur Speicherung sensibler Gesundheitsdaten zu berücksichtigen sind, findet sich in einer vom Autor dieser Arbeit separat getätigten Publikation [465]. Die MHMD-Lösung weist den verschiedenen Teilnehmern des Netzwerks ihre Identität durch Zertifikate nach, die von einer Zertifizierungsstelle ausgestellt werden. Nach dem die Daten vom Dateneigentümer in das Netzwerk eingespeist worden sind, kann ein Datenkonsument die Daten auf der Grundlage dieser Zertifikate anfordern. In Kombination mit benutzerdefinierten Regeln kann ein Dateneigentümer den Datenaustausch und Zugang zu seinen eigenen Gesundheitsdaten verwalten und autorisieren, um nachvollziehen zu können, wer die Daten wann und unter welchen Bedingungen verwendet hat. Die Daten können von den Datenkonsumenten anhand von Metadaten gefunden werden, die in einem zentralen Katalog zusammengefasst sind. William J. Gordon et al. erörtern die Herausforderungen und Grenzen des Einsatzes der Blockchain-Technologie in Gesundheitsdiensten [421]. Insbesondere werden die Themen Skalierbarkeit, Anreizmechanismen für die Weiterentwicklung und Wartung der Interoperabilität, Patienten-Schlüsselmanagement und -Beteiligung angesprochen. Um die Harmonisierung des Datenschutzes (Privacy-by-Design) zu gewährleisten, werden Werkzeuge zur Verschlüsselung und Deidentifizierung mit der Blockchain-basierten Lösung in MHMD kombiniert [47, 235]. Das übergeordnete Ziel von MHMD ist es, einen vertrauenswürdigen Informationsmarktplatz zu schaffen, der Bürgern, Krankenhäusern, Forschungszentren und Unternehmen einen Mehrwert bietet. Jaremko et al. gehen auf die Vertrauensfrage bezüglich der Datenübermittlung und -verarbeitung sowie des Datenkonsumenten ein, indem sie die Daten vollständig anonym und nur für die vorgesehenen Zwecke verarbeiten werden. Der Bezug von Daten zu deren Verarbeitungszweck geht verloren, wenn dieser nicht direkt daran geknüpft ist. Daher sind Blockchain- in Kombination mit Containerisierungs-Technologien sowie validierten Methoden zur Deidentifizierung von Daten vorgesehen, um diese Verknüpfung zu kapseln [207]. In der Informatik fasst das Konzept der Kapselung Mechanismen zusammen, die es Programmierern erlauben Daten und darauf bezogene Verarbeitungsschritte an einem Ort zu gruppieren und somit schützenswerte oder irrelevante Informationen von Nutzern zu abstrahieren [370]. Hierauf beruht auch das Konzept der Containerisierung, welches voneinander isolierte und gebündelte virtualisierte Umgebungen, die ihre eigene Software, -Bibliotheken, Daten und Konfigurationsdateien beinhalten (vgl. Absatz 2.4.2). Über definierte Schnittstellen können diese miteinander kommunizieren. Alle Container werden von einem einzigen Betriebssystemkern ausgeführt und benötigen daher weniger Ressourcen als virtuelle Maschinen. Das Projekt *OwnYourData*<sup>4</sup> greift auf dieses Konzept zurück. Hierbei sollen Personen dabei unterstützt werden, die eigenen Daten zu sammeln, sicher zu speichern, zu organisieren und weiterzugeben. Das

<sup>4</sup> <https://www.ownyourdata.eu/>, zuletzt abgerufen am 21.02.2022.

übergeordnete Ziel ist das Teilen von Daten, ohne die Kontrolle über die Nutzung der Daten vollständig aufzugeben [315]. Hierfür wird eine schnelle, transparente und standardisierte Open-Source-Lösung für die Bereitstellung von Daten angestrebt [315]. Qualität, Vollständigkeit und Herkunft der geteilten Daten sollen dadurch ersichtlich sein, um deren Nutzung zu erleichtern [315]. Um dieses Ziel umzusetzen werden sogenannte *Semantic Container* zur Sammlung und Verteilung von Daten durch eine Non-Profit-Organisation entwickelt.

Zusammenfassend befinden sich Infrastrukturen und Rahmenwerke zur Realisierung von Datenmobilität im Gesundheitswesen zum Zeitpunkt der Anfertigung dieser Arbeit noch in der Forschung und Entwicklung. Der aktuelle Stand nationaler und internationaler Bemühungen sowie Herausforderungen und Potentiale lassen sich aus diesem Abschnitt entnehmen. Teile dieser bestehenden Konzepte werden innerhalb der Arbeit aufgegriffen und auf deren Anwendbarkeit hinsichtlich des beschriebenen Anwendungsfalls hin evaluiert.

## **3.2 Datengetriebene Ansätze in der medizinischen Versorgung**

Gesundheitsdaten dienen als Grundlage für medizinische Entscheidungen und dadurch bezüglich präventiver, diagnostischer oder therapeutischer Maßnahmen. Bei der Bewältigung von chronischen Erkrankungen sind die Betroffenen selbst mit der kontinuierlichen Erfassung und Verwaltung ihrer konditionalen Daten konfrontiert. Hierfür entwickelte patienten-zentrierte Anwendungen für das Selbstmanagement gelten daher für die Verbesserung der Diagnose und Behandlung chronischer Krankheiten als besonders wichtig, da sie die Ferndatenerfassung und -überwachung ermöglichen [399]. Selbstmanagementsysteme und assistive Technologien verfolgen das Ziel Menschen zu unterstützen, die beispielsweise an chronischen Krankheiten, wie Diabetes [399], schweren psychischen Erkrankungen [410] oder Demenz [245] leiden. Zwei Säulen für datengetriebene Ansätze in der Medizin bilden somit die digitale Erfassung und der Austausch von Daten woraus einerseits das gemeinsame Treffen von Entscheidungen und andererseits personalisierte Behandlungsansätze resultieren können.

### **3.2.1 Datenaustausch für personalisierte Medizin**

Um sich dem Ziel von personalisierten medizinischen Behandlungsmöglichkeiten zu nähern, werden in der wissenschaftlichen Literatur identifizierte und vom Autor bereits publizierte Ansätze zum Datenaustausch bezüglich chronischer Erkrankungen nochmals zusammengefasst [473]. Als Herausforderungen bei der Einbettung von digitalen Lösungen zur Datenerfassung und dem -austausch im wissenschaftlichen, medizinischen und patientenbezogenen Umfeld werden von Lawler et al. ethische, rechtliche und regulatorische Fragen gesehen [253]. Hierbei wird insbesondere die Kombination von klinischen Daten und Genom-Daten als ein wichtiger Aspekt für die gemeinsame Nutzung von Daten für personalisierte Medizin am Beispiel der Krebsforschung betrachtet. Mit dem Argument, dass Genom-Daten als schwer zu anonymisierende Daten zählen, spielen die zuvor identifizierten Fragen eine zentrale Rolle [316]. Hu et al. sehen die kontinuierliche Erfassung von Gesundheitsdaten durch das Monitoring von chronisch erkrankten Menschen im ambulanten Umfeld als notwendige Grundlage von neuen datenbasierten Therapiemöglichkeiten [438]. Hierfür vergleichen sie unterschiedliche Cloud-Deployment-Modelle anhand der von ihnen identifizierten Kategorien Sicherheit, Privatsphäre, Skalierbarkeit, Funktionalität, Adaptierbar-

keit, Einrichtungs- und Wartungskosten sowie rechtliche Fragen mit der Schlussfolgerung, dass hybride Cloud-Deployment-Modelle passend für dafür einzusetzende Systeme sind [438]. Darauf aufbauend stellen sie eine hybride cloudbasierte Lösung vor, um die in dieser Umgebung gesammelten Daten sowohl mit Gesundheitsdienstleistern über eine private Cloud als auch mit Gesundheitsdiensten und Forschern über eine öffentliche Cloud auszutauschen [438]. Dazu notwendige mobile Anwendungen müssen laut Al-Tae et al. die erfassten sensiblen Daten von Grund auf sicher verwalten, um den Schutz der Privatsphäre zu gewährleisten [268]. Dementsprechend identifizieren sie Anforderungen an die Sicherheit und die persönliche Privatsphäre, die in mobilen Gesundheitsinformationssystemen erforderlich sind [268]. Der Hauptfokus ihres Ansatzes liegt auf dem Einsatz von Zugangskontroll-, Datenschutzmechanismen, kryptographischen Methoden und Auditverfahren, die in diesem Fall den Patienten und Angehörigen von Gesundheitsberufen die Möglichkeit geben sollen, die Offenlegung von identifizierbaren Gesundheitsdaten zu kontrollieren. Ma et al. stellen eine Open-Source-Cloud-Plattform für den Austausch von Gesundheitsdaten und zur Unterstützung der ambulanten Versorgung chronisch erkrankter Menschen vor [206]. Insbesondere die Aspekte der Interoperabilität, Sicherheit und Datenschutz werden als die zentralen Herausforderungen für einen effektiven Austausch von Gesundheitsdaten identifiziert. Die Bereitschaft, der Forschung und den Zentralregistern patientengenerierte Daten zur Verfügung zu stellen, hängt auch davon ab, ob es möglich ist, im Gegenzug Informationen und Wissen daraus zu erhalten. In einer Studie von Smith et al. wird anhand von ehemals an einer Krebserkrankung leidenden Menschen festgestellt, dass die Teilnehmer bereit waren, Informationen über die langfristigen Auswirkungen und die Behandlung von Krebs auszutauschen, jedoch mit dem Wunsch verknüpft von den Zentralregistern weitere evidenzbasierte Informationen zu erhalten, die für ihr Selbstmanagement und die gemeinsame Versorgung über die Krebserkrankung hinaus hilfreich sein könnten [205, 385]. Eine Umfrage von O’Kane et al. unter Diabetes-Patienten und ihrem Gesundheitsnetzwerk ergab, dass es Diskrepanzen zwischen dem Informationsbedarf von Patienten und Klinikern gibt [309]. Kliniker sind vor allem an den Gesundheitsdaten selbst und den Bedürfnissen der Patienten interessiert. Im Gegensatz dazu wollen Patienten aus den verfügbaren Daten Erkenntnisse gewinnen, um ihren Gesundheitszustand besser überwachen und verstehen zu können. Entsprechend muss auch die Visualisierung an diese Bedürfnisse angepasst werden. Da Selbstverwaltungssysteme mit der kognitiven Aufnahme großer Mengen an Gesundheitsinformationen und der Kommunikation mit den Leistungserbringern des Gesundheitswesens einhergehen, um fundierte Entscheidungen zu ermöglichen, benötigen beide Akteure visuelle Hilfsmittel, um die Verwaltung komplexer Informationen zu vereinfachen [340]. Rajwan et al. liefern einen Leitfaden für die Gestaltung und Umsetzung solcher Visualisierungen [340]. Neben der Visualisierung muss der gesamte Datenaustausch- und Spendeprozess einer rechtskonformen Einwilligung der Dateneigentümer entsprechen (DSGVO), weshalb ein Einwilligungsmanagement notwendig ist. O’Kane et al. zeigen, dass Einverständniserklärungen zur gemeinsamen Nutzung von Daten sich im Laufe eines Lebens dem sich ändernden Gesundheitszustand und der technologischen Erfahrung anpassen [308]. Ihnen zufolge sollte für eine Lösung die Möglichkeit in Betracht gezogen werden, eine gegebene Einwilligung im Hinblick auf die eigene aktuelle Datenschutzsensibilität des Patienten zu modifizieren. Die Studie von Grando et al. überprüft die Meinungen von Leistungsempfänger und -erbringer im Bereich der Verhaltensmedizin zu Einwilligungspräferenzen und der gewünschten Granularität beim Austausch von verhaltensmedizinischen Daten für die Versorgung oder Forschung [169]. Die Studie kommt zu dem Schluss, dass Patienten eine granulare Einwilligungskontrolle für ihre persönlichen Gesundheitsdaten wünschen. Aus der Sicht des Gesundheitsdienstleisters sollten die Patienten jedoch darin geschult werden, solche Entscheidungen mit Hinblick auf die Auswirkungen in der Versorgung informiert zu treffen. Die Herausforderungen bezüglich der Erhebung gemeinsamer Datengrundlagen für die Findung informierter Entscheidungen bildet deshalb den Ausgangspunkt des nächsten Abschnitts.

### 3.2.2 Datengrundlage zur gemeinsamen Entscheidungsfindung

Gegenwärtig sehen Odisho und Gore eine gemeinsame Entscheidungsfindung zwischen Patient und Leistungserbringer auf der Grundlage von Gesundheitsdaten als blockiert, da Wissensbarrieren die Patienten beeinträchtigen, ein effektiver Teil des eigenen Versorgungsprozesses zu sein [37]. Sie schlagen deshalb einen strukturierten Ansatz für eine patientenzentrierte Aufbereitung von Gesundheitsinformationen vor, um den Patienten zu helfen, ihre Informationen besser zu verstehen. Jourquin et al. sehen auch den Einbezug und die Übertragung der Kontrolle an die Patienten selbst als Herausforderungen für optimierte Therapieentscheidungen im Zusammenhang mit Brustkrebs [215]. Diesen stehen auch Chancen gegenüber, die in der Entwicklung und Implementierung von Entscheidungsunterstützungssystemen zur Ermöglichung einer personalisierten medizinischen Behandlung gesehen werden [270, 436, 437]. Durch die Verarbeitung von enormen Datenmengen unter Anwendung neuester technologischer Fortschritte und Methoden, können Entscheidungen auf einer anderen Grundlage als bisher getroffen werden. Wang et al. stellen ein gemeinsames Rahmenwerk für die Entscheidungsfindung vor, welches EHR-Daten verwendet, um anhand eines Klassifikationsmodells eine Liste möglicher Diabetes-Medikamente zu erstellen [437]. Dieser empfohlene Medikationsplan soll als Grundlage für die Beratung zwischen Arzt und Patient dienen. Goletsis et al. stellen eine Entscheidungsunterstützungsplattform vor, die auf Datenclustering-Verfahren basiert, welche Patienten mit ähnlichen klinischen Merkmalen und genetischen Veranlagungen identifizieren [436]. Sie verwenden ihre Ergebnisse zur Auswahl eines personalisierten Behandlungsplans, basierend auf der Annahme, dass Patienten mit einem ähnlichen Profil auch einen ähnlichen Behandlungsplan benötigen. Semantische Informationen, die mit Hilfe von Ontologien beschrieben werden, dienen als Input für deren Clustering-Methoden. Sqalli et al. stellen ein konzeptionelles KI-gestütztes Gesundheitscoaching-Modell und einen Prozess vor, der darauf abzielt, den Patienten bei der langfristigen Bewältigung seiner chronischen Erkrankung zu unterstützen, indem personalisierte Gesundheitspläne unter Verwendung von Sensordaten, maschinellem Lernen und Visualisierungswerkzeugen erstellt werden [270]. Krithara et al. stellen ein Analysesystem vor, welches Daten aus klinischen Notizen, Genomik, ePA und passender Literatur kombiniert, wodurch Muster identifiziert werden können, die dabei helfen sollen, die Gesundheitspolitik und individuelle Diagnose- und Behandlungsstrategien für Demenz und Lungenkrebs anzupassen [18]. Ihr Ziel ist es, Wissen für politische Entscheidungsträger abzuleiten. Sui-nesiaputra et al. konzentrieren sich auf die Erforschung von Herzkrankheiten [19]. Sie zeigen ebenfalls auf, wie wichtig die gemeinsame Nutzung und der offene Zugang zu Daten ist, um die Herzfunktionen besser zu verstehen und Therapien gegen Herzerkrankungen zu entwickeln. In diesem Zusammenhang verwenden Seth et al. Daten von einem Body Sensor Network (BSN), um die Herzfunktion für eine personalisierte Behandlung in Form von prädiagnostischen Verschreibungen zu modellieren [103]. Für das Training von Entscheidungsmodellen und den Aufbau solcher Entscheidungsunterstützungssysteme werden nach Li et al. große Datenmengen aus verschiedenen Gesundheitseinrichtungen benötigt, wobei nach Jaremko et al. aufgrund der Sensibilität der Daten dem Recht auf Privatsphäre Vorrang eingeräumt werden muss. Dafür müssen KI-Methoden so angepasst werden, dass sie funktional sicher und dennoch verständlich sind [207, 260]. Sajedi et al. überprüfen diesbezüglich verschiedene Algorithmen im Bereich der Kryptographie, digitalen Wasserzeichen und Steganographie für deren Anwendung in Gesundheitssystemen [186]. Sie heben den Zielkonflikt zwischen dem Verbergen und Analysieren von Informationen hervor und beschreiben die Herausforderungen der Sicherheit in Gesundheitssystemen. In Bezug auf die Herausforderung des Datenschutzes stellen Li et al. zwei, die Privatsphäre wahrende, verteilte Ansätze vor, die ohne Offenlegung sensibler Daten auf Patientenebene verwendet werden können, indem sie zunächst aus Daten von Einrichtungen lernen und dann diese lokal trainierten Modelle nutzen, indem diese zu integrierten Datenmodellen kombiniert werden [260]. Paddock et al.

konnten nachweisen, dass mittels homomorph verschlüsselter Daten das Echtzeit-Lernen in der personalisierten Krebsmedizin unterstützt werden kann, da es die Analyse der Daten ermöglicht und diese dabei jederzeit verschlüsselt bleiben [316]. Ihrer Ansicht nach ist anzumerken, dass die homomorphe Verschlüsselung nicht alle Datenschutzbedenken auflösen kann, da es insbesondere im Hinblick auf die Deidentifizierung möglich ist, Personen selbst aus gruppierten Proben von Genom-Daten zu identifizieren. Shahbaz et al. schlagen einen Anonymisierungsalgorithmus für DICOM-Bilder und textuelle Patienten- und Kommunikationsinformationen vor, um den Datenschutz und die Sicherheit der Daten in einem Cloud-basierten System zu gewährleisten [356]. Die Anonymisierung wird mit einer Zugangsrichtlinie kombiniert, so dass sie entsprechend der Rolle des Benutzers angewendet wird. Die Lösung eines flexiblen, standardisierten und datenschutzfreundlichen Verwaltungsansatzes für Benutzerprofile auf der Grundlage einer adaptiven erweiterten Merkle-Struktur von Sánchez-Guerrero et al. ermöglicht es den Patienten, Identitätsinformationen selektiv offenzulegen und die Privatsphäre zu schützen [338]. Diese Ansätze können die Wahrnehmung von Risiken für die Privatsphäre reduzieren und den Nutzen für den Einzelnen hervorheben, was neben der Zusicherung von politischen Verantwortlichkeiten und Bedingungen auch als wichtiger Aspekt für den Aufbau von Rahmenwerken für die Verwendung von Gesundheitsdaten erscheint [321].

### 3.3 Demenz-spezifische Potentiale und Lösungen

In der datengetriebenen Demenzforschung wird ein großes Potential für präzisionsmedizinische Ansätze gesehen, um im Gesundheitssystem mittels personalisierter Dienste mehr Transparenz und Effizienz zu realisieren [344, 428]. Einerseits zeigen bisherige pharmakologische Therapieansätze in klinischen Studien nicht die gewünschte Wirkung, andererseits bilden beispielsweise die Identifikation von reversiblen Risikofaktoren und Biomarkern eine Grundlage auf der Präzisionsmedizin aufbauen kann [176, 262, 344]. In computergestützten Ansätzen wird die Möglichkeit gesehen ein besseres Verständnis von den zugrundeliegenden Ursachen zu erhalten [428]. Damit kann der pflegerische Bedarf und die Behandlung entsprechend angepasst werden [262, 428]. Eine Standardisierung der Beurteilung einer Demenz zur objektiven Einschätzung wird hierbei als Ziel genannt [428]. Als Konsequenzen werden geeignete Dateninfrastrukturen gefordert, die ebenfalls computerisierte Ansätze aus der Psychiatrie und Neurologie integrieren sowie zur Entwicklung klinischer Entscheidungsunterstützungssysteme beitragen [428]. Herausforderungen hierfür sind von struktureller, technischer sowie ökonomischer Natur, liegen aber auch in der politischen Entscheidungsfindung, die mit solchen Implementierungen einhergehen [428]. Für die Translation dieser Ansätze wird nachfolgend zwischen den Ansätzen zu Früherkennung von Demenzen unterschieden, deren Notwendigkeit und Anforderungen an die Digitalisierung herausgestellt sowie das Potential computergestützter Automatisierung und Optimierung in diesem Bereich aufgezeigt.

#### 3.3.1 Abwägung der Mechanismen zur Früherkennung und Behandlung in der Primärversorgung

Obwohl sich für Demenz, insbesondere für die AD, bisher keine Behandlungsmaßnahmen etabliert haben, die eine Verbesserung der Symptomatik herbeiführen könnten [412], zeigen aktuelle Forschungsergebnisse neue Perspektiven auf. Der mit dem Fortschreiten der AD einhergehende Abbauprozess gesunder Neuronen führt zu Unterbrechungen der neuronalen Struktur, weshalb versucht wird durch die Entwicklung von Medikamenten, das weitere Absterben zu verhindern, indem die Ablagerung von Amyloid-Plaques

im Gehirn unterbunden werden soll, welche hypothetisch mit der Ursache des Prozesses verbunden werden (vgl. Unterunterabschnitt 2.2.4). Ein Beispiel hierfür liefert das Antikörperpräparat Aducanumab der Firma Biogen, welches zum Zeitpunkt der Anfertigung dieser Arbeit als eines der vielversprechendsten Medikamente gilt [439]. Ergebnisse einer randomisierten, doppelverblindeten und placebokontrollierten Studie in Phase 1b zeigten, dass die Konzentration der Plaques damit reduziert werden kann [375]. Für Patienten, bei denen eine starke Reduktion der Plaques verzeichnet werden konnte, wurde eine Stabilisierung der kognitiven Verschlechterung erreicht [375]. Darauf folgende Studien wurden bis in Phase 3 durchgeführt, jedoch im März 2019 vorzeitig abgebrochen, weil die Wirkung des Medikaments nicht statistisch signifikant nachgewiesen werden konnte. Aufgrund einer Ende 2019 durchgeführten Neuauswertung der Daten durch den Hersteller wurde teilweise Signifikanz nachgewiesen und eine Zulassung des Medikaments bei der FDA beantragt [369]. Im Juni 2021 wurde das Medikament vorläufig zugelassen, ein evidenzbasierter positiver Wirkungsnachweis ist aber noch zu erbringen. Da sich die bereits verlorenen neuronalen Verbindungen nicht wiederherstellen lassen, verbessert sich jedoch dadurch auch nicht die Symptomatik [412]. Deshalb stehen präventive Maßnahmen, die frühestmöglich ergriffen werden sollten, ebenfalls im Fokus [153, 262]. Für rund ein Drittel der weltweiten AD-Fälle werden bestimmte Risikofaktoren verantwortlich gemacht, die als potenziell reversibel gelten [203, 262]. Beispiele hierfür sind langfristig zu adressierende Faktoren wie der Bildungsgrad. Kurz- bis mittelfristig umzusetzende Interventionen, wie die Durchführung regelmäßiger physischer und geistiger Aktivitäten oder die pharmakologische Behandlung physiologischer Risikofaktoren, wie Bluthochdruck werden ebenfalls als wirksam betrachtet, um das Einsetzen einer Demenz hinauszögern zu können [203, 412]. Für die Realisierbarkeit dieser risikominimierenden Intervention fehlt aktuell jedoch ein klinischer Rahmen [203]. Insgesamt stellt sich dementsprechend die Frage, welche Diagnose- und Behandlungsverfahren zu welchem Zeitpunkt einsetzbar sind (vgl. Unterabschnitt 2.2.4). Um Demenzen möglichst früh zu erkennen, ist als erster Ansprechpartner im Gesundheitswesen die Primärversorgung einzubinden. Eine Studie aus den USA zeigt, dass jedoch bei klinischen Diagnosen von AD in der Primärversorgung die Sensitivität und Spezifität zwischen 71% und 97% beziehungsweise 44% und 71% liegt [48, 428]. Als Hauptgrund wird mangelndes Wissen von Medizinern über die Erkrankung und deren Diagnose angeführt, aber auch der Mangel an standardisierten Tests, welche ein routinemäßiges Screening innerhalb einer Sprechstunde ermöglichen [48, 428]. Die Wahl der Werkzeuge, die in der Primärversorgung ihre Anwendung finden, hängen laut Wong-Lin et al. einerseits von deren Sensitivität, Spezifität und Genauigkeit, andererseits vom zeitökonomischen und kostengünstigen Umgang sowie derer Gebrauchstauglichkeit ab [428]. Deshalb stehen neben Anamnesegesprächen, die kognitiven Beurteilungen, neuropsychologische Testungen sowie körperliche Untersuchungen und Überprüfungen der Medikamente als Instrumente zu Auswahl [428]. Verfahren wie die neurologische Bildgebung oder Liquordiagnostik, sind zumindest zum Zeitpunkt der Anfertigung dieser Arbeit, dem Zwecke der Abklärung beim Spezialisten in der Sekundärversorgung vorbehalten und nicht für die zeitökonomische Anwendung bei Allgemeinmedizinern oder gar im ambulanten Umfeld bestimmt [428].

### 3.3.2 Digitalisierung

Die zuvor gezeigten Abwägungen setzen die Möglichkeit voraus für die Diagnose und Behandlung relevante Daten effizient zu erfassen und zu verarbeiten. Um evidenzbasierte Lösungen zu ermöglichen und in den Alltag zu integrieren, spielt die Qualität dieser Daten eine gewichtige Rolle [428]. Hierbei ist ein Unterschied zwischen der Datenqualität aus der klinischen Forschung und dem medizinischen Alltag zu beachten [428]. Aufgrund der Vielfalt an Ausprägungen und mangelnden Standardisierung

von Demenzdiagnose und -behandlungsprozessen entstehen für die Demenzforschung multimodale Datenquellen [289, 428]. Hierfür wird ein Rahmenwerk als notwendig erachtet, welches das Thema der Datenverwaltung und die informierte Zustimmung regelt, wodurch eine effektive und ethische Nutzung der Daten ermöglicht und die Interessen der Datenspender berücksichtigt werden [289]. Für die Integration der dadurch möglichen longitudinalen Datensammlungen sind Verfahren notwendig, die eine Bereinigung, Aufbereitung und Harmonisierung dieser Datenbestände vor Analysen ermöglichen [289, 322]. Als Komplement für diese Ansätze, die den Umgang mit großen Datenmengen beschreiben, werden ebenfalls Forderungen nach der Erforschung des sogenannten „Small Data“ Paradigmas gestellt, um auf individueller Ebene für eine Person auf Grundlage ihrer Daten spezifische Verbesserungen zu erzielen [137, 187, 289]. Im Abgleich zwischen den Zielen der Forschung und Praxis wird darin der Vorteil gesehen, dass für multifaktorielle, individuelle und dynamische Zusammenhänge, die beispielsweise bei chronischen Erkrankungen vorliegen, durch agiles und schnelles Lernen mit weniger Daten, eine bessere Verwaltung dieser realisiert werden kann [187]. Auch im Zuge der Demenz rückt dieser Ansatz in den Fokus, da sich individuelle Profile aufgrund von unterschiedlichen Eigenschaften der Umgebung und des Lebensstils feststellen lassen, was einen individuellen Blick auf die Zusammenhänge vor Ort notwendig macht [428]. Insgesamt fordern Wong-Lin et al. eine systematische digitale Erfassung von Gesundheits- und Krankenakten, die eine zuverlässige und praktische Dateninfrastruktur voraussetzt, um Datenmodellierung und -analyse zu realisieren [428].

### **Digitales Demenzscreening**

Die Erzeugung der hierfür notwendigen Daten obliegt ebenfalls digitalen Werkzeugen, mit deren Hilfe Daten generiert, integriert und analysiert werden müssen. In Bezug auf die zuvor getroffene Abwägung spielen in der Primärversorgung die Anamnesegespräche, die kognitiven Beurteilungen, neuropsychologische Testungen sowie körperliche Untersuchungen und Überprüfungen der Medikamente als Instrumente eine Rolle (vgl. Unterabschnitt 3.3.1). Sogenannte digitale Marker könnten zusätzlich eingesetzt werden, um die Diagnose zu unterstützen [428]. Insbesondere die kognitiven Beurteilungen durch neuropsychologische Testungen sind daher in den letzten Jahren in den Vordergrund der Digitalisierungsbemühungen gerückt [402]. Hierbei werden verschiedene Schwerpunkte durch die Kombination digitaler Technologien und psychometrischen Verfahren gesetzt. Primär wird versucht, die zuvor vorgestellten Testformate von Stift und Papier in digitale Formate zu übertragen (vgl. Unterabschnitt 2.2.5). Einen Überblick über bestehende Werkzeuge geben Thabtah et al., die für den englischen Sprachraum im Jahr 2020 über die Distributionsplattformen von iOS und Android verfügbare mHealth Anwendungen systematisch gegenüberstellen, die nach Angaben der Entwickler für das Screening von Demenz verwendet werden können [401, 402]. Deren Fazit ist, dass keine Screening-Anwendung zur Früherkennung einer Demenz alle angelegten Evaluationskriterien, wie beispielsweise die Sensitivität und Spezifität aber auch die abgedeckten kognitiven Domänen, erfüllt. Insbesondere diese werden in keiner Anwendung den Leitlinien der DSM-V folgend komplett erfüllt, wodurch nicht die gesamte Bandbreite der entscheidenden Domänen abgebildet wird [402]. Darüber hinaus bestehen wissenschaftliche Untersuchungen und Konzepte bezüglich solcher Screening-Werkzeuge, deren Umsetzungen entweder nicht distribuiert sind oder nicht in englischer Sprache verfügbar und daher nicht in dem zuvor genannten Überblick erscheinen. Ein Beispiel hierfür bildet der Mobile Cognitive Screening von Zorluoglu et al. [444]. Nach Aussagen der Autoren wurde dieser für 72- bis 82-jährige Personen entworfen und stellt eine digitale neuropsychologische Testung bestehend aus 33 Fragen mit 14 Untertests dar. Insgesamt werden acht verschiedene kognitive Domänen getestet [444]. Hierfür sind bereits vergleichende Studien mit dem etablierten MoCA verfügbar (vgl.

Unterabschnitt 2.2.5). Demnach soll eine Unterscheidung zwischen gesunden und kognitiv beeinträchtigten Personen anhand eines reinen Scoring-Verfahrens statistisch signifikant möglich sein [444]. In anderen Ansätzen wurde beispielsweise das Ziel verfolgt, die MMSE zu digitalisieren (vgl. Unterabschnitt 2.2.5). Ob Demenztests mobil auf Tablet-Computern durchgeführt werden können, untersuchen Polgar et al., indem sie die Akzeptanz der Tests im Vergleich zu Tests mit Stift und Papier vergleichen [329]. Dabei werden in allen untersuchten Kategorien keine signifikanten Unterschiede zwischen analogen und digitalen Formaten festgestellt [329]. Prinzipiell scheint also ein mobiles Format auf Tablets geeignet zu sein. Ein vergleichbares Ergebnis schildern Lunardini et al., die unter anderem die Eignung des TMT in digitalisierten Versionen untersuchten (vgl. Unterabschnitt 2.2.5) [265]. Einen weiteren Ansatz verfolgen Lancaster et al., bei denen mit der Anwendung *GalleryGame* das Langzeitgedächtnis über mehrere Tage hinweg getestet werden soll, um die Hypothese zu prüfen, dass diese Domäne als früh betroffener Indikator für eine Demenz gilt [247].

Um ein Gefühl für den Markt und die kommerzialisierte Verwertung solcher Screening-Ansätze zu bekommen, werden nachfolgend als Beispiele der *BrainTest* und *Demenz Screening Test (DST)* angeführt. Diese digitalen Anwendungen wurden mit dem Ziel entwickelt, ein Demenz-Screening ohne Fachpersonal alleine von Zuhause aus durchführen zu können. Die für den englischsprachigen Raum realisierte Anwendung *BrainTest* enthält zwölf Aufgaben zu verschiedenen kognitiven Funktionen und wurde auf der Grundlage des Self-administered Gerocognitive Examination (SAGE) für die Früherkennung von Demenz und MCI entwickelt [66, 366]. Dieser eSAGE erreicht als digitalisierte Version in einer Studie mit 69 Teilnehmern eine 90-%ige Spezifität und 71-%ige Sensitivität in der Differenzierung von kognitiv eingeschränkten und normalen Probanden [367]. *BrainTest* ist dafür ausgelegt von Patienten selbstständig und ohne Unterstützung alle sechs Monate bearbeitet zu werden. Vor Beginn des Tests werden dem Nutzer einige Fragen zu dessen Wohlbefinden und gesundheitlichem Zustand gestellt. Anschließend muss der Nutzer unter anderem das aktuelle Datum eintragen, Bilder benennen und einen Würfel zeichnen. Die Durchführung der Tests dauert 10-15 Minuten und endet mit der Übermittlung von aufgezeichneten Testdaten zur Auswertung durch Mediziner [66]. Nach spätestens 48 Stunden erhält der Nutzer einen ausführlichen Bericht zu den Ergebnissen. Die Anwendung kann auf Tablet-Computern mit Android und iOS ausgeführt werden [66, 67]. Nach einem Probezeitraum von einem Monat fallen je nach gewähltem Preismodell Kosten in Höhe von 19,99 € (einmalige Testung) bis 39,99 € (zwei Testungen pro Jahr) an (Stand: Mai 2021) [66, 67].

Die erste Anwendung, die zum Zeitpunkt der Anfertigung dieser Arbeit als CE-zertifiziertes Medizinprodukt existiert, ist der DST [372]. Nach Herstellerangaben erreicht die Anwendung eine Spezifität von 91% sowie eine Sensitivität von 96% bei einer Durchführungsdauer von 10-15 Minuten, Studiendaten in Bezug auf die Anwendung liegen nicht vor, auf Rückfrage gibt es jedoch ein Literaturverweis auf den Computerized Self Test (CST) von Dougherty et al., an dem sich der DST orientiert [125]. Die Anwendung soll zur Erkennung verschiedener Demenzerkrankungen sowie deren Schweregraden eingesetzt werden [372]. Die Verarbeitung erfolgt auf dem Gerät und die Nutzer können ihre Ergebnisse speichern, um den Verlauf einer möglichen Demenz zu dokumentieren sowie bei Bedarf dem behandelnden Arzt ausgedruckt oder via E-Mail zu übermitteln [372]. Zu den Aufgaben in der Anwendung gehören beispielsweise das Zeichnen einer Uhr und die Eingabe des aktuellen Datums. Auf Grundlage der Antworten wird eine Punktzahl aufgrund eines hinterlegten Scoring-Modells zwischen 0 bis 30 errechnet. Punktzahlen von 29 oder 30 Punkten gelten als normal, 24 bis 28 Punkte als leicht eingeschränkt. Werden weniger als 23 Punkte erreicht, wird von demenzbedingten Einschränkungen der kognitiven Leistungen ausgegangen [402]. Die Anwendung ist für Android- und iOS-Geräte sowohl in deutschsprachigen als auch in englischsprachigen Ländern verfügbar und kostet einmalig 5,49 € [372, 373].

## Altersbedingte Auswirkungen auf die Gestaltung digitaler Werkzeuge

Die zuvor beschriebenen Anwendungen fokussieren sich auf die Übertragung der ursprünglich analogen Darstellung psychometrischer Tests in deren digitales Äquivalent. Da sich die Zielgruppe, aufgrund des dominanten Risikofaktors Alter, primär auf Menschen mit altersbedingten Beeinträchtigungen bezieht, sind diese Besonderheiten bei der Gestaltung digitaler Werkzeuge zu berücksichtigen. In der wissenschaftlichen Literatur finden sich hierzu seniorengerechte Gestaltungsempfehlungen, die sich mit der Gesamtheit einer Anwendung beschäftigen und solche, die einzelne Aspekte in diesem Zusammenhang beleuchten. Darvishy et al. identifizieren Faktoren, die von Senioren bei mobilen Smartphone-Anwendungen als wichtig eingestuft werden und worin die größten Schwierigkeiten bei deren Nutzung liegen [105]. Hierbei stellen sich sowohl die Übersichtlichkeit der Anwendung als auch die Einfachheit und sichere Bedienung als die wichtigsten Aspekte für Senioren heraus [105]. Um die Bedürfnisse und Anforderungen von Senioren an Smartphones und Tablets zu erforschen sowie Richtlinien für eine Verbesserung der Gebrauchstauglichkeit zu entwickeln, präsentieren Sano et al. und Zaphiris et al. Gestaltungsempfehlungen, die vergleichbar aufgebaut sind [361, 442]. Eine Erkenntnis ist, dass eine für Senioren benutzerfreundlich gestaltete Anwendung insgesamt die Usability für alle Nutzer steigern kann [134]. Inwiefern sich Gestaltungsempfehlungen einer Benutzerschnittstelle für ältere Nutzer mit den Gestaltungsrichtlinien der verschiedenen Betriebssysteme vereinen lassen untersuchen Friedrich und Kunze, indem für ausgewählte Interaktionselemente, wie das Scrollen, verschiedene Varianten implementiert und evaluiert werden [154]. Dabei konnte festgestellt werden, dass Flat Design, das mit flachen anstelle von räumlichen Elementen arbeitet und teilweise in den Gestaltungsrichtlinien der Betriebssysteme empfohlen wird, zu vermeiden ist [154]. Des Weiteren untersuchen Noichl et al., welche Symbolik von Senioren intuitiv erkannt wird und wie zielgruppenspezifische Tastaturdarstellungen sowie adaptive Einstellungsmenüs gestaltet sein müssen, um deren Bedürfnisse zu berücksichtigen [304, 305].

### 3.3.3 Computergestützte Automatisierung und Optimierung

Neben den, zu diagnostischen Zwecken üblichen reinen Scoring-Modellen, sehen Thabtah et al. auch in der Anwendung von Verfahren des ML und der KI zur Auswertung dieser Tests ein großes Potenzial für die Verbesserung der Genauigkeit und Effizienz dieser Werkzeuge, das in künftigen Forschungsarbeiten ausgeschöpft werden soll [401]. Dieser potenzielle Nutzen wird von weiteren Autoren gesehen und eruiert. Maroco et al. werten die Daten von neuropsychologischen Tests mithilfe von ML-Verfahren aus und erzielen damit eine höhere Diagnosegenauigkeit als durch herkömmliche Diskriminanzanalysen [277]. Ähnliche Untersuchungen stellten auch Yin et al., Ang et al. und Zhu et al. an [36, 440, 443]. Für diese Vorgehensweise werden neben der reinen Betrachtung neuropsychologischer Verfahren auch Potentiale in der Kombination mit Daten aus weiteren medizinischen Bereichen, wie beispielsweise der Neurologie für die Entwicklung computergestützter Werkzeuge zur Diagnose gesehen [428]. Durch diese datengetriebenen Modelle erhofft man sich objektivere Methoden zur Erkennung und Vorhersage einer Demenz durch zusätzliche Marker. Potential liefern hierbei insbesondere Verfahren der neurologischen Bildgebung, welche durch Deep Learning oder durch die Kombination verschiedener Bildgebungsverfahren, menschliche Experten bei der Erkennung des Schweregrads von Demenz in der Genauigkeit übertrumpfen [336, 428, 441]. Diese Ansätze können zur Merkmalsselektion und Dimensionsreduktion eingesetzt werden und somit zur Identifikation und Eingrenzung von effektiven Markern verwendet werden [255, 359]. Orrù et al. sehen Verfahren des ML zur Analyse psychologischer Experimente als komplementär zu Methoden der Inferenzstatistik, um biologisch sinnvolle Schlüsse zu ziehen [312]. Sie

sehen insbesondere den Vorteil, dass die Wahl des statistischen Verfahrens mit dem Wissen über das zu modellierende System einhergeht, während die Wahl des ML-Verfahrens auf deren empirischen Fähigkeiten im Umgang mit Daten sowie deren Vorhersagegenauigkeit basiert [312]. Des Weiteren sehen sie in der Verwendung klassischer Inferenzstatistik durch ein falsches Verständnis von p-Werten die Gründe für die Replikationskrise innerhalb der psychologischen Forschung, die durch die Verwendung von in der ML-Entwicklung etablierten Kreuzvalidierungsverfahren, adressiert werden kann [312]. Ansart et al. geben einen Überblick zu Literatur, die sich mit Prognosemodellen für die Entwicklung von MCI mittels ML-Verfahren befassen [38]. Hierbei werden neben einer quantitativen Analyse der Auswirkungen von Entscheidungen auf die Modellleistung auch eine Übersicht über die Probleme der dieser zugrundeliegenden Studien identifiziert [38]. Als Ergebnis stellt sich heraus, dass insbesondere auch die Einbeziehung kognitiver Variablen, beispielsweise im Gegensatz zu Bilddaten aus einem MRT die Vorhersageleistung der Modelle signifikant verbessern [38]. Auf dieser Grundlage plädieren sie für eine weitere Ausgestaltung domänenspezifischer Beurteilungen der kognitiven Leistungen aber auch eine Verbesserung der wissenschaftlichen Qualitätssicherung, da diese Studien als Grundlage zur Unterstützung medizinischer Entscheidungen verwendet werden [38].

Insgesamt können computergestützte Automatisierungs- und Optimierungsansätze relevante von irrelevanten Parametern durch Merkmalsselektion trennen und genauere Modelle zur Diagnose und Prädiktion bei Demenz liefern, jedoch sind dem potentiellen klinischen Mehrwert der damit ermittelten Marker und neu etablierter Verfahren eine Kosten-Nutzen-Analyse gegenüberzustellen [38, 428].

### **3.3.4 Vertrauenswürdigkeit computergestützter Modelle**

Um Ungewissheiten bezüglich ML-gestützten Entscheidungen auszuräumen, wird es als notwendig erachtet, die Merkmale, welche ein Modell für die Vorhersage verwendet, transparent darzulegen [384]. Darauf aufbauend wird eine ethisch vertretbare und zuverlässige Verwendung dieser Verfahren nur dann gewährleistet, wenn auch die Algorithmen erklärt und damit deren Entscheidungen nachvollzogen werden können [343]. Durch Erfüllung der beiden Eigenschaften Transparenz und Erklärbarkeit entsteht Vertrauenswürdigkeit, welche somit als die Grundvoraussetzung zur Anwendung dieser Verfahren im Gesundheitswesen gilt [343, 384]. Diese werden wiederum ebenso als Grundlage zur Erfüllung der Rechenschaftspflicht gemäß der DSGVO gefordert [192] (vgl. Absatz 2.3.2).

Um die zuvor beschriebenen computergestützten Modelle für die Anwendung in der medizinischen Praxis zu nutzen, werden die in Unterunterabschnitt 2.4.3 beschriebenen Verfahren zur Interpretierbarkeit von Modellen notwendig. Der diesbezügliche Stand der Wissenschaft wurde angelehnt an die Taxonomie von Molnar durch eine strukturierte Literaturrecherche erschlossen [292]. Hierbei wurden verschiedene medizinische Anwendungen von XAI untersucht und Teilergebnisse durch den Autor bereits publiziert [468]. Insbesondere modell-agnostische Verfahren sind in den letzten Jahren im Diskurs der wissenschaftlichen Forschung. Hierbei treten in der Literatur Local Interpretable Model-agnostic Explanations (LIME) und Shapley Values Additive Explanations (SHAP) als am häufigsten genutzte Verfahren auf, weshalb diese beiden im Folgenden genauer erläutert werden sollen. Des Weiteren wird auf modell-spezifische Verfahren eingegangen, die zur Erklärung konkreter Algorithmen entworfen werden.

## Local Interpretable Model-agnostic Explanations

Mit LIME wird das Ziel verfolgt Erklärungen auf lokaler Ebene für einzelne Beobachtungen eines zu erklärenden Modells heuristisch zu generieren [345]. Hierfür werden interpretierbare Modelle als Surrogaten zur Vereinfachung eingeführt, um die von dem Black-Box-Modell generierten Vorhersagen zu approximieren. Die Entscheidungslogik des zu erklärenden Modells wird dabei nicht betrachtet [345]. Ausgangspunkt für das Verfahren stellt die zu erklärende Beobachtung dar. Anhand dieser wird für jedes Attribut des Datensatzes eine Verteilung angenommen, womit zusätzliche synthetische Daten als Beobachtungen generiert werden [292, 345]. Mittels dieser Datenpunkte lassen sich durch das zu erklärende Modell Ausgangswerte prognostizieren. Daraus entsteht ein permutierter Datensatz, indem die Datenpunkte entsprechend ihrer Distanz zur zu erklärenden Beobachtung gewichtet werden [345]. Damit wird ein interpretierbares Modell als Surrogat trainiert, um eine lokale Näherung an das Black-Box-Modell zu erreichen [345]. Diese Näherung hängt maßgeblich von der gewählten Gewichtungsfunktion ab [292]. Schematisch lässt sich das Verfahren wie in Abbildung 3.1 darstellen [345]. Der zu erklärende Datenpunkt  $x$  ist umgeben von gewichteten und permutierten Beobachtungen. Um die Lokalität von  $x$  zu definieren wird  $\pi_x$  als Maß verwendet, um die Nähe zwischen einer generierten Instanz und  $x$  zu gewichten. Hierbei stellt die schwarz-grau vertikal schraffierte Fläche die Entscheidungsfunktion  $f$  zwischen zwei Klassen des zu erklärenden Black-Box-Modells dar. Die Approximation dieser Funktion wird in diesem Fall durch eine lineare Regressionsgerade  $g$  dargestellt. Damit lässt sich eine lokale Entscheidungsfunktion für die zu erklärende Beobachtung finden, wobei diese auf globaler Ebene nicht auf die Logik des zu erklärenden Modells eingeht [345]. Gleichung 3.1 beschreibt den Ansatz mathematisch [345]:

$$\varepsilon(x) = \underset{g \in G}{\operatorname{argmin}} L(f, g, \pi_x) + \Omega(g) \quad (3.1)$$

Ziel ist es eine Erklärung  $\varepsilon(x)$  von der Vorhersage des Black-Box-Modells  $f$  für die Beobachtung  $x$  zu finden. Dabei soll die Verlustfunktion  $L$  in Abhängigkeit von einer Familie der erklärbaren Modelle

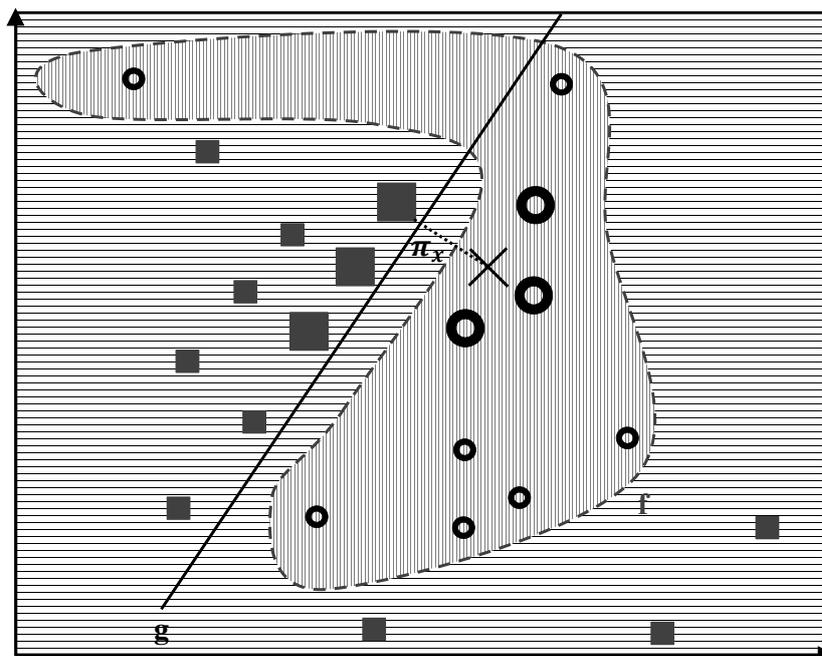


Abbildung 3.1: Schematische Darstellung von LIME nach Ribeiro et al. [345].

$G$  bei möglichst niedriger Modellkomplexität  $\Omega(g)$  minimiert werden. Die Komplexität wird durch den Anwender bestimmt, indem beispielsweise die maximale Anzahl von einzubeziehenden Features definiert wird [292].

Im Kontext medizinischer Anwendungen wurde LIME beispielsweise zur Erklärung von durch ein Convolutional Neural Network (CNN) interpretierten histologischen Befunden eingesetzt und zeigte eine menschenähnliche Erklärbarkeit [317]. Auch in einer Fallstudie von Kutawal et al. zur Vorhersage der Sterblichkeit von Patienten auf Intensivstationen konnte positiver Nutzen aus der Anwendung des Verfahrens gezogen werden [225].

### Shapley Values Additive Explanations

Das Verfahren der Shapley-Werte (Shapley Values) basiert auf der Koalitionsspieltheorie [292, 378]. Hierbei werden anhand einer Koalitionssfunktion Auszahlungen an Spieler zugewiesen, deren Höhe an ihrem Anteil von der Gesamtauszahlung bemessen wird [378]. Übertragen auf ML lässt sich damit für jedes Merkmal eines Modells eine Aussage über dessen Einfluss auf die Vorhersage treffen [292]. Die Vorhersage einer zu erklärenden Beobachtung wird somit mit einem Spiel gleichgesetzt, welches von koalierenden Spielern in Gestalt von Werten unterschiedlicher Merkmale und dem Gewinn als Delta zwischen der Vorhersage einer zu erklärenden Merkmalsausprägung und der durchschnittlichen Vorhersage aller bekannten Instanzen modelliert wird [292]. Über mehrere Beobachtungen hinweg lassen sich damit die Einflüsse der Merkmale aggregieren, um deren globalen mittleren Einfluss zu bestimmen [292]. Exemplarisch bedeutet das in einem Merkmalsraum  $A, B, C$  für ein zu erklärendes Merkmal  $A$ , dass alle möglichen Koalitionen der übrigen Merkmale  $B$  und  $C$  gebildet werden [292]. Die hieraus entstehenden vier Koalitionsmöglichkeiten  $\{\emptyset, B, C, BC\}$  werden verwendet, um anhand des Black-Box-Modells weitere Beobachtungen zu generieren [292]. Hierbei werden alle Werte für Merkmale, die nicht Teil einer Koalition sind durch zufällige Werte des entsprechenden Merkmals einer Instanz aus dem Datensatz ersetzt [292]. Für das zu erklärende Merkmal  $A$  wird in dieser Konstellation zunächst der wahre Ursprungswert angenommen und dieser danach ebenfalls durch einen zufälligen Wert einer Instanz ersetzt [292]. Durch die Differenz der Prognosen zwischen dem zu erklärenden Ursprungswert und der durch einen zufälligen Wert generierten Prognose ergibt sich der marginale Beitrag des zu erklärenden Merkmals [292]. Über alle möglichen Merkmalswertkombinationen in  $S$  hinweg entsteht der, durch die in Gleichung 3.2 und Gleichung 3.3 gewichtete und summierte, Shapley-Wert für das Merkmal  $j$  [292]:

$$\phi_j(val) = \sum_{S \subseteq \{1, \dots, p\} \setminus \{j\}} \frac{|S|! (p - |S| - 1)!}{p!} (val(S \cup \{j\}) - val(S)) \quad (3.2)$$

$$val_x(S) = \int \hat{f}(x_1, \dots, x_p) d\mathbb{P}_{x \notin S} - E_X(\hat{f}(X)) \quad (3.3)$$

Als nachteilig wird für dieses Verfahren der exponentiell zunehmende Rechenaufwand bei zunehmender Zahl von Merkmalen gesehen sowie die Verrechnung korrelierender Merkmale durch das zufällige Ersetzen von Merkmalsausprägungen, was zu Kombinationen führen kann, die so nicht möglich sein können [292]. Dem gegenüber steht neben der Flexibilität modell-agnostischer Ansätze durch die Shapley-Werte ein mathematisch fundiertes Verfahren als beweisbare Grundlage von Erklärungen [292], da die damit einhergehende Koalitionssfunktion als einzige alle Axiome und damit die Eigenschaften der Null-Spieler, Symmetrie, Pareto-Effizienz und Additivität für eine gerechte Zuteilung erfüllt [292].

SHAP baut auf der Erklärung durch Shapley-Werte auf und stellt ebenfalls ein modell-agnostisches Verfahren zur Suche nach einer lokalen Erklärung eines Black-Box-Modells dar [292]. Hierbei werden

die Shapley-Werte durch ein lineares Modell repräsentiert und wie beim Ansatz von LIME (vgl. Unterunterabschnitt 3.3.4) durch ein interpretierbares Modell dargestellt [292, 371]. Diese Vorgehensweise wird Additive Feature Attribution Method genannt [371]. Grundlage bildet das interpretierbare Modell  $g$ , welches das Modell  $f$  erklären soll [292]. Hierbei stellt  $z' \in \{0, 1\}^M$  den Koalitionsvektor dar, mit  $M$  wird die maximal mögliche Koalitionsgröße angegeben sowie durch  $\phi_j \in \mathbb{R}$  der Shapley-Wert eines Merkmals  $j$  bestimmt (vgl. Gleichung 3.4) [292]. Durch den binären Vektor  $z'$  wird entschieden, welche Merkmale und damit deren Shapley-Werte in einem interpretierbaren Modell  $g$  aufgenommen werden [292]. Dadurch liefert  $g$  ein lineares Modell der Koalitionen worin die Werte von  $\phi$  wie in Gleichung 3.2 berechnet werden [292].

$$g(z') = \phi_0 + \sum_{j=1}^M \phi_j z'_j \quad (3.4)$$

Um die zu erklärende Instanz  $x$  zu berechnen, werden alle Merkmale aktiviert und erhalten in  $x'$  den Wert 1 (vgl. Gleichung 3.5) [292].

$$g(x') = \phi_0 + \sum_{j=1}^M \phi_j \quad (3.5)$$

Der für SHAP zusätzlich zu den Eigenschaften der Shapley-Werte notwendige Umgang bezüglich des Fehlens von Werten und dessen potentiellen Einfluss auf das Ergebnis wird von Lundberg und Lee durch die Eigenschaft  $x'_j = 0 \Rightarrow \phi_j = 0$  definiert, um für fehlende Merkmale immer einen Shapley-Wert von 0 zu erhalten [292, 371]. Für die Umsetzung von SHAP schlagen Lundberg und Lee KernelSHAP und TreeSHAP als Berechnungsmethoden vor [266, 371]. Molnar fasst den KernelSHAP-Algorithmus in den folgenden fünf Schritten zusammen [292]:

1. Systematische oder zufällige Auswahl der Koalitionen über  $z'_k \in \{0, 1\}^M$ ,  $k \in \{1, \dots, K\}$ . Im Fall systematischer Vorgehensweise werden hauptsächlich Koalitionen mit besonders wenigen oder besonders vielen Koalitionspartnern gebildet, da diese die Schätzungen verbessern [292].
2. Generierung der Vorhersagen für alle  $z'_k$  durch die Transformation dieser auf den ursprünglichen Merkmalsraum zur Anwendung des Originalmodells  $\hat{f} : \hat{f}(h_x(z'_k))$ , wobei  $h_x(z') = z$ ,  $h_x : \{0, 1\}^M \rightarrow \mathbb{R}^p$  um die ausgewählten Merkmale in den Koalitionen auf die korrespondierenden Werte der zu erklärenden Instanz  $x$  abzubilden [292]. Nicht ausgewählte Merkmale erhalten Werte einer anderen Instanz aus dem Datensatz [292].
3. Berechnung der Gewichtungen für alle  $z'_k$  mit der Kernelfunktion in Gleichung 3.6.

$$\pi_x(z') = \frac{(M-1)}{\binom{M}{|z'|} |z'| (M-|z'|)} \quad (3.6)$$

Für eine lineare Regression mit diesem Kernelgewicht ergeben die geschätzten Koeffizienten die Shapley-Werte [292, 371].

4. Erstellung eines linearen Modells aus Gleichung 3.4 anhand der Optimierung der Verlustfunktion in Gleichung 3.7.

$$L(\hat{f}, g, \pi_x) = \sum_{z' \in Z} [\hat{f}(h_x(z')) - g(z')]^2 \pi_x(z') \quad (3.7)$$

Hierbei sind  $Z$  die Trainingsdaten und die geschätzten Koeffizienten des Modells sind die Shapley-Werte [292].

### 5. Rückgabe der Shapley-Werte $\phi_k$ des linearen Modells.

Ein grundlegender Unterschied im Vergleich zu LIME stellt die Gewichtung der Instanzen des Surrogatmodells dar [292]. Während LIME die generierten Instanzen über die Distanz zur erklärenden Instanz gewichtet, nutzt KernelSHAP die Gewichtung, welche die Koalition nach der Schätzung der Shapley-Werte erhält [292]. Die Shapley-Werte besitzen eine solide mathematische Grundlage, wovon letztlich das SHAP-Verfahren ebenso profitiert [292]. Außerdem kann SHAP potentiell auch für globale Erklärungen verwendet werden [292]. Als nachteilig wird der hohe Rechenaufwand gesehen sowie bei allen permutationsbasierten Verfahren, dass die Struktur der Abhängigkeiten zwischen den Merkmalen nicht berücksichtigt wird, wodurch unrealistische Kombinationen entstehen können, welche einen Einfluss auf die Ergebnisse haben [292]. Außerdem wird bei dieser Art von Verfahren der Zugang zu den Trainingsdaten des Black-Box-Modells vorausgesetzt [183].

In der Literatur finden sich bereits Anwendungen von SHAP im medizinischen Kontext. Tseng et al. erklären damit die Ergebnisse von ML-Modellen, welche die Auswirkung von intraoperativen Variablen bei akuten Nierenschädigung aufgrund herzchirurgischer Eingriffe darstellen und damit Ursachen identifizieren, die durch andere Verfahren nicht ersichtlich wurden [407]. Zur Erklärung ML-gestützter Diagnosen von AD setzen El-Sappagh et al. ebenfalls das SHAP Verfahren ein [130]. Hierbei werden Daten von psychometrischen Tests und systembiologischen Diagnoseverfahren verwendet um einen RF zu trainieren und dessen Vorhersagen anschließend unter anderem mittels SHAP zu erklären [130]. Die Erklärungen, die SHAP den Ärzten liefert, werden als konsistent und zuverlässig bewertet [130].

## Modell-spezifische Erklärungen

Im Gegensatz zu modell-agnostischen Verfahren sind modell-spezifischen Verfahren auf bestimmte Arten von Modellen zugeschnitten (vgl. Unterunterabschnitt 2.4.3). So schlagen beispielsweise Lundberg et al. für ML-Verfahren mit Baumstrukturen eine modellspezifische Alternative zu dem zuvor beschriebenen KernelSHAP vor [266]. Hierfür präsentieren sie eine angepasste Variante namens TreeSHAP, welche zwar die kritisierte Rechenkomplexität im Vergleich zu KernelSHAP reduziert, jedoch teilweise nicht intuitive Zuweisungen von Merkmalen generiert [266, 292]. Daneben werden dieser Gruppe von Verfahren insbesondere intrinsisch interpretierbare Modelle zugeordnet, welche per Definition modell-spezifisch sind [292]. Ein Beispiel hierfür sind Entscheidungsbäume, welche durch ihre Struktur als selbsterklärendes Modell konstruiert sind und deren Erklärungen in Form von Wenn-Dann-Regeln abgeleitet werden können [343]. Da diese, die in Absatz 2.4.3 beschriebenen Nachteile aufweisen, werden mehrere zusammen in einem Ensemble als RF verwendet, wodurch die Übersichtlichkeit der Ergebnisse von Entscheidungsbäumen verloren geht [292]. Hatwell et al. schlagen diesbezüglich ein modell-spezifisches Verfahren namens Collection of High Importance Random Path Snippets (CHIRPS) vor, welches auf einem Verfahren zur Erklärung einer AdaBoost-Klassifikation basiert [182, 183]. Hierfür wird die Eigenschaft genutzt, dass sich RF aus Entscheidungsbäumen zusammensetzen, indem aus jedem dieser ein Entscheidungspfad extrahiert wird, der zum Mehrheitsvotum im Rahmen der Klassifikationsentscheidung beigetragen hat [183]. Hierzu kommt FPM zum Einsatz (vgl. Absatz 2.4.3) [182, 183]. Die Ergebnisse werden im Anschluss analysiert und versucht Erklärungen abzuleiten, weshalb eine zu erklärende Instanz  $x$  entsprechend klassifiziert wurde [183]. Die Vorgehensweise wird grundsätzlich in folgende vier Schritte (vgl. Abbildung 3.2) unterteilt [183]:

1. Klassifikation von Instanz  $x$  und Identifikation aller Entscheidungspfade anhand der Knoten jedes Entscheidungsbaums [183]. Eine Entscheidung lässt sich pro Knoten als Ergebnis einer Unglei-

chung  $x_j < v$  darstellen [183]. Dabei ist  $x_j$  die Ausprägung des Merkmals  $j$  und  $v \in \mathbb{R}$  der entsprechende Schwellwert [183]. Als  $\tau \in \{0, 1\}$  wird die binäre Entscheidung der Auswertung definiert [183]. Jeder Knoten lässt sich durch das Tripel  $(j, v, \tau)$  beschreiben [183]. Wobei deren Interpretation von der entsprechenden Domäne von  $\chi_j$  abhängt. Damit lässt sich eine Filterung aller Entscheidungspfade, die eine Schnittmenge mit der Mehrheitsentscheidung bilden durchführen [183].

2. Durchführung von FPM (vgl. Absatz 2.4.3) [183]. Rückgabe der High Importance Random Path Snippets (HIRPS) als Itemsets und deren *support*-Werte [183].
3. Berechnung der Rangfolge anhand der heuristischen Bewertungsvorschrift in Gleichung 3.8 [183].

$$Score(ps, w, sup, \alpha) = w \times sup \times \frac{(|ps| - \alpha)}{|ps|} \tag{3.8}$$

Hierbei ist  $ps$  das HIRPS,  $|ps|$  die Kardinalität des HIRPS,  $sup$  dessen *support*-Wert und  $\alpha$  ein Tuning-Parameter, um den *support*-Wert entsprechend der Kardinalität der HIRPS mittels einer Konstante anzupassen [183].  $w$  ist eine weitere Gewichtung, die den Standardwert 1 hat, eine verbesserte Leistung wurde von Hartwell et al. jedoch beobachtet, wenn für  $w$  in Abhängigkeit von  $ps$  die Kullback-Leibler-Divergenz verwendet wird. Die Rangfolge ergibt sich aus den absteigend sortierten Scores [183].

4. Generierung einer Assoziationsregel mittels Greedy-Algorithmus [182] und Rückgabe der damit einhergehenden Qualitätsmetriken (Confidence und Support, vgl. Absatz 2.4.3) [183].

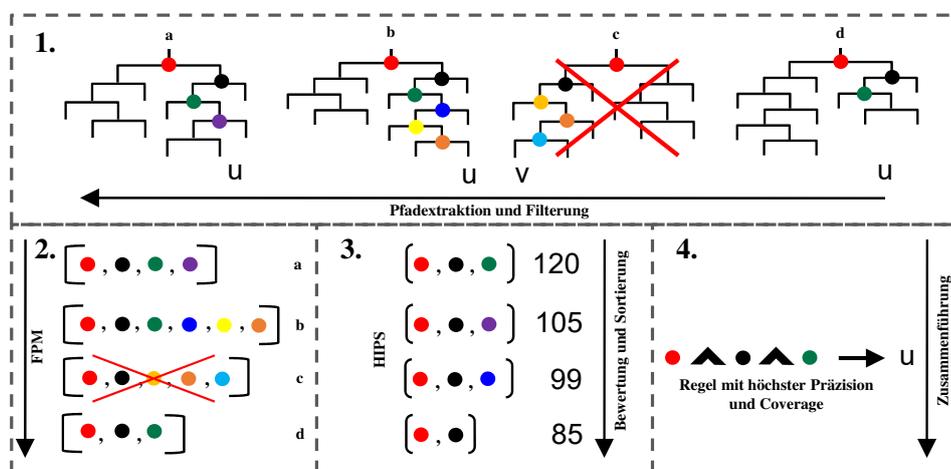


Abbildung 3.2: Schematische Darstellung von CHIRPS nach Hatwell et al. [183].

In verschiedenen medizinischen Bereichen zeigt die Anwendung von RF eine gute Performance. Wang et al. nutzen diese beispielsweise zur Erkennung epileptischer Anfälle anhand von Signalen einer Untersuchung mittels Elektroenzephalografie (EEG) [416]. Auch bezüglich der Prognose von Depressionen bei Menschen mit AD konnten damit in einer Forschungsarbeit bereits Verbesserungen erzielt werden [93]. Daleen et al. zeigen wie ein Decision Tree zur Vorhersage von AD bei 17 Patienten erfolgreich eingesetzt werden konnte und welche Merkmale dabei relevant sind [22]. Jedoch wurde keine Evaluation der Erklärung durchgeführt. Das CHIRPS-Verfahren könnte hiermit auch im Kontext von Demenz nützlich sein.

### 3.3.5 Translation in die Versorgung

Aus biomedizinischer Perspektive definiert die European Society for Translational Medicine (EUSTM) die Translationale Medizin (TM) als einen interdisziplinären Zweig, der sich mit Fragestellungen im Kontext der Transition und darüber hinaus der Rückkopplung medizinischen Wissens aus dem Labor über das Krankenbett bis in die Gesellschaft beschäftigt [98]. Hierbei sollen Methoden und Werkzeuge kombiniert und etabliert werden, um den Zielen der Präventions-, Diagnose- und Therapieoptimierung folgend, letztlich zur Verbesserung der Gesundheitspolitik insgesamt beizutragen [98]. Nachfolgend werden die in dieser Arbeit erörterten Ansätze der digitalen Medizin mit der ursprünglich auf den Bereich der Biomedizin bezogenen Definition von TM verknüpft. Um den Nutzen jüngster Forschungsergebnisse zeitnah in die Praxis zu transferieren ist ein Katalysator notwendig. Die Vorteile einer frühen Diagnose von Demenz stellen neben der rein medizinischen Verbesserung für Patienten auch aus ökonomischer Sicht eine Entlastung der Pflege- und Krankenkassen dar (vgl. Unterabschnitt 2.2.3 und Unterabschnitt 2.2.4). Da Demenz häufig nicht erkannt wird, hat die Regierung Großbritanniens für Allgemeinmediziner in der Primärversorgung eine Fallpauschale für jede diagnostizierte Demenz als Anreiz gesetzt [332, 428]. Als Folge sind negative Patientenerfahrungen aufgrund von falsch-positiv Diagnosen, wachsende Wartelisten in Gedächtniskliniken und die Inanspruchnahme nicht notwendiger medizinischer Leistungen zu beobachten [428]. Daher werden Werkzeuge benötigt, um Mediziner in der Primärversorgung informierte klinische Entscheidungen zu ermöglichen. Hierfür werden Systeme der Kategorie Clinical Decision Support (CDS) als Möglichkeit gesehen, diese Abwägungsentscheidungen evidenzbasiert zu unterstützen [381]. Vor allem das Potential telemedizinischer Anwendungen durch die Verbreitung mobiler und auch tragbarer Technologien wird im Kontext dementieller Erkrankungen gesehen [236, 428]. Neben diesem Potential werden auch Herausforderungen in ethischen und rechtlichen Aspekten rund um die Freigabe von Daten durch kognitiv eingeschränkte Personen sowie Anforderungen an die Datensicherheit gesehen [199]. Die Vorteile, welche durch die Identifikation zusätzlicher digitaler Biomarker, die Standardisierung der Diagnose- und Behandlungspfade sowie die Verbesserung der Zugänglichkeit zu Diagnose- und unterstützenden Pflegedienstleistungen gesehen werden, überwiegen jedoch [199, 236, 299, 428]. Allerdings werden hierfür longitudinale Beobachtungsstudien notwendig, um die Variabilität zwischen und innerhalb von Probanden zu berücksichtigen [236]. Die Förderung des Aufbaus dafür geeigneter Dateninfrastrukturen ist eine Grundvoraussetzung dafür und bildet die Grundlage für den Vorschlag einer nächsten Stufe von datengetriebenen Demenzklassifikationen mittels computergestützten Modellen anstelle von konsentierten Leitlinien [199, 428]. Durch die in Abschnitt 2.3 beschriebenen gesetzlichen Änderungen bezüglich der Monetarisierung digitaler Gesundheitsanwendungen in der Regelversorgung und Verwendung damit generierter Gesundheitsdaten in Forschung und Entwicklung entsteht die Chance, diese Potenziale in Deutschland tatsächlich zukünftig zu realisieren.

## 3.4 Kritik

In den Abschnitten zuvor wird der zum Zeitpunkt der Anfertigung dieser Arbeit aktuelle Stand der Wissenschaft und Technik dargestellt. Nachfolgend werden Forschungslücken herausgearbeitet, welche durch die definierten Forschungsfragen beantwortet und zur angestrebten Problemlösung abgegrenzt werden. Generell lässt sich in Unterabschnitt 3.1.1 und Abschnitt 3.3 ein Mangel an Standardisierung feststellen. Bemühungen diesen zu beheben, äußern sich in zahlreichen Initiativen, welche die interoperable Verwaltung von Gesundheitsdaten und deren nachhaltige Verwertung anstreben (vgl. Unterabschnitt 3.1.2). Die

Berücksichtigung der Anforderungen zukünftiger Gesundheitsdateninfrastrukturen und regulatorische Rahmenbedingungen werden in die Entwicklung von digitalen Gesundheitsanwendungen einzubeziehen sein, weshalb die Identifikation und Integration dieser auch Teil von Lösungsincrementen sein muss. Von den dadurch zu erwartenden datengetriebenen Ansätzen werden neben dem Potential zur gemeinsamen Entscheidungsfindung auch Möglichkeiten für die Entwicklung personalisierter Diagnose- und Behandlungsansätze gesehen (vgl. Abschnitt 3.2). Hierbei kann vor allem die Primärversorgung profitieren, welche durch digitale Unterstützungslösungen aus der Sekundärversorgung geleitete Behandlungsprozesse begleiten oder gar initiieren kann. Für Demenzen spielt vor allem deren Früherkennung eine Rolle, die Studien zufolge derzeit mit stark variierenden Erkennungsraten erfolgt, da letztlich auch Wissen über die korrekte Durchführung auf Seiten der Leistungserbringer fehlt [48, 428]. Hierbei kann die Digitalisierung und computergestützte Auswertung eine sinnvolle Unterstützung zur Standardisierung des Diagnoseprozesses liefern (vgl. Abschnitt 3.3). Tabtah et al. stellen außerdem fest, dass keine digitale Gesundheitsanwendung existiert, welche alle zur Demenz-Diagnose relevanten Kriterien nach der Leitlinie DSM-V prüft und im ambulanten Umfeld oder der Primärversorgung eingesetzt werden kann [402]. Daraus ergibt sich ein spezifisches Ziel für die Entwicklung digitaler Gesundheitsanwendungen, welches in dieser Domäne als sinnvoller Anwendungsfall zu verfolgen ist. Bei der Übersetzung bestehender Diagnose- und Behandlungsansätze aus der klinischen Praxis hin zu einem im ambulanten Umfeld funktionierenden robusten Werkzeug sind weitere Anforderungen an die Digitalisierung sowie der Bedarf an geeigneten Technologien zu identifizieren. Hierbei steht der potentielle Patient als Nutzer im Zentrum der Anwendungsentwicklung. Die Verknüpfung zwischen den entsprechenden Normen und den zielgruppenspezifischen Gestaltungsempfehlungen stellt jedoch eine Lücke in der wissenschaftlichen Literatur dar, weshalb ein Leitfaden notwendig wird, der einerseits allgemein anerkannte Richtlinien befolgt und andererseits auf die Zielgruppe abgestimmt ist. Des Weiteren stellt Unterabschnitt 3.3.3 heraus, dass die Potentiale von ML und KI zur computergestützten Automatisierung nicht ausreichend berücksichtigt werden. Diese können als eine sinnvolle Ergänzung oder gar perspektivisch als Ersatz für bestehende Leitlinien-Konzepte durch computerisierte Modelle gesehen werden. Um Vertrauen in diese Modelle herzustellen, werden Methoden benötigt, welche ein gewisses Maß an Transparenz ermöglichen, um algorithmische Entscheidungen nachvollziehen zu können und damit Risiken zu minimieren sowie Wissen zu generieren (vgl. Unterabschnitt 3.3.4). Eine anwendungsfallspezifische Evaluation dieser Verfahren muss daher durchgeführt werden. Außerdem ist die Realisierung der zuvor genannten Vorteile nur dann möglich, wenn die adressierten Nutzergruppen dafür entwickelte und integrierte digitale Werkzeuge anwenden [428]. Um diese Vision zu transferieren und Entscheidungen bezüglich der Entwicklung dafür notwendiger digitaler Werkzeuge zu stützen, werden Untersuchungen zur Bestimmung der Adhärenz bei Leistungsempfängern für die entsprechende Anwendungsdomäne notwendig. Darüber hinaus sind insbesondere die Akzeptanzkriterien bei Leistungserbringern zu identifizieren, damit einerseits datengetriebene Algorithmen mit entsprechenden Referenzdaten versorgt werden können und andererseits ein Transfer der entstandenen Artefakte in die Versorgung sichergestellt werden kann. Aus beiden Perspektiven sind im Kontext dementieller Syndrome zum Zeitpunkt der Anfertigung dieser Arbeit keine validierten Aussagen in der wissenschaftlichen Literatur zu finden.

Zusammenfassend werden IT-Artefakte benötigt, die dabei unterstützen, die Grundidee der Präzisionsmedizin in eine personalisierte ambulante Versorgung zu transferieren. Für diese Transition sind Methoden und Werkzeuge zu identifizieren, welche die dafür notwendige datengetriebene Entwicklung von digitalen Gesundheitsanwendungen unterstützen. Hierbei spielen unterschiedliche Schwerpunkte eine Rolle. Insbesondere zur Unterstützung der Standardisierung des Diagnose- und Behandlungsprozesses werden Interoperabilitätseigenschaften erforderlich, die den Datenfluss durch zukünftige Datenmobilitätskonzepte unter Beachtung der regulatorischen Anforderungen ermöglichen. Außerdem müssen Methoden identifi-

ziert werden, welche den Transfer der Digitalisierungspotentiale durch digitale Gesundheitsanwendungen sowie deren gebrauchstauglichen Gestaltung unterstützen. Die Untersuchung von Automatisierungs- und Optimierungsansätzen sowie die direkte Einbindung der Interessenvertreter soll Aufschluss über die Translation medizinischer Innovationen in die Versorgung geben. Übergeordnet muss somit das Ziel verfolgt werden, ein Konzept für die personalisierte Versorgung durch integrierte Diagnose-, Präventions- und Behandlungsprogramme im Kontext von Demenz zu erschließen.

# 4 Anforderungsanalyse und Konzeption des Rahmenwerks für eine datengetriebene Entwicklung

**D**IE Entwicklung digitaler Gesundheitsanwendungen setzt die Berücksichtigung verschiedener Rahmenbedingungen voraus. Hierfür lassen sich Anforderungen und Informationsbedarfe aus dem zuvor erörterten Stand der Wissenschaft und Technik sowie den Ergebnissen von eingebetteten empirischen Untersuchungen innerhalb des wissenschaftlichen Umfelds erheben (vgl. Abschnitt 1.4 und Kapitel 3). Aufgrund der Interdisziplinarität der Problemstellung findet eine Unterteilung der Anforderungen statt. Die Anforderungsanalyse gliedert sich in zwei Phasen. Unter Abschnitt 4.1 werden generische Anforderungen an den Umgang mit Gesundheitsdaten anhand der in Kapitel 3 diskutierten Ansätzen aus der Literatur abgeleitet, welche in einem Rahmenwerk für die datengetriebene Entwicklung digitaler Gesundheitsanwendungen im Kontext entstehender Gesundheitsdatenökosysteme münden (vgl. Abschnitt 4.1). Nach der Zieldefinition des für diese Arbeit skizzierten Anwendungsfalls werden diese Anforderungen spezialisiert (vgl. Abschnitt 4.2 und Unterabschnitt 4.2.2). Anschließend werden die hierfür notwendigen Schritte auf dem Weg zu einer digitalen Gesundheitsanwendung, dem abgeleiteten Vorgehensmodell folgend, durchlaufen (vgl. Unterabschnitt 4.1.2).

## 4.1 Anforderungen an den Umgang mit Gesundheitsdaten im Kontext entstehender Gesundheitsdatenökosysteme

Dieser Abschnitt bezieht sich primär auf eine bereits durch den Autor getätigte Publikation [473]. Im Kontext entstehender Gesundheitsdatenökosysteme wird eine Übersicht von grundlegenden Anforderungen an den Umgang mit und Austausch von Gesundheitsdaten in Tabelle 4.1 dargestellt und detailliert in Tabelle A.1 beschrieben. Diese erweiterbare Liste ist in vier Dimensionen unterteilt. Unter strukturellen Anforderungen (Strukt.) werden notwendige Voraussetzungen auf ethischer, juristischer und organisatorischer Ebene zusammengefasst. Aus ethischer Sicht bestehen Fragestellungen, die bei der Entwicklung von Systemen im Gesundheitssektor betrachtet werden müssen. Diese gehen mit den juristischen Rahmenbedingungen einher, welche für eine Gesellschaft festgelegt sind und in die Entwicklung mit einbezogen werden müssen. Dadurch entstehen weitere Anforderungen aus organisatorischer Sicht, welche sich in dieser Kategorie zusammenfassen lassen. Die ökonomischen Anforderungen (Ökon.) stellen den wirtschaftlichen Bezug und Anreize zur Weitergabe von medizinischen Daten her. Die medizinischen Anforderungen (Med.) stellen die grundlegenden Anwendungsfelder und Schwerpunkte aus medizini-

scher Sicht dar. Diese drei Dimensionen münden in technischen Anforderungen (Techn.) an Systeme, die bei der Erhebung, dem Austausch und der Verarbeitung von Gesundheitsdaten eine Rolle spielen.

<b>Strukt.</b>	<b>S1 - Austausch und gemeinsame Nutzung von Gesundheitsdaten</b> <i>S1.1 - Die Freigabe von Daten an Dritte</i> <i>S1.2 - Organisatorische Interoperabilität</i>
	<b>S2 - Souveräne Gesundheitsdatenverwaltung</b> <i>S2.1 - Verwaltung von Zustimmungen</i> <i>S2.2 - Vertrauenswürdigkeit</i> <i>S2.3 - Rückverfolgbarkeit des Datenlebenszyklus</i>
	<b>S3 - Datenspende für die Forschung</b> <i>S3.1 - Post-Mortem-Datenspende</i> <i>S3.2 - Gemeinsame Nutzung von Forschungsergebnissen</i> <i>S3.3 - Integration von patientengenerierten Längsschnittdaten</i>
	<b>S4 - Transparenz und Reproduzierbarkeit von Forschungsergebnissen</b> <i>S4.1 - Antrag auf Verwendung in der Forschung</i> <i>S4.2 - Sekundäre Nutzung von medizinischen Forschungsdaten</i> <i>S4.3 - Registrierung von Studien</i>
<b>Ökon.</b>	<i>E1 - Wahrnehmbarer Nutzen der Datenspende</i>
	<i>E2 - Finanzierung der Weitergabe von Forschungsdaten</i>
<b>Med.</b>	<i>M1 - Medizinische Relevanz</i>
	<i>M2 - Datengetriebene personalisierte Medizin</i>
	<i>M3 - Kollaborative Entscheidungsfindung unterstützen</i>
<b>Techn.</b>	<b>T1 - Datenschutz</b> <i>T1.1 - Datenlöschung</i> <i>T1.2 - Deidentifizierung</i> <i>T1.3 - Datenschutzwahrende Analysemöglichkeiten</i> <i>T1.4 - Nutzungs- und Zugriffsverwaltung</i> <i>T1.5 - Nutzerauthentifizierung und -autorisierung</i>
	<b>T2 - Datensicherheit</b> <i>T2.1 - Verschleierung</i> <i>T2.2 - Genauigkeit</i> <i>T2.3 - Unveränderlichkeit</i> <i>T2.4 - Verfügbarkeit</i> <i>T2.5 - Beweisbarkeit</i>
	<b>T3 - Benutzerfreundlichkeit</b>
	<b>T4 - Datenauffindbarkeit</b>
	<b>T5 - Datenqualität</b>

<b>Techn.</b>	<b>T6 - Nachhaltigkeit</b>
	<b>T6.1 - Domänengetriebene Datenmodellierung</b>
	<b>T6.2 - Vermeidung von Lock-in-Effekten</b>
	<b>T6.3 - Semantische Interoperabilität</b>
	<b>T6.4 - Datenaggregation und -harmonisierung</b>
	<b>T6.5 - Strukturelle Interoperabilität</b>
	<b>T6.6 - Grundlegende Interoperabilität</b>
	<b>T7 - Dezentralität</b>
<b>T8 - Prädiktive Analysemöglichkeiten</b>	
<b>T9 - Datenanreicherung</b>	
<b>T10 - Skalierbarkeit und Leistung</b>	
<b>T11 - Flexibilität</b>	
<b>T12 - Erweiterbarkeit</b>	

Tabelle 4.1: Zusammengefasste Anforderungen an den Umgang mit Gesundheitsdaten in Gesundheitsdatenökosystemen nach Schinle et al., 2021 [473]. Details in Tabelle A.1.

### 4.1.1 Der Datenkreislauf als Rahmenwerk

Die iterative Entwicklung von digitalen Gesundheitsanwendungen ermöglicht eine kontinuierliche Integration neu erhobener Daten in den Entwicklungsprozess [473]. Hierdurch entstehen Produkte, Dienstleistungen oder Verfahren, die wiederum neue Daten generieren, welche zur kontinuierlichen Verbesserung bestehender Lösungen oder zur Entwicklung innovativer Ansätze durch die Kombination verschiedener Datenquellen genutzt werden können. Um einen Rahmen für diesen Zusammenhang zu schaffen, wird nachfolgend ein Datenkreislauf dargelegt, der das Ziel einer nachhaltigen Nutzung von Gesundheitsdaten verfolgt. Die zuvor abgeleiteten Anforderungen werden darin verknüpft und deren primären Interessenvertretern zugeordnet (vgl. Abbildung 4.1). Der Nutzer steht dabei als Dateneigentümer im Zentrum und benötigt die Möglichkeit, über ihn erhobene Daten von den sammelnden sowie speichernden Parteien anzufordern, um selbst zu bestimmen, welche Daten an Dritte weitergegeben werden, wer darauf Zugriff hat und was mit diesen geschehen darf (S1, S1.1, S2). Dies erfordert Werkzeuge zur Erfassung, Visualisierung und Verwaltung der vorhandenen Gesundheitsdaten durch den Dateneigentümer selbst, welche benutzerfreundlich gestaltet sind (T3). Eine forschungskompatible, institutionenübergreifende elektronische Gesundheitsakte kann diese Rolle übernehmen und zusätzlich zu digitalen Gesundheitsanwendungen, Daten integrieren, welche an die Bedürfnisse medizinischer Entscheider angepasst sind (M1), um einerseits die Entscheidungsfindung bestmöglich zu unterstützen und andererseits das Verständnis des Patienten für seinen Gesundheitszustand zu verbessern (M3). Neben der Qualität dieser Daten (T5) spielt die grundlegende und strukturelle Interoperabilität durch Abbildung und Weiterleitung der Daten in international akzeptierten Standard- und Schnittstellenformaten eine Rolle (T6.5, T6.6). Die semantische Interoperabilität kann durch standardisierte Codierung dieser Ressourcen gewährleistet werden und verhindert Lock-in-Effekte (T6.1, T6.2, T6.3). Außerdem muss sichergestellt werden, dass nicht nur Daten, sondern auch Prozesse, die damit verbundenen Identitäten und Zugriffsrechte interorganisatorisch interoperabel sind (S1.2). Neben einer möglichen rechtlichen Notwendigkeit für Hersteller durch gesetzliche Verordnungen, muss der damit verbundene Mehraufwand auch durch Kompensationsmechanismen für

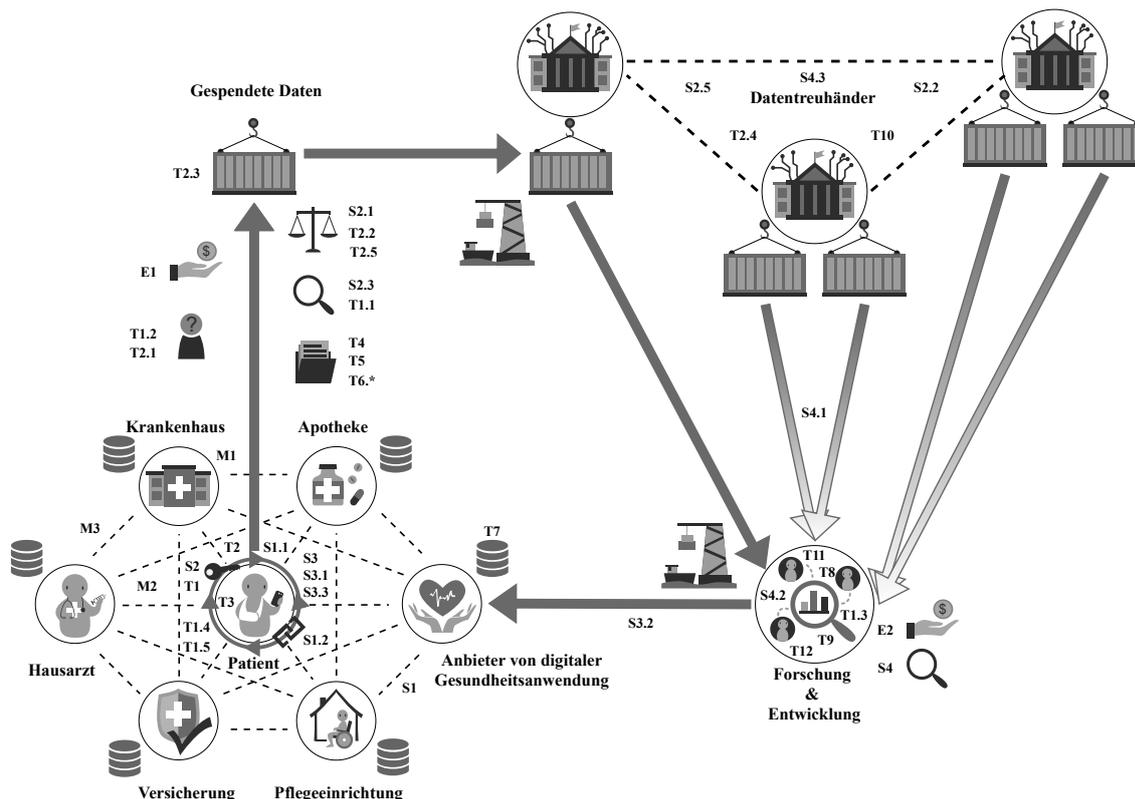


Abbildung 4.1: Schematische Darstellung des Datenkreislaufs nach Schinle et al., 2021 [473].

die Daten erhebenden Personen und Einrichtungen sowie für Forscher und Forschungsprojekte ausgeglichen werden, um nachhaltig Daten zur Verfügung stellen zu können (E2). Die gesammelten Daten selbst können von den Besitzern entweder kostenlos (philanthropischer Ansatz) oder im Gegenzug für Dienstleistungen beziehungsweise monetäre Werte (kapitalistischer Ansatz) zur Verfügung gestellt werden (S3, E1). Diese Ansätze gehen zusätzlich mit ethischen Fragestellungen und Diskussionen einher, welche aufgrund des Fokus der Arbeit an dieser Stelle nicht weiter thematisiert werden. Durch die erfolgreiche Anonymisierung personenbezogener Gesundheitsdaten lassen sich auch komplette Gesundheitsverläufe für Forschungszwecke zur Verfügung stellen (S3.1) oder einzelne Datensätze, die kombiniert mit vergleichbaren Datensätzen eine ganzheitliche Sicht auf einen bestimmten Gesundheitszustand ermöglichen. Die Pseudonymisierung ermöglicht eine kontinuierliche Erhebung und Überwachung von Längsschnittdaten (S3.3), erfordert jedoch die sichere Verwaltung der wahren Identität durch den Dateneigentümer selbst und Daten verwahrender Institutionen. In diesem Fall handelt es sich um personenbezogene Daten, die es rechtlich erforderlich machen, einen Zweck für die Datenverarbeitung zu definieren, also festzulegen von wem und wofür die Daten verwendet werden dürfen (S2.1). Diese Einwilligung und Nutzungsbedingungen müssen digital, in einem automatisch verarbeitbaren elektronischen Format, verfügbar gemacht werden. Dabei muss das Vertrauen und damit die Transparenz gewährleistet sein, dass die Daten ausschließlich für die genannten Zwecke verwendet werden (S2.2). Ein weiterer Vorteil der Pseudonymisierung ist die Möglichkeit, weitere Daten für Studien anzufordern, wenn die betroffenen Personen einer Kontaktaufnahme ausdrücklich zugestimmt haben (S4.1, S4.2). Die Kombination von Gesundheitsdaten, ihren Metadaten und der Einwilligung lässt sich kapseln, um die Integrität (T2.3) und Nachvollziehbarkeit der Datenverarbeitungsschritte zu gewährleisten. Diese Datenkapseln müssen entweder von der erhebenden Stelle, vom Dateneigentümer selbst oder von einem Datentreuhänder aufbewahrt werden und über Logik verfügen, die transparente Zugriffskontroll- und Validierungsmechanismen in

Kombination mit einer vertrauenswürdigen Infrastruktur implementiert. Um die Daten für andere interessierte Parteien freizugeben, wird ein Register benötigt, welches die Daten auffindbar macht (S4.3, T4). Dieses und die Speicherung der Daten selbst sollte dezentralisiert werden (T7), wobei eine hohe Verfügbarkeit gewährleistet sein muss (T2.4). Informationen über die Richtigkeit der Daten (T2.2) durch den Datenurheber sowie Freigaben müssen in Form von manipulationssicheren Zugriffsrechten (T1.4) auf vertrauenswürdige Akteure (T1.5) oder in einem dezentral verwalteten Register abgebildet werden. Autorisierte Organisationen (T1.5), die ein Interesse und die Kompetenz haben, aus den Daten einen Mehrwert zu generieren, sollten unter Einhaltung der Datenschutz- und Sicherheitsanforderungen (T2, T6, T6.3, T6.4) zu den Entwicklungsprozessen beitragen, indem sie vorhandene Datensätze kombinieren. Durch den Einsatz flexibler (T11), erweiterbarer (T12), die Privatsphäre und Vertraulichkeit wahrender (T1, T2.5) Datenanalysewerkzeuge können epidemiologische Vorhersagen in Echtzeit getroffen (T8, T10) oder neue Erkenntnisse zur Unterstützung der Gesundheitsversorgung gewonnen werden (T9). Es muss transparent und persistent gespeichert sein, wofür die Datenkapseln verwendet wurden und inwieweit neue Erkenntnisse auf sie zurückgeführt werden können (S2.3, S4, T2.5). Diese Informationen müssen dem Dateneigentümer durch den Hersteller zur Verfügung gestellt werden (S3.2). Die neuen Erkenntnisse fließen als Artefakte in die Entwicklung neuer Behandlungsmethoden (M2) sowie medizinischer Produkte und Dienstleistungen (S1.1) ein, die wiederum neue Daten generieren, deren Auswirkungen analysiert werden müssen.

## 4.1.2 Abgeleitetes Vorgehen

In Bezug auf den zuvor skizzierten Datenkreislauf zeigt der in Abbildung 4.2 dargestellte Produktlebenszyklus den klassischen Ablauf in der Medizinproduktentwicklung. Ziel ist es, die Lücke zwischen den

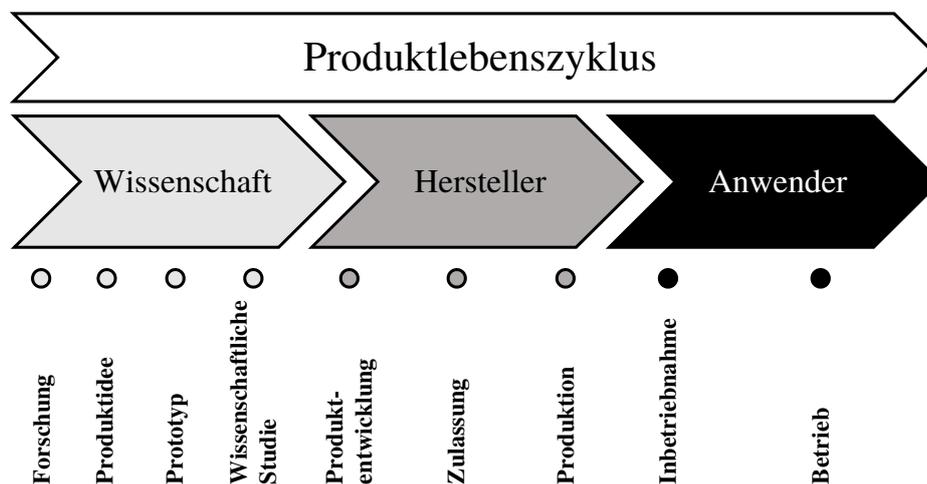


Abbildung 4.2: Medizinproduktlebenszyklus nach Hastenteufel und Renaud [180].

dargestellten Phasen zu verringern und daraus ein zyklisches Vorgehen zu initiieren, indem entstehende Gesundheitsdatenökosysteme mit diesem Prozess verzahnt werden. Da die Entwicklung digitaler Gesundheitsanwendungen voraussetzt, dass die Datengenerierung und Informationsbereitstellung nutzerzentriert gestaltet wird, lässt sich das Vorgehen auf den Grundsätzen der agilen Softwareentwicklung aufbauen, um auf neu gewonnene Anforderungen und Prioritäten durch eine iterative Vorgehensweise reagieren zu können. Dieser lässt sich nun durch die, während der Produktentwicklung bis zum Betrieb gewonnenen Daten erweitern, indem diese wiederum Bestandteil des gesamten Entwicklungsprozesses werden.

Je nach Art und Umfang der neu gewonnenen Daten kann hieraus eine neue Produktidee angestoßen oder das bestehende Produkt zielgerichtet weiterentwickelt werden. Um das Vorgehen entsprechend dem zuvor beschriebene Rahmenwerk zu operationalisieren, wird hieraus nachfolgend ein Vorgehensmodell abgeleitet (vgl. Abbildung 4.3).

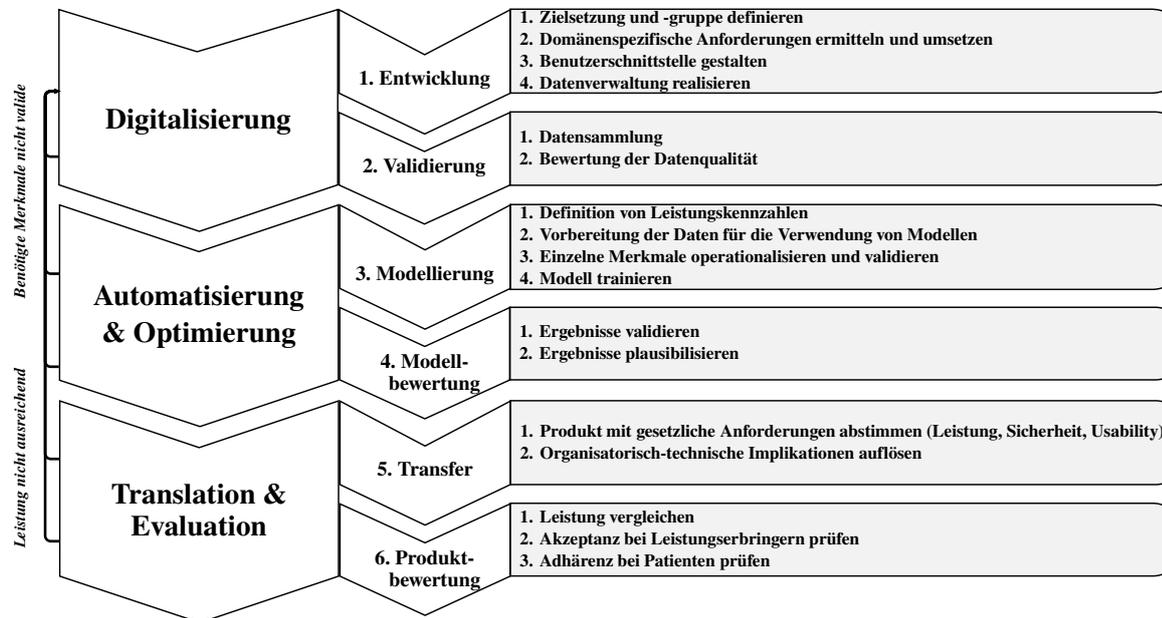


Abbildung 4.3: Abgeleitetes Vorgehen basierend auf Schinle et al., 2021 [471].

Zusammenfassend lässt sich das Vorgehen in die drei Phasen Digitalisierung, Automatisierung & Optimierung und Translation & Evaluation sowie in folgende sechs Schritte mit entsprechenden Leitfragen untergliedern:

#### Schritt 1 - Entwicklung:

1. Welches gesundheitsrelevante Ziel wird angestrebt?
2. Welche Informationsbedarfe und domänenspezifischen Anforderungen bestehen darin?
3. Wie muss eine digitale Lösung gestaltet sein, um in der medizinischen Versorgung akzeptiert zu werden?
4. Wie werden damit erhobene Daten verwaltet?

#### Schritt 2 - Validierung:

1. Lassen sich mit der Anwendung zur Erreichung des gesundheitsrelevanten Ziels genügend Daten sammeln?
2. Genügt die Qualität dieser Daten?

#### Schritt 3 - Modellierung:

1. Welche Metriken werden benötigt, um Vorgänge und Entscheidungen zu optimieren und automatisieren?
2. Wie lassen sich einzelne Merkmale operationalisieren und validieren?

3. Welche Modelle und Verfahren kommen zum Einsatz?

**Schritt 4 - Modellbewertung:**

1. Welche Aussagekraft haben die Modelle?
2. Wie plausibel sind die Ergebnisse und wie transparent müssen diese sein?

**Schritt 5 - Transfer:**

1. Welche Anforderungen spielen für die Einführung dieser Anwendung als Produkt eine Rolle?
2. Welche Implikationen erwachsen daraus und wie lassen sich diese auflösen?

**Schritt 6 - Produktbewertung:**

1. Lässt sich die Leistung mit bestehenden Lösungen vergleichen?
2. Welche Kriterien spielen bezüglich der Akzeptanz von Leistungserbringern eine Rolle?
3. Welche Kriterien spielen bezüglich der Adhärenz bei der leistungsempfangenden Zielgruppe eine Rolle?

Als Voraussetzung für die Entwicklung gilt es in **Schritt 1** ein gesundheitsrelevantes Ziel zu definieren, welches mit Hilfe einer digitalen Lösung erreicht werden soll. Hierfür relevante Informationen lassen sich identifizieren, indem für die Interessenvertreter notwendige Parameter und für die Digitalisierung geeignete Verfahren zur Erfassung dieser bestimmt werden. Wo, wann und durch wen diese Daten erhoben und verwaltet werden ist ebenfalls Teil dieses Schritts. Durch Datenerhebungsstudien müssen Daten gesammelt werden, um die Machbarkeit des zuvor definierten Konzepts inklusive der getroffenen Annahmen zu prüfen. Die Phase der Digitalisierung befasst sich somit primär mit der technischen Umsetzung von zuvor identifizierten Erfassungsmöglichkeiten sowie Vernetzungs- und Automatisierungspotentialen. **Schritt 2** beschäftigt sich mit der Aus- und Bewertung der Wirksamkeit einer entwickelten Lösung. Hierin muss mit Hilfe festgelegter Metriken überprüft werden, ob die mittels digitalisierter Verfahren erfassten Parameter in Bezug auf das definierte Ziel aussagekräftige Ergebnisse liefern können. Auf Grundlage dieser Daten werden in **Schritt 3** die Potentiale der Digitalisierung durch Optimierungs- und Automatisierungsmöglichkeiten ausgelotet. Diese Phase der Wissensentdeckung erfolgt in Anlehnung an den CRISP-DM (vgl. Unterabschnitt 2.1.3). Um Lösungen zur Hebung der gefundenen Potentiale zu bewerten werden Metriken benötigt, welche eine Aussage über den Einsatz von Verfahren im Kontext der Anwendungsdomäne ermöglichen. Die hierfür erforderliche Operationalisierung einzelner Merkmale hat Einfluss auf die Gestaltung der Modelle. Durch **Schritt 4** werden diese letztlich anhand der Zielvorgabe bewertet und gegenübergestellt. Dieser Schritt liefert Erkenntnisse über den Einfluss und Implikationen des Einsatzes der entsprechenden Verfahren. Außerdem muss transparent gemacht werden, welche Merkmale einen Beitrag für die zukünftige Produktentwicklung liefern sowie ob diese plausibel bezüglich der Vorgaben potenzieller Leitlinien sind. Um Daten kontinuierlich zu erfassen und die Erkenntnisse nachhaltig zur Verfügung zu stellen, muss der Transfer von erforschten digitalen Lösungen in den Markt ermöglicht werden. In der dritten Phase werden deshalb die gewonnenen Erkenntnisse verwendet, um dies zu ermöglichen. Davor werden mehrere Iterationen zwischen **Schritt 2** bis **Schritt 5** durchlaufen, um einen sinnvollen Übergang in **Schritt 6** einzuleiten. Hierbei muss zunächst überprüft werden, inwiefern das potenzielle Produkt regulatorische Anforderungen erfüllen muss (vgl. Unterabschnitt 2.3.1). Der Entwicklungs- und entsprechende Konformitätsbewertungsprozess orientiert sich an den daraus abzuleitenden Vorgaben. Des Weiteren muss für die Überführung in ein Produkt, das möglichst optimale

Verhältnis von Wirksamkeit und Dauer, also das richtige Aufwand-Nutzen-Verhältnis gefunden werden. Dieser Punkt spielt mit dem Punkt zusammen, dass um Daten in ausreichender Menge und Qualität zu erhalten, auch Anreizmechanismen konzipiert werden müssen, welche die Erfassung und Bereitstellung medizinischer Daten über Studien hinaus motivieren und insbesondere bei einer notwendigen aktiven Selbsterhebung von Daten durch den Nutzer kontinuierliche Datenflüsse gewährleistet sind. Auch nach erfolgreicher Durchführung von **Schritt 6** werden weitere Zyklen des Vorgehensmodells durchlaufen, um das Produkt zu erweitern und optimieren.

Für die Umsetzung dieses Ablaufs werden Werkzeuge und Methoden benötigt, welche die Erfüllung der zugrundeliegenden Anforderungen unterstützen und den weiteren Aufbau dieser Arbeit widerspiegeln. Für den in dieser Arbeit adressierten Anwendungsfall stehen insbesondere die Gestaltung einer Benutzerschnittstelle und die Verwaltung von digital erhobenen Daten vor allem aber diesbezügliche Datensouveränität, -interoperabilität und -mobilität im Vordergrund. Anschließend stehen Verfahren zur Validierung und Plausibilisierung der damit erhobenen Daten für den entsprechenden Anwendungskontext im Fokus. Eine darauf folgende Automatisierung von Verarbeitungsvorgängen kann unter Einsatz entsprechender Analysewerkzeuge und Algorithmen stattfinden. Nachfolgend wird auf Basis der zuvor skizzierten Anwendungsdomäne das erarbeitete Vorgehensmodell durchlaufen, um die Ausgestaltung weiterer Methoden und Werkzeuge zu definieren sowie domänenspezifische Artefakte diesbezüglich zu implementieren und evaluieren.

## 4.2 Zieldefinition einer digitalen Gesundheitsanwendung

In der Phase der Digitalisierung lässt sich gemäß **Schritt 1** die Idee und damit das gesundheitsrelevante Ziel für die digitale Anwendung wissensbasiert aus der Versorgungslandschaft und entsprechender Literatur erschließen. Aufgrund der mit AD verbundenen datenintensiven Früherkennungs- und Diagnoseprozesse, bilden diese zur Erreichung des adressierten Ziels dieser Arbeit den notwendigen Anwendungskontext. Wie in Unterunterabschnitt 2.2.3 dargestellt, hat AD, neben dem negativen Einfluss auf die Lebensqualität der Betroffenen und Angehörigen selbst, ökonomische Auswirkungen auf die Gesundheitssysteme weltweit. Aufgrund des demographischen Wandels hin zu einer überalterten Gesellschaft wird zukünftig von weiteren Implikationen ausgegangen [332]. Eine möglichst frühe Erkennung von AD und ursächlicher Risikofaktoren bis hin zu einer kontinuierlichen Überprüfung des Verlaufs kann in Kombination mit geeigneten Maßnahmen zu einer Verlangsamung des kognitiven Abbaus und damit zu geringeren medizinisch-pflegerischen Aufwänden führen (vgl. Unterunterabschnitt 2.2.4). Die stigmatisierende Wirkung neurodegenerativer Erkrankungen und mangelnde Verfügbarkeit von Spezialisten führen dazu, dass Betroffene und deren Angehörige häufig nicht oder zu spät professionelle Hilfe aufsuchen [201]. Als Zielgruppe gelten bezüglich der in Unterunterabschnitt 2.2.3 dargestellten weltweiten Inzidenz- und Prävalenzraten sowie einem damit verbundenen Bedarf einer möglichst frühzeitigen Erkennung, Menschen mittleren und höheren Alters. Ab 60 Jahren steigt die Inzidenzrate exponentiell, weshalb dieses Alter aus Sicht der Erkrankten als späteste Einstiegsschwelle für mit der Prävention, Diagnose und Behandlung einhergehende Systeme definiert wird [262, 332]. Aus dem persönlichen Umfeld und der medizinischen Versorgung von erkrankten Personen stammen sekundäre Nutzergruppen, woraus Anforderungen an den Umgang und die Interpretation von damit erhobenen Informationen entstehen. Innerhalb einer ersten Iteration der Digitalisierungsphase wird auf die Gestaltung der digitalen Gesundheitsanwendung aus Sicht der Betroffenen und damit der primären Zielgruppe eingegangen, zusätzliche

Anforderungen aus Perspektive der sekundären Nutzergruppen werden in Abschnitt 7.1 weitergehend untersucht und deren Einfluss thematisiert. Daher sollen in technische Systeme eingebettete Ansätze entwickelt werden, welche ohne professionelle Hilfe zu bedienen sind und durch eine zeitökonomische Aufnahme von Indikatoren aus dem ambulanten Umfeld heraus eine Einschätzung des Krankheitsbildes ermöglichen. Darauf aufbauend können geeignete Präventions- und Behandlungsmaßnahmen abgeleitet sowie Behandlungserfolge überprüft werden. Hierdurch soll einer stigmatisierenden Wirkung und der damit verbundenen Angst vor einer Diagnose mit einem niederschweligen Angebot und Informationen über Behandlungsmöglichkeiten begegnet werden. Eine digitale Gesundheitsanwendung kann somit einen Beitrag zur Implementierung präzisionsversorgerischer Ansätze in den Bereichen der Demenzprävention, -diagnose und des -behandlungsmanagements leisten. Des Weiteren sind aufgrund der steigenden Fallzahlen in Ländern mit geringen bis mittleren Einkommen, kostengünstige Lösungsansätze, deren Verfügbarkeiten gewährleistet sind, zu präferieren (vgl. Unterunterabschnitt 2.2.3). Nachfolgend liegt demnach der Fokus auf einer ambulanten Erfassung von Parametern, um ein möglichst vollständiges Bild des multifaktoriellen kognitiven Abbauprozesses einer AD zu erhalten und dabei die Kombination mit zusätzlichen Daten aus unterschiedlichen Fachbereichen zu ermöglichen. Daraus resultierende Anwendungen werden nachfolgend mit dem Namen *DemPredict* abgekürzt. Demnach handelt es sich um die Entwicklung eines Medizinproduktes, welches gemäß seiner Zweckbestimmung einer entsprechenden Klasse zuzuordnen ist (vgl. Unterunterabschnitt 2.3.1).

### 4.2.1 Identifikation und Kombination relevanter Parameter

Da insbesondere dem psychischen Gesundheitszustand unterschiedliche Faktoren zugrunde gelegt werden, bildet eine Informationsbedarfsanalyse zur Identifikation der relevanten Parameter und Anforderungen die Wissensgrundlage der Digitalisierungsphase. Nachfolgend werden die zur Erreichung des zuvor skizzierten Ziels relevanten Informationen identifiziert sowie dafür zu erfassende Parameter und die Verfahren zu deren Erfassung bestimmt. Dieser Prozessschritt setzt die Einbindung medizinischer Fachexperten voraus. Der Früherkennungs- und Diagnoseprozess bei einer AD folgt dem Ausschlussprinzip (vgl. Unterabschnitt 2.2.4). Da für die Diagnose einer Demenz laut ICD die damit einhergehenden Symptome mindestens sechs Monate vorliegen müssen, ist eine einmalige Untersuchung in einem sehr frühen Stadium nicht ausreichend (vgl. Unterunterabschnitt 2.2.4). Eine kontinuierliche Aufzeichnung der damit verbundenen Parameter erscheint daher als eine wichtige Voraussetzung. Für eine gelingende Früherkennung muss außerdem aufgrund des progressiven Krankheitsprozesses die Möglichkeit für eine regelmäßige Parametererfassung bestehen. Dabei lassen sich, wie in Abbildung 4.4 illustriert, die nachfolgenden drei Arten von geeigneten Verfahren zur Erkennung einer dementiellen Erkrankung auf Grundlage des zuvor beschriebenen medizinischen Hintergrundwissens identifizieren und bewerten (vgl. Unterabschnitt 2.2.4). Einerseits existieren systembiologische Verfahren, welche mithilfe von Biomar-

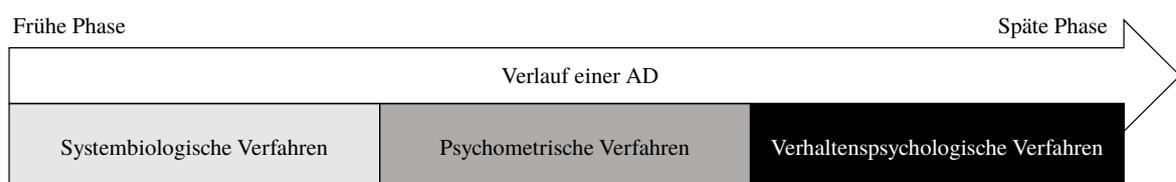


Abbildung 4.4: Verfahren zur Erkennung in Abhängigkeit des Verlaufs nach Ritchie et al. [347].

kern, den Nachweis über das Vorliegen der Erkrankung in einer frühen Phase erbringen können. Wie

in Unterabschnitt 2.2.4 skizziert, fallen hierunter invasive Verfahren, wie Liquoruntersuchungen, welche jedoch aufgrund der Liquorentnahme und damit ihrer Durchführungsvoraussetzungen für ein ambulantes Screening als ungeeignet erscheinen. Auch wenn derzeit in der Forschung mobile Ansätze beispielsweise für EEG erforscht werden, wird für bildgebende Verfahren wie PET der Einsatz von stationären Gerätschaften in medizinischen Einrichtungen vor Ort notwendig, womit finanzielle und logistische Aufwände verbunden sind. Verfahren zur Erkennung von im Blut nachweisbaren Biomarkern könnten eine regelmäßige Kontrolle gegebenenfalls auch im ambulanten Umfeld ermöglichen, jedoch sind diese zum Zeitpunkt der Anfertigung dieser Arbeit noch Gegenstand der Forschung. Da das Ziel einer Früherkennung im ambulanten Umfeld möglichst ohne professionelle Hilfe unter Berücksichtigung der definierten Zielgruppe verfolgt wird, müssen auch mit körperlichen Untersuchungen einhergehende Ansätze ausgeschlossen werden. Insgesamt lassen sich die systembiologischen Verfahren selbst im aktuellen klinischen Alltag lediglich unter Akzeptanz der damit verbundenen Aufwendungen als geeignet einstufen. Dem gegenüber stehen die psychologischen Methoden, deren weitere Unterteilung hier basierend auf Psychometrie und Verhaltenspsychologie vorgenommen wird.

Die Anwendung von verhaltenspsychologischen Verfahren dient dazu Verhaltensänderungen zu erkennen, die aus dem kognitiven Abbauprozess resultieren. Um subjektive Auffälligkeiten im Verhalten einer Person festzustellen, sind Beobachtungen von und Interaktionen mit den Betroffenen und deren Angehörigen beispielsweise durch Anamnesegespräche selbst notwendig. Weitere Ansätze zur Objektivierung dieser Messgrößen können ihre Anwendung im ambulanten Umfeld beispielsweise durch die Aufnahme und automatisierte Detektion von Verhaltensanomalien mittels technischer Systeme finden und eine Möglichkeit schaffen, zusätzliche Parameter zu erfassen, die Hinweise auf psychische Erkrankungen liefern ohne dabei auf die Kooperationsbereitschaft von Betroffenen angewiesen zu sein [476]. Solche Ansätze zur passiven Verhaltensbeobachtung mittels ambienter Monitoring-Systeme wurden im Rahmen dieser Arbeit ebenfalls untersucht [476]. Da Verhaltensänderungen erst in einer späten Phase der Demenz auftreten, werden diese aufgrund der zuvor getätigten Zieldefinition nicht weiter ausgeführt.

Mit Hilfe von psychometrischen Tests wird versucht die Leistungseinbußen in unterschiedliche kognitiven Domänen zu quantifizieren (vgl. Unterabschnitt 2.2.5). Da psychometrische Tests in der Regel ebenfalls bei Spezialisten vor Ort durchgeführt werden und hier noch häufig analoge Formate ihre Anwendung finden, ist eine engmaschige Testung ebenfalls mit personellen und logistischen Aufwänden verbunden [152]. Hierbei könnte jedoch die Digitalisierung und Gestaltung dieser Verfahren in einem durch Betroffene eigenständig durchführbaren Format (PRO) Abhilfe schaffen. Dies setzt eine aktive Rolle und dadurch Kooperationsbereitschaft der Betroffenen voraus (vgl. Unterabschnitt 3.3.2). Mittels einer nutzerzentrierten Digitalisierung dieser aktiven Erfassung nicht-invasiver psychometrischer Verfahren und deren Transfer soll eine ambulante Erhebung von Parametern möglich werden, die zur Früherkennung einer AD dienen. Welche Parameter relevant sind und wie eine diesbezügliche Testauswahl gestaltet werden muss, hängt von den Erkenntnissen der Forschung und entsprechenden Fachgesellschaften ab. Die DSM gruppiert mögliche Auswirkungen auf die kognitive Leistungsfähigkeit in Klassen, die überprüft werden müssen, um mögliche Defizite zu identifizieren (vgl. Abbildung 2.8). Um diese verschiedenartigen Ausprägungen der kognitiven Beeinträchtigung zu testen, ist jede durch die DSM-V vorgegebene Domäne zumindest durch die Prüfung der Leistungsfähigkeit einer Subdomäne zu erfassen (vgl. Unterabschnitt 2.2.3). Als Herausforderung hierfür gilt es in Zusammenarbeit mit Domänenexperten die Tests zu identifizieren, die in ein digitales Format überführt werden können und möglichst zeiteffizient durchführbar und auswertbar sind. In Tabelle 4.2 sind die im Rahmen des wissenschaftlichen Umfelds aus bestehenden Testbatterien ausgewählten und adaptierten Tests. Kombiniert mit daran angelehnten und selbst konstruierten psychometrischen Tests stellen diese die Grundlage dar, um die jeweils adressierten kognitiven Domänen abzubilden (vgl. Unterabschnitt 2.2.5). Diese erfassen unterschiedliche Metriken, welche eine Aussage

<b>Kürzel</b>	<b>Ursprung</b>	<b>Ausgewählte Aufgaben</b>	<b>Adressierte DSM-Domänen</b>	<b>Adressierte Subdomänen</b>
<b>A</b>	MMSE [147]	Fragen zur aktuellen Zeit und zum Aufenthaltsort	Komplexe Aufmerksamkeit	Anhaltende Aufmerksamkeit (Orientierung)
<b>B</b>	FCSRT [171]	Benennen, Lernen und Erinnern von gesehenen Objekten	Lernen und Gedächtnis Sprache	Freier und wiederholter Abruf Wiedererkennung Objekte benennen
<b>C</b>		Verbinden von Zahlenreihen		
		(TMT-A)	Komplexe Aufmerksamkeit	Verarbeitungsgeschwindigkeit
<b>D</b>	CERAD [364]	Verbinden von Zahlen- und Buchstabenreihen (TMT-B)	Exekutivfunktion	Flexibilität
<b>E</b>		Lernen und	Lernen und Gedächtnis	Implizites Lernen
<b>F</b>	GMLT [326, 386]	Erinnern eines Pfades innerhalb eines Labyrinths	Exekutivfunktion	Wiederholtes Erinnern
			Komplexe Aufmerksamkeit	Wiedererkennung
<b>G</b>		Umwandlung von Zahlensymbolen in Wörter (N2L) und umgekehrt (L2N)	Exekutivfunktion	Arbeitsgedächtnis
<b>H</b>	DemTect [220]	Gesehene Zahlenreihen erinnern	Exekutivfunktion	Arbeitsgedächtnis
<b>I</b>		Rückwärts zählen	Exekutivfunktion	Arbeitsgedächtnis
<b>J</b>		Würfel abzeichnen		Visuelle Wahrnehmung
<b>K</b>	RUDAS [395]	Verschiedene Körperteile zeigen	Wahrnehmungsmotorische Funktion	Visuell-konstruktives Denken Visuelle und räumliche Orientierung
<b>L</b>		Zeigen und Bestätigen von angezeigten		
<b>M</b>	Eigenentwicklung	Symbolen und Wörtern (Go/No-Go-Aufgabe)	Komplexe Aufmerksamkeit	Selektive Aufmerksamkeit
<b>N</b>	SET [121]	Vervollständigung von Bildergeschichten	Soziale Kognition	Wahrnehmung von Emotionen Theorie des Geistes (Theory of Mind) Einsicht

Tabelle 4.2: Übersicht über verwendete psychometrische Testverfahren nach Schinle et al., 2018 [478].

über den Leistungsstand einer Domäne zulassen. Der ausgewählte Test A prüft angelehnt an den MMSE die kognitive Domäne der Orientierung (vgl. Unterunterabschnitt 2.2.5). Hierzu werden Fragen gestellt, welche die zeitliche und räumliche Orientierungsfähigkeit der Nutzer prüfen. Im konkreten Fall sind das zum Zeitpunkt der Testung, sowohl das aktuelle Datum und der Wochentag nach dem gregorianischen Kalender sowie die Jahres- und Uhrzeit als auch das Bundesland, indem sich die Testperson befindet. Als Variablen lassen sich hierbei die Anzahl richtiger und falscher Antworten aufzeichnen, deren Summen sowie die Abweichungen von den angegebenen Werten des Datums und der Uhrzeit zum wahren Wert. Über Test B werden die Bereiche Lernen, Gedächtnis und Sprache geprüft. Dies verlangt vom Nutzer das Benennen, Lernen und Erinnern von gezeigten Objekten (vgl. Unterunterabschnitt 2.2.5). Dazu werden in vier Durchgängen durch insgesamt sechzehn Abbildungen visualisierte Objekte in Vierergruppen dargestellt und deren korrekte Bezeichnung nacheinander über deren Objektkategorie abgefragt. Anschließend müssen die vier neu gelernten Begriffe anhand der Objektkategorie ohne visuelle Unterstützung aus dem Gedächtnis abgerufen werden. Nach den vier Durchgängen sollen alle sechzehn gelernten Begriffe in freier Reihenfolge wiedergegeben werden. Die gegebenen Antworten werden alle notiert, um eine Unterscheidung zwischen vergessenen Objekten und erinnerten Wörtern, die nicht gelernt wurden (Intrusionen) zu ermöglichen. Die Fähigkeit zur Aufmerksamkeit und Exekutivfunktionen werden über Test C erfasst. Dieser ist auch unter der Bezeichnung TMT-A bekannt (vgl. Unterunterabschnitt 2.2.5). Hierbei müssen auf einer Fläche zufällig angeordnete Zahlen einer Zahlenreihe in aufsteigender Reihenfolge nacheinander verbunden werden. Als Bewertungskriterium der Leistungsfähigkeit dient die Bearbeitungsdauer [405]. Zusätzlich wird die Anzahl fälschlicherweise ausgewählter Zahlen während der Durchführung aufgezeichnet. Diese Art von Fehler impliziert eine verlängerte Bearbeitungsdauer. In Test D findet eine Erweiterung des Testkonzepts von C statt. Als TMT-B ist dieser Test ebenfalls Teil des MoCA und der CERAD-Testbatterie und unterscheidet sich darin, dass nicht ausschließlich Zahlen, sondern abwechselnd Zahlen und Buchstaben alphabetisch in aufsteigender Reihenfolge miteinander verbunden werden müssen (vgl. Abbildung 2.12). Wie auch beim TMT-A werden die Bearbeitungsdauer und die Anzahl der Fehler gespeichert. Im Zuge des Test E wird das Gedächtnis, insbesondere das wiederholte Erinnern geprüft (vgl. Unterunterabschnitt 2.2.5). Auf einer quadratischen Fläche als Matrix nebeneinander angeordnete Kacheln verdecken einen versteckten Pfad, der in diesem Test durch Versuch und Irrtum schrittweise erkundet und eingepägt werden muss. Lediglich Start- und Zielfeld sind vorgegeben. Die Kacheln lassen sich nur einzeln aufdecken und werden danach wieder verdeckt. Der Pfad ist somit niemals komplett sichtbar. Nachdem das Zielfeld zum ersten Mal erreicht wurde, muss der eingepägte Pfad zwei Mal direkt hintereinander aus dem Gedächtnis abgerufen werden. Als Fehlermaß wird zwischen der Anzahl von drei verschiedenen Fehlerarten unterschieden (vgl. Absatz 2.2.5). Dieser Test, wird von Test F vervollständigt, indem der zuvor erlernte Pfad nach der Durchführung weiterer Tests (Interferenzaufgaben), erneut aus dem Gedächtnis abgerufen und aufgezeigt werden muss. Damit wird die Wiedererkennungsfähigkeit überprüft. Alle drei Fehlertypen (ER, EL, EP) werden mit ihren Häufigkeiten nach den Testdurchläufen getrennt gespeichert. Ebenso wird die Summe aus allen Fehlern eines Testdurchlaufs vermerkt. Durch die Tests G, H und I werden ebenfalls die Exekutivfunktion, insbesondere aber das sogenannte Arbeitsgedächtnis getestet (vgl. Unterunterabschnitt 2.2.5). In G müssen zunächst als Ziffern dargestellte Zahlen in Wörter umgewandelt werden und anschließend die in Wörtern dargestellten Zahlen in deren äquivalente Form in Ziffern. Dabei gilt die Anzahl der falsch umgewandelten Zahlen als Fehlermaß. In H wird eine mehrstellige Zahl als Ziffernfolge mit Einzelziffern vorgeführt. Anschließend muss diese in umgekehrter Reihenfolge aus dem Gedächtnis abgerufen werden können. Der Test startet mit einer Ziffernfolgenlänge von zwei. Diese wird pro Durchgang iterativ um eine Ziffer verlängert, bis eine maximale Länge von sechs erreicht wird. Insgesamt gibt es zwei Durchgänge. Falls ein Fehler gemacht wird bevor die Maximallänge erreicht ist, wird ein Durchgang beendet. Als Bewertungskriterium gilt die Länge der längsten erinnerten

Zahlenfolge. Durch I wird die Rechenfähigkeit durch das kontinuierliche Bilden der Differenz getestet. Beginnend bei der ersten vorgegebenen Zahl, muss eine zweite vorgegebene Zahl davon und dem entsprechenden Zwischenergebnis schrittweise abgezogen werden. Um die wahrnehmungsmotorische Funktion zu testen, werden in Test J und K die visuelle Wahrnehmung, das visuell-konstruktive Denken und die visuelle und räumliche Orientierung getestet (vgl. Unterunterabschnitt 2.2.5). Bei Test J handelt es sich dabei um die Aufgabe, einen abgebildeten Würfel abzuzeichnen. Dieser muss dabei so exakt wie möglich in ein Zeichenfeld übertragen werden. Anhand einer Vier-Punkte-Skala wird eine Qualitätsmetrik wie in Animation 4.5 dargestellt aus vier Kriterien zusammengesetzt:

1. Parallelität (*PP*): Parallelität der Kanten und Seiten.
2. Dreidimensionalität (*PD*): Dreidimensionalität der gezeichneten Figur.
3. Frontseite (*PF*): Korrekte Darstellung der Frontseite.
4. Innere Kanten (*PI*): Korrekte Darstellung der inneren Kanten.

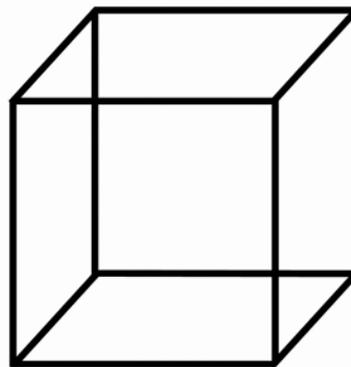


Abbildung 4.5: Würfel-Test<sup>1</sup>.

K testet die visuelle und räumliche Orientierung, indem vorgegebene Körperteile bei einer dargestellten Figur richtig erkannt und verortet werden müssen. Dabei zählt die Anzahl der falsch verorteten Körperteile. Um die kognitive Funktionsdomäne der komplexen Aufmerksamkeit zu testen, wird durch Test L und M mittels sogenannter Go/No-Go-Tests die selektive Aufmerksamkeit adressiert (vgl. Unterunterabschnitt 2.2.5). Dabei unterscheiden sich die beiden Tests lediglich durch die Art der verwendeten visuellen Reize. In L werden Symbole nacheinander in beliebiger Reihenfolge eingeblendet und der Nutzer ist dazu aufgefordert lediglich ein gewisses Symbol zu identifizieren und dieses zu bestätigen, falls es erscheint. Andernfalls sollte nicht mit dem Test interagiert werden. Demgegenüber werden in M anstelle von Symbolen beliebige Wörter gezeigt und der Nutzer dazu aufgefordert zu reagieren, sobald ein Wort zu einer vorgegebenen Kategorie passt. Dabei werden die Reaktionszeiten gemessen und die Anzahl an Fehlern aufgezeichnet, das heißt, ob ein Symbol oder Wort fälschlicherweise beziehungsweise eben fälschlicherweise nicht identifiziert und bestätigt wird. Mit dem Test N wird die soziale Kognition getestet, insbesondere die Wahrnehmung von Emotionen, der sogenannten Einsicht und Theory of Mind

<sup>1</sup> Die Animation wird nur in Adobe Acrobat Reader bei aktiviertem JavaScript dargestellt.

(vgl. Unterunterabschnitt 2.2.5). Dabei wird davon ausgegangen, dass ein sozialer Rückzug, der mit einer dementiellen Erkrankung einhergehen kann, ebenfalls Auswirkungen auf Fähigkeiten im Kontext der sozialen Wahrnehmung hat. Hierzu werden Bildergeschichten in Form von drei gezeichneten Abbildungen gezeigt, die eine Handlungssequenz darstellen. Eine Auswahl von drei weiteren Bildern stellen unterschiedliche Folgesequenzen der dargestellten Handlung dar, von denen nur eine von gesunden Menschen als richtig gewertet werden sollte. Der Nutzer muss das aus seiner Sicht passende Ende dieser Sequenz wählen. Die Anzahl an falsch gewählten Enden wird dabei aufgezeichnet. Zusammenfassend lassen sich die entsprechend relevanten Parameter durch das in Abbildung 4.6 dargestellte latente Variablenmodell repräsentieren.

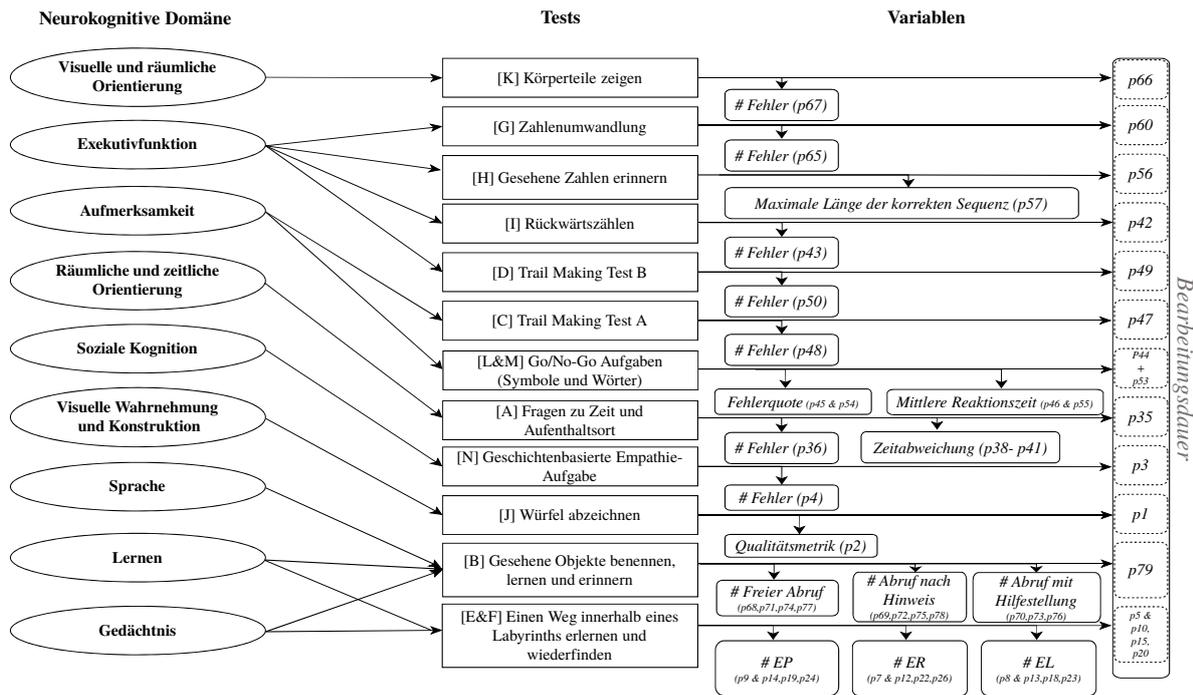


Abbildung 4.6: Modell der latenten Variablen (links) und der Tests (mittig) sowie relevante Leistungskriterien (rechts) nach Schinle et al., 2021 [471].

## 4.2.2 Grundsätzliche Anforderungen an eine digitale Gesundheitsanwendung

Aufgrund des Ziels und der zu dessen Erreichung identifizierten Parameter wird der im Kontext entstehender Gesundheitsdatenökosysteme zuvor abgeleitete Anforderungskatalog zum Umgang mit Gesundheitsdaten um grundsätzliche Anforderungen an digitale Gesundheitsanwendungen erweitert. In Tabelle 4.3 werden diese demnach induktiv anhand der definierten Anwendungsdomäne erschlossen und ebenfalls nach den in Abschnitt 4.1 abstrahierten Kategorien unterteilt. Dabei stehen vor allem die Technologien, Werkzeuge und Methoden zur Erfüllung dieser, empirisch innerhalb des wissenschaftlichen Umfelds, erhobenen funktionalen Anforderungen (FA) im Fokus des späteren Verlaufs der Arbeit (vgl. Abschnitt 1.4). Da diese bei der Entwicklung von Anwendungen, insbesondere im regulierten medizinischen Sektor, auch von nicht-funktionalen Anforderungen (NFA) flankiert werden, stehen ebenfalls Mechanismen zur Erfüllung dieser zur Diskussion. Aus medizinischer Perspektive sind Anforderungen aus verschiedenen Disziplinen zu berücksichtigen (vgl. Unterabschnitt 2.2.3, Unterabschnitt 2.2.4 und Abschnitt 3.3). Einem agilen Vorgehen folgend dient das durch die Kollaboration mit Experten aufzubauende und aus

den, durch Fachgesellschaften konsentierten, Leitlinien zu erschließende domänenspezifische Wissen als Grundlage für die Auswahl der eingesetzten Verfahren (DFAM1). Unter Berücksichtigung von Anforderungen der zuvor identifizierten Zielgruppe müssen digitale Lösungen realisiert werden, um eine frühzeitige Einschätzung bezüglich des Vorliegens von Symptomen zu ermöglichen und nachvollziehbar darzustellen (DFAM2, DFAM3). Grundsätzlich wird eine Unterscheidung zwischen zwei Rollen bei der Parametererfassung aus Sicht des Betroffenen vorgenommen. Einerseits existieren Verfahren, die eine aktive Rolle der Betroffenen in Form von Interaktionen mit technischen Systemen, deren Angehörigen oder Mediziner voraussetzen. Andererseits kommen Verfahren zum Einsatz, die eine passive Rolle für die Betroffenen zulassen, wobei auf Grundlage von Beobachtungen durch Mediziner, Angehörige oder technische Systeme eine Parametererfassung stattfindet. Um die Validität (DNFAM2, DNFAM2.1, DNFAM2.2, DNFAM2.3) der entwickelten Ansätze zu gewährleisten, setzt die Entwicklung dafür geeigneter technischer Systeme den kontinuierlichen Zugang zu Daten in entsprechender Menge und Qualität voraus (DNFAM1, DNFAM1.1, DNFAM1.2), was insbesondere bei Menschen mit AD, eine zu adressierende Herausforderung darstellt. Auch die Sicherheit der Nutzer im Umgang mit der Anwendung und den daraus resultierenden Ergebnissen muss gewährleistet sein (DNFAM3). Die strukturellen Anforderungen orientieren sich an der geltenden gesellschaftlichen Werteordnung und Rechtsprechung der Länder im Umgang mit Gesundheitsdaten und Produkten zu deren Erfassung, in denen die Anwendung in Verkehr gebracht wird. Auf europäischer Ebene bezieht sich dies insbesondere auf die EU-weit geltende MDR und DSGVO. Die Erfüllung dieser bildet auch die Grundlage für Zertifizierungen in Ländern und Kontinenten außerhalb der EU. Zur Sicherstellung einer nachhaltigen Verwertung von Ergebnissen sind ebenfalls ökonomische Aspekte mit in die Systementwicklung einzubeziehen. Hier steht vor allem die Frage nach möglichen Geschäftsmodellen und Anreizmechanismen im Vordergrund (DNFAE1, DNFAE3), insbesondere der Einfluss von gesetzlichen Änderungen bezüglich der Abrechenbarkeit von Leistungen gegenüber der GKVen (vgl. Unterunterabschnitt 2.3.1). Zusätzlich dazu sollte eine Freiwilligkeit der Nutzung bestehen bleiben (DFAS1). Um die Akzeptanz bei den jeweiligen Nutzergruppen zu steigern, besteht eine zeitökonomische Durchführbarkeit der Datenaufnahme und -interpretation ebenfalls als Anforderung an die Anwendung (DNFAE1). Die technischen Anforderungen resultieren aus den Kategorien zuvor und berücksichtigen entsprechend die primäre Zielgruppe bei der Umsetzung (DFAT1-DFAT5, DNFAT1-DNFAT4).

<b>Strukt.</b>	<b>DFAS1 - Freiwilligkeit:</b> Der Anwender darf zu keiner Zeit dazu gezwungen sein, Daten erheben und extern zur Verfügung stellen zu müssen.
<b>Ökon.</b>	<b>DNFAE1 - Ökonomische Tragfähigkeit:</b> Integration in tragfähiges Geschäftsmodell um einen nachhaltigen Betrieb sicherzustellen.
	<b>DNFAE2 - Zeitökonomische Durchführbarkeit:</b> Zeitökonomische Erfassung relevanter Parameter, um die Motivation aufrechtzuerhalten.
	<b>DNFAE3 - Anreizmechanismen:</b> Zusätzliche Anreize für eine kontinuierliche Datenerfassung und -überlassung.
<b>Med.</b>	<b>DFAM1 - Erfassung relevanter Parameter durch geeignete Messverfahren:</b> Erfassung der relevanten Parameter entsprechend offizieller Leitlinien und medizinischer Standards. Auswahl geeigneter Verfahren zur Bestimmung der Parameter.
	<b>DFAM2 - Reduktion stigmatisierender Wirkung:</b> Adressierung von Ängsten durch gezielte Information potentiell betroffener Menschen.

<b>Med.</b>	<b>DFAM3 - Übersicht über Messungen:</b> Darstellung der spezifischen Messergebnisse und des zeitlichen Verlaufs mehrerer Messungen.
	<b>DNFAMI - Qualität:</b> Für den Anwendungskontext ausreichende Zuverlässigkeit in der Erfassung und Auswertung der gesammelten Daten. <b>DNFAMI.1 - Korrekte Erfassung der Parameter:</b> Zuverlässigkeit in der Datenaufnahme. <b>DNFAMI.2 - Korrekte Auswertung der Messdaten:</b> Zuverlässigkeit in der Messdatenanalyse und Ableitung der Parameter.
	<b>DNFAM2 - Validität:</b> Ausreichende Validität und Nutzen der eingesetzten Messverfahren sowie der Anwendung insgesamt. <b>DNFAM2.1 - Genauigkeit der Verfahren:</b> Für den Anwendungskontext ausreichende Sensitivität und Spezifität der digitalisierten Verfahren. <b>DNFAM2.2 - Wiederholbarkeit:</b> Wiederholbarkeit der eingesetzten Messverfahren ohne Qualitätsverlust in den Ergebnissen. <b>DNFAM2.3 - Nachweisbarkeit:</b> Nachweisbarkeit des positiven Nutzens der digitalen Gesundheitsanwendung.
	<b>DNFAM3 - Nutzersicherheit:</b> Dem Anwender darf auch bei fehlerhaften oder unvollständigen Eingaben kein Schaden entstehen.
	<b>DNFAM4 - Transparenz:</b> Transparente Darstellung und Sicherstellung der Erklärbarkeit von Analyseergebnissen insbesondere beim Einsatz computerisierter Modelle.
	<b>Techn.</b>
<b>DFAT2 - Automatisierte Auswertung:</b> Automatisierbare Auswertung der Ergebnisse.	
<b>DFAT3 - Datenschutz:</b> Eine rein lokale und anonyme Auswertung muss möglich sein.	
<b>DFAT4 - Benutzerverwaltung:</b> Mehrere Nutzer müssen abbildbar sein.	
<b>DFAT5 - Nutzerzentrierte Zugriffs- und Nutzungsrechteverwaltung:</b> Die Vergabe von Zugriffs- und Nutzungsrechten für medizinische Daten kann durch den Patienten gesteuert werden.	
<b>DFAT6 - Nutzerautonomie:</b> Die Anwendung muss möglichst selbsterklärend sein, sodass im Umgang damit keine Unterstützung Dritter notwendig ist.	
<b>DNFAT1 - Zielgruppengerechte Gestaltung:</b> Intuitive und zielgruppengerechte Bedienbarkeit sowie Darstellung der Ergebnisse.	
<b>DNFAT2 - Zugänglichkeit:</b> Möglichst viele Personen müssen Zugang zur technischen Lösung bekommen können, um damit relevante Parameter zu erfassen.	
<b>DNFAT3 - Ortsunabhängigkeit:</b> System für den Einsatz auch außerhalb von klinischen Einrichtungen, wie beispielsweise im ambulanten Umfeld.	
<b>DNFAT4 - Interoperabilität:</b> Kombination mehrerer Systeme zur Erfassung verschiedener Parameter durch Interoperabilität zwischen digitalen Gesundheitsanwendungen. <b>DNFAT4.1 - Standardisierte Schnittstellen:</b> Implementierung standardisierter Schnittstellen. <b>DNFAT4.2 - Standardisiertes Datenformat:</b> Nutzung eines standardisierten und etablierten Datenstandards.	

Tabelle 4.3: Identifizierte Anforderungen an die Umsetzung digitaler Gesundheitsanwendungen.

# 5 Nutzerzentrierte Digitalisierung am Beispiel psychometrischer Tests

**D**AS zuvor abgeleitete Vorgehensmodell für eine datengetriebene Entwicklung digitaler Gesundheitsanwendungen setzt Werkzeuge und Methoden voraus, um die jeweiligen Schritte umzusetzen (vgl. Unterabschnitt 4.1.2). Zunächst steht die Transformation der vorgestellten, bisher analogen, Erfassungsmöglichkeiten im Vordergrund der Digitalisierungsphase und bildet somit den Kern dieses Kapitels (vgl. Abschnitt 5.1). Hierfür wird ein Architekturmodell entworfen (vgl. Unterabschnitt 5.1.3), bei dem die Erfüllung der in Kapitel 4 abgeleiteten Anforderungen im Fokus steht. Darin abgeleitete Komponenten bilden die Grundlage für den weiteren Aufbau dieser Arbeit. Als ein zentraler Aspekt wird die zielgruppengerechte Gestaltung von technischen Lösungen betrachtet. Demnach erfolgt neben der Optimierung einer Benutzerschnittstelle (vgl. Abschnitt 5.2), die Wahl von Technologien und Mechanismen zur Verwaltung darüber erhobener Daten (vgl. Abschnitt 5.3). Indem gezeigt wird, was bei einer nutzerzentrierten Umsetzung zu beachten ist, schließt dieses Kapitel, mit einer Untersuchung von Einflüssen der im Kontext einer agilen und iterativen Entwicklung stattfindenden Überführung der analogen Parametererfassung in deren digitales Äquivalent, ab (vgl. Abschnitt 5.4).

## 5.1 Koordination und Modellierung der Erfassungsmöglichkeiten

Um die aktive Erfassung von psychometrischen Leistungsdaten im ambulanten Umfeld zu ermöglichen, stellt sich in **Schritt 1** ebenso die Frage, wie eine digitale Lösung gestaltet sein muss, um in der medizinischen Versorgung akzeptiert zu werden. Dies macht eine bedarfsgerechte und damit nutzerzentrierte Digitalisierungsphase notwendig. Die zuvor skizzierten 14 Tests können Hinweise auf unterschiedliche Beeinträchtigungsformen der kognitiven Leistungsfähigkeit im Bereich einer AD liefern, um geeignete Optionen als Gegenmaßnahmen eruieren zu können. In Gestalt einer digitalen Gesundheitsanwendung kann eine diesbezügliche Einschätzung bereits beim Hausarzt oder zu Hause ermöglicht werden. Des Weiteren kann damit der Verlauf engmaschig durch eine regelmäßige Überprüfung der Leistungsfähigkeit kontrolliert werden [131]. Auch als Behandlungselement kann eine digitale Applikation einerseits die Verwaltung von Gegenmaßnahmen zur Adressierung von Risikofaktoren ermöglichen, andererseits als Trainingsanwendung für die mentale Gesundheit genutzt werden, was jüngsten Erkenntnissen zufolge ebenfalls einen positiven Effekt auf den Verlauf dieser Erkrankung haben kann [412].

### 5.1.1 Koordination der Parametererfassung

Um die Auswahl der zuvor getroffenen Parameterkombination zu eruieren, muss ein Datenbestand aufgebaut werden, der hinsichtlich der Zieldefinition valide Aussagen zulässt (vgl. Abschnitt 4.2). Dabei muss eine digitale Anwendung entstehen, welche die zuvor beschriebene und von Experten konsentrierte Auswahl (vgl. Unterabschnitt 4.2.1) integriert, um ein Profil der kognitiven Leistungsfähigkeit der Nutzer zu erhalten. Dabei ist zu beachten, wie die in Unterabschnitt 2.2.5 beschriebenen Verfahren umgesetzt werden, damit der Einfluss der Digitalisierung auf die klinisch validierten Verfahren möglichst gering gehalten wird. Um dies zu gewährleisten werden die Tests hintereinander in einer Sitzung durchgeführt (vgl. Abbildung 5.1). Des Weiteren müssen, zur Sicherstellung einer autonomen Durchführbarkeit der Tests durch die Anwender selbst (DFAT6), Instruktionen und Beispiele integriert werden, welche die Ergebnisse der Testdurchführung nicht beeinflussen.

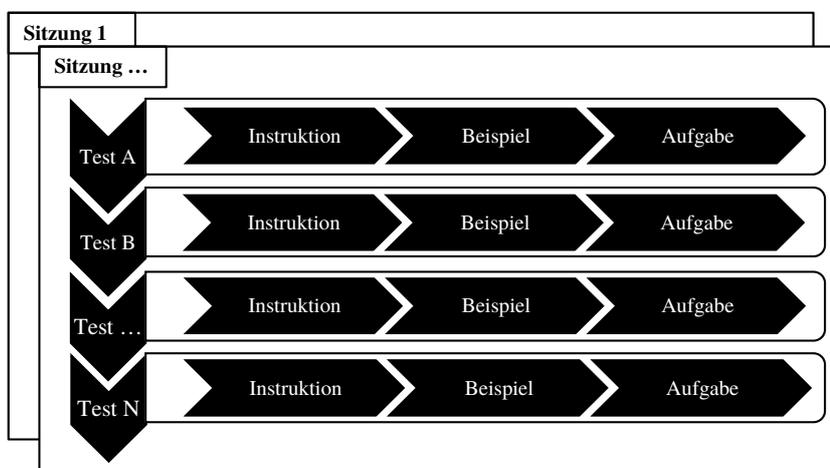


Abbildung 5.1: Aktive Parametererfassung.

Der Reihenfolge, in welcher einzelne Tests hintereinander durchgeführt werden, wird dabei kein Gewicht zugemessen. Hierbei wird lediglich beachtet, dass beim Ablauf einer Sitzung Interferenzaufgaben zwischen einzelnen Tests mit eingeplant werden müssen. Damit dies realisiert werden kann sowie um Vergleichbarkeit der damit generierten Daten herzustellen, muss allen Nutzern zunächst die gleiche Reihenfolge der Tests innerhalb einer Sitzung angeboten werden.

### 5.1.2 Modellierung einzelner Testabläufe

Die Prozessmodelle der einzelnen Testabläufe zur Erfassung der ausgewählten Parameter dienen als Grundlage für die Modellierung des zu digitalisierenden Gesamtablaufs der Parametererfassung. Nachfolgend werden diese modelliert, um die Ziele und den Testablauf der zuvor beschriebenen kognitiven Leistungstests zu verdeutlichen. Dies geschieht am Beispiel des FCSRT (Test B), dessen Inhalte komplett in eine digitale Version transformiert werden (vgl. Unterabschnitt 2.2.5). Nach einführender Anleitung und Beispielaufgabe, findet eine Lernphase mit drei Erinnerungszyklen statt, worin die Lernfähigkeit und Gedächtnisleistung quantifiziert werden (vgl. Abbildung 5.2).

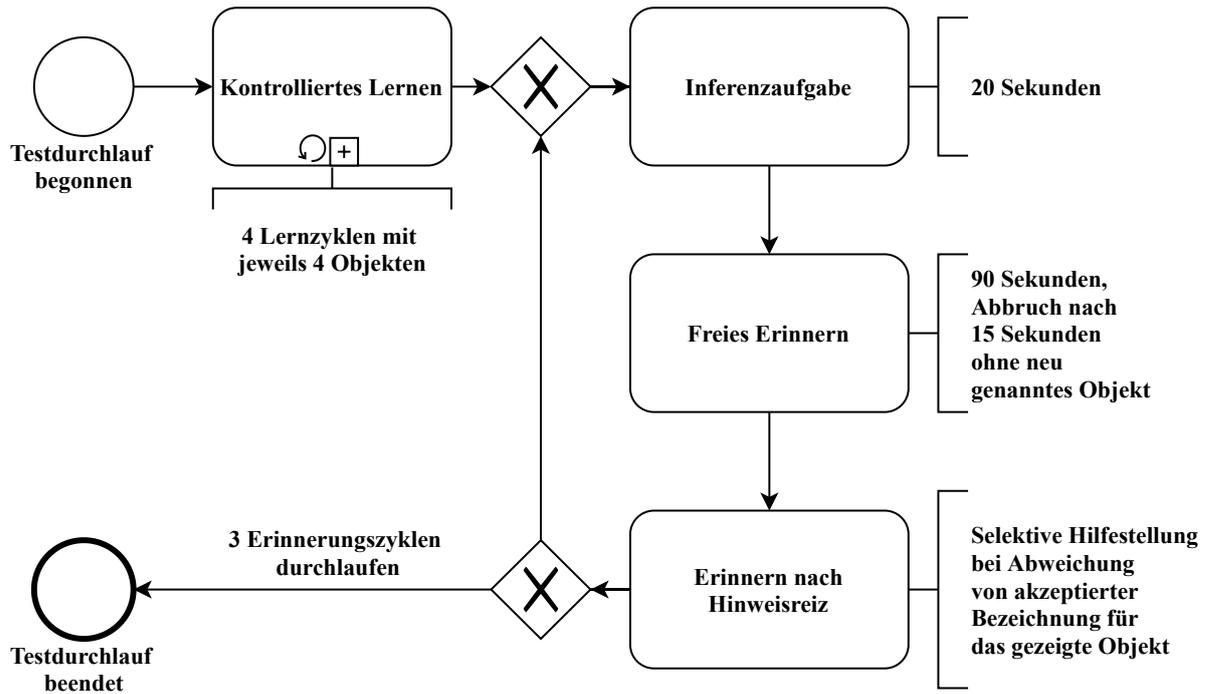


Abbildung 5.2: Testdurchlauf FCSRT aus Testteilnehmersicht.<sup>1</sup>

**Kontrolliertes Lernen:** Pro Durchlauf wird zunächst eine kontrollierte Lernphase mit Hilfestellung durchgeführt (vgl. Abbildung 5.3). Hierin muss der Testteilnehmer auf Grundlage von durch den Testleiter gegebenen Kategorie-Hinweisreizen, wie zum Beispiel „Zeigen Sie auf das Sportgerät und benennen Sie dieses“ auf abgebildete Objekte zeigen und diese konkret zum Beispiel durch die Antwort „Tennis-schläger“ korrekt benennen (vgl. Abbildung 5.4).

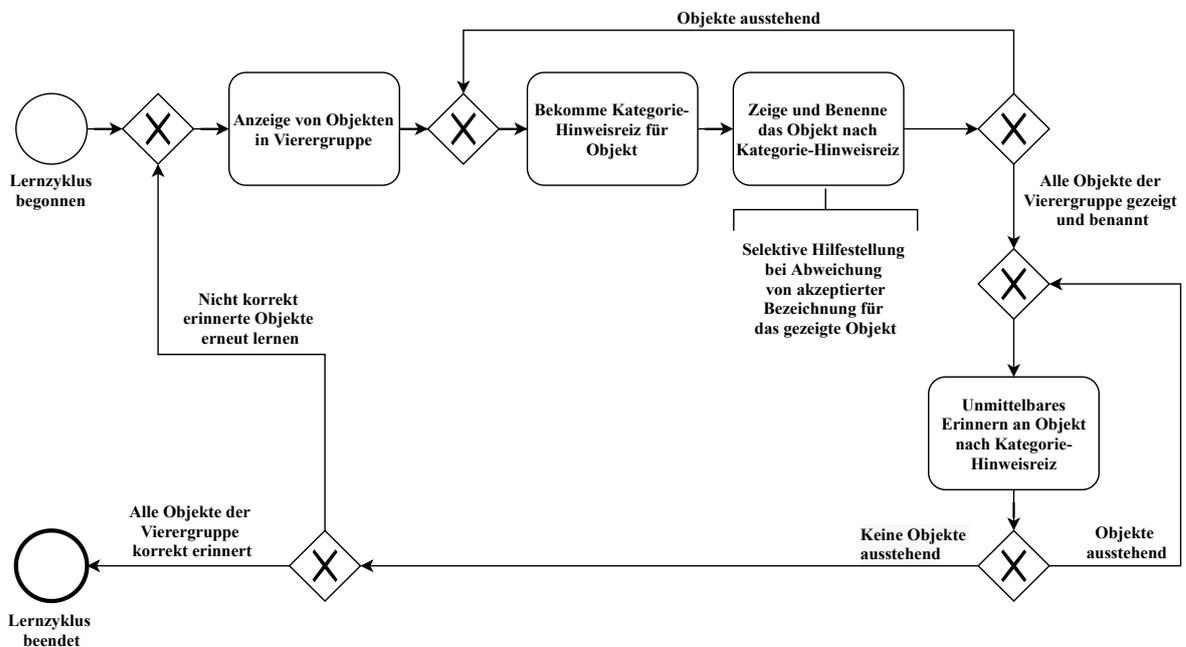


Abbildung 5.3: Unterprozess - Kontrolliertes Lernen aus Testteilnehmersicht.<sup>2</sup>

<sup>1</sup> modelliert nach Business Process Model and Notation (BPMN) 2.0.

<sup>2</sup> modelliert nach BPMN 2.0

Dieser Lernprozess findet stufenweise statt, das heißt es werden in einem Testdurchlauf insgesamt 16 Objekte gelernt, wobei vier Lernzyklen mit jeweils vier Objekten pro Abbildung gleichzeitig angezeigt und nacheinander eingepägt werden müssen, indem der Testteilnehmer auf die den entsprechend genannten Kategorie-Hinweisen zuzuordnenden Objekte zeigt und diese benennt. Jede Objektfigur lässt sich eindeutig einer Kategorie aus der vordefinierten disjunkten Menge von Kategorien zuordnen. Entscheidend ist, ob der Testteilnehmer einerseits auf das richtige Objekt zeigen und es andererseits korrekt benennen kann. Falls nicht das richtige Objekt gezeigt oder benannt wird, muss der Testleiter korrigierend eingreifen, indem er einen standardisierten Hinweis gibt (z.B. das Sportgerät heißt Tennisschläger). Alternative Bezeichnungen der Objekte, wie Synonyme oder regional akzeptierte Namen, die inhaltlich als korrekt zu klassifizieren sind, jedoch von der exakten Antwort abweichen, werden aus Gründen der Standardisierung des Vorgehens mit dem Hinweis versehen, wie das Objekt in der Folge benannt werden muss. Einzige Ausnahme bildet die Verwendung von akzeptierten Begriffsalternativen, welche sich auf die Anzahl beschränken, wie zum Beispiel der Plural von Begrifflichkeiten.

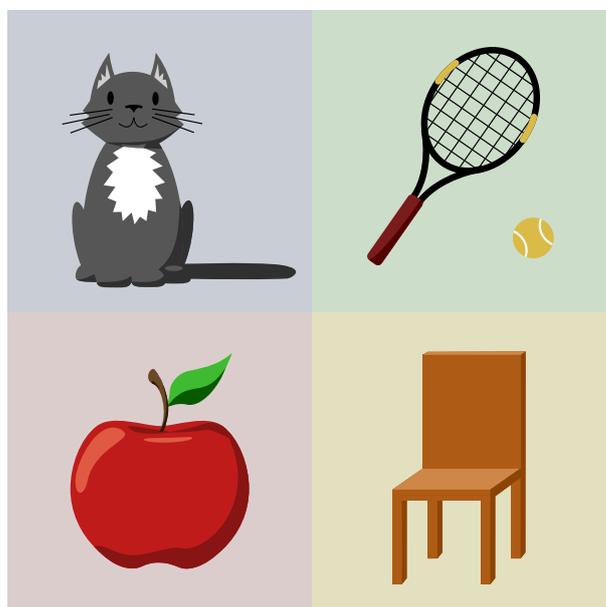


Abbildung 5.4: Beispielhafte Vierergruppe von Objekten für kontrolliertes Lernen im FCSRT.

Im Anschluss an jede Lernphase werden alle vier abgebildeten Objekte pro Gruppe in der gelernten Reihenfolge abgefragt. Hierbei werden lediglich die Kategorie-Hinweise nacheinander gegeben. Die gesehenen Objekte müssen aus dem Arbeitsgedächtnis frei abgerufen werden. Hierbei wird vermerkt, welche Objekte nach gegebenem Hinweisreiz unmittelbar abgerufen werden können. Für nicht erinnerte Objekte wird erneut die Abbildung der Vierergruppe gezeigt und gezielt in gleicher Weise wie zuvor auf diejenigen Objekte eingegangen, die nach den Kategorie-Hinweisen nicht richtig beantwortet wurden. Das kontrollierte Lernen endet nachdem alle 16 Objekte in vier Runden entsprechend der zuvor beschriebenen Vorgehensweise gelernt wurden.

**Interferenzaufgabe:** Nach der kontrollierten Lernphase findet eine Interferenzaufgabe statt. Diese dient dazu, den Lernvorgang mit einer ablenkenden Aktivität für 20 Sekunden zu unterbrechen. Dabei muss der Testteilnehmer in Dreierschritten von 100 rückwärts zählen.

**Freies Erinnern:** Nach der Interferenzaufgabe folgt die Phase des freien Erinnerns aller 16 Objekte ohne Hinweisreiz. Hierbei hat der Testteilnehmer 90 Sekunden Zeit diese in beliebiger Reihenfolge aufzuzählen. Der Testleiter notiert alle genannten Objekte. Wenn mehr als 15 Sekunden lang keine neuen Objekte genannt werden, welche zuvor nicht bereits erinnert wurden, wird diese Phase vorzeitig

abgebrochen.

**Erinnern nach Hinweisreiz:** Alle in der Phase des freien Erinnerns nicht genannten Objekte werden mit den bereits im kontrollierten Lernen verwendeten Kategorie-Hinweisreizen erneut abgefragt. Pro Hinweisreiz hat der Testteilnehmer zehn Sekunden Zeit die richtige Objektbezeichnung zu nennen. Bei korrekt benannten Objekten wird vermerkt, dass diese Objekte durch einen Hinweisreiz abgerufen werden konnten. Falls ein Objekt nicht abgerufen werden kann, erfolgt eine selektive Hilfestellung, indem die Bezeichnung des Objekts durch den Testleiter genannt wird.

Am Beispiel des FCSRT wird deutlich, wie umfangreich die notwendigen Interaktionen allein für einen Test sind. Welchen Einfluss dies auf die Digitalisierung hat, wird nachfolgend anhand des Systemkonzepts deutlich.

### 5.1.3 Systemarchitektur einer digitalen Gesundheitsanwendung

Zur Aufnahme und Verwaltung von Daten wird die Systemarchitektur einer digitalen Gesundheitsanwendung entworfen, welche die zuvor als relevant identifizierten Parameter digital aufnehmen und verarbeiten kann (DFAM1, DFAT1, DFAT2). Um Flexibilität gegenüber Erweiterungen durch spezifische Module in Abhängigkeit des Anwendungsfalls oder bezüglich weiterer aktiver oder passiver Parametererfassungen zu erreichen, wird eine modulare Systemarchitektur angestrebt (T11, T12). Diese setzt sich daher grundlegend aus den in Abbildung 5.5 dargestellten Komponenten zusammen.

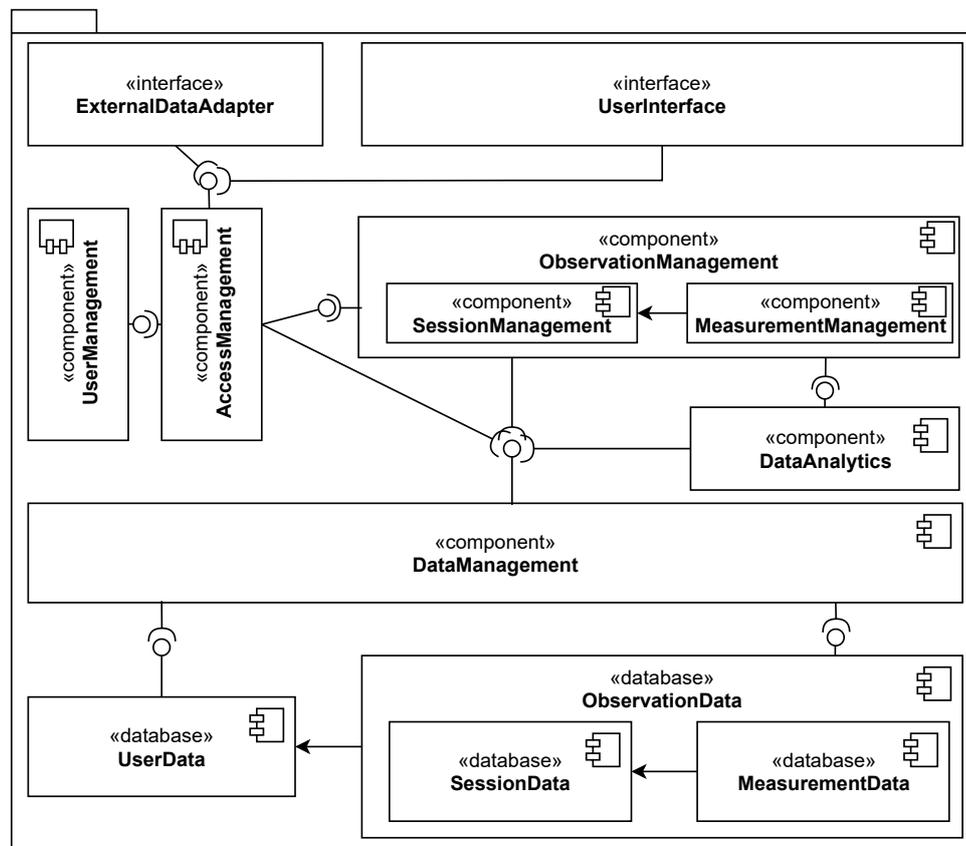


Abbildung 5.5: Komponentendiagramm der grundlegenden Systemarchitektur.<sup>3</sup>

Da es sich hierbei um ambulant einsetzbare Systeme handelt, werden möglichst geringe infrastrukturelle

<sup>3</sup> modelliert nach Unified Modeling Language (UML) 2.0.

Abhängigkeiten vorausgesetzt, die sich nicht negativ auf die Zuverlässigkeit und Validität des Systems auswirken (DNFAT1, DNFAT2, DNFAT3, DNFAM2, DNFAM4). Der Betrieb des Systems, insbesondere die damit verbundene Datenspeicherung und -verarbeitung, muss daher möglichst komplett auf der dafür eingesetzten Hardware stattfinden können (DFAT3). Aufgrund der Zielgruppe ist die potentielle Verwendung eines Geräts für mehrere Nutzer und somit der Personenbezug von erfassten Parametern eine zu adressierende Herausforderung. Hierfür muss die Zuordnung zwischen digitalen Identitäten und den generierten Daten stattfinden können (DFAT4). Die Komponente *UserManagement* ist für die Nutzerverwaltung zuständig und stellt daher eine Grundvoraussetzung für die nutzerzentrierte Umsetzung digitaler Gesundheitsanwendungen dar. In Kombination mit einer Komponente für das *AccessManagement* muss eine Zugriffsverwaltung bestehen (T1.4), die Datenschutzerfordernungen erfüllt indem Zugriffsrechte und Zweckbestimmungen zur Verarbeitung der nutzerspezifisch erhobenen Daten verwaltet werden (DFAT5). Diese, mit der digitalen Gesundheitsanwendung gesammelten Daten, müssen ebenfalls persistiert werden. Hierfür steht eine *DataManagement*-Komponente zur Verfügung, welche die Verwaltung der Daten *UserData* und *ObservationData* übernimmt. Für das Speichern und Verwalten von gesammelten Daten existieren etablierte Konzepte, je nach Entwicklungsstand und Einsatzort der Anwendung sind unterschiedliche Konzepte zu realisieren (vgl. Abschnitt 5.3). Hierbei ist insbesondere auf die Erfüllung der Interoperabilitätseigenschaften (DNFAT4.1, DNFAT4.2) und strukturellen Anforderungen zu achten (DFAS1). Zur kontrollierten Erfassung gesundheitsrelevanter Parameter wird eine Komponente benötigt, die diesen Vorgang steuert, da Fehleingaben zu einer Verfälschung der Ergebnisse im Kontext eines potentiell gesundheitsrelevanten Problems führen. Hierbei müssen auch durch externe Systeme und Erweiterungen erhobene Daten über Schnittstellen (*ExternalDataAdapter*) integrierbar sein (DNFAT4). Die Qualität der erfassten Parameter sowie der darauf basierenden Auswertung ist sicherzustellen (DNFAM1, DNFAM1.1, DNFAM1.2, DNFAM3). Da es sich im konkreten Anwendungsfall primär um eine aktive Parametererfassung durch den Betroffenen selbst handelt und eine zeitliche Komponente in Form von Sitzungen eine Rolle spielt, wird das *ObservationManagement* um eine Sitzungsverwaltung (*SessionManagement*) erweitert, welche die Durchführung der Sitzungen kontrolliert und erfasst. Die einzelnen Tests setzen sich aus mehreren Komponenten zusammen, welche ineinander greifen und in einer zeitlich sequentiellen Reihenfolge aufgerufen werden (vgl. Unterabschnitt 5.1.1). Hierfür ist die Komponente *MeasurementManagement* zur Verwaltung der konkreten Ausprägung einer Parametererfassung notwendig, welche in diesem Fall die korrekte Durchführung und Darstellung der Tests sowie Testergebnisse überwacht. Basierend auf den gesammelten Daten besteht gemäß der Zieldefinition aus Abschnitt 4.2 die Notwendigkeit für eine automatisierte Auswertung und Interpretation (DFAT2). Hierfür wird eine Komponente notwendig, welche dieser Anforderung nachkommt (*DataAnalytics*). Die dazu erforderlichen Algorithmen und Modelle können auf Grundlage von Studiendaten oder medizinischen Leitlinien erstellt werden (vgl. Kapitel 6). Diese sind in kompatible Formate zu überführen und in die digitale Gesundheitsanwendung zu integrieren. Da bei einer aktiven Parametererfassung mit digitalen Medien die Mensch-Technik-Interaktion von der Zielgruppe abhängt, steht die Gestaltung dieser Benutzerschnittstelle (*UserInterface*) und deren Inhalte im Vordergrund der Digitalisierung (DFAM2, DFAM3). Insbesondere im Hinblick auf den Anwendungsfall, müssen spezifische Anforderungen und Kriterien erfüllt sowie in die Systemlösung eingearbeitet werden, damit diese der jeweiligen Zielgruppe gerecht wird.

### 5.1.4 Technologieauswahl

Digitale Erfassungsmöglichkeiten gehen mit der Auswahl geeigneter Technologien einher, welche sich am angestrebten Ziel orientiert. Hierbei liegt der Schwerpunkt vor allem auf dem Einsatz weltweit etablierter

Technologien, welche neben den Digitalisierungs- auch Vernetzungs- und Automatisierungspotentiale realisierbar machen, wofür deren Zugänglichkeit und Verfügbarkeit zu eruieren ist (DNFAT2, DNFAT3). Die entsprechende Leistungsfähigkeit muss ebenso gewährleistet sein, wie ein Grad an Benutzbarkeit und Flexibilität (T3, T10, T11). Bezüglich der Zielgruppe und des, dem Anwendungskontext geschuldeten, ambulanten Einsatzgebietes stehen daher webbasierte und mobile Technologien im Fokus (vgl. Unterabschnitt 2.4.2). Zur Entwicklung der zuvor beschriebenen Software-Komponenten werden Formate, Protokolle und Plattformen bevorzugt, welche kostenlos zur Verfügung stehen und aktiv weiterentwickelt werden. Hierbei spielt die Vernetzung mittels interoperabler Protokolle und Datenformate (T6) zur Kombination dezentral (T7) verwalteter Datenquellen eine wesentliche Rolle, um den Detailgrad der Gesundheitsinformationen (M3) weiter zu erhöhen sowie übergeordnetes Wissen abzuleiten (M2). Da im Anwendungsfeld Gesundheitswesen der Datenschutz und -sicherheit (T1,T2) sowie Datensouveränität (S2) eine Rolle spielen, müssen hierfür ebenfalls Mechanismen identifiziert und implementiert werden. Um die zuvor beschriebenen Anforderungen zu erfüllen, bieten sich für den beschriebenen Anwendungsfall zur digitalen Abbildung klassischer *Pen&Paper*-Tests, mobile Geräten wie Smartphones oder Tablet-Computer als einzusetzende Hardware-Komponenten an. Hierbei bestehen durch die Integration von Sensoren und Schnittstellen, wie Funkprotokolle, berührungsempfindliche Bildschirme, digitale Stifte, Spracheingabe oder zusätzliche externe Peripherie, unterschiedliche Interaktionsmöglichkeiten. Im weiteren Verlauf der Arbeit werden die verwendeten Technologien und deren Eigenschaften im entsprechenden Abschnitt genauer spezifiziert. Die prototypische Implementierung der digitalen Gesundheitsanwendung für diese Arbeit erfolgt mittels mobilen Geräten mit Android<sup>4</sup>-Betriebssystem.

## 5.2 Gebrauchstauglichkeit der Benutzerschnittstelle

Die Gebrauchstauglichkeit digitaler Gesundheitsanwendungen hängt von der Wahl der eingesetzten Technologien und Interaktionsmöglichkeiten sowie den Charakteristiken der Benutzerschnittstelle ab (vgl. Absatz 2.4.2). Für den gewählten Anwendungsfall liegt die damit interagierende Zielgruppe bei Menschen höheren und mittleren Alters (vgl. Abschnitt 4.2). Mit dem Altern einhergehende körperliche und kognitive Veränderungen können sich auf die Interaktion mit und Akzeptanz von digitalen Lösungen auswirken und diese teilweise einschränken [44]. Dieser Umstand war in der Vergangenheit bereits ebenfalls Bestandteil wissenschaftlicher Studien und muss demnach in den Entwurf einer benutzerfreundlichen Schnittstelle (T3) in Form von weiteren Anforderungen einfließen (vgl. Unterunterabschnitt 3.3.2).

### 5.2.1 Anforderungen an eine grafische Benutzerschnittstelle

In Tabelle 5.1 werden die durch den natürlichen Alterungsprozess entstehenden Auswirkungen auf den Entwurf einer GUI zusammengefasst. Neben den zielgruppenspezifischen Anforderungen an eine Benutzerschnittstelle bestehen bei medizinischen Anwendungen ebenfalls gesetzliche Anforderungen an die Gebrauchstauglichkeit (vgl. Unterunterabschnitt 2.3.1). Als Anhaltspunkte, wie diese Anforderungen spezifiziert werden sollen, dienen die Normen DIN EN ISO 9241-110 und DIN EN ISO 9241-112 [116, 117, 180]. In ihrer Anwendung auf den beschriebenen Anwendungsfall fließen diese mit den Anforderungen durch zielgruppenspezifische Einschränkungen aus Abschnitt 5.2 in Unterabschnitt 5.2.2 zusammen. Da es sich bei der agilen Entwicklung um einen iterativen Prozess handelt, wird nachfol-

<sup>4</sup> <https://www.android.com/>, zuletzt abgerufen am 14.11.2021.

<b>Kategorie</b>	<b>Domäne</b>	<b>Veränderung</b>	<b>Auswirkung</b>
Körper	Sehen	Schärfe & Fokussierung	Die Visualität nimmt im Alter ab, was dazu führt, dass Darstellungen von Linien, Texten und Schattflächen abhängig von deren Größe leichter oder schwerer erkannt werden. Die Lesegeschwindigkeit und Genauigkeit in der Navigation innerhalb einer Anwendung nehmen bei kleiner werdenden Darstellungen deshalb ab.
		Textwahrnehmung	Je kleiner die Schriftgrößen desto schwerer werden diese wahrgenommen.
		Hell-Dunkel-Adaption	Das menschliche Auge reagiert in höherem Alter träger auf wechselnde Helligkeit und mit einer reduzierten Lichtaufnahme. Insgesamt wird dadurch helleres Licht benötigt, um gut sehen und lesen zu können. Die Kontrastempfindlichkeit nimmt ebenso ab, wodurch Informationen, die nicht in deutlichem Kontrast zu ihrem Hintergrund stehen schwerer erkannt werden.
Hören	Hörfunktion & Frequenzwahrnehmung	Farbwahrnehmung	Die Augenlinse trübt mit höherem Alter ein, Farben erscheinen dadurch verändert. Farbunterscheidungen im grünen, blauen, violetten und roten Spektrum werden dadurch erschwert.
			Durch eine eingeschränkte auditive Wahrnehmung nehmen die Fähigkeiten ab, Hintergrundgeräusche entsprechend herauszufiltern und leise Töne in Frequenzen aus den höheren beziehungsweise tieferen Bereichen wahrzunehmen.
Tastsinn	Feinmotorik	Abnehmende Geschicklichkeit und Fertigkeiten, die mit der taktilen Wahrnehmung einhergehen, erschweren die Bedienung von berührungsempfindlichen Bildschirmen insbesondere die Interaktion mit klein dargestellten Objekten durch Gesten.	
Kognition	Gedächtnis	Lernvermögen	Die kognitive Leistungsfähigkeit nimmt im Alter ab (vgl. Abbildung 2.7). Eine Verlangsamung bei der Aufnahme und Verarbeitung von Informationen sowie beim Abruf bereits gelernten Wissens, wirkt sich daher auch auf die Interaktion mit Technik aus. Hierzu zählt sowohl die abnehmende Fähigkeit sich in virtuellen Räumen, wie Menüstrukturen zurechtzufinden als auch die nachlassende Konzentrationsfähigkeit.
		Orientierung	Das nachlassende Langzeitgedächtnis führt ebenfalls zu einer abnehmenden Fähigkeit sich in virtuellen Räumen, wie Menüstrukturen zurechtzufinden.
		Aufmerksamkeit	Die natürlich nachlassende kognitive Leistungsfähigkeit führt zu höheren Reaktionszeiten. Die Konzentrationsfähigkeit lässt nach, was sich ebenfalls negativ auf die Fähigkeit irrelevante Stimuli zu ignorieren auswirkt.
Erfahrung	Technikertätigkeit	Mangelnde Erfahrung älterer Menschen mit neuen Technologien kann zu Angst vor Fehlern führen. Ebenso führt fehlendes Wissen bezüglich alternativer Lösungswege dazu, dass Aufgaben schneller abgebrochen werden.	

Tabelle 5.1: Auswirkungen zielgruppenspezifischer Einschränkungen auf die Gestaltung einer Benutzerschnittstelle nach [23, 44, 105, 107, 132, 133, 154, 264, 419].

gend die Grundlage zur Erstellung eines Leitfadens für das Design der Benutzerschnittstelle und deren Implementierung sowie der Planung und Durchführung von Evaluationsstudien pro Iteration gelegt.

<b>Elemente</b>	<b>UI1 - Deutlich:</b> Elemente umranden und schattieren, um zu verdeutlichen, dass mit diesen interagiert werden kann [105, 132, 259].
	<b>UI2 - Selbsterklärend:</b> Einsatz selbsterklärender Elemente, Beschriftungen und Symbolik [44, 107].
	<b>UI3 - Unterscheidbar:</b> Sich ähnelnde Elemente, die mit unterschiedlichen Aktionen verknüpft sind, deutlich durch Farbgebung, Beschriftung oder Symbolik von anderen unterscheiden.
	<b>UI4 - Positioniert:</b> Größe und Abstände der Elemente auf visuelle und motorische Beeinträchtigungen anpassen [105, 132, 264]. Empfohlene Mindestgrößen für berührungsempfindliche Flächen von 12 mm und Mindestabstand zwischen einzelnen Elementen von 6 mm, was in etwa einer Fingerkuppe entspricht [105]. Mindestabstände und visuelle Zugehörigkeit von Elementen stehen in einem Spannungsfeld, in Ausnahmefällen ist es deshalb sinnvoll von den Mindestabständen abzuweichen. Die berührungsempfindlichen Flächen sollten jedoch immer der Fläche des Elements entsprechen.
	<b>UI5 - Bekannt:</b> Insbesondere für technikaverse ältere Menschen erscheinen Icons nicht intuitiv, deshalb sind Beschriftungen zu bevorzugen [132, 134]. Falls dennoch Icons eingesetzt werden sollen, ist darauf zu achten, dass diese auf allgemein bekannte Symbolik zurückzuführen sind [442].
<b>Navigation</b>	<b>UI6 - Logisch:</b> Um fehlerhafte Bedienung und dadurch ausgelöste Frustration zu vermindern, sollte die Navigation innerhalb einer Anwendung konsistent und in logischen Schritten gestaltet sein [442].
	<b>UI7 - Eindeutig:</b> Navigationselemente sollten konsistent, gut sichtbar und in ihrer Funktion eindeutig interpretierbar gestaltet sein. Dafür müssen diese sich ebenso optisch von anderen Steuerelementen unterscheiden sowie allgemein bekannte Symbolik oder eindeutige Bezeichnungen für Navigationselemente verwenden [44, 105, 132, 442].
	<b>UI8 - Interpretierbar:</b> Der Einsatz von Gestensteuerung erschwert die Navigation für eine ältere Zielgruppe (vgl. Tabelle 5.1). Daher sind für die Navigationselemente eindeutige oder vom Betriebssystem vorgegebene bekannte Elemente zu wählen [44, 105, 132, 442].
	<b>UI9 - Flach:</b> Relevante Funktionen sollten auf dem Startbildschirm angezeigt werden. Dieser dient als inhaltliche Funktionsübersicht der Anwendung sowie Einstiegs- und Rückkehrpunkt. Flache Navigationshierarchien sind gegenüber tiefen zu bevorzugen [105, 107, 261].
	<b>UI10 - Reversibel:</b> Dem Nutzer muss zu jeder Zeit deutlich gemacht werden, an welcher Stelle er sich in der Anwendung befindet [105, 132]. Unbeabsichtigte Aktionen müssen über ein Navigationselement wieder rückgängig gemacht werden können [105, 133].
<b>Layout (Anordnung)</b>	<b>UI11 - Übersichtlich:</b> Beim Layout der Anwendung muss auf die Übersichtlichkeit, Klarheit und Konsistenz der Darstellung geachtet werden [105, 264, 442].
	<b>UI12 - Adaptiv:</b> Betriebssystemspezifische Interaktions- und Darstellungsparadigmen sollten beibehalten werden, sofern kein Widerspruch zu anderen Gestaltungsempfehlungen besteht [132]. Geräteübergreifende Konsistenz der Inhalte ist jedoch sicherzustellen, dementsprechend muss auf verschiedene Bildschirmgrößen, -ausrichtungen und Betriebssysteme adaptiv reagiert werden [105, 132].
	<b>UI13 - Platziert:</b> Navigationsmenüs und Suchfelder müssen gut sichtbar und an derselben Stelle platziert sein, wie bei anderen Anwendungen. Präferiert wird entweder oben mittig oder oben rechts [105, 132]. Steuerelemente, die häufig verwendet werden und relevante Aktionen auslösen, sind im oberen Bereich des Bildschirms zu platzieren, um versehentliches Auslösen einer Aktion weitestgehend zu vermeiden [23].

<b>Layout (Anordnung)</b>	<b>UI14 - Zugehörig:</b> Für die Bestätigung von Aktionen und Inhalten verwendete Elemente, sollten unter dem zu bestätigenden Inhalt platziert werden [133].
	<b>UI15 - Relevant:</b> Irrelevante Informationen sind zu vermeiden, relevante Informationen sollten, um schnell gefunden zu werden, sinnvoll strukturiert primär mittig angeordnet werden. Für mehrseitige Inhalte gilt es eindeutig erkennbare Unteransichten zu gestalten, die sich in die Gesamtstruktur einfügen [44, 105, 132, 264, 442].
<b>Farbkonzept</b>	<b>UI16 - Kontrastreich:</b> Ausreichender Kontrast muss zwischen den Elementen vorhanden sein. Farben sollten sparsam eingesetzt werden, insbesondere ist der Verzicht auf Komplementärfarben bei nebeneinander angeordneten Elementen zu beachten [44, 105, 132, 133, 264].
	<b>UI17 - Konsistent:</b> Empfehlung von schwarzer Schrift auf einem hellen Hintergrund. Um Ermüdungserscheinungen der Augen entgegenzuwirken sind gemusterte oder rein weiße Hintergründe zu vermeiden. Die Hintergrundfarbe sollte über alle Ansichten hinweg konsistent sein, um Helligkeitsänderungen zu verhindern [44, 105, 132, 133, 264, 442].
<b>Sprache und Schrift</b>	<b>UI18 - Zielgruppenorientiert:</b> Die Sprachgestaltung muss sich an der Zielgruppe orientieren, eine einfache und konsistente Sprache ist zu bevorzugen. Fachbegriffe oder Fremdwörter sind insbesondere bei Navigationselementen zu vermeiden [105, 132, 259, 264, 442].
	<b>UI19 - Standardisiert:</b> Prinzipiell sollte auf Standardschriftarten des Betriebssystems zurückgegriffen werden, Schriftarten aus der Sans-Serif-Familie sind bei selbst gewählten Schriften zu bevorzugen. Als Mindestgröße wird das typographische Maß 12 Punkt empfohlen, Kursiv- und Majuskelschrift (Großbuchstaben) sollten vermieden werden [44, 105, 133, 264].
	<b>UI20 - Komprimiert:</b> Bezüglich der Darstellung von Texten wird eine linksbündige Ausrichtung empfohlen. Hierbei sollen Informationen komprimiert, die wichtigsten Punkte hervorgehoben und mit eindeutigen Überschriften versehen werden [105, 442].
<b>Interaktionsdesign</b>	<b>UI21 - Minimalistisch:</b> Manuelle Eingaben über virtuelle Tastaturen sollten minimiert werden. Falls Eingaben notwendig sind, sollte das Tastaturlayout entsprechend dem Inhalt des auszufüllenden Feldes (Zahlen/Text) angepasst sein. Eingabefelder sollten die Inhalte gut leserlich darstellen [105, 107, 132, 133].
	<b>UI22 - Fehlertolerant:</b> Eine fehlertolerante Gestaltung der Eingabemöglichkeiten sollte durch eindeutige Fehlermeldungen und Anweisungen, wie damit umzugehen ist sowie diese zu korrigieren sind intuitiv gestaltet werden [105, 132, 264, 442].
	<b>UI23 - Multi-sensorisch:</b> Multi-sensorische Rückmeldungen an den Nutzer durch Vibration oder Tastenton sollten das Auslösen von Aktionen unterstützen. Ungewollte Aktionen sollten abgebrochen werden können sowie bei destruktiven Aktionen Hinweise gegeben werden, welche Auswirkung diese haben. Eine explizite Aufforderung zur Bestätigung destruktiver Aktionen wird empfohlen [105, 132, 264].
	<b>UI24 - Unterstützend:</b> Hilfestellungen bezüglich der Navigation und Inhalte der Anwendung sollten deren Bedienung nicht beeinträchtigen. Animationen sollten hierbei nur dann zum Einsatz kommen, wenn diese einen Mehrwert darstellen und sollten nicht automatisch im Hintergrund gestartet werden. Hilfestellungen müssen klar erkenntlich gemacht und vom Nutzer übersprungen werden dürfen [105].
	<b>UI25 - Informierend:</b> Der Zugriff auf zusätzliche Informationen über referenzierte Hyperlinks sollte einheitlich und eindeutig dargestellt werden [105].

<b>Interaktionsdesign</b>	<b>UI26 - Kontextsensitiv:</b> Interaktionsformen, die in Verbindung mit Gesten stehen, wie beispielsweise das Wischen zum Scrollen auf einer Seite, sollten auf ein Minimum reduziert beziehungsweise nur als zusätzliche Alternative angeboten werden. Um auf diese Interaktionsmöglichkeiten aufmerksam zu machen, sollten kontextsensitive Hinweise eingeblendet werden [44, 105, 132, 264].
---------------------------	--

Tabelle 5.2: Leitfaden mit zielgruppenspezifischen Anforderungen an die Benutzerschnittstelle.

## 5.2.2 Entwicklung einer grafischen Benutzerschnittstelle

Der Zusammenhang zwischen den Anforderungen aus Abschnitt 5.2 und den zuvor beschriebenen Normen ist in Abbildung 5.6 dargestellt. Dadurch werden die darin enthaltenen Grundsätze mit Nutzungsanforderungen sowie Gestaltungsleitlinien verknüpft und damit tangibel, um in die Gestaltung der grafischen Benutzerschnittstelle mit einfließen zu können.

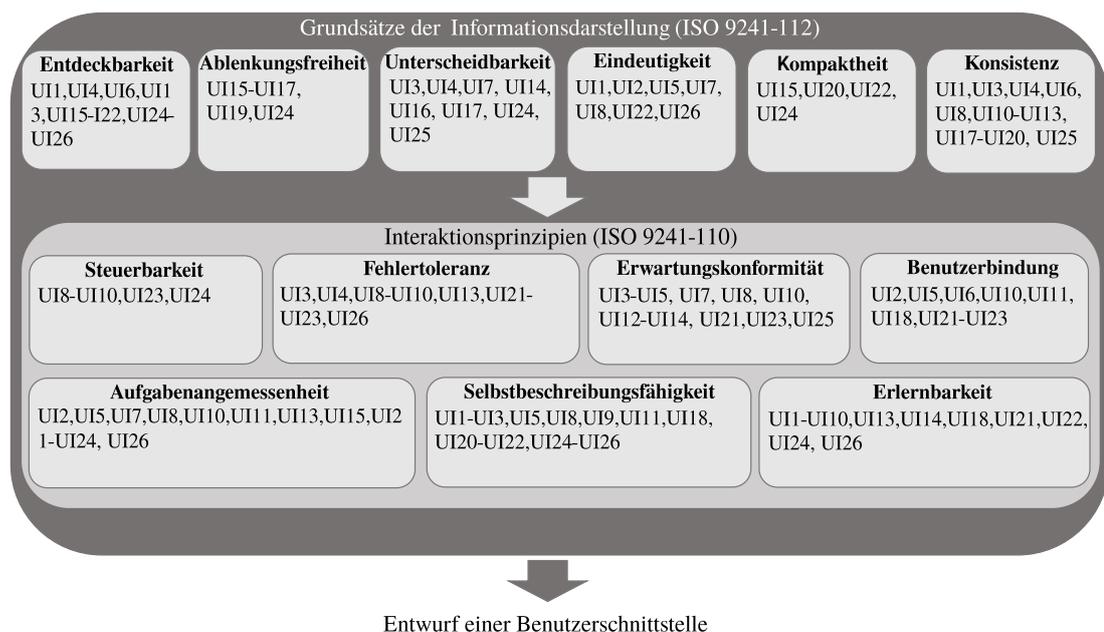


Abbildung 5.6: Verknüpfung von zielgruppenspezifischen Anforderungen und Normen für eine gebrauchstaugliche Benutzerschnittstelle.

In der Norm 62366-1:2015 der International Electrotechnical Commission (IEC) wird die Planung und Durchführung von formativen und summativen Bewertungen verlangt. Pro Iteration findet eine formative Evaluation in Form von empirischen Untersuchungen durch Interviews, Gebrauchstauglichkeitstests und Fragebögen statt, um Erkenntnisse zu sammeln, damit eine Weiterentwicklung des Artefakts entsprechend der neu gewonnenen Erkenntnisse vollführt werden kann. Im Gegensatz zur summativen Bewertung wird dadurch keine abschließende Bewertung als Nachweis einer sicheren Nutzerschnittstelle erbracht [180]. Qualitative Aussagen durch die zuvor benannten Gebrauchstauglichkeitstestformate sind für diesen Anwendungsfall aufgrund von individuellen Einschränkungen der teilweise multimorbiden Zielgruppe, wie

beispielsweise deren Verfügbarkeit, Teilnahmebereitschaft und eingeschränkter Mobilität möglicherweise aussagekräftigeren quantitativen Tests im Labor vorzuziehen. Auch der Einsatz digitaler Werkzeuge zur Unterstützung bei der Durchführung dieser kann aufgrund von infrastrukturellen Defiziten im ambulanten Umfeld nicht garantiert werden. Daher gliedert sich das Vorgehen nachfolgend in die im Rahmen dieser Arbeit erreichten vier Iterationen, von denen sich jeweils zwei in die Bereiche Wissenschaft und zwei in die Phase der Produktentwicklung des Medizinproduktlebenszyklus einordnen lassen (vgl. Abbildung 4.2). In den ersten beiden Iterationen der Wissenschafts-Phase wird aufgrund der stigmatisierenden Wirkung von AD das medizinische Personal als Beobachter eingesetzt. Für jede Iteration werden zunächst die geplanten Änderungen an der Benutzerschnittstelle mittels Mockups dargestellt und durch diese evaluiert (vgl. Absatz 2.4.2). Anhand der im Anschluss implementierten Version der Anwendung werden, parallel zur Testdurchführung durch die Studienteilnehmer, der gewählten TAM inhärenten Protokolle, Informationen bezüglich der Wahrnehmung und zu den jeweiligen Gestaltungselementen der Tests erhoben (vgl. Unterabschnitt 2.1.2). Über einen Fragebogen zur Technik-Erfahrung werden Kontrollvariablen eingeführt, die eine Abhängigkeit zur Einstellung gegenüber Technik generell sowie der persönlichen Anwendungserfahrungen der Studienteilnehmer mit digitalen Technologien darstellt (vgl. Abschnitt A.7). Um die Anwendung auf Normen-Konformität zu prüfen und gleichzeitig Informationen über die Entwicklungsschritte bezüglich der Gebrauchstauglichkeit zu erhalten, wird ein evaluierter Fragebogen mit 21 Fragen zu den Richtlinien der ISO Norm 9241-110 eingesetzt (vgl. Abschnitt A.11) [144, 334, 335]. Darin werden die sieben Interaktionsprinzipien jeweils mit drei Fragen auf einer siebenstufigen Skala von sehr negativ (- - -) bis sehr positiv (+++) bewertet und für die Auswertung in ganzzahlige Werte zwischen -3 und +3 kodiert sowie pro Interaktionsprinzip arithmetisch gemittelt. Als Richtwert für eine gute Bewertung gilt +1, Werte darunter stellen einen Indikator für Optimierungsbedarf dar [334]. Um das Delta des Versionsprungs zu bestimmen, findet die Beantwortung des Fragebogens durch die Teilnehmer zweimal statt, einmal auf Grundlage der bestehenden Anwendungsversion und danach basierend auf Mockups des Designs der neuen Version. Daher folgen qualitative Fragen zu einzelnen Tests, die sich verändert haben, um konkrete Verbesserungsvorschläge zu generieren. Für jeden Test wird dazu erfragt, ob er verständlich aufbereitet ist und um eine Begründung dafür gebeten, falls aus Sicht des Befragten etwas fehlt. Außerdem werden Probleme bei der Durchführung ebenfalls mit Bitte um Begründung erfragt. Offene Fragen zum direkten Vergleich zwischen der aktuellen und nachfolgenden Version der Anwendung mit einer Erläuterung von Dingen die für die bestehende Version besser gefallen als für das Design der neuen Version und umgekehrt werden ebenso wie weitere Anmerkungen und Verbesserungsvorschläge erfasst. Hierzu ist zu ergänzen, dass Iteration I zur generellen Überprüfung des angestrebten Ziels, der Funktionalität und Praktikabilität der Anwendung als Vorstudie dient. Die hieraus erschlossenen Erkenntnisse fließen in Iteration II ein, welche als Grundlage der Pilotstudie zur wissenschaftlichen Überprüfung der Erreichbarkeit des medizinischen Ziels dient. Zur Integration von Produkthanforderungen findet ab Iteration III auf Grundlage der zuvor gesammelten Erfahrungen eine Fokussierung auf die Darstellung einzelner Tests und Elemente auf unterschiedlichen Bildschirmgrößen sowie die Anreicherung um Kontextinformationen zum Thema AD statt. Die Aufteilung der Gesundheitsanwendung in zwei verschiedene Versionen und damit unterschiedliche Zweckbestimmungen spielt dabei eine Rolle. Einerseits steht die Entwicklung einer Gesundheitsanwendung zur möglichst frühzeitigen Diagnose einer AD im Vordergrund der Wissenschaftsphase. Andererseits, aufgrund der im Rahmen eines Forschungsvorhabens nicht vollumfänglich adressierbaren gesetzlichen Anforderungen an die Medizinproduktentwicklung, die Gestaltung eines Informationsproduktes für Angehörige und Betroffene, um die Öffentlichkeit über diese Krankheit und den damit zusammenhängenden Abläufen informieren zu können und deren stigmatisierenden Wirkung entgegenzutreten. Die für diese beiden Versionen einzunehmende Gestaltungsperspektive wird im ambulanten Umfeld mit Studienteilnehmern formativ evaluiert. Am Beispiel des Labyrinth-Tests (E und

F) lässt sich das Kondensat der damit erreichten Evolution nachvollziehen (vgl. Abbildung 5.7). Um einen

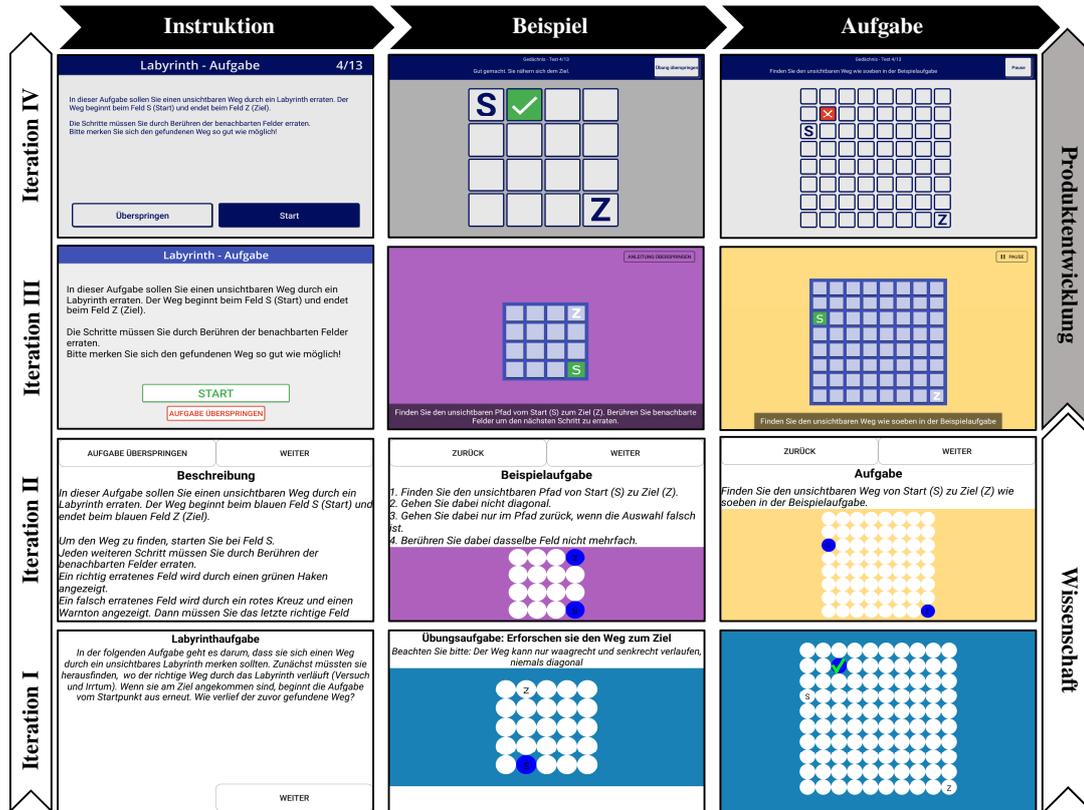


Abbildung 5.7: Evolution der GUI am Beispiel des Test E in der Tablet-Version.

Eindruck davon zu bekommen, wie die zuvor aufgestellte Leitlinie hierfür angewandt wird und dennoch den Umfang dieser Arbeit nicht überzustrapazieren, wird dieser Transformationsvorgang am Beispiel der Transition von Iteration III auf Iteration IV nachfolgend dargestellt. Das kumulative Ergebnis der damit verbundenen Evaluationsstudie wird in Abschnitt 8.1 ausgeführt, diskutiert und bewertet. Der Erkenntnisprozess und dessen Auswirkung auf die GUI wird deshalb lediglich anhand der Kategorie „Elemente“ (UI1-UI5) skizziert. Bezüglich der Umgestaltung der Kategorie „Elemente“ bedeutet dies für *DemPredict* die Umrandung, Schattierung (UI1) und beschreibende Gestaltung aller Bedienelemente (UI2, UI3). Als ein Beispiel können hier die Schaltflächen bei der Registrierung als Nutzer über die Zugriffsverwaltung angeführt werden, um das Geschlecht anzugeben. Auch die angepasste Beschriftung und Symbolik soll zu einem selbsterklärenden Umgang mit diesen Elementen dienen (vgl. Abbildung 5.8).

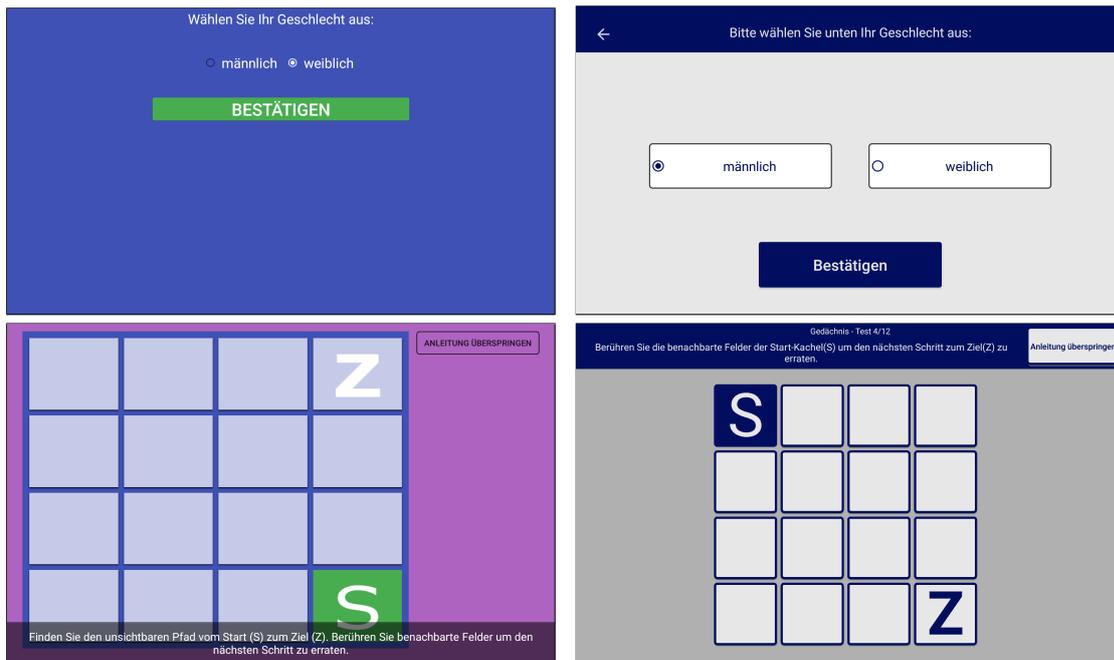


Abbildung 5.8: Beispiel neu gestalteter Bedienelemente in *DemPredict* im Vergleich zwischen Iteration III (links) und Iteration IV (rechts) der Tablet-Version.

Um eine bessere Unterscheidung des Start- und des Zielfeldes zu gewährleisten und das Problem des versehentlichen Vertauschens von Start und Ziel durch farbliche Abhebung zu reduzieren, wird für den Test E sowohl ein blauer Hintergrund für das Startfeld (UI3) gewählt als auch die Abstände zwischen den Eingabefeldern erweitert (UI4) (vgl. Abbildung 5.8). Für die Interaktion mit der Anwendung werden zur Vereinfachung an den Stellen, an denen die Bedienung der Interaktionselemente zuvor nicht eindeutig war, zusätzliche Schaltflächen eingefügt (UI1). Beispielsweise gilt dies für die Auswahl der Frequently Asked Questions (FAQ) und die Bedienung von Drehreglern (vgl. Abbildung 5.9). Schaltflächen, die destruktive

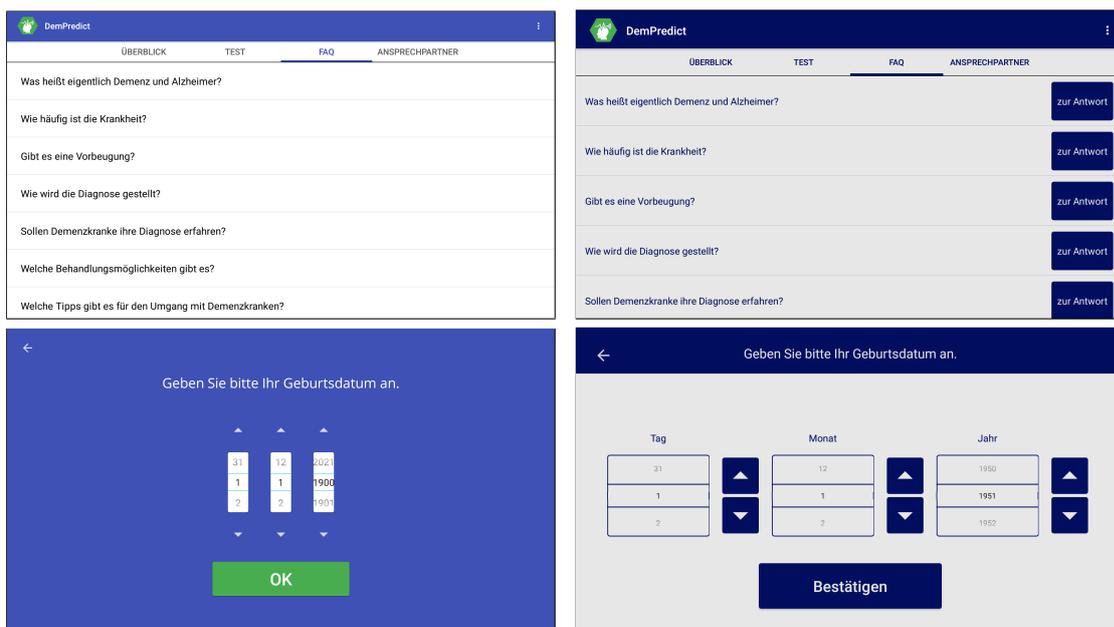


Abbildung 5.9: Beispiel hinzugefügter Bedienelemente in *DemPredict* im Vergleich zwischen Iteration III (links) und Iteration IV (rechts) in der Tablet-Version.

Aktionen ausführen, werden dem Farbschema folgend in einer differierenden Farbe dargestellt (UI5) und über die Anwendung hinweg konsistent an Stellen auf dem Bildschirm positioniert, an denen sie nicht versehentlich ausgelöst werden können. Beispielsweise werden diese bei den Erklärungen der Tests anschließend unten links platziert (vgl. Abbildung 5.10).

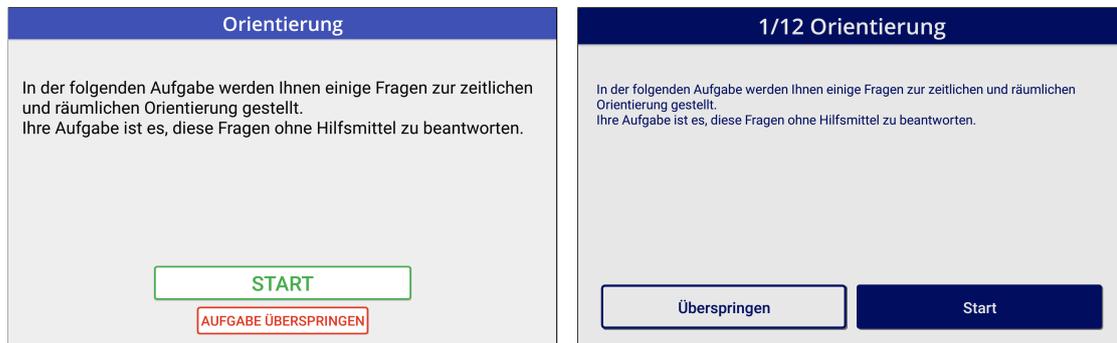


Abbildung 5.10: Beispiel hinzugefügter Bedienelemente in *DemPredict* im Vergleich zwischen Iteration III (links) und Iteration IV (rechts) in der Tablet-Version.

## 5.3 Selbstbestimmte Datenverwaltung

Während die grafische Benutzerschnittstelle als Komponente zur Aufnahme relevanter Parameter dient, bildet die Verwaltung dieser Parameter in Form von digitalen Daten die Aufgabe der Datenverwaltungskomponente (vgl. Abbildung 5.5). Hierin laufen sowohl die Daten der Nutzer als auch der damit verbundenen Beobachtungen ihrer Gesundheitszustände zusammen. Digitale Gesundheitsanwendungen sollen zukünftig einen direkten Beitrag zur Verbesserung der Versorgung leisten (vgl. Unterunterabschnitt 2.3.1), können aber auch indirekt dazu genutzt werden Daten zu Forschungszwecken zu erheben, weshalb nachfolgend sowohl die Aspekte aus Forschungs- als auch Produktentwicklungssicht tangiert werden. Hier zeigt sich während der iterativen Entwicklung, welche unterschiedlichen Konzepte in den jeweiligen Phasen zu berücksichtigen sind (vgl. Unterabschnitt 5.3.1). Aufgrund regulatorischer Vorgaben der DSGVO und des PDSG steht das Datensubjekt im Mittelpunkt der Gesundheitsdatenverwaltung, weshalb diese Selbstbestimmtheit (S2) auch bei der Entwicklung digitaler Gesundheitsanwendungen eine Rolle spielt (vgl. Unterabschnitt 2.3.2). Damit einhergehende Anforderungen und technische Möglichkeiten zur Umsetzung von Datennutzungsbedingungen werden in Unterabschnitt 5.3.2 weiter untersucht. Im Laufe der Entwicklung wird ebenso ein Übergang von proprietären Datenformaten und -strukturen hin zu standardisierten Formaten notwendig, um Produktkonformität und -interoperabilität zu gewährleisten (vgl. Unterunterabschnitt 2.4.2). Die in Unterabschnitt 2.4.1 vorgestellten Entwicklungen dienen als Entscheidungsgrundlage bei der Abbildung von gesammelten Gesundheitsdaten auf entsprechende Standards sowie deren organisatorische Umsetzung (S1.2, T6.3, T6.4, T6.5, T6.6), welche in Unterabschnitt 5.3.3 genauer untersucht werden. Insbesondere diese letzte Stufe wird unter Berücksichtigung individueller Nutzungsbedingungen unter dem Begriff der Gesundheitsdatenmobilität (vgl. Unterabschnitt 3.1.2) weiter ausgeführt. Weitere Rahmenbedingungen für die Forschungsdatenverwaltung werden durch die, zum Zeitpunkt der Anfertigung dieser Arbeit entstehenden, nationalen Forschungsdateninfrastrukturen gesetzt und in entsprechenden Projekten wie beispielsweise der MII ausgearbeitet. Daher werden nachfolgend lediglich Teilaspekte untersucht, welche bei der nutzerzentrierten Digitalisierung im definierten Anwendungskontext zu beachten sind.

### 5.3.1 Stufenweise Realisierung der Datenverwaltung

Eine Orientierungshilfe für die Phasen der Datenverwaltung sowie daran zu koppelnde Anforderungen und Konzepte zu deren Umsetzung stellt der sogenannte Datenlebenszyklus dar [400]. Dieser dient in der Forschung dazu den Umfang und die Bedeutung der Datenverwaltung sowie Datenflüsse nachzuvollziehen (S2.3). Er lässt sich nach Surkis und Read in folgende Schritte unterteilen [400]:

- Daten erzeugen und sammeln.
- Daten verarbeiten.
- Daten analysieren.
- Daten sichern.
- Daten freigeben.

Weitere Anforderungen, die an eine Datenverwaltung für forschungsrelevante Daten gestellt werden, entstehen durch die Einhaltung der Findability, Accessibility, Interoperability, and Reusability (FAIR)-Prinzipien (S1, S1.2, T2.4, T4, T6.3, T6.4, T6.5, T6.6) [420]. Die Beschreibung der durch digitale Gesundheitsanwendungen erzeugten und gesammelten Daten muss demnach in einem verständlichen und interoperablen Format erfolgen damit Forscher die Validität (DNFAM2) der Ergebnisse prüfen (S4) und weitere Analysen darauf ausführen können [400]. Hierzu müssen Daten ebenfalls gesichert und an weitere Forscher freigegeben (S1.1, S3.2) werden können [400]. Je nach Art und Umfang der erhobenen Daten existieren unterschiedliche Datenbankkonzepte, wie beispielsweise Zeitreihen-, relationale oder dokumentenorientierte Datenbanken. Auch entsprechende Werkzeuge und Methoden zur Analyse, Sicherung und Freigabe der Daten werden demnach für die jeweiligen Phasen notwendig. Für den beschriebenen Anwendungsfall bedeuten die ersten beiden Phasen des Datenlebenszyklus das Sammeln und Erzeugen von Daten sowie die Verarbeitung der für die in Unterabschnitt 4.2.1 beschriebenen Parametern notwendigen Rohdaten. Im Rahmen der Wissenschaftsphase ist aufgrund des anwendungsspezifischen Einsatzes im ambulanten Umfeld sowie in pflegerischen und medizinischen Einrichtungen die Verfügbarkeit der Daten durch die Abbildung auf ein datei- oder datenbankbasiertes Speicherkonzept auf dem Interaktionsgerät selbst sicherzustellen (T2.4). Diese dezentrale Art der Speicherung (T7) bietet initial, neben der Unabhängigkeit von telematischen Infrastrukturen, den Vorteil, dass sowohl die Datensicherheits- als auch Datenschutzerfordernungen (T1, T2) nicht zusätzlich für die notwendige Infrastruktur realisiert werden müssen. Für die Phase der Analyse von generierten Daten spielt die Prüfung der Datenqualität (T5) eine Rolle, um den Einfluss der Digitalisierung auf die darüber zu gewinnenden medizinischen Erkenntnisse zu verstehen, welche wiederum Aussagen über die Notwendigkeit (M1) der Speicherung dieser zulassen und den Datenminimierungsgrundsatz stützen, indem irrelevante Erhebungen identifiziert werden (vgl. Absatz 2.3.2 und Abschnitt 5.4). Für die Realisierung der letzten beiden Phasen des Datenlebenszyklus müssen die erhobenen und gespeicherten Daten sicher gespeichert, archiviert und für die Weiterverarbeitung durch Dritte freigegeben werden können. Letzteres birgt neben den Vorteilen der Telemedizin das Potential für eine personalisierte Behandlung durch die Kombination bestehender Daten mit zusätzlichen Daten, setzt jedoch den Aufbau entsprechender Infrastrukturen voraus (vgl. Unterabschnitt 3.3.3). Auch im Kontext der Produktentwicklungsphase steht die nutzerorientierte rein lokale Datenverarbeitung dem potentiellen Interesse aus Sicht der Hersteller an den über mehrere Nutzer hinweg gesammelten Daten zur Produktoptimierung und Gewinnung weiterer Erkenntnisse gegenüber. Zur Aggregation und Integration der dezentral gesammelten Daten lassen sich in webbasierte Client-Server-

Architekturen integrierte Datenbanksysteme verwenden, die in Kombination mit Cloud-Infrastrukturen ohne eigene Server-Hardware betrieben werden kann (vgl. Absatz 2.4.2). Dem Stand der Technik entsprechende Übertragungsprotokolle und Client-Technologien ebnet damit ebenso wie virtualisierte und skalierbare Serverumgebungen den Weg zu einer vernetzten Lösung (T10). Um den Sicherheitsaspekten ebenfalls Rechnung zu tragen, ist die Verbindung zwischen Daten-, Nutzer- und Zugriffsverwaltung notwendig. Kryptographische Verfahren sowie organisatorische Mechanismen zum Schutz dieser Daten müssen eingesetzt werden (T1, T2). Komplementär dazu wird für dieses Anwendungsumfeld in einer separaten Publikation durch den Autor gezeigt, welche Entscheidungspunkte und Herausforderungen bei der Wahl geeigneter De-Identifizierungsverfahren (S3.3) existieren [470]. Aus Sicht der Nutzer hängt daran auch die Bestimmung der Verwendungsmöglichkeiten ihrer Daten für fremde Dritte. Die Frage danach wer, wann, zu welchem Zweck darauf zugreifen darf, stellt eine Voraussetzung des gesamten Datenlebenszyklus dar.

### 5.3.2 Selbstbestimmtheit durch Einwilligung

Eine nutzerzentrierte Digitalisierung setzt die Perspektive des Nutzers auf den Umgang mit seinen personenbezogenen Daten voraus. Insbesondere bei der Überlassung von gesundheitsrelevanten Daten an die Forschung (S1.1, S3) besteht die Sorge, dass die Nutzer die Verwendung ihrer eigenen Daten nicht ausreichend kontrollieren und sich darüber informieren können [219]. Da die Offenlegung persönlicher Daten den Betroffenen schaden kann, resultiert dies in Vorsicht, welche verhindert, dass Daten der Forschung zugänglich gemacht werden [214]. Gesundheitsbezogene Daten werden deshalb als besondere Art personenbezogener Daten rechtlich geschützt (vgl. Unterabschnitt 2.3.2). Die Abwägung zwischen offenem und geschlossenem Umgang mit personenbezogenen Daten muss daher durch den Nutzer selbst über ein souveränes Datenaustauschmodell entschieden werden (S2). Hierfür sind Einwilligungsmechanismen erforderlich (DFAS1), welche den Willen des Nutzers für den Umgang mit seinen Daten formalisiert erfassen. In der Literatur wird prinzipiell zwischen verschiedenen Varianten dieser Einwilligungen, insbesondere bezüglich dem Umgang mit biomedizinischen Daten unterschieden. Als pauschale Einwilligung oder „blank consent“ wird die Zustimmung zu jeder Form zukünftiger Forschung verstanden [417]. Hierbei gibt der Nutzer die Kontrolle über die Verwendung seiner Daten komplett ab. Eine offene Einwilligung oder „open consent“ ist ein auf Transparenz beruhender Ansatz, welcher beispielsweise die Überführung der Daten in frei zugängliche Datenbanken erlaubt [267]. Datenschutzgarantien werden hier den Nutzern nicht gegeben. Über eine dynamische Einwilligung oder „dynamic consent“ bestehen Wahlmöglichkeiten für die Zustimmung zur Datennutzung sowie die kontinuierliche Kommunikation zwischen Datenspendern und Datennutzer über die damit erzielten Ergebnisse. Deren Umsetzung erfolgt webbasiert damit Datenspendern diese Informationen zur Verfügung stehen und sich ändernde Präferenzen im Laufe der Zeit berücksichtigt werden können [79]. Als flexible Umsetzung dieses Ansatzes, wird die sogenannte „meta consent“ beziehungsweise Meta-Einwilligung gesehen [328]. Hierbei legt der Datenbesitzer fest, wann und für welchen Zweck er nach einer spezifischen Einwilligung gefragt werden will. Über eine IT-gestützte Plattform wird eine Einwilligungsanfrage basierend auf der zuvor getroffenen Auswahl des Datenbesitzers erzeugt [328]. Dieses Konstrukt setzt für den Datenbesitzer selbst eine Identität (ID), welche ihn eindeutig identifiziert und eine Möglichkeit Kontakt mit ihm aufzunehmen, voraus [328]. Über die Meta-Einwilligung soll ein Gleichgewicht zwischen der Notwendigkeit von Datenspenden an die Forschung und dem individuellen Schutz des Datenbesitzers sowie deren Bereitschaft sich mit dem zeitintensiven Einwilligungsprozess auseinanderzusetzen, hergestellt werden [327, 328]. Aus Sicht der medizinischen Forschung in Deutschland setzt die MII auf das Konzept der breiten Einwilligung oder

auch „broad consent“, weil neue medizinische Fragestellungen auch nach dem Zeitpunkt der Einwilligung entstehen können. Hierbei wird argumentiert, dass zum Zeitpunkt der Erhebung von medizinischen Daten noch nicht alle Fragestellungen, zu deren Beantwortung diese Daten zukünftig beitragen könnten, bereits feststehen können. Aus juristischer Sicht ist hierbei der Erwägungsgrund 33 der DSGVO und § 363 SGB V, Abs. 8 erfüllt (vgl. Abschnitt 2.3). Die Einwilligung kann demnach entsprechend breit, für bestimmte Forschungsbereiche oder -projekte, in einem zweckbezogenen Maße erteilt werden. Diese Variante implementiert dennoch Restriktionen bezüglich der Verwendung von Daten. Am Beispiel der MII sind diese abgestuft und mit identischen Optionen für den Patienten vereinheitlicht damit eine Nachnutzung möglich wird. Während für die wissenschaftliche Forschung Erwägungsgründe vorgesehen sind, gelten für die Datenerhebung im Zuge der Nutzung und Entwicklung digitaler Gesundheitsanwendungen die Grundsätze der DSGVO. Diese Einwilligungsmodelle gehen somit von den breiten Einwilligungen in spezifische Einwilligungen über, die gegebenenfalls dynamisch nach sich ändernden Zweckbestimmungen erneut eingeholt werden müssen. Durch die Möglichkeit elektronische Einwilligungen zu erzielen, entstehen in den letzten Jahren Ansätze zur Formalisierung von digitalen Datenzugriffs- und Nutzungsrechten, welche die Absichten von Datenbesitzer und -verarbeiter in einem maschinenlesbaren Format abbilden [43, 64, 197]. Aus konzeptioneller Sicht werden diese Nutzungsbedingungen verwendet um festzulegen welche Verarbeitungsvorgänge autorisiert sind. Jedoch ist damit derzeit nicht die gleiche rechtliche Qualität wie bei sogenannten Volltextverträgen möglich, sondern lediglich die Bereitschaft zur Zusammenarbeit in bestimmten Formen ohne rechtlich durchsetzbare Konsequenzen [43]. Die formale Beschreibung der Inhalte durch entsprechende technische Formate hat zwar Vorteile bezüglich der Suche und Vermittlung passender Daten, bisherige Zugriffs- und Nutzungskontrollsysteme können jedoch nicht mit solchen unscharfen oder vagen Konstrukten arbeiten [43]. Die aus einem Vertrag hervorgehenden Anforderungen und Verpflichtungen müssen dadurch so wenig Interpretationsspielraum wie möglich zulassen, ob jedoch ein so formalisierter Vertrag einer rechtlichen Prüfung standhalten würde oder nicht ist unbekannt [43]. Bader et al. sprechen diesbezüglich eine Reihe von Hindernissen an, die sich insbesondere an der Gesetzgebung in der EU orientiert und beispielsweise das Verbot einer automatisierten Entscheidungsfindung sowie die Notwendigkeit von Auslegungen einer Vertragsklausel und Vertragstexten beinhaltet [43]. Zusätzliche, diese Nutzungsbedingungen betreffende, rechtliche Anforderungen entstehen im Kontext einer Demenz durch den Verlust der Geschäfts- und Einwilligungsfähigkeit, welche je nach Fortschritt der Erkrankung im Einzelfall festgestellt werden muss oder gemäß § 1901a BGB mittels Vorsorgevollmacht durch einen Betreuer stellvertretend wahrgenommen wird. Im Rahmen dieser Arbeit können die damit einhergehenden juristischen Herausforderungen nicht aufgelöst werden, jedoch soll auf ihre Auswirkungen aufmerksam gemacht und anhand einer beispielhaft formalisierten Abbildung solcher Bedingungen dargestellt werden, wie eine maschinenlesbare Nutzungsbedingung für digitale Gesundheitsanwendungen zukünftig aussehen kann. Weil darüber auch festgelegt wird, welche Rechte und Pflichten im Umgang mit den diesbezüglichen Daten über Organisationsgrenzen hinweg bestehen, ist die Entwicklung eines Standards auch die Grundvoraussetzung für eine organisatorische Interoperabilität (S1.2). Die Umsetzungen der zuvor angedeuteten Datensouveränitätstechnologien, insbesondere der formalisierten Nutzungsbedingungen, bedienen sich an Konzepten, die sich bestehende Ontologien und Vokabulare aus den Bereichen des Semantic Webs und Linked-Data-Prinzipien zu nutze machen (vgl. Absatz 2.4.2). Esteves et al. vergleichen hierzu unterschiedliche Policy Languages zur Umsetzung der juristischen Anforderungen der DSGVO [49]. Als Ausgangspunkt für den Vergleich dienen die Möglichkeiten zur Abbildung der einzelnen Rechte und Pflichten, die in der DSGVO verankert sind und durch Informationsflüsse zwischen Akteuren geltend gemacht werden [49]. Hierzu wird bezüglich der Interaktionen zwischen den fünf Gruppen, den Datensubjekten, den verantwortlichen Stellen für die Datenverarbeitung, den Daten verarbeitenden Stellen, den Empfängern und den Aufsichtsbehörden

bezüglich Datenschutz, unterschieden [49]. Als ein beispielhaftes Grundgerüst für Bedingungen zur Verwaltung und der transparenten Nutzung von Gesundheitsdaten wird das Minimum Core Model von Scalable Policy-aware Linked Data Architecture For Privacy, Transparency and Compliance (SPECIAL)<sup>5</sup> vorgestellt, welche durch die SPECIAL Usage Policy Language (SPL) und das SPECIAL Policy Log Vocabulary (SPLog) in OWL 2 kodiert ist und auf bestehenden Ontologien wie Open Digital Rights Language (ODRL) und Platform for Privacy Preferences Project (P3P) aufbauen [49, 102, 198, 229, 230]. Wie in Abbildung 5.11 dargestellt besteht eine Nutzungsbedingung demnach grundsätzlich aus fünf Di-

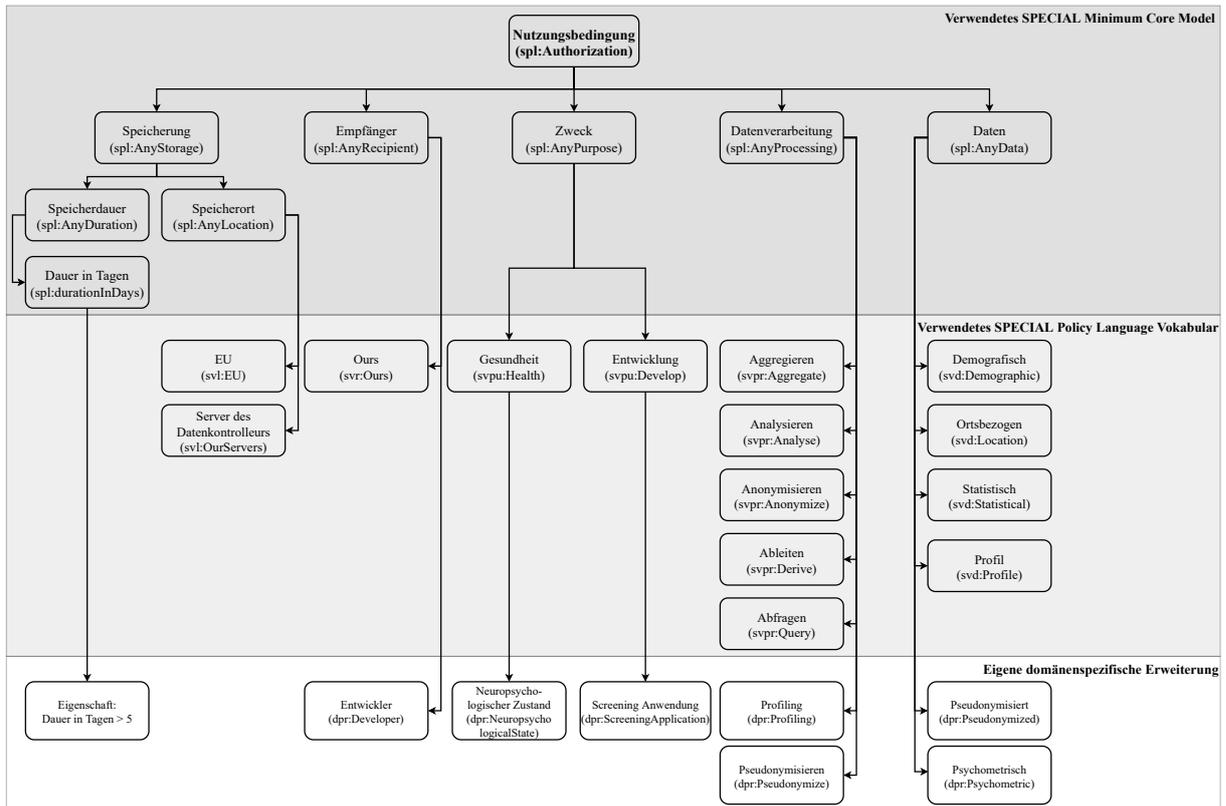


Abbildung 5.11: Erweitertes SPECIAL-Modell nach Kirrane et al. [229].

mensionen, welche sich an den Vorgaben der DSGVO ausrichten und domänenspezifisch spezialisieren lassen [64, 229]. Zum einen sind die Kategorien von personenbezogenen Daten (*Data*) zu beschreiben, welche vom Datenbesitzer gesammelt werden. Andererseits sind die Verarbeitungsschritte (*Processing*), die auf den Daten ausgeführt werden, durch eine passende Ontologie zu definieren. Als weitere Komponente wird die Angabe des Verarbeitungszwecks (*Purpose*) modelliert, welcher je nach Kontext der Datenverarbeitung kategorisch zu unterscheiden ist. Wo die Daten gespeichert werden und für wie lange (*Storage*) ist ebenso anzugeben wie an wen (*Recipients*) die Ergebnisse der Datenverarbeitung gerichtet sind beziehungsweise mit wem die Daten geteilt werden. Um einen autorisierten Datenverarbeitungsvorgang zu spezifizieren müssen aus diesen fünf Kernelementen bestehende Tupel instanziiert werden [49]. Mittels SPLog lassen sich die damit einhergehenden Verarbeitungsvorgänge beschreiben, die durch den Dateneigentümer erteilt wurden. SPLog basiert auf dem Vokabular von PROVO, welches für den Austausch von Herkunftsdaten im Internet entwickelt wurde und ist mit SPL kompatibel [49, 254]. Damit lassen sich Protokolle erstellen, welche die zeitliche Entwicklung der Einwilligung durch Richtlinienbeiträge und der Datenverarbeitungsschritte durch Datenereignisse nachvollziehbar machen [49, 231]. Eine weiterführende Spezialisierung dieser Vokabulare hinsichtlich der DSGVO findet seit dem Jahr 2018

<sup>5</sup> <https://specialprivacy.ercim.eu/>, zuletzt zugegriffen am 21.02.2022.

durch das W3C Data Privacy Vocabularies and Controls Community Group (DPVCG) als Data Privacy Vocabulary (DPV) statt [63, 318]. Am Beispiel der angestrebten Zieldefinition erfolgt auf Grundlage von psychometrischen Daten dem Anwendungsfall entsprechend, eine Erweiterung der mittels der SPECIAL Policy Language ausgedrückten Nutzungsbedingungen durch Zusatzvokabeln in Form von neuen Klassen. Die Klasse *svd:Health*, welche die Kategorie der personenbezogenen Daten (*Data*) der Daten-Ontologie erweitert, lässt sich somit durch zusätzliche Unterklassen wie beispielsweise *dpr:Psychometric* spezialisieren (vgl. Abbildung 5.11). Weitere für die Entwicklung zu erhebende Datenkategorien werden ebenfalls aus dem bereits bestehenden Vokabular eingefügt oder entsprechend ergänzt. Eine Nutzungsbedingung der Anwendung *DemPredict* könnte also umfassen, dass psychometrische, ortsbezogene, statistische sowie demografische, Profil-Daten von Entwicklern auf eigenen Servern oder in der EU gesammelt werden dürfen (vgl. Listing A.1). Die Verarbeitung darf in der Gestalt von Abfragen, Analysen, Aggregation, Ableitungen, Pseudonymisierung und Profilbildung erfolgen beispielsweise um Modelle oder ein Nutzerprofil zu erstellen, welche zum Zweck einer Screening-Anwendung genutzt werden. Hiermit lässt sich ebenfalls der Zweck, einen neuropsychologischen Zustand abzuleiten, verbinden. Die notwendigen Daten werden für mindestens fünf Tage gespeichert, zum Beispiel ausschließlich für ein Modell-Training, und nicht an Drittparteien weitergegeben.

Esteves et al. kommen im Jahr 2021 zu dem Schluss, dass ODRL, DPV und GDPRtEXT den höchsten Reifegrad der Ressourcen erreicht haben sowie weitere Voraussetzungen erfüllen, um die Rechte und Pflichten der DSGVO zu modellieren [49]. In Kombination können diese drei Sprachen 39 der 57 identifizierten Informationselemente modellieren und damit eine Mehrheit an Informationsflüssen formal abbilden [49]. Aus Sicht von Herstellern wird sich demnach zukünftig daran orientieren zu sein, welche gesetzlichen Vorgaben insbesondere für die Integration der Daten aus einer DiGA in die ePA unter Beachtung der Selbstbestimmtheit für eine interoperable elektronische Einwilligungserklärung zu erfüllen sind. Ob es hierzu ein ähnliches Standardisierungskonzept wie bei den MIO geben wird, ist zum Zeitpunkt der Anfertigung dieser Arbeit noch offen. Die Definition solcher formaler Bedingungen setzt neben der Überwindung von den zuvor skizzierten technischen und regulatorischen Hürden, eine an deren Notwendigkeit geknüpfte Akzeptanz dieser von Seiten potentieller Nutzer voraus. Auch die Bereitschaft deren Daten zum Zwecke der Forschung und Produktentwicklung öffentlichen und privaten Einrichtungen zur Verfügung zu stellen muss im Anwendungskontext untersucht werden. Laut einer repräsentativen Umfrage des Branchenverbands Bitkom in Deutschland geben fast 90% aller Befragten an, dass sie ihre Gesundheitsdaten privaten Forschungseinrichtungen zur Verfügung stellen würden [61]. Knapp 83% geben demnach an, dass sie ihre Daten zur Verbesserung ihrer eigenen Behandlung auch privaten Unternehmen überlassen würden [61]. Unter Angabe des genauen Leidens würden sich 76% dazu veranlasst fühlen ihre Daten für andere zu spenden, 48% sagen, dass sie bei seltenen Erkrankungen ihre Daten zu privaten Forschungszwecken spenden würden, um damit weiteren Patienten mit der gleichen Diagnose helfen zu können [61]. 47% der Befragten würden sogar unabhängig von einem persönlichen Mehrwert einer Datenspende an private Unternehmen zustimmen [61]. Inwiefern diese Bereitschaft zu Datenspenden und die Wahrnehmung von Nutzungsbedingungen auch für die Entwicklung digitaler Gesundheitsanwendungen im Anwendungskontext dementieller Syndrome eine Rolle spielt wird nachfolgend anhand einer Umfrage untersucht. Darin sollen ein diesbezügliches Stimmungsbild umrissen und weitere Anforderungen identifiziert werden. Diese umfasst Fragen durch welche, neben demografischen Informationen, das Interesse an der Selbstverwaltung gesundheitsbezogener Daten und die Bedingungen einer Datenspende sowie die Bereitschaft dazu eruiert werden (vgl. Abschnitt A.3). Im Zeitraum von Februar bis März 2020 wurden 238 Probanden im Alter zwischen 18 und 79 Jahren befragt (vgl. Abschnitt A.4). Die Umfrage wurde primär von Menschen aus dem universitären Umfeld in Baden-Württemberg beantwortet, was den Altersdurchschnitt von 31,6 bei einem Median von 25 erklärt, wovon wiederum 85% zumindest das Abitur als

höchsten Bildungsabschluss angeben. Die Ergebnisse sind dieser demografischen Struktur folgend zu interpretieren und zeigen eine entsprechende Zukunftsperspektive. Nachfolgend wird ausschließlich auf die für diese Arbeit relevanten Aspekte eingegangen, weitere Ergebnisse befinden sich unter Abschnitt A.5. Mehr als 85% der Befragten hätten gerne einen besseren Überblick über ihre Gesundheitsdaten, indem sie diese und deren Speicherort einsehen können (vgl. Abbildung 5.18). Auf die Frage wie hoch die Bereitschaft ist selbst generierte Gesundheitsdaten anonymisiert an Forschungseinrichtungen zu Forschungszwecken zu spenden, geben mehr als die Hälfte der Befragten sehr wahrscheinlich oder ganz sicher an (vgl. Abbildung 5.13). Knapp ein Drittel ist sich unsicher, circa 15% würden wahrscheinlich nicht oder keinesfalls ihre Gesundheitsdaten spenden. Im Vergleich mit der oben zitierten repräsentativen Studie der Bitkom<sup>6</sup> fällt die Bereitschaft somit geringer aus. Bezüglich der Transparenz im Umgang mit Gesundheitsdaten halten es 73% der Befragten für wichtig oder eher wichtig nachvollziehen zu können, wo, von wem und zu welchem Zweck anonymisierte Gesundheitsdaten verarbeitet werden (vgl. Abbildung 5.14). Die Erfüllung dieser Anforderung würde die Bereitschaft zu einer Datenspende erhöhen (vgl. Abbildung 5.13 und Abbildung 5.15). Den Krankenkassen wird mit 42% am häufigsten zugetraut eine Plattform zu betreiben, die Gesundheitsdaten verwaltet (vgl. Abbildung 5.12).

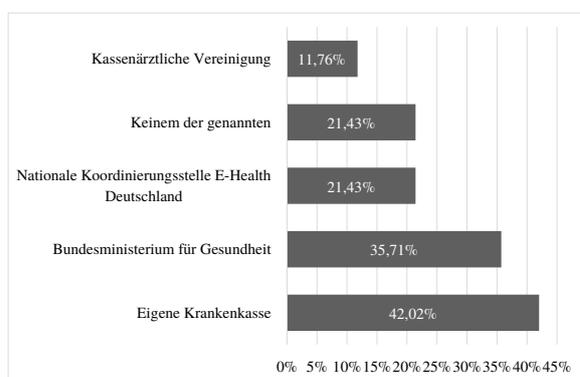


Abbildung 5.12: Antworten auf Frage 3 (vgl. Abschnitt A.3).

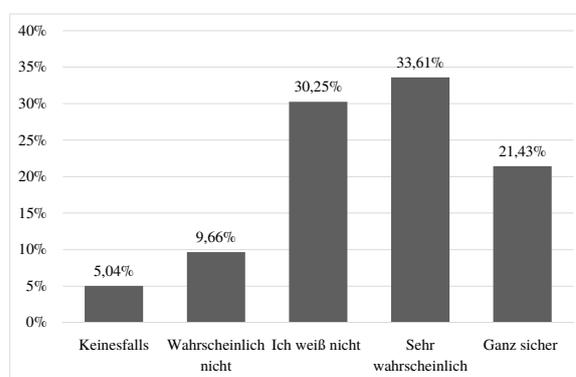


Abbildung 5.13: Antworten auf Frage 6 (vgl. Abschnitt A.3).

Als Antwort auf die Frage, ob selbst generierte Gesundheitsdaten an die Krankenkasse übertragen würden, um den Beitrag zu senken geben knapp 20% ganz sicher und 29% sehr wahrscheinlich an. Mehr als 17% wissen es nicht, 15% würden wahrscheinlich nicht und 19% keinesfalls ihre Daten für diese Gegenleistung hergeben (vgl. Abbildung 5.17). Als Anreizmechanismus (DNFAE3) ist diese Maßnahme somit nur für knapp die Hälfte aller Befragten ansprechend.

<sup>6</sup> Nach Rückfrage des Autors bei den Verantwortlichen der Bitkom wurde bestätigt, dass die Ergebnisse der Studie lediglich in der referenzierten Pressemitteilung veröffentlicht werden.

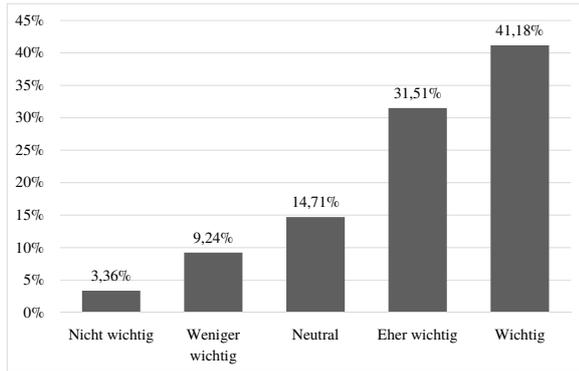


Abbildung 5.14: Antworten auf Frage 9 (vgl. Abschnitt A.3).

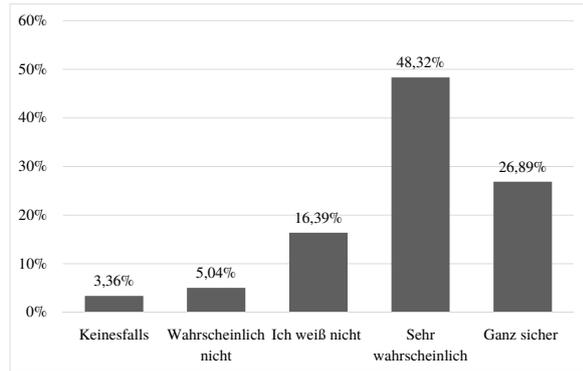


Abbildung 5.15: Antworten auf Frage 15 (vgl. Abschnitt A.3).

In Abbildung 5.16 werden die Antworten, auf die Frage nach der Bereitschaft die erfassten Daten aus einer App zur Diagnose einer Demenz wie *DemPredict* zur Verfügung zu stellen, dargestellt. In diesem Fall erklären sich 83,61% dazu bereit diese Daten zu Forschungszwecken zu spenden, dem gegenüber stehen 7,98% die nicht spenden würden sowie 8,40% die keine Angabe dazu machen. Insgesamt also ähnlich hohe Werte wie in der referenzierten Bitkom-Studie.

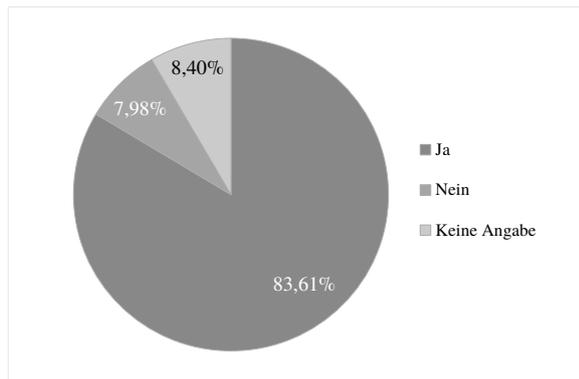


Abbildung 5.16: Antworten auf Frage 10 (vgl. Abschnitt A.3).

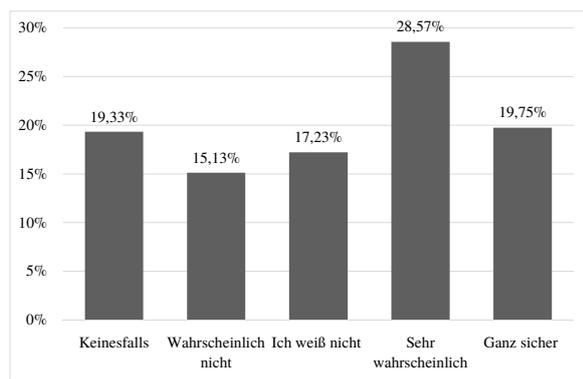


Abbildung 5.17: Antworten auf Frage 7 (vgl. Abschnitt A.3).

Am Beispiel der Dimensionen, welche nach dem angeführten SPECIAL-Minimum-Core-Model abgedeckt sind, wird in Abbildung 5.19 die Einschätzung ersichtlich, für wie wichtig diese gehalten werden. Alle Dimensionen werden überwiegend als wichtig angesehen, die größte Rolle spielen dabei die Art der Daten, der Zweck der Datenerfassung sowie der Datenempfänger. Am wenigsten stark gewichtet wird was mit den Daten geschieht (Datenverarbeitung) sowie die Dauer und der Ort der Aufbewahrung.

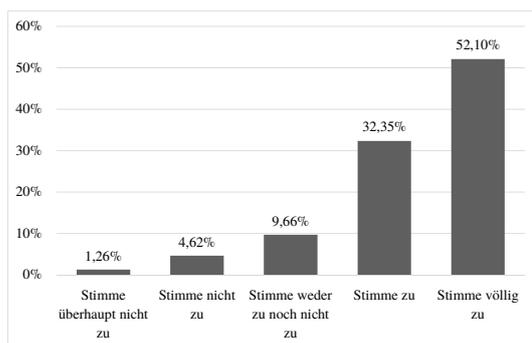


Abbildung 5.18: Antworten auf Frage 1 (vgl. Abschnitt A.3).

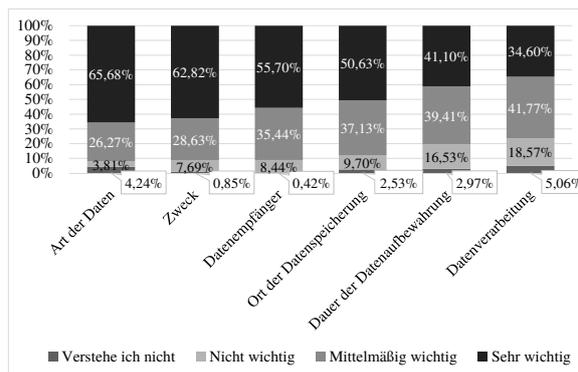


Abbildung 5.19: Antworten auf Frage 13 (vgl. Abschnitt A.3).

Faktisch scheint die Bereitschaft zur Gesundheitsdatenspende, um damit einen Beitrag für die weitere Erforschung und Entwicklung digitaler Werkzeuge für den Umgang mit chronischen Erkrankungen wie Demenzen zu leisten gegeben, solange Transparenz und Selbstbestimmtheit über deren Verwendung vorliegt. Welche Standards und Formate letztlich die angesprochenen Herausforderungen in Zukunft überwinden, muss im Zuge der Entwicklung digitaler Gesundheitsanwendungen verfolgt werden.

### 5.3.3 Notwendige Interoperabilitätseigenschaften

Für eine kollaborative Entscheidungsfindung zwischen Ärzten und Patienten (M3) dienen, auf der Ebene individueller Patienten, zu einem gesamtheitlichen Gesundheitsprofil integrierte partielle Gesundheitsdaten. Auch die gesetzlich geplante Möglichkeit einer Datenspende setzt interoperable Informationssysteme im Gesundheitswesen voraus. Zu deren Realisierung sind Datenstandards notwendig, worüber ein möglichst vollständiges Bild des Gesundheitszustands und der Versorgungssituation eines Patienten dargestellt werden kann. Die Entwicklung der ePA legt dafür einen Grundstein. Für das Gesundheitssystem bergen über diese Gesundheitsprofile von mehreren Datensubjekten hinweg harmonisierte, aggregierte und angereicherte Datenbestände (T6.4, T9) das Potential, eine datengetriebene personalisierte Versorgung zu realisieren (M2), welche mittels prädiktiver Analysemöglichkeiten (T8) eine frühere Erkennung und zielgerichtete Behandlung von Erkrankungen ermöglichen soll. Dies setzt auch bei der Entwicklung digitaler Gesundheitsanwendungen den nachhaltigen Umgang mit Gesundheitsdaten (T6) voraus. Wie bereits in Abbildung 4.1 dargestellt, kann ein daraus resultierender zyklischer Datenfluss zwischen Organisationen zur Entwicklung neuer und Optimierung bestehender Anwendungen genutzt werden. Die Sicherstellung von allen Ebenen der Interoperabilität zwischen Systemen im Gesundheitswesen (vgl. Abbildung 2.19) stellt daher eine Grundvoraussetzung dar, weshalb Lock-In-Effekte bei Herstellern (T6.2) verhindert werden müssen. In den Vereinigten Staaten von Amerika (USA) existiert hierfür seit dem Jahr 2020 die *Cures Act Final Rule*, welche das sogenannte „Information Blocking“ durch Hersteller sogar unter Strafe stellt und jedem Datensubjekt garantiert seine gesundheitsbezogenen Daten abfragen und teilen zu können [108]. In Deutschland erarbeitet die gematik zum Zeitpunkt der Anfertigung dieser Arbeit einen verbindlichen Schnittstellenstandard auf Basis von FHIR, der nach § 373 SGB V durch KIS-Anbieter verpflichtend umgesetzt werden muss, da ansonsten deren Betriebserlaubnis erlischt. Somit wird eine grundlegende Interoperabilität (T6.6) Schritt für Schritt gesetzlich verordnet (vgl. Unterunterabschnitt 2.4.2). Daneben spielen zur Wahrung der Nachhaltigkeit solcher Systeme weitere Interoperabilitätseigenschaften (DNFAT4) eine Rolle. Um die Verknüpfung der Daten mit einer dazugehörigen Repräsentation des Datensubjekts zu erreichen, wird eine strukturelle Interoperabilität (T6.5) durch die

Verwendung eines akzeptierten und einheitlichen Formats notwendig. Deren Aussagekraft muss durch semantische Interoperabilität (T6.3) unterstützt werden. Diese erfordert die Verwendung von Metadaten, welche in standardisierten Datenformaten hinterlegt werden können. Mittels domänengetriebener Datenmodellierung (T6.1) wird durch die Konsensbildung interdisziplinärer Fachgruppen sichergestellt, dass Sachverhalte in Datenmodellen inhaltlich richtig codiert und abgebildet werden. In Deutschland werden diese in Form von MIO gemäß § 355 SGB V durch die KBV syntaktisch und semantisch festgelegt. Dies erfolgt durch die Definition von Datenelementen und deren Verknüpfung mit entsprechend standardisierten Terminologien und dazugehörigen Ontologien (vgl. Unterabschnitt 2.2.2). Abschließend steht die Ebene der organisatorischen Interoperabilität (S1.2), welche einerseits durch den in Unterabschnitt 5.3.2 erörterten Schutz der Rechte des Datensubjekts mittels Festlegung von Datennutzungsbedingungen gestützt wird. Andererseits werden, um die Integration der Daten in ein Datenintegrationswerkzeug, wie beispielsweise eine ePA vorzunehmen, systemübergreifende Prozesse und Identitäten notwendig. Diese können an national oder international etablierte digitale Identitäten geknüpft werden. In Deutschland ist dies beispielsweise durch den elektronischen Personalausweis oder für die Nutzung der ePA durch die eGK realisierbar (vgl. Unterabschnitt 2.4.1 und Unterabschnitt 2.3.1). Für die Anwendungsdomäne AD sind zur Diagnose gemäß den Leitlinien die Ergebnisse verschiedener Untersuchungen relevant (vgl. Unterabschnitt 2.2.4). Aus Sicht von Ärzten liegt daher die zielgruppengerechte sowie benutzerfreundliche Gestaltung und Verwaltung (DNFAT1, T3) der Ergebnisse dieser Untersuchungen durch eine Integration der dabei anfallenden Daten in deren Praxisverwaltungssystem (PVS) nahe. Über Standardschnittstellen muss die Übertragung der Daten in diese Systeme erfolgen. Der Einfluss, der zuvor erörterten rechtlichen Anforderungen führt zu einer marktorientierten Auswahl des FHIR-Standards. Wie die Transformation der psychometrischen Daten am Beispiel von *DemPredict* aus einem proprietären Datenformat in FHIR und den zugrundeliegenden Konzepten erfolgen kann, wird nachfolgend beschrieben. Der modulare Ansatz dieses Standards durch die Kombination und Verknüpfung vielfältiger vordefinierter Ressourcen ermöglicht eine flexible Abbildung von wachsenden Gesundheitsdatenprofilen und deren Anreicherung um weitere Daten unterschiedlicher Herkunft und Ausprägung (vgl. Unterabschnitt 3.1.1). Als Ausgangspunkt für darin zu integrierende Daten dient ein Stammdatensatz der Patienten. Diese lassen sich als Ressource der Klasse *FHIR Patient* abbilden und müssen sich mit den Angaben der Nutzerverwaltung einer jeweiligen digitalen Gesundheitsanwendung verknüpfen lassen (vgl. Abbildung 5.20). Die Zuord-

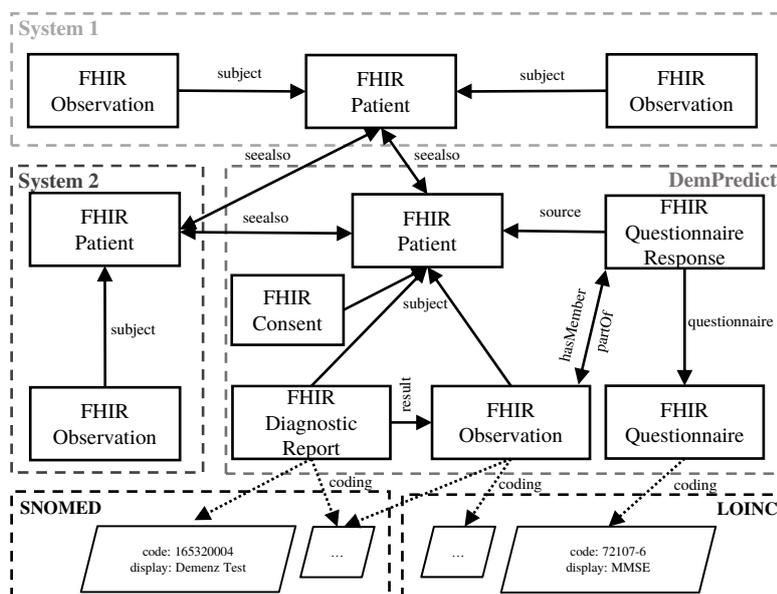


Abbildung 5.20: Schematische Darstellung der Datenintegration am Beispiel von *DemPredict* auf FHIR.

nung des *identifier* Attributs zeigt auf das entsprechende System, aus welchem die Daten bezogen werden und identifiziert den entsprechenden Nutzer darin. Die Daten zur Durchführung des Screenings via *Dem-Predict* lassen sich mit dem Ergebnis der Untersuchung in einem *FHIR DiagnosticReport* abbilden (vgl. Listing A.2). Zusammengefasst werden darin wiederum die Testdaten einer Testsitzung, die sich über eine *FHIR Observation* Ressource abbilden lassen, referenziert. Den Sitzungen kann darüber ein Status, der beschreibt inwiefern diese vollständig beendet wurden, zugewiesen werden. Die dabei erreichte Punktzahl und Dauer sowie die inkludierten einzelnen Testergebnisse (*hasMember*) können referenziert werden. Die einzelnen Tests lassen sich ebenfalls als *FHIR Observation* oder im Falle eines Fragebogenbasierten Tests, wie dem MMSE (Test A), als *FHIR QuestionnaireResponse* abbilden. Als Komponenten des Tests werden die gesammelten Parameter als Bestandteil der *Observation* hinzugefügt. Für den Test I stellen diese beispielsweise die Bearbeitungszeit, Fehleranzahl, die Benutzereingabe, sowie die Dauer, die der Benutzer den Test pausiert hat, dar. Über bereits existierende Codierungen von SNOMED CT und LOINC lassen sich beispielsweise sowohl die Versionsangabe der verwendeten digitalen Gesundheitsanwendung als auch die darin umgesetzten Testkonzepte, damit erfassten Leistungsparameter und Einheiten semantisch interoperabel annotieren. Darüber hinaus bietet FHIR zum Zeitpunkt der Anfertigung dieser Arbeit in Version 4.6 erste Ansätze über *FHIR Consent* auch die Verwaltung der Einwilligungserklärungen bezüglich der Datenverarbeitung innerhalb einer Einrichtung als Ressourcen abzubilden (vgl. Unterabschnitt 5.3.2). Insgesamt lässt sich das technische Konzept zur Erreichung der Interoperabilität, wie in Abbildung 5.21 zusammenfassen.

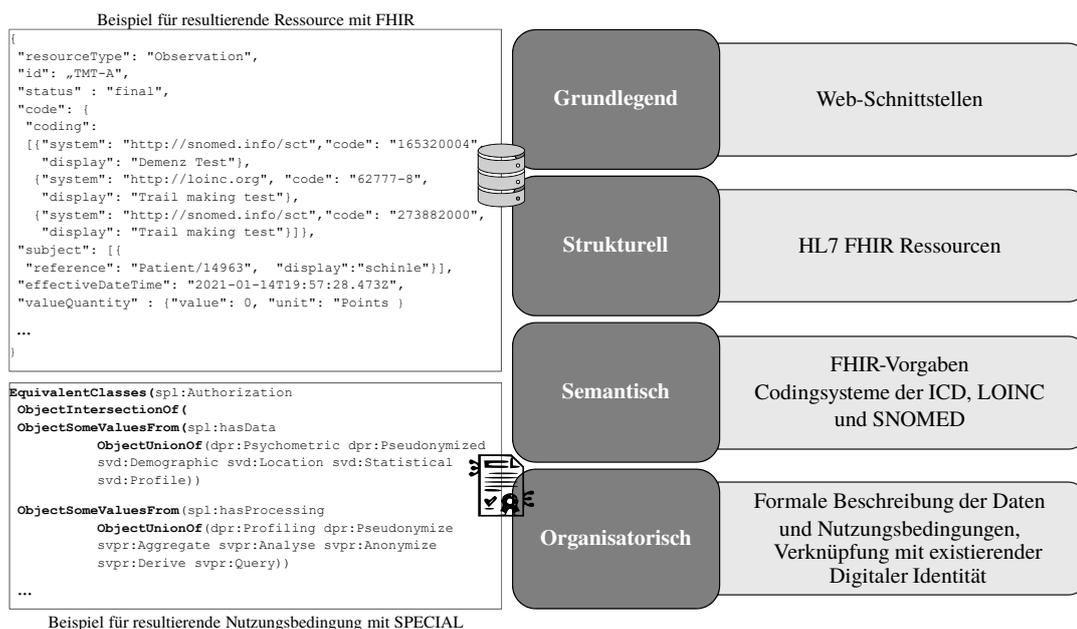


Abbildung 5.21: Technisches Konzept zur Erreichung der Interoperabilität.

## 5.4 Validierung und Einflüsse der Digitalisierung

Auch wenn bei der Entwicklung einer digitalen Gesundheitsanwendung nicht auf komplett neue digitale Verfahren gesetzt wird, ist zur Überprüfung der Validität (DNFAM2) dieser entwickelten digitalen Lösung, neben der Berücksichtigung medizinischer Sachverhalte, auch die kontinuierliche Einbindung der unterschiedlichen Interessenvertreter erforderlich [78, S. 37 ff.]. Hierfür werden mehrere Evaluationsrunden in Form von Studien durchgeführt, um die Lösung deren Bedürfnissen (DNFAE2) entsprechend zu

gestalten und die Akzeptanz zu erhöhen. Einer agilen Vorgehensweise folgend sowie zur Untersuchung von Einflüssen der Digitalisierung werden zum Ende einer jeden Iteration Beobachtungsstudien mit unterschiedlichen Schwerpunkten, wie Benutzerschnittstellendesign, Nutzerakzeptanz, Funktionsumfang oder Konstruktvalidität notwendig (vgl. Unterabschnitt 4.1.2). Jeder dieser beaufsichtigten Evaluationschritte verlangt direkte Rückmeldung von den Studienteilnehmern, welche beispielsweise durch die TAM bei der Nutzung einer Anwendung geschehen kann (vgl. Unterabschnitt 2.1.2). Hierdurch lassen sich durch Missverständnisse auftretende Fehlerquellen ausschließen und Zusatzinformationen bezüglich UI- und UX-Design auch im ambulanten Umfeld ableiten. Weitere Beobachtungsstudien werden notwendig, damit der Erfüllungsgrad der angestrebten medizinischen Zieldefinition und damit des Zwecks der Gesundheitsanwendung (DNFAM2.3), wie beispielsweise die angestrebte Genauigkeit (DNFAM2.1), geprüft werden können (vgl. Unterabschnitt 5.4.4). Dieser Schritt erfordert ein Vorgehensmodell, welches die strukturierte Analyse von Datenbeständen und dadurch Wissensentdeckung anleitet, um die Einflüsse der Digitalisierung und die dadurch verfügbar werdenden digitalen Daten zu verstehen (vgl. Unterabschnitt 2.1.3). Im Zuge dessen finden statistische Verfahren und Metriken ihre Anwendung, welche in Unterabschnitt 2.1.4 zusammengefasst sind und in Unterabschnitt 5.4.1 kontextualisiert werden. Um den Einfluss zu erfassen, der durch ein verändertes Datenerfassungsmedium entsteht, wird insbesondere auf die Vergleichbarkeit der Daten bezüglich analoger und digitaler Versionen getestet. In Bezug auf die zu digitalisierende Domäne wird daher partiell geprüft, ob die Wirksamkeit der bereits klinisch validierten Testverfahren in der digitalen Welt erhalten bleibt (vgl. Unterabschnitt 5.4.2). Des Weiteren ist zu untersuchen, inwiefern der Erfahrungsschatz von Anwendern im Umgang mit digitalen Medien einen Einflussfaktor auf den erfolgreichen Einsatz des digitalen Werkzeugs darstellt. Auch hierzu sind Kontrollvariablen einzuführen, welche die Technik-Erfahrung erfassbar und deren Einfluss quantifizierbar machen (vgl. Unterabschnitt 5.4.3).

### 5.4.1 Datengrundlage und Merkmale

In Bezug auf die durchgeführten Beobachtungsstudien wird nachfolgend erläutert, auf welcher Datengrundlage weitere Entwicklungsschritte geplant werden können und welchen Einfluss die Datenvorverarbeitung darauf hat. Dieser Validierungsschritt lässt sich am Beispiel von *DemPredict* zum Abschluss von Iteration II (vgl. Abschnitt A.9) zeigen, indem eine Feldstudie im Zeitraum von Juli 2017 bis Februar 2018 mit  $n = 55$  Studienteilnehmern aus der Zielgruppe stattfand (vgl. Tabelle A.3). Deren klinische Diagnosen sind von Psychologen und Geriatern der Klinik für Psychiatrie, Psychotherapie und Psychosomatik in Reutlingen (PP.r) gestellt. Es wird nach Mistridis et al. zwischen gesund (G), mit leichten kognitiven Beeinträchtigungen (MCI) und dement (D) differenziert [290]. Diese Referenzdiagnosen werden anhand der Durchführung einer analogen Version der CERAD-Plus-Testbatterie auf kognitionspsychologischer Ebene gestützt (vgl. Unterabschnitt 2.2.5). Pro Studienteilnehmer sind bis zu zwei Testungen in einem Abstand von circa 6 Monaten (T1 und T2) vorhanden. Um die Vergleichbarkeit der Ergebnisse zu gewährleisten, erfolgt die Datenerhebung für *DemPredict* mittels einheitlichem Modell eines Tablet-Computers bei einer Bildschirmdiagonale von zehn Zoll sowie eines damit kompatiblen Eingabestifts. Neben der Aufnahme von demografischen und psychometrischen Daten über die digitale Gesundheitsanwendung sowie der CERAD-Testbatterie als Referenz werden die Kontrollvariablen anhand von Fragebögen erfasst. Die Verarbeitung der digital erhobenen Datenmengen folgt den Prinzipien des DM und orientiert sich daher an den Phasen des Industriestandards CRISP-DM (Phase). Die Umsetzung der Datenanalyse erfolgt in der Programmiersprache  $R^7$ . Um eine Bewertung der Qualität und Plausibilität der mit dem

---

<sup>7</sup> <https://www.r-project.org>, zuletzt abgerufen am 10.03.2021.

digitalen Instrument gesammelten Daten (T2.2, T5) durchzuführen, diese zu verstehen (Data Understanding), zu filtern (Data Preparation) und die Trennschärfe der verschiedenen Variablen zu vergleichen, wird nach den geltenden Prinzipien der Statistik vorgegangen. Der Umfang des zur Verfügung stehenden Datensatzes fällt hinsichtlich der Anwendung von statistischen Verfahren relativ gering aus, daher steht nachfolgend das „Wie“ und nicht das „Was“ im Vordergrund. Mittels explorativer Analyse der empirischen Daten werden Hypothesen aufgestellt und diese unter Anwendung deskriptiver und induktiver Statistik untersucht. Eine perfekte Unterscheidung zwischen Diagnosegruppen wird erreicht, wenn sich die Daten der Gruppen nicht überschneiden. Die Frage, welche induktiven statistischen Verfahren eingesetzt werden können, hängt von der Verteilung der Daten ab (vgl. Unterunterabschnitt 2.1.4). Bei der Durchführung statistischer Tests muss daher zwischen parametrischen und nicht-parametrischen Verfahren in Abhängigkeit der zugrundeliegenden Verteilungsannahme unterschieden werden. Wie zuvor bereits angedeutet ist zu Beginn der Entwicklung, insbesondere im Bereich der psychischen Gesundheit, von einer eingeschränkten Menge von verfügbaren Daten auszugehen. Der zentrale Grenzwertsatz und die Normalverteilung sind daher bei konservativer Schätzung zunächst nicht anzunehmen. Mithilfe des SWTs kann untersucht werden, ob eine Normalverteilung vorliegt [358]. Wenn diese Annahme getroffen werden kann, ist die Anwendung parametrischer Verfahren möglich. Falls jene für eine oder mehrere Variablen verworfen werden muss, wird für alle Daten ein nicht-parametrisches Verfahren verwendet, um eine bessere Vergleichbarkeit der Ergebnisse zu erreichen. Aufgrund der statistischen Power gilt der SWT gegenüber anderer Testverfahren zur Überprüfung einer Normalverteilungsannahme als vorteilhaft (vgl. Absatz 2.1.4). Mit dem nicht-parametrischen KWT kann untersucht werden, ob  $k$  unabhängige Stichproben aus der gleichen Grundgesamtheit stammen. Als parametrische Alternative könnte in diesem Zusammenhang die Varianzanalyse (ANOVA) verwendet werden [358]. Um die Diagnosegruppen auch paarweise zu vergleichen, kann der nicht-parametrische WRT als Post-hoc-Test durchgeführt werden. Eine parametrische Alternative ist der t-Test [358]. Um Modelle zu entwerfen (Modeling), welche beispielsweise die Stärke des linearen Zusammenhangs zweier Variablen ermitteln, stellt der Korrelationskoeffizient nach Pearson ein geeignetes Maß dar (vgl. Unterunterabschnitt 2.1.4). Hierbei ist die Annahme einer Normalverteilung erforderlich, um von einer beobachteten Korrelation auf die Grundgesamtheit schließen zu können [358]. Als nicht-parametrische Alternative existiert hierfür nach Spearman der Rang-Korrelationskoeffizient, welcher die Ränge von Beobachtungen der Variablen vergleicht. Hierdurch lässt dieses Maß, unabhängig von Verteilungsannahmen, Rückschlüsse auf die Grundgesamtheit zu [358]. Durch die Anwendung nicht-parametrischer Verfahren auf normalverteilte Daten entsteht kein falsches Ergebnis, lediglich die statistische Power der nicht-parametrischen Verfahren fällt im Vergleich zu deren parametrischen Pendanten etwas geringer aus [358]. Dadurch kann das Testurteil unter Umständen etwas konservativer ausfallen, als bei der entsprechenden Datenlage nötig wäre (Evaluation). Als Signifikanzniveau wird für alle statistischen Verfahren der Wert von  $\alpha = 5\%$  gewählt.

Aufgrund der zielgruppenspezifischen Altersstruktur wird die Einstellung der Studienteilnehmer zu Technik und deren diesbezügliche Erfahrungswerte erfasst. Die persönliche Einstellung soll Aufschluss über die Gründe mangelnder Technikerfahrung geben und Aussagen über die Akzeptanz einer digitalen Lösung bei der Zielgruppe zulassen. Die Einstellung gegenüber Technik und Erfahrung mit technischen Geräten wird ebenfalls durch Fragen eines validierten Fragebogens bestimmt (vgl. Abschnitt A.7). Zur Quantifizierung der sich darin befindenden Aussagen werden zwei verschiedene Skalen verwendet, welche sich auf einen zweiteiligen Fragebogen beziehen. Der erste Teil erfragt die Einstellung gegenüber Technik und wird auf einer Likert-Skala von „Trifft überhaupt nicht zu“ (0) bis „Trifft voll zu“ (4) bewertet. Diese Fragen sind so konzipiert, dass erkennbar wird, ob Studienteilnehmer Technologien zugeneigt, abgeneigt oder neutral gegenüberstehen. Diese Zuneigung kann abgelesen und mit der Häufigkeit der Techniknutzung korreliert werden. Die Frage, ob die Nutzung von Technik frustriert oder entmutigt, wird bezüglich

der Aussagenlogik der Likert-Skala umgekehrt dargestellt und bepunktet. Die persönliche Einstellung gegenüber Technik (PE) wird somit aus der Beantwortung des ersten Teils des Fragebogens erschlossen (vgl. Gleichung 5.1), also maximal mit 24 Punkten bewertet und über die Anzahl der beantworteten Fragen gemittelt ( $NAP = \text{Anzahl gestellter Fragen} - \text{Anzahl unbeantworteter Fragen} = 6 - X_{PE}$ ).

$$PE = \frac{\sum_{i=1}^6 PE_i}{NAP} \quad (5.1) \quad TE = \frac{\sum_{i=1}^5 HG_i + \sum_{i=1}^7 HZ_i}{NAT} \quad (5.2)$$

Der zweite Teil des Fragebogens erfragt Erfahrung im Umgang mit Technik (TE) über die Häufigkeit der Nutzung technischer Geräte und reicht von „täglich“ (5) bis „nie“ (0) (vgl. Abschnitt A.7). Zusätzliche qualitative Anmerkungen sowie Aussagen über die Häufigkeit von Nutzungen technischer Geräte wie Fernseher oder Mobiltelefone sind ebenfalls erfassbar, werden aber aufgrund des Schwerpunkts der zu überprüfenden technischen Anwendung und des gewählten Interaktionsmediums nicht in die quantitative Bewertung mit eingeschlossen. Die Technikerfahrung wird mit maximal 20 Punkten durch die Häufigkeit der Nutzung technischer Geräte (HG) und zusätzlichen maximal 35 Punkten als Kombination von Häufigkeiten verschiedener Nutzungszwecke (HZ) erfasst. Die Aggregation dieser Bewertungen wird nachfolgend als Kontrollvariable zur Untersuchung des Einflusses der Technikerfahrung eingeführt. Um der verzerrenden Wirkung unbeantworteter Fragen ( $NAT = \text{Anzahl gestellter Fragen} - \text{Anzahl unbeantworteter Fragen} = 12 - X_{TE}$ ) zu begegnen wird diese Summe ebenfalls entsprechend der Anzahl gegebener Antworten relativiert (vgl. Gleichung 5.2). Mittels TE soll untersucht werden, ob Technikerfahrung im Zusammenhang mit den Ergebnissen der digitalen Tests steht und wie sich diese im Vergleich mit deren analogen Version verhalten. Eine Korrelationsanalyse der erreichten Punktzahlen von PE und TE ergibt nach Spearman ein Korrelationskoeffizient von  $\rho = 0,119$  bei einem p-Wert von 0,341. Es existiert somit aufgrund des gewählten Signifikanzniveaus kein signifikanter Zusammenhang zwischen PE und TE [97].

Zusätzlich werden die in Unterabschnitt 4.2.1 identifizierten domänenspezifischen Parameter, die Auskunft über den kognitiven Zustand eines Nutzers als Teil einer digitalen Gesundheitsanwendung liefern, validiert. Somit dienen diese Daten als Grundlage für zielgerichtete Analysen und müssen zunächst entsprechend aufbereitet und gefiltert werden (Data Preparation). Im Rahmen der Datenakquise im Feld können Fehler bei der Durchführung einer Testsitzung nicht ausgeschlossen werden (vgl. Abbildung 5.1). Neben unvollständigen ist ebenso mit unerwarteten Benutzereingaben, Pausen bei der Durchführung, dem Abbruch von einzelnen Tests oder gesamter Sitzungen zu rechnen. Auch die Teilnahmebereitschaft an der sich über mehrere Messzeitpunkte hinweg erstreckende Longitudinalstudie nimmt über die Dauer ab und führt zu fehlenden Messungen (vgl. Tabelle A.3). Ungeeignete Testformate oder missverständliche Instruktionen sind über diese Metadaten im Rahmen der iterativen Entwicklung schrittweise auszuschließen. Die visuelle Aufbereitung der einzelnen Parameter zeigt Ausreißer, welche Hinweise auf eine inkorrekte Durchführung einzelner Tests liefern. Im Falle von *DemPredict* kann sich das in verhältnismäßig kurzen Bearbeitungsdauern, komplett fehlenden Eingaben beziehungsweise fehlenden Eingabe-Reaktionen der Nutzer niederschlagen. Um die Vergleichbarkeit mit den Referenzdaten und die Diskriminationsfähigkeit zwischen Diagnosegruppen zu gewährleisten, werden diese aus den statistischen Analysen entfernt. Aus den für die Zieldefinition als relevant identifizierten Parametern lassen sich neben den durch die Digitalisierung zusätzlich erfassbaren Variablen, weitere Maße bilden, welche auf die theoretischen Beziehungen der Messkonzepte zwischen den einzelnen Tests eingehen. Hier wird beispielsweise die Leistungsdifferenz zwischen den Tests E und F als ein zusätzlicher Parameter interpretiert, der Information über die Veränderung zwischen den Durchgängen darstellt. Ein weiteres Beispiel hierfür ist der Quotient von den

Bearbeitungsdauern der Tests D und C, welcher als geeignetes Maß zur Beurteilung der exekutiven Funktion dienen kann [152]. Da einzelne Tests oder Teile davon aus bestehenden Testbatterien entnommen sind, werden Metriken notwendig, welche insbesondere die Vergleichbarkeit mit den Referenzdaten sicherstellen. So ist beispielsweise Test A mit vergleichbaren inhaltlichen Fragen in der CERAD-Testbatterie enthalten. Der Ausschluss von Erhebungsmethoden kann auch aufgrund von verwendeten digitalen Interaktionsparadigmen vorzunehmen sein, über welche nicht das gewünschte Ergebnis erzielt werden konnte. Hier ist beispielsweise der Einsatz von Spracherkennungssoftware als Eingabemethode für Test B zu nennen, welcher aufgrund des damit erzielten limitierten Datenumfanges und der mangelnden Datenqualität komplett von der Auswertung ausgeschlossen werden muss [152]. Eine detaillierte Erklärung aller 79 für *DemPredict* spezifischen digitalen Parameter, nachfolgend auch Merkmale oder Features genannt, kann aus Tabelle A.4 entnommen werden. Für den Umgang mit fehlenden Werten gibt es unterschiedliche Strategien [240, S. 202-204]. Fehlende Eingaben können dazu führen, dass Studienteilnehmer, abhängig vom jeweiligen Untersuchungsgegenstand, komplett ausgeschlossen werden müssen. Für die Anwendung von statistischen Modellen und ML-Algorithmen führt diese Vorgehensweise aufgrund der limitierten Datenmengen zu einer weiteren Reduktion von Trainingsdaten und damit zu potentiellen Performanzproblemen. Eine Alternative stellt deshalb der Ersatz von nicht angegebenen Daten durch Füllwerte dar. Hier kann für fehlende Werte beispielsweise der Mittelwert über die beobachtete Gesamtpopulation oder gewisser Subpopulationen dienen. Am Beispiel von *DemPredict* können die Mittelwerte pro Diagnosegruppe zur Substitution der fehlenden Werte dienen, allerdings muss bei der späteren Wahl des Verfahrens und dafür relevanter Features dieser Umstand ausreichend berücksichtigt werden. Ebenso ist je nach Art des Verfahrens eine Standardisierung aller Features sinnvoll, da sich die Gewichtung dieser aufgrund deren Lage unterschiedlich auf das Modell auswirken kann [249]. Üblich ist hier die z-Transformation (vgl. Gleichung 2.1). Bei der statistischen Analyse zur Bewertung der Plausibilität und Qualität der Daten wird nach den zuvor skizzierten Prinzipien der Statistik vorgegangen. Da die Normalverteilungsannahme nicht für alle domänenspezifischen Variablen getroffen werden kann (vgl. Abschnitt A.11), finden nachfolgend entsprechend nicht-parametrische Verfahren ihre Anwendung. Bei den Studienteilnehmern, für die zu den Testzeitpunkten T1 und T2 Daten vorliegen, kann ein Vergleich der individuellen Leistung pro Testzeitpunkt Aufschluss darüber geben, wie sich die Werte entwickeln. Bei vorliegender Konstruktvalidität der Tests und gleich bleibender Diagnose der Studienteilnehmer über den entsprechenden Zeitraum ist keine statistisch signifikante Abweichung zwischen den beiden Testzeitpunkten zu erwarten. Eine Leistungssteigerung zu späteren Testzeitpunkten kann als Hinweis auf Lerneffekte interpretiert werden, eine Verschlechterung auf eine abnehmende kognitive Leistungsfähigkeit. Letzteres müsste in einer Neubefundung und gegebenenfalls Änderung der Referenzdiagnose resultieren. Für die vorliegenden Daten lässt sich dieser Vergleich mittels zweiseitigem WVRT-Test ziehen. Aufgrund der Anzahl an Features, wofür jeweils der Test durchgeführt werden muss, findet eine Korrektur der  $\alpha$ -Fehlerkumulierung, wie in Absatz 2.1.4 beschrieben, für das zuvor definierte Signifikanzniveau  $\alpha = 5\%$  und die Anzahl an Testdurchführungen  $n_{Testinstanzen} = 66$  statt. Die aufsteigend sortierten  $p$ -Werte werden paarweise mit den lokalen  $p$ -Werten verglichen (vgl. Abschnitt A.12). Sobald bei einem Feature keine Signifikanz anzunehmen ist, kann auch für alle nachfolgenden Features keine Signifikanz mehr angenommen werden. Diese Bedingung trifft bereits für den Parameter mit dem niedrigsten Rang zu und bezieht sich somit auf alle Features (vgl. Abschnitt A.12). Dies lässt darauf schließen, dass sich die Ränge und demnach auch die zentralen Tendenzen zwischen den Testzeitpunkten nicht signifikant ändern. Aussagen hinsichtlich der realen Veränderung kognitiver Leistungsfähigkeit über die Zeit sind, auf Grundlage der Ergebnisse dieser Gegenüberstellung und der Tatsache, dass sich für das bestehende Studienteilnehmerkollektiv keine Änderung der Diagnosen zwischen T1 und T2 ergibt, mittels *DemPredict* möglich (DNFAM2.2).

## 5.4.2 Vergleich zwischen analoger und digitaler Lösung

Da Teile des analog erhobenen Referenzdatensatzes auf den gleichen Testkonzepten beruhen, wie sie auch in der digitalen Anwendung zum Einsatz kommen, lässt sich zumindest anhand dieser Tests ein direkter Vergleich zwischen digitalen und analogen Umsetzungen ziehen. Um damit auch den potentiellen Einfluss der Digitalisierung zu erkennen werden nachfolgend zwei Ansätze verfolgt. Einerseits findet der Vergleich in aggregierter Form, das heißt über alle Testdurchführungen hinweg, statt. Andererseits ein individueller Vergleich von den Leistungen zwischen analogen und digitalen Subtests pro Nutzer. Die vergleichbaren Tests sind A, C, D sowie J, da sie sowohl analog in der CERAD-Testbatterie als auch digital in *DemPredict* abgebildet sind.

### Aggregierter Vergleich

Um die Ergebnisse der Tests auf Parameterebene zu vergleichen und die mit dem digitalen Instrument erhobenen Daten bezüglich dessen Ziels einordnen zu können, werden nachfolgend die analog erhobenen Ergebnisse, den digitalen Ergebnissen gegenübergestellt. Ein Vergleich zwischen den Daten aus *DemPredict* mit denen der CERAD-Testbatterie setzt vollständige Daten der beiden Erhebungen voraus. Da für sechs Studienteilnehmer aufgrund der verteilten Durchführung zwar die Diagnosen aber nicht die Ergebnisse der damit korrespondierenden CERAD-Testungen vorhanden sind, werden diese von allen weiteren Untersuchungen bezüglich des Einflusses der Digitalisierung ausgeschlossen (vgl. Abbildung A.3). Die Zahl der Studienteilnehmer reduziert sich dadurch auf  $n = 49$ . Einen Überblick über die digitalen Daten lässt sich visuell verschaffen, um auf aggregierter Ebene die Ergebnisse aller Studienteilnehmer gegenüberzustellen. In Histogrammen werden die Häufigkeitsverteilungen der individuellen Abweichungen

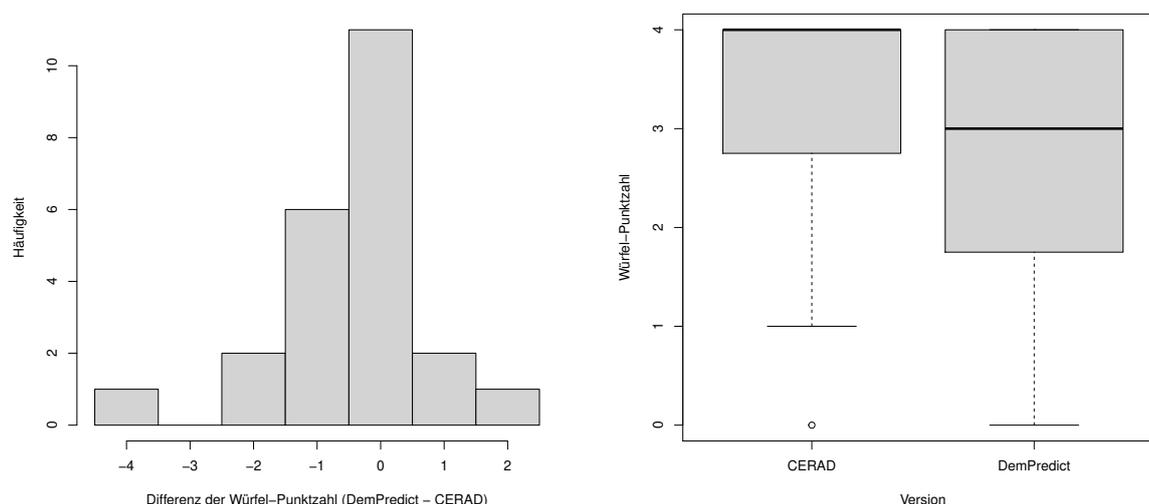


Abbildung 5.22: Vergleich zwischen Ergebnissen der CERAD- und *DemPredict*-Version für Test J.

zwischen digitalen und analogen Referenzdaten sichtbar. Hierdurch wird deutlich, ob und wie stark die erzielten Ergebnisse einzelner Studienteilnehmer zwischen analoger und digitaler Methoden voneinander abweichen (vgl. Abschnitt A.13). Boxplots visualisieren in einer direkten Gegenüberstellung der Lage- und Streuungsmaße, wie die Test-Versionen pro Leistungsparameter abschneiden (vgl. Abschnitt A.13). Anhand der Würfel-Punktzahl des Tests J und der Ergebnisse beider Versionen findet nachfolgend eine

exemplarische Gegenüberstellung statt. Die Bewertung der erreichten Punktzahlen beider Test-Versionen ist hierbei vom Ermessen menschlicher Entscheider abhängig und kann somit auch subjektiven Schwankungen unterliegen. Verblindete Bewertungen durch mehrere unabhängige Personen versuchen diese Schwächen auszugleichen schließen diese jedoch nicht komplett aus. Für die mit *DemPredict* erzielten Ergebnisse wird pro Studienteilnehmer die durchschnittliche Würfel-Punktzahl über T1 und T2 unter der Voraussetzung einer gleich bleibenden Diagnose angenommen, um Verzerrungen die durch die Bewertung oder Tagesform abhängen können vorwegzunehmen. Bezüglich der vergebenen Punkte lässt sich einerseits visuell feststellen, dass der größte Teil an Studienteilnehmern keine abweichenden Ergebnisse zwischen der analogen und digitalen Testversion erreicht hat (vgl. Abbildung 5.22). Die linksschiefe Verteilung zeigt jedoch, dass von den Studienteilnehmern im Vergleich eine höhere Punktzahl in der analogen als in der digitalen Version erreicht wurde. Während die Test-Versionen im Median lediglich einen Punkt voneinander abweichen, zeigen die Streuungsmaße eine deutlich stärkere Abweichung in der digitalen Testversion. Hierdurch liegt die Vermutung nahe, dass der Digitalisierungsschritt eine zusätzliche Hürde für die Studienteilnehmer darstellt und die analogen Testergebnisse nicht eindeutig reproduziert werden können. Die Anzahl gesunder Studienteilnehmer wirkt sich auf die Punkteverteilung aus. Während niedrige Punktzahlen im CERAD-Datensatz als Ausreißer klassifiziert werden, liegen deren Entsprechungen in *DemPredict* im unteren Whisker. Diese beziehen sich auf Studienteilnehmer mit starker Demenz und dementsprechend der potentiell eingeschränkten Fähigkeit diese Aufgabe den Leistungsparametern von Test J entsprechend zu reproduzieren.

### Individueller Vergleich

Um die Testergebnisse der einzelnen Studienteilnehmer gegenüber zu stellen, werden zur Visualisierung der jeweiligen Ergebnisse Streudiagramme eingesetzt, welche die Daten aus der analogen Testung mit den Daten aus der digitalen Testversion pro Studienteilnehmer verknüpfen. Die Abszisse stellt hierbei die Dimension der analogen Testergebnisse des CERAD dar, die Ordinate trägt digitale Daten von *DemPredict* auf (vgl. Abbildung 5.23 und Abbildung A.7). Mittels Korrelationsanalyse wird überprüft, ob

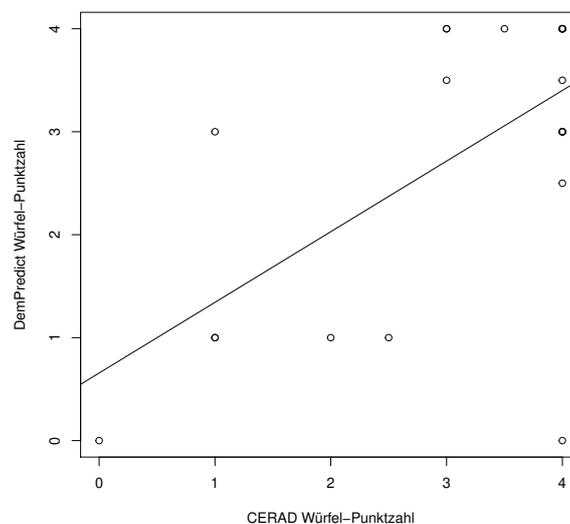


Abbildung 5.23: Korrelation zwischen Würfeltest-Punktzahl in CERAD- und *DemPredict*-Version nach Pearson.

und in welchem Zusammenhang die Ergebnisse der beiden Tests stehen. Aufgrund der abzulehnenden

Normalverteilungsannahme (vgl. Tabelle A.5) dient der Rang-Korrelations-Koeffizient nach Spearman zur Überprüfung der Stärke und Richtung des Zusammenhangs. Durch die Betrachtung der Ränge liefert dieser Korrelationskoeffizient ein robusteres Ergebnis, bildet jedoch die im Streudiagramm dargestellte lineare Regression etwas weniger gut ab als der Korrelationskoeffizient nach Pearson. Am Beispiel des Würfeltests zeigt sich ein positiver linearer Zusammenhang (vgl. Abbildung 5.23). Die Korrelationsanalyse zeigt mit  $\rho = 0.44$  und einem p-Wert von 0.03 für Test J wie deutlich dieser Zusammenhang ist, nach Cohen geht man bei diesen Werten von einem mittleren bis starken Zusammenhang aus [97]. Nachfolgend werden tabellarisch die Korrelationskoeffizienten der einzelnen vergleichbaren Testparameter nach Spearman zusammengefasst (vgl. Tabelle 5.3). Für alle Parameter gelten die Zusammenhänge als signifikant. Bei Test A wird über die Anzahl richtiger Antworten (p37) eine Vergleichbarkeit des im CERAD enthaltenen MMSE mit den in *DemPredict* integrierten Fragen zur räumlichen und zeitlichen Orientierung hergestellt. Dieser Zusammenhang ist ebenfalls mittel bis stark. Der Test C zeigt eine starke Korrelation zwischen den Bearbeitungszeiten der beiden Versionen und damit dem für diese Domäne relevanten Leistungsparameter. Der Test D zeigt sogar eine sehr starke Korrelation zwischen den Bearbeitungszeiten der beiden Versionen. Die individuelle Gegenüberstellung der Leistungen in analogen und digitalen Formaten zeigt, dass die Ergebnisse, bis auf geringfügige Abweichungen, auf einem vergleichbaren Niveau sind und sich damit digitale psychometrische Testkonzepte prinzipiell auch im ambulanten Umfeld als valide einstufen lassen. Weitere Rückschlüsse bezüglich der Eignung der zusätzlichen Tests von *DemPredict*, für die keine direkte analoge Referenz besteht, können über die durchgeführten Beobachtungen nicht gezogen werden.

Test (Kürzel)	Parameter	p-Wert	$\rho$
A	p37	0,026	0,474
C	p47	0,0004	0,601
D	p49	$9,357 \cdot 10^{-8}$	0,847
J	p2	0,034	0,444

Tabelle 5.3: Ergebnisse der Rang-Korrelationsanalyse nach Spearman für vergleichbare Parameter.

### 5.4.3 Untersuchung weiterer Einflussfaktoren

Der Inhalt einer digitalen Gesundheitsanwendung, insbesondere die Validität damit erzielter Ergebnisse bezüglich des Verwendungszwecks, spielt eine Rolle für deren Akzeptanz. Damit verbundene multidimensionale Einflüsse können über die Erhebung zusätzlicher Variablen während den Evaluationsphasen überprüft werden. Als ein Beispiel hierfür kann die Technikerfahrung der potentiellen Nutzer dienen, welche einen Einfluss auf die zu erhebenden gesundheitsrelevanten Informationen haben kann. Die in Unterabschnitt 5.4.1 dargestellte Kennzahl für Technik-Erfahrung (*TE*) dient daher zur Bestimmung dieses Einflussfaktors (vgl. Gleichung 5.2). Um den Einfluss der Technik-Erfahrung auf den Digitalisierungsschritt zu quantifizieren, werden Korrelationen zwischen den Ergebnissen der *DemPredict*- und CERAD-Testungen sowie dem Delta in deren Bearbeitungszeiten bestimmt. Hierzu muss zunächst Aufschluss darüber gegeben sein, ob Technik-Erfahrung mit der kognitiven Leistungsfähigkeit des Studienteilnehmers und deren Diagnosen korreliert.

Ein Vergleich der Mediane pro Diagnosegruppe zeigt, dass die Erfahrung im Umgang mit Technik bei gesunden Studienteilnehmern ( $\overline{TE} = 1,5$ ) höher ausfällt als bei denen mit diagnostizierten MCI ( $\overline{TE} = 0,616$ ) oder Demenz ( $\overline{TE} = 0,5$ ). Der WRT gibt Auskunft darüber, ob in einem paarweisen Vergleich signifikante Unterschiede in der Technikerfahrung zwischen den Diagnosegruppen vorliegen (vgl. Absatz 2.1.4). Die p-Werte zeigen, dass davon nicht auszugehen ist (vgl. Tabelle 5.4).

	D	G
G	0,58	-
MCI	0,75	0,21

Tabelle 5.4: Ergebnisse des WRTs zum paarweisen Vergleich der Diagnosegruppen und deren  $TE$ .

Das Maß  $TE$  gilt somit als unabhängig zwischen den Diagnosegruppen. Um herauszufinden, ob die Testformate Menschen mit weniger Technik-Erfahrung benachteiligen, wird die Korrelation zwischen den Parametern der Formate und der Technik-Erfahrung untersucht (vgl. Abschnitt A.7). Bei positiver Korrelation liegt die Vermutung von einer Benachteiligung der weniger erfahrenen Nutzern nahe, bei negativer von einem Vorteil. Als Maß zur Untersuchung der Unterschiede zwischen analogen und digitalen Formaten sowie der Frage, ob zwischen Technik-Erfahrung und gemessener kognitiver Leistungsfähigkeit in den jeweiligen Formaten ein signifikanter Zusammenhang besteht, wird die Differenz ( $DemPredict - CERAD$ ) der jeweiligen Leistungsparameter eingeführt. Sollte hierbei ebenfalls eine signifikante Korrelation feststellbar sein, ist von einem systematischen Vor- oder Nachteil in dem jeweiligen Format auszugehen.

Am Beispiel des Test J lässt sich eine solche Gegenüberstellung exemplarisch abbilden (vgl. Abbildung 5.24). Die Würfel-Punktzahl ( $p_2$ ) ist in diesem Fall der relevante Leistungsparameter in beiden Formaten. Für die mit  $DemPredict$  erzielten Ergebnisse wird hier wie bereits zuvor pro Studienteilnehmer die durchschnittliche Würfel-Punktzahl über T1 und T2 unter der Voraussetzung einer gleich bleibenden Diagnose verwendet. Visuell lässt sich eine positive Korrelation zwischen der Anzahl er-

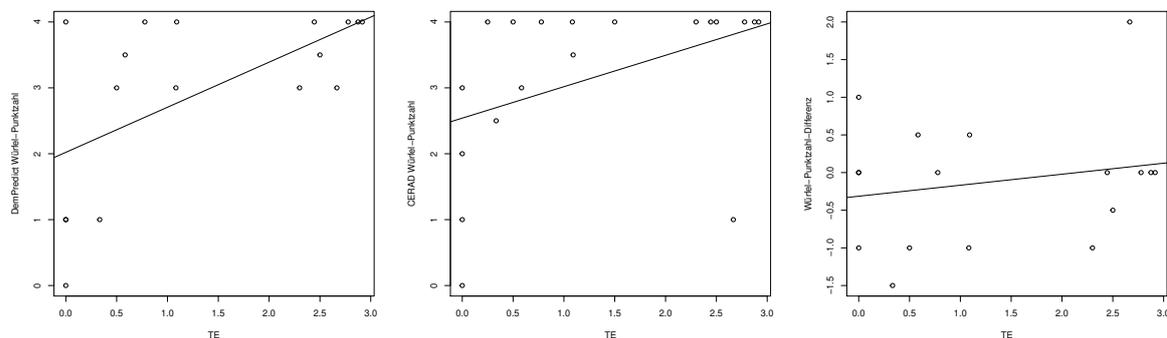


Abbildung 5.24: Korrelationsanalyse zwischen den Würfeltest-Ergebnissen der CERAD- und  $DemPredict$ -Version mit der  $TE$ .

reichter Punkte und Technik-Erfahrung feststellen. Die Korrelationsstärke liegt im  $DemPredict$ -Format bei  $\rho = 0,6$  und  $p = 0,011$ , was nach Cohen zur Annahme eines bedeutenden Zusammenhangs mit hoher Signifikanz führt [97]. Die analogen Ergebnisse aus dem CERAD zeigen mit  $\rho = 0,571$  und  $p = 0,011$  ebenfalls einen bedeutenden Zusammenhang mit hoher Signifikanz. Diese Beobachtung kann darauf zurückzuführen sein, dass kognitiv leistungsfähigere Studienteilnehmer wie zuvor bereits durch

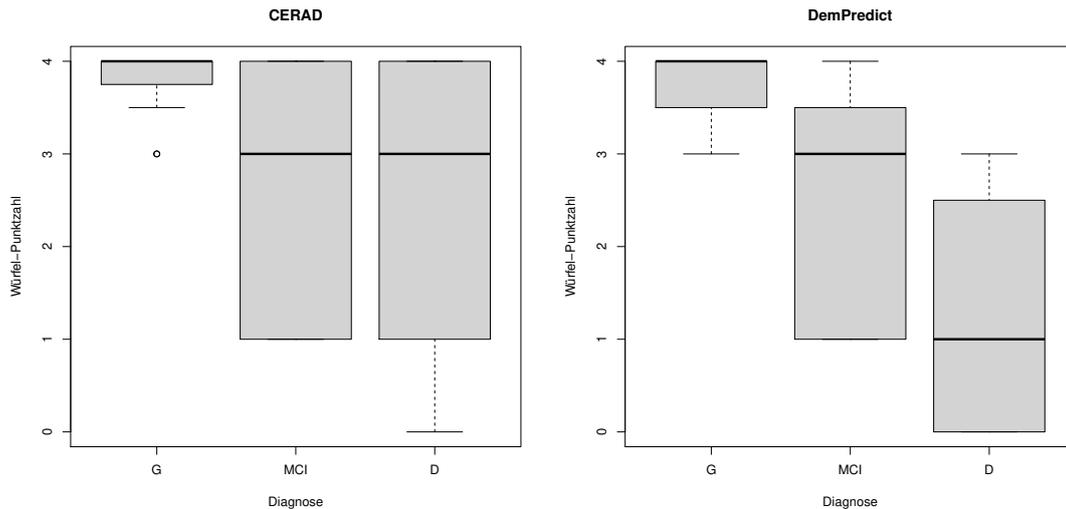
die Mediane dargestellt, eine etwas höhere Technik-Erfahrung besitzen und dadurch zumindest einen Teil der beobachteten positiven Korrelation zwischen erreichter Punktzahl und Technik-Erfahrung ausmachen. Um diese Hypothese zu untermauern, kann die Differenz zwischen den erreichten Punktzahlen in *DemPredict* und CERAD betrachtet werden. Bei Betrachtung der Abweichung ist davon auszugehen, dass sich dieser Effekt dadurch aufhebt. Die Korrelationsstärke dieser Differenz liegt mit  $\rho = 0,155$  und  $p = 0,552$  ebenfalls im leicht positiven Bereich, was einen gewissen Vorteil von Studienteilnehmern mit Technik-Erfahrung ausweist, der allerdings ohne statistische Signifikanz ist. Die detaillierten Ergebnisse der Korrelationsanalyse für die erhobenen Parameter finden sich in Abschnitt A.15 und werden nachfolgend zusammengefasst.

Insgesamt fällt der Einfluss des Faktors Technik-Erfahrung bis auf einzelne Parameter gering aus. Die Korrelation gibt lediglich Aufschluss darüber, dass bei gewissen Formaten dieser Einflussfaktor eine Rolle spielt. Die Ursachen bleiben jedoch Bestandteil weiterer Untersuchungen in nächsten Iterationen des Entwicklungszyklus. Insbesondere Auswirkungen auf die Bearbeitungsdauer der Tests können von Eingabemethoden abhängig sein, welche gegebenenfalls von den Nutzern mit weniger Erfahrung langsamer durchgeführt werden. Solange die Dimension Dauer der Durchführung aber nicht als Leistungsparameter für die jeweilige Domäne anzusehen ist, ist der Ausschluss des Tests nicht erforderlich.

#### 5.4.4 Diskriminationsfähigkeit

Nachdem für die digitalisierten Versionen der Tests gezeigt werden konnte, dass diese im Vergleich mit deren analogen Äquivalenten ein vergleichbares Niveau erreichen, muss hinsichtlich der in Abschnitt 4.2 festgelegten Zieldefinition, die Eignung der gesamten digitalen Gesundheitsanwendung zur ambulanten Früherkennung einer Demenz geprüft werden. Hierbei stellt sich die Frage, ob die Qualität, welche die mittels *DemPredict* gesammelten Daten aufweisen, ausreicht, um eine möglichst präzise Unterscheidung zwischen den Diagnosegruppen zu erreichen. Eine eindeutige Zuordnung des zu klassifizierenden Nutzers zu einer Diagnosegruppe setzt eine große Interklassendistanz bei geringer Intra-Klassendistanz voraus. Um diese Diskriminationsfähigkeit der digitalen Parameter zu bewerten sind die Unterschiede zwischen deren Ausprägungen und den Diagnosegruppen zu untersuchen. Eine detaillierte Untersuchung erfolgt diesbezüglich mittels statistischer Verfahren. Hierfür ist es notwendig herauszufinden, ob die Daten in Abhängigkeit der Diagnosegruppe derselben Verteilung folgen. Im Falle signifikanter Unterschiede in den Verteilungen wird von einer gewissen Unterscheidbarkeit zwischen den Gruppen ausgegangen [358]. Wie in Unterabschnitt 5.4.1 beschrieben, lässt sich unter Berücksichtigung der zu verwerfenden Normalverteilungsannahme der KWT zur Prüfung von  $k$  unabhängigen Stichproben und deren Zugehörigkeit bezüglich einer Grundgesamtheit verwenden (vgl. Abschnitt A.11). Hiermit lässt sich feststellen, ob es zwischen den Gruppen signifikante Unterschiede gibt, jedoch nicht, ob zwischen allen Gruppen diese Unterschiede bestehen. Als Post-Hoc-Test wird daher für einen paarweisen Vergleich zwischen den Diagnosegruppen der WRT durchgeführt.

Am Beispiel des Tests J ist in Abbildung 5.25 zu erkennen, wie sich die Lage- und Streuungsmaße für den relevanten Leistungsparameter Würfel-Punktzahl ( $p_2$ ) bezüglich der unterschiedlichen Diagnosegruppen im Vergleich zwischen der CERAD- und *DemPredict*-Version verhält. Besonders auffällig an den Ergebnissen der Studiendaten ist, dass die Unterscheidung zwischen den Gruppen auch für den CERAD nicht eindeutig ausfällt. Die digitalisierte Version in *DemPredict* zeigt, dass es für diesen Parameter stärkere Unterschiede zwischen den einzelnen Gruppen gibt. Vor allem zwischen MCI und D existiert jedoch eine starke Überschneidung der Intervalle in beiden Versionen. Auffällig ist auch, dass alle gesunden Studienteilnehmer im Durchschnitt zwischen drei und vier Punkten erzielen. Eine zuverlässige Unterscheidung

Abbildung 5.25: Würfeltest-Ergebnisse der CERAD- und *DemPredict*-Version pro Diagnosegruppe.

<i>CERAD</i>		<i>DemPredict</i>			
	D	G		D	G
G	0,290	-	G	0,006	-
MCI	0,900	0,200	MCI	0,113	0,033

Tabelle 5.7: Ergebnisse des WRTs zum paarweisen Vergleich der Diagnosegruppen und Würfel-Punktzahl pro Version.

zwischen den drei einzelnen Klassen scheint daher alleine durch Betrachtung dieses Leistungsparameters nicht möglich zu sein. Auf eine detaillierte textuelle Interpretation der visuellen Darstellung wird an dieser Stelle zugunsten von Ergebnissen tiefer gehender statistischer Untersuchungen verzichtet. Die  $p$ -Werte des KWTs liegen für die beiden Versionen bei  $p_{CERAD} = 0,1336$  und  $p_{DemPredict} = 0,002$ . Dies deutet darauf hin, dass die analoge Version, im Gegensatz zur digitalen Version, keine signifikante Unterscheidung zwischen den Diagnosegruppen zulässt. Da sich die analoge Version auf einen standardisierten Test bezieht, sind Einflüsse der limitierten Stichprobengröße auch an dieser Stelle nicht von der Hand zu weisen.

Als nächster Verarbeitungsschritt folgt der paarweise Vergleich der Diagnosegruppen, um genauer differenzieren zu können, zwischen welchen signifikante Unterschiede vorliegen. Hierfür liefert der WRT die entsprechenden Ergebnisse (vgl. Tabelle 5.7). Während der CERAD keine signifikante Unterscheidbarkeit zwischen den Diagnosegruppen liefert, erreicht *DemPredict* diese mit  $p_{DemPredict-G-D} = 0,006$  und  $p_{DemPredict-MCI-G} = 0,033$  zwischen den gesunden Studienteilnehmer und denen mit Demenz beziehungsweise MCI. Die Diskriminationsfähigkeit für den digitalen Leistungsparameter Würfel-Punktzahl ( $p_2$ ) deutet also eine generelle Unterscheidbarkeit zwischen gesund und ungesund (D oder MCI) an. Die Ergebnisse dieser Untersuchungen für die anderen Tests sind in Abschnitt A.16 zu finden. Insgesamt lässt sich feststellen, dass die unterschiedlichen Parameter sehr stark hinsichtlich ihrer Aussagekraft bezüglich der Diagnosegruppe divergieren. Die Kombination aus Boxplots und statistischen Tests vermittelt einen ersten Eindruck und dient als Maß zur Beurteilung ihrer Trennschärfe. Die Untersuchungen zeigen, dass es für jeden der implementierten Tests mindestens einen Parameter gibt, der eine signifikante diagnostische Differenzierung zwischen mindestens zwei Gruppen abbildet. Ob der Nutzen all dieser Merkmale ausreicht und welche Modelle dazu geeignet sind, um mit in die digitale Gesundheitsanwendung aufgenommen zu werden, bildet den Gegenstand weiterer Untersuchungen in den folgenden Kapiteln.



# 6 Digitale Automatisierungs- und Optimierungsansätze zur Testauswertung

NEBEN den zuvor untersuchten Einflüssen der Digitalisierung bei der Aufnahme medizinischer Parameter im Vergleich mit analogen Interaktionsformen entstehen dadurch insbesondere Herausforderungen und Potentiale im Zuge derer Auswertung (DFAT2) (vgl. Unterabschnitt 3.3.3). In den letzten Jahren werden gesundheitsrelevante Entscheidungen zunehmend computergestützt getroffen [94, 212]. Digitale Gesundheitsdaten gelten dabei als Rohstoff, um diese Verfahren zu kalibrieren (vgl. Unterabschnitt 2.4.3). Die agile Natur moderner Softwareentwicklung ermöglicht durch die Integration einer zunehmenden Menge an Daten in den Entwicklungsprozess auch eine kontinuierliche Weiterentwicklung der Auswertelogik von digitalen Gesundheitsanwendungen (Phase Automatisierung & Optimierung, vgl. Unterabschnitt 4.1.2). Hierbei stellt das Ziel der zu entwickelnden Anwendung die Leitschnur diesbezüglicher Entscheidungen dar. Welche Kategorien von Optimierungs- und Automatisierungsansätzen am konkreten Beispiel von *DemPredict* existieren wird nachfolgend dargestellt. Insbesondere entsteht für die autonome Nutzung einer digitalen Gesundheitsanwendung durch den Nutzer selbst die Möglichkeit zur automatisierten Erfassung und Auswertung der erhobenen Daten, welche algorithmisch abgebildet werden muss. Diesbezügliche Möglichkeiten, die im Rahmen der Erfassung einzelner Parameter notwendig sind, werden in Abschnitt 6.1 untersucht. Da die Früherkennung eines dementiellen Syndroms einem Klassifikationsproblem zuzuordnen ist, besteht primär der Bedarf ein Modell zur verlässlichen Vorhersage einer Diagnoseklasse zu entwickeln. Inwiefern dafür eine für den menschlichen Entscheider verständliche Bewertungsmetrik abgeleitet werden kann, wird in Abschnitt 6.2 untersucht. Zur Optimierung dieser Datenauswertung über alle Parameter hinweg können in einem nächsten Schritt Verfahren eingesetzt werden, um eine frühestmögliche Vorhersage des Verlaufs treffen zu können und die besten Behandlungsoptionen zu finden [38]. Hierfür werden nachfolgend unterschiedliche Ansätze identifiziert, die durch ML unterstützt werden können. Die Wahl der hierzu geeigneten Algorithmen basiert auf einem Suchprozess, der in Abschnitt 6.3 beschrieben ist. Zur Plausibilisierung der Entscheidungen und Risikominimierung bezüglich des Systemverhaltens bei der Verwendung intransparenter Modelle werden in Abschnitt 6.4 abschließend XAI-Verfahren evaluiert, um zu untersuchen, welche sich für das beschriebene Anwendungsfeld eignen.

## 6.1 Automatisierte Extraktion einzelner Merkmale am Beispiel von Test J

Bei der Erfassung von Merkmalen, bei denen menschliche Interpretationsfähigkeiten zur Auswertung benötigt werden, ist ebenfalls die Eignung von Ansätzen aus dem Bereich des ML zu deren Automatisierung

zu prüfen. Um eine autonome Nutzung einer digitalen Gesundheitsanwendung im ambulanten Umfeld zu ermöglichen, muss die Erfassung und Auswertung der digitalen Parameter unabhängig von infrastrukturellen Voraussetzungen erfolgen können (DNFAT3). Daher muss die Transformation wissenschaftlicher Ergebnisse hin zu performanten und in realen Umgebungen einsetzbaren technischen Lösungen betrachtet werden. Letzteres fällt zwischen die Modellierungs- und Transfer-Schritte der abgeleiteten Vorgehensweise (vgl. Abbildung 4.3). In welcher Form eine Automatisierung durch ML gestützt erfolgen kann, hängt von der multimedialen Erfassung einzelner Parameter ab. Am Beispiel von *DemPredict* kommt für den Test B eine Sprachanalysesoftware zum Einsatz, um die relevanten Parameter entsprechend zu transkribieren und inhaltlich auswerten zu können. Durch die Erhebung im ambulanten Umfeld mussten aufgrund der mangelnden Qualität der Daten und Algorithmik diese Ergebnisse ausgeschlossen werden (vgl. Unterabschnitt 5.4.1). Aktuelle Forschungsansätze zeigen konträr zu den im Rahmen dieser Arbeit gesammelten Erfahrungen, dass für neurologische Erkrankungen, insbesondere auch für Demenz digitale Biomarker im Bereich der Sprache existieren und für Diagnosen genutzt werden können [352, 397]. Derartige digitale Parameter können einen ganzheitlichen Ansatz zur Früherkennung unterstützen und deuten weitere Potenziale einer digitalen Erfassung an, setzen jedoch auch Erfahrung bei den Nutzern im Umgang mit neuen Technologien und damit verbundenen Interaktionsparadigmen voraus. Bei der Entwicklung solcher Verfahren stellen der Umfang und die Qualität hierfür erforderlicher Daten Herausforderungen dar. Anhand des Tests J wird nachfolgend gezeigt, welche weiteren Herausforderungen und Potentiale solche Digitalisierungsbestrebungen mit sich bringen. Wie die detaillierte Problembeschreibung anhand der Würfelaufgabe aus dem RUDAS und CERAD in Unterabschnitt 2.2.5 bereits darstellt, handelt es sich bei den im Rahmen dieses Tests zu erhebenden Parametern, um die Überprüfung der visuellen Wahrnehmung sowie des visuell-konstruktiven Denkens. Diese Anforderung an die kognitive Funktionsdomäne der Nutzer lassen sich auch auf die technische Systemsicht übertragen und in digitale Bildverarbeitungsschritte herunterbrechen, wodurch ein Zusammenhang zu den Forschungsbereichen der digitalen Bilddatenerfassung und -verarbeitung besteht. Erste Lösungsansätze für eine digitale Objekterkennung in diesem Kontext existieren bereits für den Uhren-Zeichen-Test [303, 354, 387, 388]. Diese Arbeiten verfolgen unterschiedliche Ansätze bezüglich der Erfassung, Interpretation und Klassifikation der digitalen Daten hinsichtlich unterschiedlicher Merkmale, die auf verschiedene Formen einer Demenz hinweisen können. Einerseits werden Auffälligkeiten im Zeichenprozess mittels digitalem Stift erfasst und ausgewertet [387, 388]. Andererseits wird versucht das Ergebnis der durchgeführten Tests durch bestehende Bewertungskriterien zu operationalisieren und algorithmisch nachzubilden [303, 387, 388]. Da lediglich die Bewertung des Endergebnisses im ursprünglichen Testformat Berücksichtigung findet, wird diese Vorgehensweise auch nachfolgend für den Test J gewählt. Bezüglich der Entwicklung ist zu beachten, dass sich die Laufzeitumgebung auf einem eingebetteten System befindet und deshalb die Portabilität der Algorithmen nach der wissenschaftlichen Entwicklungsphase in ein produktfähiges Setting gewährleistet sein muss. Daraus muss ein Transfer aller Abhängigkeiten sowie darin verwendeter Algorithmen auf mobile Plattformen möglich sein. Je nach Verfügbarkeit offener Implementierungen auf unterschiedlichen Plattformen, muss dieser Aufwand insbesondere bei der Technologieauswahl bereits berücksichtigt werden.

### 6.1.1 Vorgehensweise

In Unterabschnitt 4.2.1 wird erläutert, anhand welcher Kriterien die Einschätzung der kognitiven Leistungsfähigkeit eines Studienteilnehmers in Test J bestimmt wird. Hierbei wird verlangt einen abgebildeten Würfel nachzuzeichnen. Dieser besteht aus zwölf Kanten, die in einer transparenten dreidimensionalen

Darstellung in äußere und innere Linien zerlegt werden können. Die Leistungsparameter zur Bewertung des Ergebnisses umfassen, wie in Abbildung 4.5 dargestellt, Parallelität (PP), Dreidimensionalität (PD) sowie die korrekte Darstellung der Frontseite (PF) und der inneren Linien (PI). Ziel ist es dieses operationalisierte Bewertungssystem nachzubilden. Welche Vor- und Nachteile zwischen klassischen Verfahren im Bereich des Computer Vision (CV) und des in diesem Bereich vermehrt auftkommenden Deep Learning (DL) bestehen wird in der wissenschaftlichen Literatur diskutiert [311]. Zusammenfassend werden DL-Modelle oft als Blackboxes bezeichnet, da deren tiefe neuronalen Strukturen eine Vielzahl an Parametern enthalten [311]. Außerdem sind diese für Overfitting bei einer zu geringen Anzahl an Trainingsdaten anfällig [311]. Die Nachvollziehbarkeit der extrahierten Merkmale und darauf basierenden Entscheidungen sind ebenfalls noch Bestandteil aktueller Forschung (vgl. Unterabschnitt 3.3.4). Dem gegenüber stehen traditionelle Ansätze, die durch einen menschlichen Entwickler formulierte Algorithmen vorsehen, welche auf das Problem zugeschnitten werden und außerhalb der Trainingsumgebung gleich funktionieren wie innerhalb [311]. An dieser Stelle wird deshalb die Extraktion verschiedener Merkmale basierend auf wissensbasierter Algorithmik geplant und umgesetzt, um einen Überblick und eine Diskussionsgrundlage für die Herausforderungen und Chancen einer automatisierten Parametereauswertung zu erhalten.

## 6.1.2 Kantendetektion

Ausgangspunkt ist die dreidimensionale Darstellung der Kanten eines transparenten Quaders mit gleicher Kantenlänge (Würfel) auf einer zweidimensionalen Fläche (vgl. Abbildung 6.1).

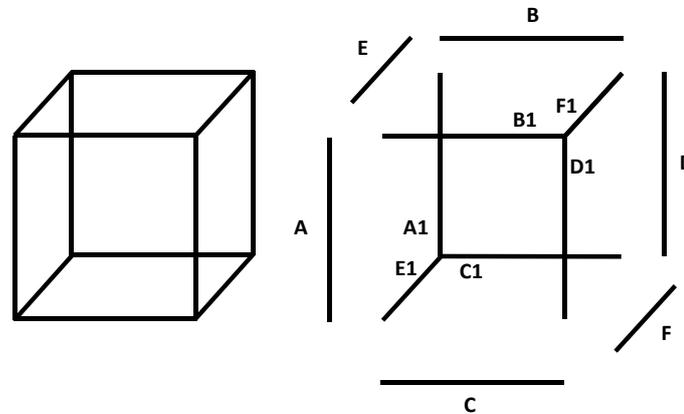


Abbildung 6.1: Visuelle Wahrnehmung und Rekonstruktion eines Würfels.

Bei dessen Reproduktion durch den jeweiligen Testteilnehmer handelt es sich um einen Würfel mit schwarzen Linien auf weißem Hintergrund. Dadurch lässt sich das Ergebnis binär in einer Matrix codieren (vgl. Abbildung 6.2). Die Zerlegung des Würfels resultiert in Suchprozessen, welche aus verschiedenen Richtungen (links, oben, unten, rechts) zunächst die äußeren Kanten identifizieren und jeweils in Matrizen für  $\mathbf{AE} = \mathbf{A} \cup \mathbf{E}$  sowie äquivalent definiert  $\mathbf{EB}$ ,  $\mathbf{CF}$  und  $\mathbf{FD}$  unterteilen. Die äußeren räumlich orientierten Kanten  $\mathbf{E}$  zwischen  $\mathbf{AE}$  und  $\mathbf{EB}$  sowie  $\mathbf{F}$  zwischen  $\mathbf{CF}$  und  $\mathbf{FD}$  stellen die jeweilige Schnittmenge, wie beispielsweise durch  $\mathbf{F} = \mathbf{CF} \cap \mathbf{FD}$  definiert, dar (vgl. Abbildung 6.3). Die äußeren horizontalen und vertikalen Kanten lassen sich daraus ebenfalls einzeln bestimmen.

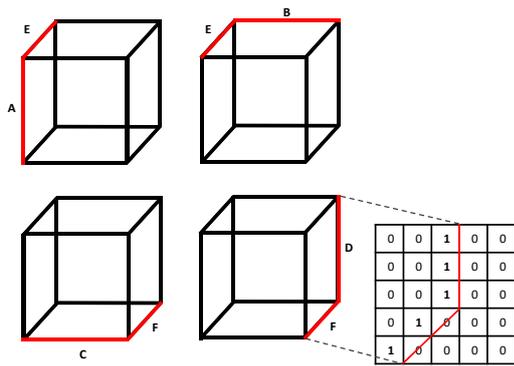
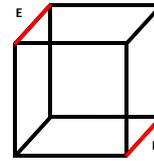


Abbildung 6.2: {AE,EB,CF,FD} in äußere Linien zerlegt.

0	0	0	0	0
0	0	0	0	0
0	0	2	1	1
0	2	0	0	0
2	0	0	0	0



0	0	1	0	0
0	0	1	0	0
0	0	2	0	0
0	2	0	0	0
2	0	0	0	0

Abbildung 6.3: Extraktion der äußeren räumlich orientierten Kanten.

Äquivalent dazu wird dieser Vorgang fortgesetzt, um die inneren Linien **A1**, **B1**, **C1**, **D1**, **E1**, **F1** zu detektieren. Als Ergebnis werden alle zwölf Kanten jeweils in einer Matrix repräsentiert. Der Schwellwert ist heuristisch gewählt (vgl. Unterabschnitt 6.1.5), um die Robustheit des Verfahrens bei Abweichung von einem idealen Würfel zu gewährleisten und sicherzustellen, dass in diesem Fall die Kanten **B**, **C** und **B1** nicht komplett in **AE** einbezogen werden (vgl. Abbildung 6.4).

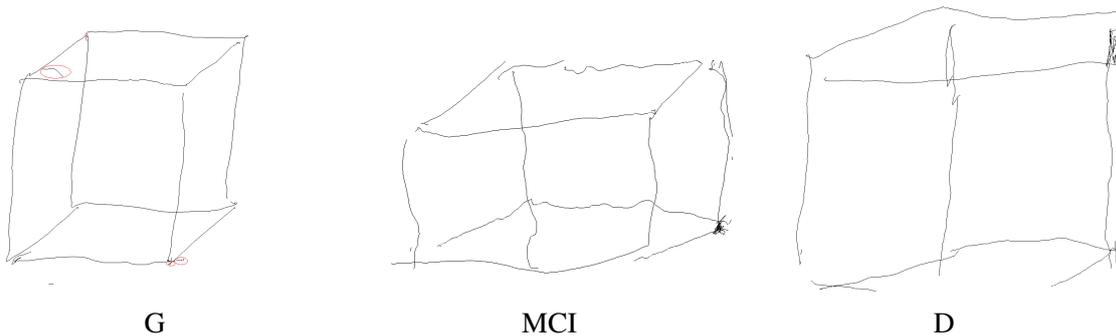


Abbildung 6.4: Ergebnis eines gesunden Testteilnehmers (rot markierte Abweichungen vom idealen Würfel), eines Testteilnehmers mit MCI und Demenz.

Weitere Abweichungen werden in Abbildung 6.4 unten rechts deutlich, indem die Kante über den eigentlichen Zeichenbereich des Würfels hinausgeht und somit als Teil der räumlich orientierten Kante **E** interpretiert werden kann. Um diese Abweichungen zu entfernen, werden verbundene Komponenten innerhalb der jeweiligen Kanten-Matrizen identifiziert (vgl. Abbildung 6.5) und bei entsprechend geringer Länge herausgefiltert [178, S. 40-48]. Auch hierbei handelt es sich um ein heuristisches Vorgehen, die Ergebnisse dieses Vorgangs werden in der Abbildung 6.6 für **AE** deutlich, indem Linien mit einer Länge unter dem Schwellwert von 5% der Abbildungsbreite entfernt werden. Um visuell zu prüfen, ob und inwiefern alle Kanten algorithmisch erkannt werden, lassen sich die einzelnen Kanten erneut zu einem Würfel zusammensetzen und der ursprünglichen Zeichnung gegenüberstellen. Die für den gesunden Testteilnehmer dargestellte Filterung zeigt, dass die Abweichungen aus der ursprünglichen Zeichnung bis auf einzelne Ausreißer entfernt werden, dadurch jedoch weitere charakterisierende Merkmale verschwinden können, die das ursprünglich gezeichnete Bild für eine spätere Auswertung verzerren. Die Optimierung der Kantendetektion durch die Anpassung der Schwellwerte oder den Einsatz weiterer Verfahren stellen eine potentielle Verbesserungsmöglichkeit dar, jedoch erscheint im Hinblick auf die Darstellungen von erkrankten Menschen eine reine Kantendetektion zur Bewertung der Leistungsparameter nicht als ausrei-

0	0	1	0	0	0	0	1	0	0	2	0	0	0	0	3
0	1	1	0	1	0	0	1	0	2	2	0	4	0	0	3
0	0	0	1	0	0	1	0	0	0	0	4	0	0	3	0
1	1	0	1	0	0	1	1	5	5	0	4	0	0	3	3
1	0	0	1	0	0	0	0	5	0	0	4	0	0	0	0
1	0	0	0	1	0	0	1	5	0	0	0	4	0	0	6
1	0	0	0	0	1	0	1	5	0	0	0	0	4	0	6
1	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0

Abbildung 6.5: Vergleich der ursprünglichen Matrix (links) und den algorithmisch erfassten zusammenhängenden Komponenten (rechts)



Abbildung 6.6: Vergleich der zusammengesetzten Kanten **AE** vor (links) und nach (rechts) der heuristischen Filterung am Beispiel des gesunden Testteilnehmers.

chend (vgl. Abbildung 6.7). Im nächsten Abschnitt wird deshalb die Erkennung charakteristischer Ecken thematisiert und eruiert inwiefern eine Kombination dieser beiden Ansätze zur Bewertung der relevanten Leistungsparameter möglich ist.

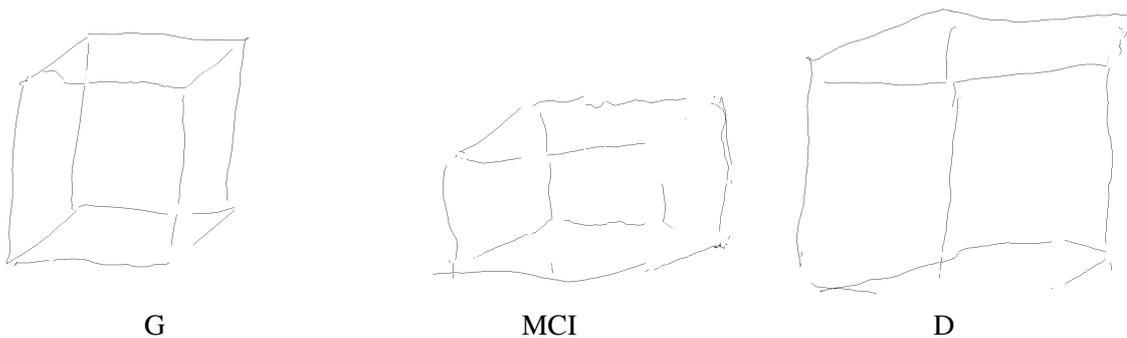


Abbildung 6.7: Vergleich der ursprünglichen Zeichnung (vgl. Abbildung 6.4) mit dem Ergebnis nach der Kantendetektion für Testteilnehmer mit den Labels G, MCI und D.

### 6.1.3 Eckendetektion

Als Eckenpunkte werden nachfolgend Verbindungsstellen von Kanten oder Linien verstanden. Neben den acht Ecken, die eine dreidimensionale Darstellung eines Würfels besitzt, treten bei idealen Würfeln, zwei weitere Schnittpunkte zwischen Kanten auf. Wie in Abbildung 6.8 dargestellt sind diese eindeutig definiert indem von jeder Ecke genau zwei Kanten ausgehen, die diese identifizieren. Nachfolgend werden

die Schritte zusammengefasst, um diese zu lokalisieren.

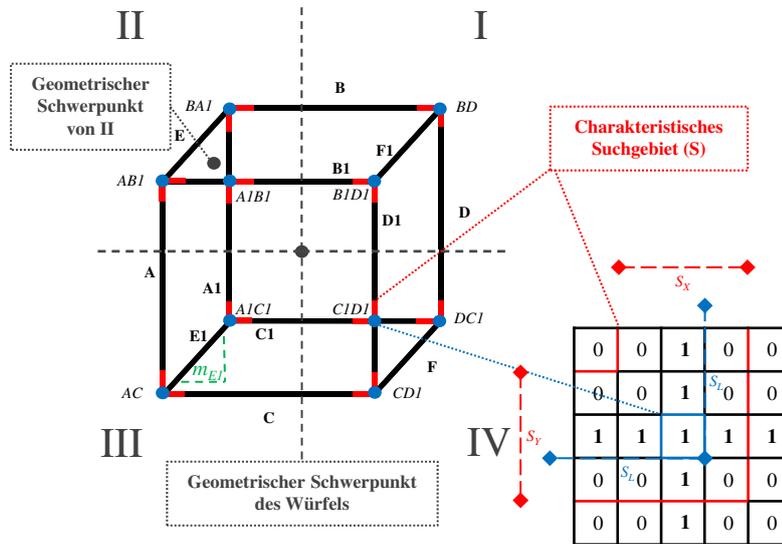


Abbildung 6.8: Definition von Parametern zur automatisierten Eckendetektion in Test J.

Ausgangspunkt ist die Annahme, dass ein Nutzer der Aufgabenstellung folgend, beabsichtigt eine ideale Abbildung des Würfels zu reproduzieren. Da es sich aus technischer Sicht um ein Schwarz-Weiß-Bild mit einer Strichstärke von einem Pixel handelt, ist der Einsatz von in der Bildverarbeitung etablierten Verfahren zu einer kantenbasierten Grauwert-Eckendetektion, wie der Harris-Stephens-Detektor, für das Ziel nicht spezifisch genug [179]. Deshalb wird nachfolgend ein Suchprozess beschrieben, welcher die zur Bestimmung der Leistungsparameter PP, PD, PF und PI notwendigen Ecken in bestimmten Bild-Bereichen erwartet. Hierfür zu bestimmende Schwellwerte werden unter Zuhilfenahme der, durch die in dieser Testmethode ausgebildeten medizinischen Fachkräfte, gelabelten Referenzdaten bestimmt. Im ersten Schritt erfolgt die Suche zunächst auf die Größe der Abbildung dynamisch angepasst, indem der geometrische Schwerpunkt aller verbundenen Komponenten innerhalb der Abbildung den Ursprung eines kartesischen Koordinatensystems bildet und dieses demnach in vier Quadranten unterteilt wird (vgl. Abbildung 6.8). Innerhalb dieser vier Quadranten werden jeweils Suchkriterien definiert, die am Beispiel von Quadrant IV, die darin enthaltenen unteren rechten Ecken charakterisieren. Entsprechend der Ausrichtung von den jeweils zwei identifizierenden Kanten für die Ecken  $CD1$ ,  $DC1$  und  $C1D1$  zählt eines dieser Fenster die Pixel nach links, das andere nach oben. Sobald an einem Ausgangspunkt der Suche ein gewählter Schwellwert für die Länge in beide charakteristischen Suchrichtungen überschritten wird, lässt sich dieser Bereich als potentielle Ecke interpretieren. Somit werden alle als Ecken zu verdächtigenden Punktwolken, welche dem Quadranten spezifischen Suchmuster folgen, identifiziert.

Unter realen Bedingungen können mehrere Ecken durch Ungenauigkeiten im Zeichenprozess entstehen. Daher werden in einem zweiten Schritt erneut die potentiellen Ecken herausgefiltert, welche wahrscheinlich nicht den gesuchten Ecken entsprechen. Hierfür werden für Quadrant I und III jeweils die potentiellen Eckpunkte gewählt, die am nächsten beziehungsweise weitesten vom geometrischen Schwerpunkt des gesamten Würfels entfernt sind. Für Quadrant II und IV wird erneut ein geometrischer Schwerpunkt für den Quadrant gebildet. Aufgrund der zu erwartenden Dreiecksform sollte sich dieser in der Mitte des Dreiecks befinden. Hier werden die Punkte mit dem jeweils größten Abstand zu dem pro Quadrant spezifischen geometrischen Schwerpunkt gewählt. Das Ergebnis der Eckendetektion für den in Abbil-

Abbildung 6.4 dargestellten Würfel eines gesunden Testteilnehmers liefert das in Abbildung 6.9 aufgeführte Ergebnis. Der Vergleich der ursprünglichen Zeichnungen in Abbildung 6.4 mit den Ergebnissen nach der Eckendetektion für einen gesunden sowie von jeweils einem Testteilnehmer mit MCI und Demenz nach Schritt 2 zeigt die Herausforderungen, die auch mit der Eckendetektion einhergehen. Der rote Punkt stellt den geometrischen Schwerpunkt der gesamten Zeichnung dar, die grünen Punkte die geometrischen Schwerpunkte der Quadranten II und IV.

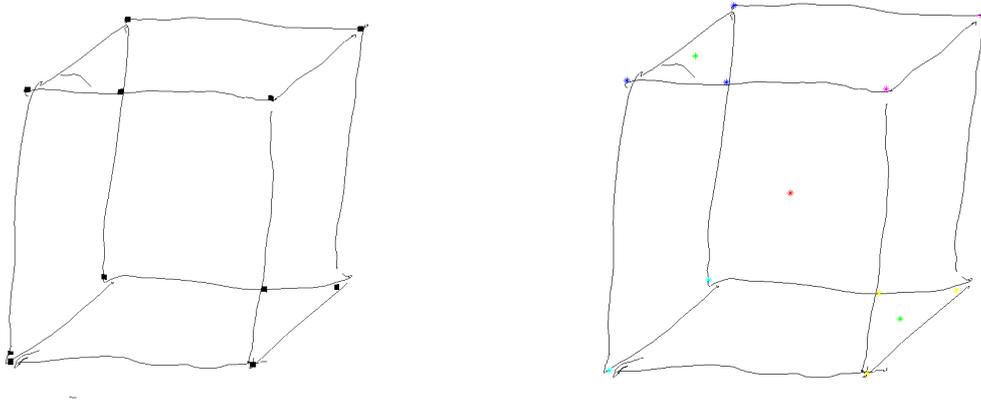


Abbildung 6.9: Ergebnis der Eckendetektion für den in Abbildung 6.4 dargestellten Würfel eines gesunden Testteilnehmers (G) nach Schritt 1 (links) und Schritt 2 (rechts).

Inwiefern sich die Eckendetektion für die unter verschiedenen Graden kognitiver Einschränkungen leidender Menschen verhält deutet Abbildung 6.10 an. Auf Grundlage der ursprünglichen Abbildung liefert das Verfahren für gesunde Testteilnehmer die entsprechenden Ecken an den zu erwartenden Stellen. Für das Beispiel der Teilnehmer mit MCI und Demenz sind nicht alle Ecken vorhanden und dementsprechend auch nicht algorithmisch zu erfassen. Auch hier bietet der Schwellwert des entsprechenden Suchgebietes weitere Optimierungsmöglichkeiten. Da auch menschliche Entscheider einen Interpretationsspielraum besitzen ist eine allumfassende Lösung nicht sicherzustellen, jedoch aufgrund der Heterogenität der möglichen Ergebnisse für solche Verfahren auch zweifelhaft.

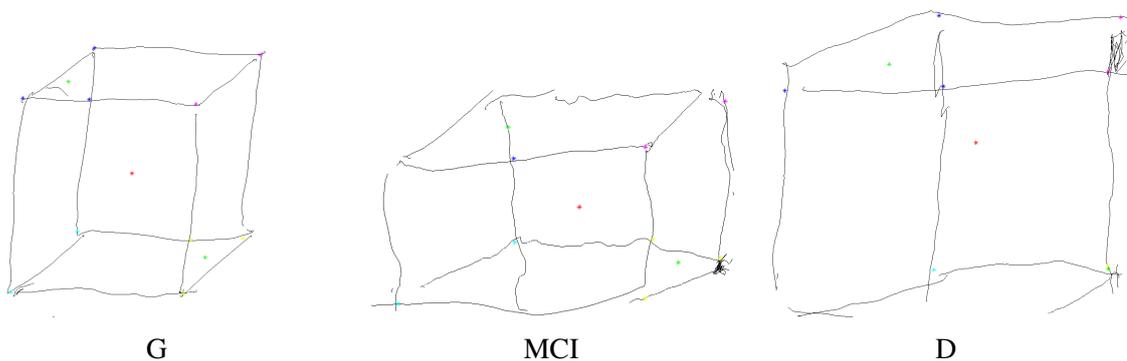


Abbildung 6.10: Ergebnis nach der Eckendetektion für einen gesunden sowie von Testteilnehmern mit MCI und Demenz nach Schritt 2

### 6.1.4 Operationalisierung der Bewertungsmetrik anhand der Eckendetektion

Die in Unterabschnitt 4.2.1 skizzierten Leistungsparameter unterliegen einer standardisierten Bewertungsmetrik, die durch einen menschlichen Testleiter ausgeführt wird. Im Zuge einer autonomen Testdurchführung des Testteilnehmers soll ebenfalls deren automatisierte Evaluation erfolgen können. Nach dem zuvor gezeigt wurde, wie eine Zerlegung des Würfels im Sinne der Leistungsparameter in die einzelnen Komponenten automatisiert erfolgen kann, wird nachfolgend ein Weg aufgezeigt, wie diese Merkmale bei der Interpretation des Ergebnisses eingesetzt werden können. Zunächst wird hierfür gegenübergestellt, welche charakteristischen Merkmale als Grundlage für die Bewertung der Leistungsparameter verwendet werden. Abbildung 6.11 zeigt die Kombination der zuvor skizzierten Ansätze zur Kanten- und Eckendetektion in unterschiedlichen Reihenfolgen. Während die linke Abbildung auf Grundlage einer durchgeführten Kantendetektion die Eckendetektion beeinflusst, wird in der mittleren Abbildung dargestellt, wie auf Grundlage einer zuvor durchgeführten Eckendetektion die Kantendetektion durchgeführt wird. Der Vergleich mit der rechten Abbildung zeigt, dass die Kantendetektion eine Fehlerquelle im Prozess darstellen kann, jedoch zwischen den erkannten Ecken auch die entsprechenden Kanten zu erwarten sind. Nachfolgend wird daher eine Metrik auf Basis der detektierten Ecken entwickelt, welche die charakteristischen Merkmale eines Würfels durch lineare Interpolation zwischen den Ecken bestimmt. Dafür wird sich auf diese Bewertungsvorschrift beschränkt, um die Herausforderungen und Potentiale der Automatisierungsphase gezielt herauszuarbeiten.

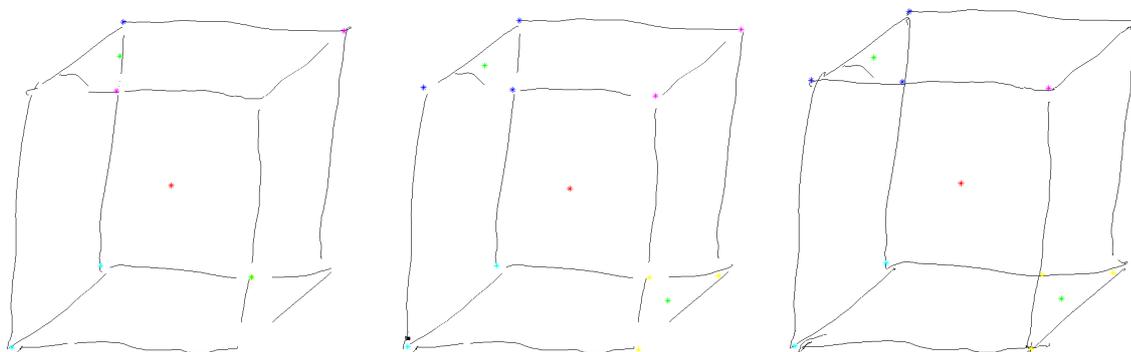


Abbildung 6.11: Ergebnis der Kombinationen aus Kanten- und Eckendetektion für den in Abbildung 6.4 dargestellten Würfel eines gesunden Testteilnehmers (G) in der Reihenfolge Kante-Ecken (links) und Ecken-Kanten (mittig) im Vergleich mit einer alleinigen Eckendetektion (rechts).

Um den gezeichneten Quader auf die Eigenschaft der Dreidimensionalität (PD) zu prüfen und zu bewerten, werden Prüfkriterien abgeleitet, die auch durch einen menschlichen Entscheider bewertet werden könnten. Als charakteristisch für die Darstellung einer dritten Dimension auf einer zweidimensionalen Fläche wird daher die Andeutung der drei Achsen des kartesischen Koordinatensystems angenommen. Diese sollten sich bei einem idealen Würfel an den acht Eckpunkten bei denen drei Kanten aufeinander treffen wiederfinden lassen. Am Beispiel von Abbildung 6.12 zeigt sich für die Ecke  $CD1$ , dass sich an diesen kartesischen Koordinaten  $(x_K, y_K)$  bei einem idealen Würfel die drei Kanten **C**, **D1** und **F** treffen. Die Steigung der Geraden ist gemäß der Zweipunkt-Kalibrierung wie folgt definiert:  $m_K = \frac{\Delta y_K}{\Delta x_K}$ . Zwischen den Punkten  $CD1$  und  $AC$  deutet sich hierbei die x-, zwischen  $CD1$  und  $C1D1$  die y- sowie zwischen  $CD1$  und  $DC1$  die z-Achse an. Sobald diese einen Schwellwert über- oder unterschreiten, wird eine stark verzerrte Darstellung des idealen Würfels angenommen. Demnach sind solche Ausreißer in

einer Metrik zu erfassen, welche sich auf die Bewertung dieses Leistungsparameters auswirkt.

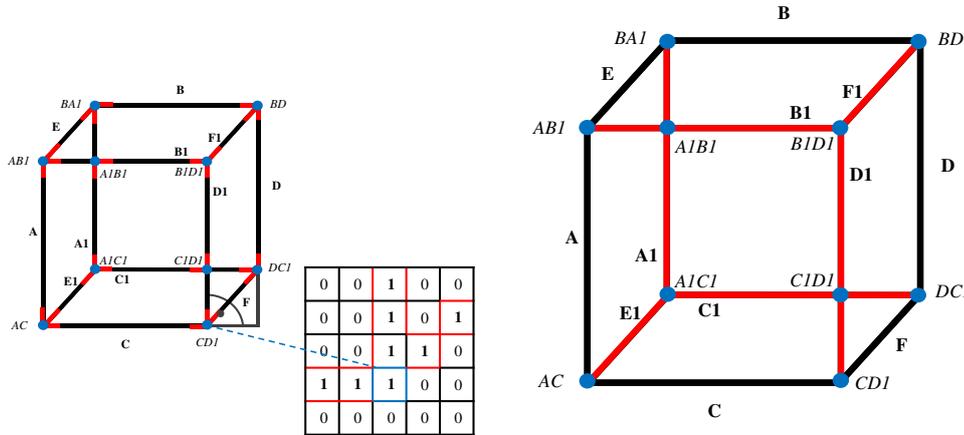


Abbildung 6.12: Charakteristiken der Dreidimensionalität.

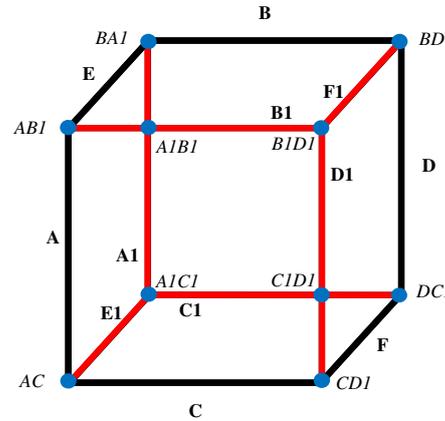


Abbildung 6.13: Charakteristiken der inneren Kanten.

Zur Bewertung des Leistungsparameters, welcher sich auf die korrekte Reproduktion der Frontseite (PF) des idealen Würfels bezieht, spielt die Extraktion der einzelnen Komponenten eine Rolle. Die Frontseite setzt sich demnach aus den vier Kanten  $\{A, B1, C, D\}$  zusammen, die sich an den vorderen Ecken  $\{AB1, A1B1, B1D1, C1D1, CD1, AC\}$  schneiden und somit durch zusätzliche Merkmale charakterisieren lässt (vgl. Abbildung 6.14). Der Leistungsparameter bezüglich der inneren Kanten (PI) prüft demnach das Vorhandensein der Kanten  $\{A1, B1, C1, D1, E1, F1\}$  beziehungsweise den damit verbundenen Ecken  $\{BA1, A1B1, BD, B1D1, AB1, A1C1, C1D1, DC1, AC, CD1\}$  und ob diese entsprechend der definierten Schwellwerte ausgerichtet sind (vgl. Abbildung 6.13).

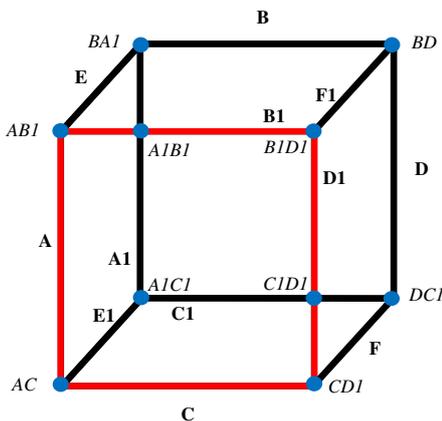


Abbildung 6.14: Charakteristiken der Frontseite.

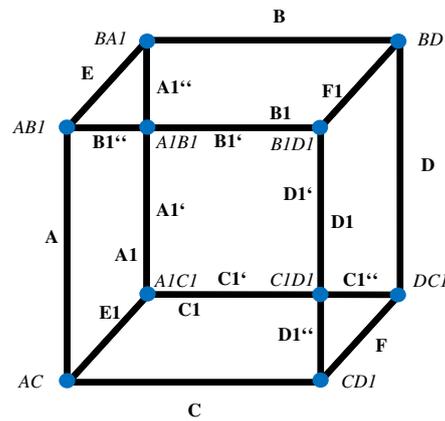


Abbildung 6.15: Charakteristiken der Parallelität.

Der Leistungsparameter zur Bewertung der Parallelität (PP) der Kanten wird durch den Vergleich der Steigungen der entsprechend gleich ausgerichteten Kanten abgebildet. Die Kanten werden hierbei auf Grundlage der erfassten Ecken linear approximiert und müssen folgende Kriterien erfüllen:

- für die waagrechten Kanten  $B, B1', B1'', C, C1', C1''$  gilt  $|m_K| \rightarrow 0$ .
- für die senkrechten Kanten  $A, A1', A1'', D, D1', D1''$  gilt  $|m_K| \rightarrow \infty$ .

- für die räumlich orientierten Kanten **E**, **E1**, **F** und **F1** gilt  $T > |m_K| > 0$ .

Hierbei liegt die Abweichung von diesen Anforderungen ebenfalls in einem zu ermittelnden Toleranzbereich (T).

Da sich die Bewertung des gesamten Tests auf einer Punkteskala von 0 bis 4 bewegt, wird eine binäre Klassifikation für jeden Parameter der vier dargestellten Leistungsparameter erforderlich. Hierfür lässt sich die Erfüllung der jeweiligen Kriterien pro Leistungsparameter zunächst selbst mit einem Punktesystem bewerten. Für PD können hierdurch die maximale Anzahl von acht relevanten Ecken und deren Ausrichtungen identifiziert werden [ $\max(N_{PD}) = 8$ ]. Für die drei weiteren Leistungsparameter gilt gemäß der daran beteiligten Ecken als Komponenten  $\max(N_{PF}) = 6$ ,  $\max(N_{PI}) = 10$  und  $\max(N_{PP}) = 16$ . Die Konsolidierung dieser einzelnen Bewertungsschritte lässt sich auf Grundlage der zuvor skizzierten Bewertungsvorschriften in einem Wert ausdrücken:

$$p2_{Eckendetektion} = \frac{N_{PD}}{8} + \frac{N_{PF}}{6} + \frac{N_{PI}}{10} + \frac{N_{PP}}{16}$$

Der Wert  $p2_{Eckendetektion}$  wird maximal 4. Um Ungenauigkeiten in der Erkennung auszugleichen und menschlichen Interpretationsspielraum zu imitieren, lassen sich die einzelnen Bewertungsparameter aufrunden, wenn sie jeweils zu X% erfüllt sind. Dies schließt einen gewissen Fehler mit ein. Eine Optimierung durch Erweiterung dieser Metrik anhand der Kombination mit zusätzlichen Parametern unter Einbezug der Kantendetektion ist denkbar.

## 6.1.5 Modelloptimierung und -validierung

Das Ziel einer möglichst frühen Erkennung von Demenz im ambulanten Umfeld geht mit einer hohen Sensitivität für ungesunde Personen einher. Eine hohe Sensitivität lässt sich prinzipiell erreichen, indem mehr Menschen als ungesund angenommen werden (vgl. Unterabschnitt 6.2.4). Bei psychischen Erkrankungen könnte bereits die Diagnose stigmatisierend wirken und weitere psychische Leiden hervorrufen oder bereits bestehende verstärken. Wie in Großbritannien geschehen, kann eine zu hohe falsch-positiv Rate auch Einfluss auf die Belastung des Gesundheitswesens durch unnötige Nachuntersuchungen haben (vgl. Unterabschnitt 3.3.5). Eine zu hohe falsch-negativ Rate schließt allerdings zu viele kranke Personen aus, denen man Hilfe zukommen lassen müsste. Deshalb werden zur Gegenüberstellung der Performanz mehrere Maße entlang der Wahrheitsmatrix berechnet, gegenübergestellt und abgewogen (vgl. Unterabschnitt 2.1.4). Durch den Cohen's Kappa lässt sich zusätzlich die Interrater-Reliabilität zwischen menschlichem und algorithmischem Entscheider abbilden (vgl. Tabelle 2.1). Dieser Trade-Off ist folglich das relevante Optimierungskriterium, um die Referenzbewertungen bestmöglich zu approximieren [96, 249].

Die iterative Entwicklung des Verfahrens erfolgt zunächst in der Entwicklungsumgebung *MATLAB*<sup>1</sup>. Daraus lässt sich in *C++* exportierter Code, beispielsweise über eine native Schnittstelle, in mobile Anwendungen integrieren und mit Bildverarbeitungsbibliotheken wie *opencv*<sup>2</sup> verknüpfen. Für die Analyse mit dem zuvor beschriebenen Verfahren stehen 64 mit *DemPredict* digital gezeichnete Würfel mit Referenzbewertung als Datengrundlage zur Verfügung (vgl. Unterabschnitt 5.4.1). Für die Verwendung des Algorithmus, müssen die zuvor benannten Algorithmusparameter definiert werden, um die Qualität des erstellten Modells bewerten zu können (vgl. Abbildung 6.8), wobei  $n_P$  der Anzahl von gezeichneten

<sup>1</sup> <https://de.mathworks.com/>, zuletzt abgerufen am 04.03.2021

<sup>2</sup> <https://opencv.org/>, zuletzt abgerufen am 14.11.2021.

Bildpunkten auf der einheitlich skalierten Zeichenfläche für die gesamte Würfelabbildung entspricht, um die Dimensionen der Zeichnung abzuschätzen:

- Breite des Suchgebiets in x-Richtung für die Eckendetektion:  $S_X = S_{X1} * n_P$
- Breite des Suchgebiets in y-Richtung für die Eckendetektion:  $S_Y = S_{Y1} * n_P$
- Länge des Suchgebiets für die Eckendetektion:  $S_L = S_{L1} * n_P$
- Steigung  $m_K$  für die waagrechten Kanten  $K = \{\mathbf{B}, \mathbf{B1}, \mathbf{C}, \mathbf{C1}\}$ :  $m_{K_{WS}} \leq m_K \leq m_K^{WS}$
- Steigung  $|\frac{1}{m_K}|$  für die senkrechten Kanten  $K = \{\mathbf{A}, \mathbf{A1}, \mathbf{D}, \mathbf{D1}\}$ :  $m_{K_{WS}} \leq \frac{1}{m_K} \leq m_K^{WS}$
- Steigungsverhalten für die räumlichen Kanten  $K = \{\mathbf{E}, \mathbf{E1}, \mathbf{F}, \mathbf{F1}\}$ :  $m_{K_A} < m_K \leq m_K^A$

Für dieses Modell werden die Algorithmusparameter  $\{S_{X1}; S_{Y1}; S_{L1}; m_{K_{WS}}; m_K^{WS}; m_{K_A}; m_K^A\}$  durch eine heuristischen Rastersuche von  $\{0,001; 0,0005; 0,006; -2,0; 0,0; -10,0; 0,0\}$  bis  $\{0,02; 0,001; 0,01; 0,0; 2,0; 0,0; 10,0\}$  optimiert und evaluiert. Eine Gegenüberstellung der Leistungsmaße für die Beurteilung der Qualität des Algorithmus basiert auf Referenzwerten menschlicher Entscheider wobei davon ausgegangen wird, dass diese korrekt gesetzt sind.

### Maximierung des Cohen's Kappa

Anhand der Zielfunktion zur Maximierung des Cohen's Kappa ergibt sich folgende Algorithmusparameterkombination:  $\{0,01997; 0,00092; 0,00994; -0,58481; 1,9380; -9,69061; 9,84803\}$ . Für die in Abbildung 6.10 abgebildeten Würfel liefert der Vergleich zwischen menschlichem und maschinellm Entscheider unterschiedliche Werte für G, MCI und D unter der a priori Annahme von  $X = 75\%$  (vgl. Tabelle 6.1). In Tabelle 6.2 werden die erreichten Werte dargestellt, die auch als Erklärung für die zuvor

Bewertungsparameter	G		MCI		D	
	Mensch	Maschine	Mensch	Maschine	Mensch	Maschine
PD	1	1	1	0 (0,50)	0	0 (0,25)
PF	1	1	1	0 (0,67)	0	0 (0,50)
PI	1	1	0	1 (0,80)	0	0 (0,70)
PP	1	1	1	0 (0,69)	1	0 (0,56)

Tabelle 6.1: Gegenüberstellung zwischen menschlicher und algorithmischer Entscheidung für Abbildung 6.10 mit dem Ziel der Maximierung des Cohen's Kappa bei  $X = 75\%$ .

an den Beispielen gezeigten Ergebnisse liefert. Insbesondere die maschinelle Bewertung der Dreidimensionalität und Parallelität scheint stark von der menschlichen Entscheidung abzuweichen. Dies lässt Raum für weitere Optimierungsbestrebungen.

Bewertungsparameter	Performanz bei X = 75%							
	Genauigkeit				Cohen's Kappa			
	G	MCI	D	gesamt	G	MCI	D	gesamt
PD	0,553	0,643	0,875	0,662	0,000	0,390	0,750	0,303
PF	0,712	0,643	0,9375	0,769	0,205	0,312	0,871	0,449
PI	0,658	0,714	0,563	0,662	0,108	0,336	0,152	0,225
PP	0,684	0,500	0,688	0,677	0,109	0,000	0,310	0,154
	Sensitivität				Spezifität			
	G	MCI	D	gesamt	G	MCI	D	gesamt
PD	0,553	0,545	0,800	0,614	1,000	1,000	1,000	1
PF	0,727	0,636	0,900	0,769	0,600	0,750	1,000	0,786
PI	0,710	0,625	0,714	0,727	0,429	0,714	0,444	0,5
PP	0,706	0,533	0,643	0,683	0,500	1,000	1,000	0,677

Tabelle 6.2: Performanz des Automatisierungsansatzes unter Verwendung der Eckendetektion über alle Diagnosegruppen hinweg mit dem Ziel der Maximierung des Cohen's Kappa.

### Minimierung des mittleren absoluten Fehlers

Dem Prinzip der zuvor durchgeführten Optimierung anhand des Cohen's Kappa folgend, lassen sich die Algorithmenparameter auch durch Minimierung des mittleren absoluten Fehlers über alle Bewertungsparameter hinweg optimieren. Hierbei werden die Algorithmenparameter so bestimmt, dass eine minimale mittlere Abweichung zur Referenzbewertung gesucht wird. Hierdurch werden die Bewertungsparameter aufgeweicht, aber die Annahme von X ausgenommen. Hierbei ergibt sich die gleiche Algorithmenparameterkombination wie in Unterunterabschnitt 6.1.5. Da in diesem Fall der Schwellwert X für die Berechnung der Optimierungsfunktion nicht a priori festgelegt wird, lässt sich dieser variieren. Die besten Ergebnisse für PF und PI ergeben sich bei X = 0,6 und für PP und PD für X = 0,5. Ein Vergleich der Performanz findet sich in Tabelle 6.3.

Bewertungsparameter	Performanz bei X = 60%			
	Genauigkeit	Cohen's Kappa	Sensitivität	Spezifität
PD	0,584	0,233	0,526	1
PF	0,662	0,209	0,692	0,571
PI	0,708	0,329	0,818	0,5
PP	0,708	0,180	0,717	0,667
Performanz bei X = 50%				
PD	0,662	0,303	0,614	1
PF	0,692	0,179	0,769	0,429
PI	0,692	0,250	0,864	0,364
PP	0,738	0,211	0,750	0,677

Tabelle 6.3: Performanz des Automatisierungsansatzes unter Verwendung der Eckendetektion über alle Diagnosegruppen hinweg mit dem Ziel der Minimierung des mittleren absoluten Fehlers bei X = 60% und X = 50%.

## Modellvalidierung

Die zuvor dargestellten Ergebnisse zeigen die Implikationen der unterschiedlichen Strategien auf. Während einerseits durch die Maximierung des Cohen's Kappa der X-Wert a priori möglichst hoch gewählt werden muss (vgl. Unterunterabschnitt 6.1.5), lässt sich durch die Minimierung des mittleren absoluten Fehlers dieser Schwellwert auf Kosten der individuellen Betrachtung der Bewertungsparameter anschließend so bestimmen, dass die zu optimierende Qualitätsmetrik maximiert wird (vgl. Unterunterabschnitt 6.1.5). Um diese Verfahren zu validieren und entsprechendes Overfitting zu vermeiden, muss ebenfalls ein geeignetes Validierungsverfahren verwendet werden. Hierfür findet die 5-fache Kreuzvalidierung entsprechend der in Unterabschnitt 6.3.2 beschriebenen Aufteilung der Test- und Trainingsdaten am Beispiel der Maximierung des Cohen's Kappa statt, welche folgende Algorithmenparameterkombination bestimmt: {0,01997; 0,00092; 0,00994; -0,81905; 1,09299; -6,77877; 10,0}. Als Ergebnis auf die gesamten Daten angewandt liefern diese Parameter die Performanz aus der nachfolgenden Tabelle 6.4.

Bewertungsparameter	Performanz bei X = 75%							
	Genauigkeit				Cohen's Kappa			
	<i>G</i>	<i>MCI</i>	<i>D</i>	<i>gesamt</i>	<i>G</i>	<i>MCI</i>	<i>D</i>	<i>gesamt</i>
PD	0,553	0,643	0,875	0,662	0,000	0,390	0,750	0,303
PF	0,658	0,643	0,938	0,738	0,151	0,312	0,871	0,401
PI	0,658	0,714	0,625	0,677	0,108	0,336	0,262	0,267
PP	0,711	0,429	0,688	0,677	0,229	0,000	0,310	0,206
	Sensitivität				Spezifität			
	<i>G</i>	<i>MCI</i>	<i>D</i>	<i>gesamt</i>	<i>G</i>	<i>MCI</i>	<i>D</i>	<i>gesamt</i>
PD	0,553	0,545	0,800	0,614	1,000	1,000	1,000	1,000
PF	0,667	0,636	0,900	0,731	0,600	0,750	1,000	0,786
PI	0,710	0,625	0,714	0,727	0,429	0,714	0,556	0,545
PP	0,710	0,467	0,643	0,667	0,750	1,000	1,000	0,833

Tabelle 6.4: Performanz des Automatisierungsansatzes unter Verwendung der Eckendetektion über alle Diagnosegruppen hinweg mit dem Ziel der Maximierung des Cohen's Kappa und 5-fach kreuzvalidierten Algorithmenparametern.

## 6.2 Entwurf einer Bewertungsmetrik

In der medizinischen Praxis werden Bewertungssysteme in Form von Scoring-Systemen eingesetzt, welche durch die Operationalisierung medizinischer Zusammenhänge eine zielgruppengerechte Unterstützung (DNFAT1) zur informierten Entscheidungsfindung darstellen (vgl. Unterabschnitt 2.2.5). Hierfür werden für gemessene Parameter bestimmte Schwellwerte als sogenannte Cut-off-Scores definiert, welche die Zuordnung der Ergebnisse zu Gesundheitszuständen ermöglichen. Deren Bestimmung kann dabei wie in Unterabschnitt 2.2.5 beschrieben mittels ROC-Analysen erfolgen (vgl. Absatz 2.1.4). Damit lassen sich Schwellwerte für binäre Schätzer finden. Für *DemPredict* bedeutet dies eine Unterteilung der gesammelten Daten in die Klassenpaare G-MCI, MCI-D und G-D oder G und nicht gesund (NG), wobei NG die Vereinigungsmenge von MCI und D darstellt. Dabei ist zu beachten, dass manche Merkmale umgekehrt skaliert sind, beispielsweise indiziert eine erhöhte Bearbeitungsdauer einen verschlechterten

kognitiven Zustand, eine größere Anzahl an richtigen Antworten stellt hingegen eine bessere kognitive Leistungsfähigkeit dar. Des Weiteren ist die Bestimmung von Schwellwerten auf Ebene der Merkmale und auf Ebene der damit getesteten kognitiven Domäne möglich. Da sich die gemessenen Werte auf Bearbeitungszeiten und Fehlerquoten beziehen, die auch in Abhängigkeit zu Eigenschaften des eingesetzten digitalen Mediums wie beispielsweise der Größe des Bildschirms stehen, ist theoretisch für jede Ausprägung eine erneute Normierung dieser notwendig. Nachfolgend werden auf der in Unterabschnitt 5.4.1 beschriebenen Datengrundlage verschiedene Varianten entworfen und gegenübergestellt.

## 6.2.1 Merkmalsselektion

Um einer Überanpassung von Klassifikationsmodellen entgegenzuwirken, kann deren Leistung positiv beeinflusst werden, wenn nur ein bestimmter Teil der verfügbaren Merkmale in die Modellbildung mit einbezogen wird (vgl. Unterabschnitt 2.4.3). An dieser Stelle sei darauf hingewiesen, dass ein Merkmalsselektionsverfahren einer leitlinienkonformen Diagnose, wie beispielsweise der DSM-V grundsätzlich widersprechen kann. Nach diesen Kriterien sollten alle relevanten kognitiven Domänen untersucht werden [195]. Die Merkmalsauswahl führt möglicherweise zu einer Verletzung dieser Anforderung, da sie zu einer Entfernung von Merkmalen führt, die eine kognitive Domäne abdecken (siehe Tabelle 4.2). Damit lassen sich jedoch möglichst häufig oder früh betroffene kognitive Eigenschaften stärker herausstellen, was zumindest zur Früherkennung mittels Screening genutzt werden kann. Welchen Einfluss diese verschiedenen Ansätze auf die Performanz von *DemPredict* haben wird nachfolgend empirisch bestimmt, indem die Leistung der daraus resultierenden Diagnosemodelle gegenübergestellt wird. Um diesen Effekt der Parameterselektion zu untersuchen, werden im Folgenden auch Modelle ohne diese Vorauswahl verglichen. Eine unabhängig von dem zu verwendenden Modell einsetzbare Methode ist die FFS. Hierbei werden die aus einer statistischen Analyse gewonnenen Erkenntnisse zur Auswahl geeigneter Merkmale verwendet. Als entscheidendes Kriterium dafür gilt am Beispiel von *DemPredict* die Diskriminationsfähigkeit (vgl. Unterabschnitt 5.4.4). Dabei wird berücksichtigt, ob ein starker Einfluss der Erfahrung im Umgang mit Technik oder der Vergleich mit den Ergebnissen aus der CERAD-Testbatterie gegen eine Aufnahme des Merkmals sprechen würde. Als Ergebnis dieses Selektionsprozesses lassen sich neben dem Alter der Nutzer die in Tabelle 6.5 zusammengefassten Merkmale identifizieren, Ergebnisse mit  $p < 0,05$  deuten auf eine statistisch signifikante Unterscheidbarkeit hin.

Test	Parameter	KWT <i>p</i> -Wert	WRT <i>p</i> -Wert (G-MCI)	WRT <i>p</i> -Wert (MCI-D)	WRT <i>p</i> -Wert (G-D)
A	p41	<b>0,01</b>	0,804	<b>0,041</b>	<b>0,041</b>
C	p47	<b>0,001</b>	<b>0,003</b>	<b>0,044</b>	<b>0,015</b>
D	p49	<b>0,009</b>	<b>0,012</b>	0,8	0,166
E	p10	<b>0,002</b>	<b>0,028</b>	0,096	<b>0,008</b>
	p11	<b>0,002</b>	0,206	0,206	<b>0,002</b>
	p13	<b>0,002</b>	0,23	<b>0,025</b>	<b>0,002</b>
	p15	<b>0,003</b>	<b>0,036</b>	0,185	<b>0,014</b>
	p20	<b>0,000</b>	<b>0,005</b>	0,328	<b>0,005</b>
F	p5	<b>0</b>	<b>0,027</b>	<b>0,027</b>	<b>0,001</b>

	p9	<b>0</b>	0,517	<b>0,001</b>	<b>0</b>
<i>I</i>	p42	<b>0</b>	0,139	<b>0,004</b>	<b>0</b>
<i>J</i>	p2	<b>0,002</b>	<b>0,033</b>	0,133	<b>0,006</b>
<i>M</i>	p54	<b>0,002</b>	0,073	0,101	<b>0,004</b>

Tabelle 6.5: Parameterselektion anhand von FFS (vgl. Abschnitt A.16).

Da die EFS verfahrensimmanent und die WFS auf einer heuristischen Exhaustionsmethode beruht, sind die Ergebnisse dieser Selektionsverfahren Teil der späteren Optimierung der Klassifikationsergebnisse (vgl. Abschnitt 6.3).

## 6.2.2 Merkmalspezifische Schwellwerte

Eine Möglichkeit der Operationalisierung stellt die Bestimmung eines Schwellwerts pro Merkmal dar. Hierbei stellt jedes Merkmal eine individuelle Diagnose, das heißt eine Einordnung in die Gruppe G, MCI oder D. Die Gewichtung der einzelnen Merkmale ist demnach gleich. Eine Möglichkeit daraus eine Diagnose abzuleiten ist es, die Summe über die Anzahl der verschiedenen Gruppenzuordnungen zu bilden und mit anschließendem Mehrheitsvotum, die häufigste Gruppe zu wählen. Die resultierenden Schwellwerte für diesen Ansatz sind auf Grundlage der gesammelten Daten ohne Merkmalsselektion in Tabelle A.9 sichtbar. Die Einflüsse verschiedener Merkmalsselektionen auf die Performanz des auf einem Mehrheitsvotum basierenden Ansatzes wird in Unterabschnitt 6.2.4 evaluiert. Eine Erweiterung dieses Ansatzes zur Vereinfachung der Interpretierbarkeit stellt die Zuordnung von Punkten pro Merkmal und Diagnosegruppe dar. Werden beispielsweise jeweils für ein Merkmal, welches zur D-Diagnose führt zwei Punkte, für eine MCI-Diagnose ein Punkt und für jede G-Diagnose null Punkte vergeben, lassen sich erneut Schwellwerte mittels ROC-Kurvenanalyse bestimmen, die eine Diagnose-Entscheidung vereinfachen. In Abbildung 6.16a ist diese Schema dargestellt. Die Hinzu- und Wegnahme von Merkmalen durch die FFS (vgl. Abbildung 6.16b) sowie die Schnittmenge mit den CERAD-Merkmalen (vgl. Abbildung 6.16c) aus Unterabschnitt 5.4.2 hat einen Einfluss auf die Punkteskala.

Punkte	Diagnose	Punkte	Diagnose	Punkte	Diagnose
$\leq 93$	G	$\leq 17$	G	$\leq 15$	G
94 – 126	MCI	18 – 28	MCI	16 – 22	MCI
$\geq 127$	D	$\geq 25$	D	$\geq 23$	D
(a) Ohne Merkmalsselektion.		(b) Mit FFS.		(c) Für CERAD-Merkmale.	

Abbildung 6.16: Merkmalspezifisches Scoring nach Punkteschema.

## 6.2.3 Auf kognitiven Domänen basierte Schwellwerte

In Tabelle 4.2 und Abbildung 4.6 werden alle Tests und damit erhobenen Parameter auf die zu prüfenden Domänen abgebildet. Hierfür werden die Parameter z-transformiert (vgl. Gleichung 2.1) und entsprechend

einheitlich skaliert, sodass der Wert verschiedener Merkmale miteinander kombiniert werden kann. Für jede kognitive Domäne wird so ein Diagnosewert bestimmt, der diesen zu testenden Merkmalen entspricht. Die Gewichtung der einzelnen Merkmale steht demnach in Abhängigkeit zu den für diese Domäne zu testenden Parameter. Falls mehrere Tests auf eine kognitive Domäne hin testen, wird auf diese Weise eine Mittelwertbildung unter Ignoranz fehlender Werte durchgeführt, um Testergebnisse, bei denen keine oder keine korrekte Testausführung erfolgt ist, auszumitteln. Auch zur Auswertung der domänenspezifischen Diagnosewerte lassen sich mittels ROC-Kurvenanalyse Schwellwerte bilden, sodass auf Grundlage jeder Domäne eine potentielle Diagnose gestellt werden kann. Die Kombination dieser einzelnen Testurteile kann auf verschiedene Weisen geschehen. Analog zu den merkmalspezifischen Schwellwerten lässt sich auch domänenspezifisch ein Mehrheitsvotum realisieren sowie das im Abschnitt zuvor eingeführte Punktesystem mit angepassten Schwellwerten verwenden. Auch hierbei findet eine Beeinflussung durch die Implementierung verschiedener Merkmalsselektionsverfahren statt. Die Vergabe der Punkte erfolgt wie für die merkmalspezifischen Schwellwerte. In Abbildung 6.17 ist dieses Schema dargestellt. Zusätzlich

Punkte	Diagnose	Punkte	Diagnose	Punkte	Diagnose
$\leq 13$	G	$\leq 8$	G	$\leq 2$	G
14 – 16	MCI	9 – 11	MCI	3 – 4	MCI
$\geq 17$	D	$\geq 12$	D	$\geq 5$	D
(a) Ohne Merkmalsselektion.		(b) Mit FFS.		(c) Für CERAD-Merkmale.	

Abbildung 6.17: Auf kognitiven Domänen basiertes Scoring nach Punkteschema.

ergibt sich die Möglichkeit aufgrund der Anzahl „auffälliger Domänen“ eine Diagnose abzuleiten. Als auffällig lassen sich alle Domänen bewerten, die nach Auswertung der jeweiligen Schwellwerte eine MCI- beziehungsweise D-Diagnose indizieren. Die Summe aller auffälligen Diagnosen mit anschließender Schwellwert-Bestimmung mittels ROC-Analyse definiert die Gesamtdiagnose. Die Schwellwerte der Summe über die Anzahl an MCI- und D-Diagnosen zur Einschätzung des Gesundheitszustandes kann nach den in Abbildung 6.18 dargestellten Schemata abgebildet werden. Als vierte Alternative wird, um den

Punkte	Diagnose	Punkte	Diagnose	Punkte	Diagnose
$\leq 3$	G	$\leq 2$	G	$\leq 1$	G
4	MCI	3	MCI	2	MCI
$\geq 5$	D	$\geq 4$	D	$\geq 3$	D
(a) Ohne Merkmalsselektion.		(b) Mit FFS.		(c) Für CERAD-Merkmale.	

Abbildung 6.18: Auf kognitiven Domänen basiertes Scoring nach Anzahl auffälliger Domänen.

Vorgaben der DSM-V gerecht zu werden, für die Diagnose einer AD das Vorliegen von Einschränkungen im Bereich Lernen und Gedächtnis sowie mindestens einer weiteren kognitiven Domäne vorausgesetzt, sodass die Alltagsfähigkeit eingeschränkt ist (vgl. Unterunterabschnitt 2.2.3). Hierfür lässt sich für jede kognitive Domäne eine Einzeldiagnose präzisieren und entsprechend der zuvor beschriebenen Logik ein Gesamturteil bilden.

## 6.2.4 Gegenüberstellung der Performanz

Wie zuvor beschrieben, lassen sich sogenannte Cut-off-Scores für jeweils unterschiedliche Ausprägungen der Performanz bestimmen. Um eine hohe Sensitivität für ungesunde Personen zu erreichen, lassen sich die Diagnosegruppen zusammenfassend auch in G und NG unterteilen. Die Berechnung erfolgt für alle Diagnosegruppen einzeln, den Durchschnitt der Gruppen, sowie für die Diagnosegruppe NG. Für die binäre Unterscheidung zwischen G oder NG gilt, mit Ausnahme des Ansatzes nach DSM-V, dass die Klassen G, MCI und D einzeln betrachtet und nur die Auswertung entsprechend zusammengefasst wird. Da sich die Performanz der unterschiedlichen Ansätze entsprechend unterscheidet, stellt Tabelle 6.15 eine Übersicht der jeweils erreichten Leistungsfähigkeit ohne Merkmalsselektion dar. Nach Anwendung der FFS resultieren die Werte in Tabelle 6.16. Für die Selektion anhand der äquivalenten CERAD-Merkmale wird die in Tabelle 6.17 dargestellten Leistungsfähigkeit erreicht.

Metrik	Merkmalspezifisches Schema		Domänenspezifisches Schema			DSM-V
	Mehrheitsvotum	Schwellwerte mit Punkteschema	Mehrheitsvotum	Schwellwerte mit Punkteschema	Schwellwerte Anzahl betroffener Domänen	
Genauigkeit	0,76	0,71	0,73	0,65	0,65	<b>0,84</b>
Cohen's Kappa	0,61	0,55	0,57	0,46	0,46	<b>0,67</b>
Sensitivität NG	0,66	<b>0,90</b>	0,83	0,83	<b>0,90</b>	0,86
Sensitivität D	<b>0,86</b>	0,64	0,79	0,57	0,36	-
Sensitivität MCI	0,33	0,60	0,53	0,53	<b>0,67</b>	-
Spezifität NG	<b>0,96</b>	0,81	0,81	0,77	0,81	0,81
Spezifität D	<b>0,93</b>	0,90	0,78	0,90	0,90	-
Spezifität MCI	<b>1,00</b>	0,78	0,98	0,75	0,70	-

Tabelle 6.15: Performanz der verschiedenen Scoring-Systeme ohne Merkmalsselektion.

Metrik	Merkmalspezifisches Schema		Domänenspezifisches Schema			DSM-V
	Mehrheitsvotum	Schwellwerte mit Punkteschema	Mehrheitsvotum	Schwellwerte mit Punkteschema	Schwellwerte Anzahl betroffener Domänen	
Genauigkeit	0,78	0,73	0,71	0,69	0,60	<b>0,84</b>
Cohen's Kappa	0,65	0,59	0,54	0,52	0,36	<b>0,67</b>
Sensitivität NG	0,79	<b>0,93</b>	0,83	0,83	0,83	0,86
Sensitivität D	<b>0,79</b>	0,71	<b>0,79</b>	0,50	0,00	-
Sensitivität MCI	0,53	<b>0,80</b>	0,47	0,73	0,73	-
Spezifität NG	<b>0,92</b>	0,69	0,81	0,77	0,85	0,81
Spezifität D	0,93	0,98	0,85	<b>1,00</b>	<b>1,00</b>	-
Spezifität MCI	<b>0,93</b>	0,70	0,88	0,70	0,58	-

Tabelle 6.16: Performanz der verschiedenen Scoring-Systeme mit FFS.

Metrik	Merkmalspezifisches Schema		Domänenspezifisches Schema		
	Mehrheitsvotum	Schwellwerte mit Punkteschema	Mehrheitsvotum	Schwellwerte mit Punkteschema	Schwellwerte Anzahl betroffener Domänen
Genauigkeit	<b>0,75</b>	0,65	0,58	0,51	0,24
Cohen's Kappa	<b>0,58</b>	0,47	0,34	0,27	-0,03
Sensitivität NG	0,66	0,83	0,69	0,86	<b>1,00</b>
Sensitivität D	0,79	0,50	0,64	0,50	<b>0,93</b>
Sensitivität MCI	0,40	<b>0,80</b>	0,27	0,53	0,00
Spezifität NG	<b>0,92</b>	0,65	0,73	0,50	0,00
Spezifität D	0,90	<b>1,00</b>	0,83	0,83	0,07
Spezifität MCI	<b>1,00</b>	0,65	0,83	0,60	0,90

Tabelle 6.17: Performanz der verschiedenen Scoring-Systeme für CERAD-Merkmale.

Über die unterschiedlichen Konzepte zur Auswertung hinweg, stellt das merkmalspezifische Mehrheitsvotum eine Alternative mit einer hohen Genauigkeit dar, wie auch durch den Kappa-Wert anhand einer entsprechend stark zu bewertenden Interrater-Reliabilität angezeigt wird. Lediglich die domänenspezifische Diagnose anhand der DSM-V liefert eine höhere Übereinstimmung, hierbei konnte jedoch aufgrund der oben beschriebenen Entscheidungslogik nur die binäre Entscheidung zwischen G und NG erfolgen, was die Vergleichbarkeit einschränkt. Da in *DemPredict* keine der CERAD-Merkmale enthalten sind, welche die Domäne Gedächtnis testen, kann hierfür dieser Merkmalsselektionsansatz nicht ausgewertet werden. Die Selektion anhand der CERAD-Merkmale liefert keine Verbesserung der Leistung (vgl. Tabelle 6.17). Wie das Beispiel des Mehrheitsvotums ohne Merkmalsselektion zeigt, kann eine Spezifität bei der Klasse NG von 96% erreicht werden, während die Sensitivität bei lediglich 66% liegt (vgl. Tabelle 6.15). Um möglichst viele NG richtigerweise zu erfassen, stellt also diese Art des Bewertungsschemas nicht die beste Option dar. Mit Hilfe der FFS lässt sich eine Sensitivität von 72% bei einer Spezifität von 92% erreichen, was die optimale Entscheidung für das Ziel einer hohen Spezifität darstellt (vgl. Tabelle 6.16). Mit 90%-iger Sensitivität stellt das Bewertungsschema der Kategorie „Schwellwerte mit Punkte-System“ ohne Merkmalsselektion eine wesentlich bessere Alternative hinsichtlich dem Ziel eines möglichst ausgeglichenen Modells dar, die Spezifität liegt hier bei 81% (vgl. Tabelle 6.15). Durch eine FFS lässt sich die Sensitivität auf 93% maximieren, jedoch auf Kosten der Spezifität, welche nur noch auf 69% zu beziffern ist (vgl. Tabelle 6.16).

Diese Beispiele verdeutlichen den Einfluss dieser Abwägungsentscheidungen, welche letztlich vom Ziel und somit dem gesundheitsökonomischen und -politischen Nutzen der digitalen Gesundheitsanwendung sowie ethischen Fragestellungen abhängen.

### 6.3 Optimierung der Klassifikationsergebnisse mittels Modellen des Maschinellen Lernens

Wie in Unterabschnitt 5.4.4 dargestellt, ist eine optimale Diskriminationsfähigkeit zwischen den Diagnosegruppen erreicht, wenn sich die Leistungsdaten der Gruppen nicht überschneiden. Da sich dieser Umstand, wie am Beispiel der Datengrundlage aus Unterabschnitt 5.4.1 ersichtlich wird, nicht für jede

psychometrische Variable realisieren lässt, werden über die unterschiedlichen Leistungsparameter hinweg Bewertungsmetriken angewandt (vgl. Abschnitt 6.2). Nachteilig bei solchen Verfahren ist, dass sich Leistungsunterschiede in unterschiedlichen Domänen gegenseitig aufwiegen können, falls diese miteinander kombiniert werden und die Gewichtung der Merkmale entsprechend vordefiniert ist. Hier soll die Anwendung von ML einen Mehrwert leisten, da implizite Gewichtungen einzelner Parameter durch unterschiedliche Ansätze gelernt und bei der Klassifikation berücksichtigt werden können (vgl. Unterabschnitt 3.3.3). Dieser Abschnitt soll einen Einblick über die Möglichkeit der Anwendung in dieser Domäne geben und wurde in Teilen bereits veröffentlicht [471]. Die Wahl der Algorithmen, Modelle und Maße bildet den Ausgangspunkt (vgl. Unterabschnitt 6.3.1). Dieser entwickelt sich von der Selektion geeigneter Merkmale bis hin zur Vergleichsimplementierung und Validierung verschiedener Verfahren (vgl. Unterabschnitt 6.3.2).

### 6.3.1 Vorgehen und Merkmalsselektion

Überwachte Lernansätze kommen im Gesundheitswesen am häufigsten zum Einsatz [212]. Hierzu zählen unter anderen Entscheidungsbäume, damit verwandte RF-Klassifikatoren, SVM, Naïve-Bayes-Klassifikatoren, der  $k$ -NN und ANN (vgl. Unterabschnitt 2.4.3). Da jedem Studienteilnehmer des in Unterabschnitt 5.4.1 beschriebenen Datensatzes eine der drei möglichen Diagnosegruppen G, MCI oder D von menschlichen Experten zugewiesen ist, kann diese Einordnung als Zielwert für diese Modelle verwendet werden. Fehlende Werte werden durch die diagnosespezifischen Mittelwerte der jeweiligen Leistungsparameter ersetzt. Zusätzlich wird eine Standardisierung der Daten mittels  $z$ -Transformation durchgeführt (vgl. Unterabschnitt 5.4.1).

Um einen Eindruck zu erhalten, was bei der Auswahl und beim Einsatz dieser Verfahren zu beachten ist sowie welche Herausforderungen und Potenziale bezüglich der Anwendungsdomäne bestehen, werden diese Modelle nachfolgend systematisch untersucht und gegenübergestellt. Hierzu werden die einzelnen Modelle auf dem gleichen Datensatz trainiert und evaluiert. Um ihre Leistung vergleichen zu können, muss davor festgelegt werden, welches Leistungsmaß durch das trainierte Modell optimiert werden soll. Je nach Anwendungsfall variieren die relevanten Maße und können zur Wahl unterschiedlicher Modelle führen (vgl. Unterabschnitt 2.4.3). Ein Bewertungskriterium für die Güte der Klassifikation stellt deren Treffergenauigkeit dar. Um bei unbalancierten Trainingsdaten einer Verzerrung des Modells durch zufällige Treffer entgegenzuwirken, wird der Kappa-Wert als zu maximierendes Zielkriterium festgelegt [96, 249]. Die Umsetzung erfolgt in  $R^3$ , primär wird das Paket Classification And REgression Training (CARET) angewandt [242, 249, 404]. Hinsichtlich der Anwendung von ML fallen die Datenmengen des zur Verfügung stehenden Datensatzes relativ gering aus, daher steht nachfolgend ebenfalls das „Wie“ nicht das „Was“ im Vordergrund.

Als Evaluationsverfahren wird eine  $k$ -mal wiederholte  $k$ -fache Kreuzvalidierung gewählt. Im Vergleich zwischen den in der Literatur etablierten 5- und 10-fachen Kreuzvalidierung, kann die Aufteilung der limitierten Gesamtmenge von Daten in zehn Untermengen zu einer deutlich größeren Varianz in der Performanz zwischen den Untermengen führen und somit potenziell das Ergebnis verfälschen (vgl. Unterabschnitt 2.4.3). Daher wird eine wiederholte 5-fache Kreuzvalidierung durchgeführt, bei der die Untermengen geschichtet werden. Sie verspricht eine robuste und zuverlässige Bewertung der Leistung auf unbekanntem Daten, obwohl in jeder Trainingsiteration eine vergleichsweise große Anzahl von Datenpunkten zum Training verwendet werden kann [242]. Dies ist aufgrund der begrenzten Datenmenge essentiell für das Training aussagekräftiger Modelle. Die insgesamt 55 Datenpunkte der Studienteilneh-

<sup>3</sup> <https://www.r-project.org>, zuletzt abgerufen am 10.03.2021.

mer werden hierzu in fünf Untermengen mit je elf Datenpunkten unterteilt. Das Training erfolgt auf vier der fünf Untermengen, also 44 pro Trainingsdatensatz, die Auswertung auf der fünften Untermenge. Unabhängig von der Evaluation muss auch ein Verfahren festgelegt werden, womit das Training der Modelle, das sogenannte Parameter-Tuning durchgeführt wird. Um möglichst viele Datenpunkte für das Training der Hyperparameter zu verwenden, wird eine Leave-One-Out-Kreuzvalidierung (LOOCV) für diesen Zweck eingesetzt. Anders als bei der Evaluation werden dabei keine Daten zurückgehalten, um die Leistung der Parameter mit geringerer Varianz zu bewerten. Eine LOOCV erzeugt zwar durch die höhere Anzahl an Iterationen mehr Aufwand beim Training, nimmt aber für das alleinige Ziel die besten Parameter zu finden eine möglicherweise erhöhte Varianz der Ergebnisse im Trainingsprozess in Kauf (vgl. Unterunterabschnitt 2.4.3).

Um die am besten geeigneten Merkmale auszuwählen, werden Verfahren aus allen drei Gruppen der Merkmalsselektionsverfahren EFS, FFS und WFS getestet sowie die erzielten Leistungen verglichen [242]. Die erste Methode stellt hierbei die EFS dar. Wenn ein Algorithmus keine solche eingebettete Parameterauswahl enthält, wird in dieser Iteration der Modellbildung keine Merkmalsauswahl angewandt. Aus der Gruppe der WFS wird eine RFE und eine heuristische Merkmalsauswahl mittels Greedy-Search implementiert, welche nachfolgend mit WFS abgekürzt wird. Diese wird, wie in Abschnitt 2.4.3 beschrieben, umgesetzt. Die Leistung der resultierenden Modelle wird nach dem Trainingsprozess verglichen. Die jeweils besten Modelle ergeben sich durch die Selektion der in Abbildung 6.19 dargestellten Merkmale und werden nachfolgend validiert.

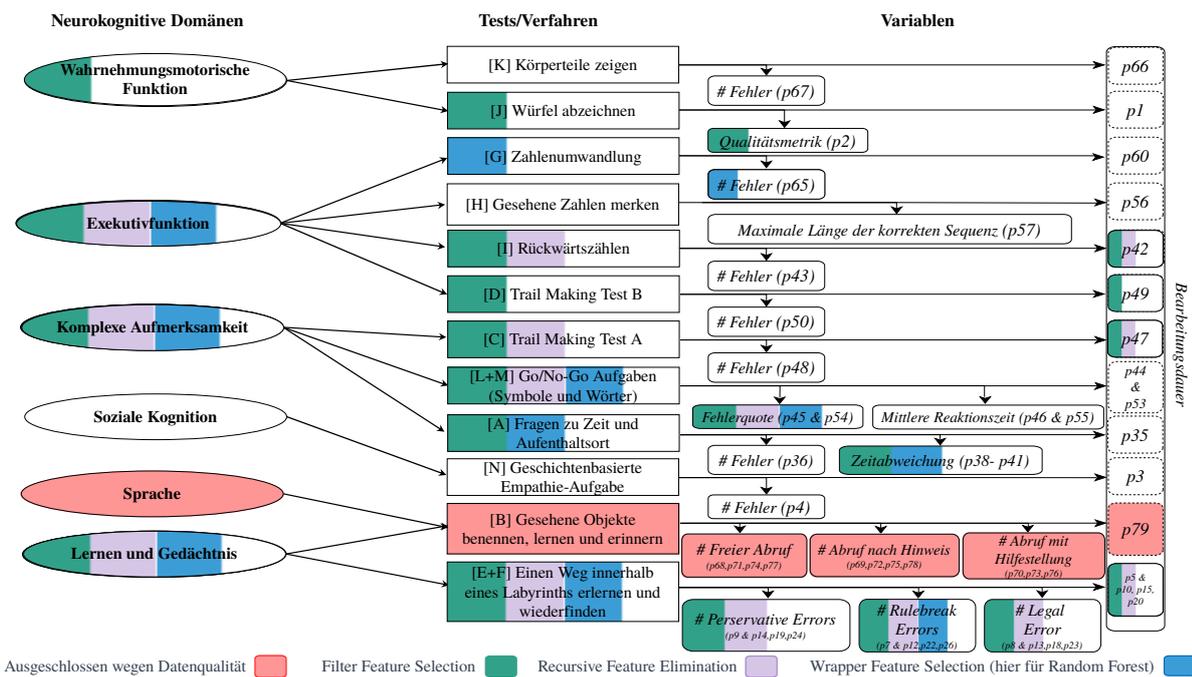


Abbildung 6.19: Modell der gefilterten latenten Variablen (links) und der Tests (Mitte) sowie relevante Leistungskriterien (rechts) nach Schinle et al., 2021 [471].

### 6.3.2 Modellvalidierung

Die Phase der Modellbewertung hat zum Ziel Aussagen über die Qualität der erstellten Modelle treffen zu können (vgl. Abbildung 4.3). Zur Validierung der Modelle wird eine fünfmal wiederholte 5-fache Kreuzvalidierung durchgeführt. Wie zuvor erörtert wird der Cohen's Kappa für das Training der Modelle als

Metrik	Decision Tree			SVM mit			Naïve Bayes	k-NN	ANN
	CART	C5.0	RF	linear kernel	RBF kernel	polynomial kernel			
Genauigkeit	0,64	0,71	<b>0,77</b>	0,63	0,73	0,72	0,76	0,70	0,63
Cohen's Kappa	0,40	0,54	<b>0,63</b>	0,41	0,58	0,55	0,61	0,50	0,42
Sensitivität NG	0,60	0,77	0,77	0,72	0,77	0,80	<b>0,82</b>	0,63	0,74
Sensitivität D	0,48	0,70	0,71	0,73	<b>0,81</b>	0,70	0,69	0,69	0,65
Sensitivität MCI	0,47	0,63	0,61	0,32	0,49	0,60	<b>0,68</b>	0,32	0,37
Spezifität NG	0,83	0,77	0,90	0,77	0,84	0,80	0,84	<b>0,93</b>	0,78
Spezifität D	0,91	0,93	0,96	0,90	0,90	0,96	0,95	<b>0,97</b>	0,89
Spezifität MCI	0,89	0,84	0,89	0,80	<b>0,91</b>	0,81	0,85	0,90	0,80

Tabelle 6.18: Durchschnittliche Modell-Leistungen ohne Merkmalsselektion erweitert nach Schinle et al., 2021 [471].

Metrik	Decision Tree			SVM mit			Naïve Bayes	k-NN	ANN
	CART	C5.0	RF	linear kernel	RBF kernel	polynomial kernel			
Genauigkeit	0,65	0,69	0,76	<b>0,77</b>	0,71	0,73	0,74	0,73	0,73
Cohen's Kappa	0,41	0,52	0,61	<b>0,64</b>	0,54	0,58	0,58	0,56	0,56
Sensitivität NG	0,57	0,76	0,75	0,79	0,73	0,76	<b>0,78</b>	0,72	0,75
Sensitivität D	0,48	0,77	0,72	<b>0,85</b>	0,84	0,78	0,73	0,69	0,81
Sensitivität MCI	0,43	0,51	0,57	<b>0,63</b>	0,48	0,56	0,53	0,51	0,51
Spezifität NG	0,87	0,77	<b>0,90</b>	0,82	0,77	0,81	0,87	0,88	0,81
Spezifität D	0,92	0,89	0,95	0,95	0,89	0,96	0,95	<b>0,98</b>	0,94
Spezifität MCI	0,90	0,86	<b>0,91</b>	0,89	0,91	0,85	0,86	0,86	0,87

Tabelle 6.19: Durchschnittliche Modell-Leistungen mit statistischem Filter für die Merkmalsselektion (FFS) nach Abschnitt 6.5.

Optimierungskriterium definiert und stellt somit ein wesentliches Beurteilungskriterium für die Qualität der trainierten Modelle dar. Die Berechnung erfolgt sowohl für alle Diagnosegruppen einzeln als auch für die Gruppe NG. Wichtig ist, dass für die binäre Unterscheidung zwischen G und NG keine modifizierten Modelle trainiert wurden, sondern nur deren Auswertung entsprechend zusammengefasst wird. Alle detaillierten Ergebnisse sind in den Tabellen 6.18 bis 6.21 zu finden. Ein direkter Vergleich zwischen der Performanz von den Bewertungsmetriken aus Unterabschnitt 6.2.4 und der durch ML gestützten Optimierung ist aufgrund der Ausgangssituation nur für die Ergebnisse der merkmalspezifischen Schemata mit Mehrheitsvotum ohne Merkmalsselektion sowie mit FFS möglich. Hierbei fällt auf, dass die Genauigkeit ohne Anwendung der Merkmalsselektion durch den RF von 76% auf 77% und der Kappa-Wert von 61% auf 62,7% gesteigert werden kann (vgl. Tabelle 6.15 und Tabelle 6.18). Bei Anwendung der durch die FFS bestimmten Merkmale wird die Leistung der klassischen Bewertungsmetrik von keinem Modell erreicht (vgl. Tabelle 6.16 und Tabelle 6.19). Insgesamt zeigt die Anwendung der WFS die höchste Performanz für alle Maße. Hierfür werden pro Modell unterschiedliche Merkmale ausgewählt, die auf Grundlage der Testdaten die besten Ergebnisse bei der Anwendung des Modells erzielen (vgl. Abbildung 6.19). Die jeweiligen Ergebnisse divergieren insofern, dass es nicht das eine optimale Modell gibt, sondern verschiedene Modelle individuelle Stärken besitzen. Auch hier sind je nach Anwendung und Zielsetzung unterschiedliche Modelle als optimal zu betrachten. Wenn *DemPredict* als Instrument zur Früherkennung verwendet wird, bieten der k-NN- und der RF-Klassifikator die Modelle mit der höchsten Sensitivität für nicht gesunde Menschen. Für Menschen mit Demenz (D) erreicht der RF-Klassifikator mit durchschnittlich 90% Sensitivität und 96,22% Spezifität die beste Performanz nach der WFS unter Berücksichtigung der Parameter p7, p39, p45 und p64. Für Menschen mit MCI werden durch den k-NN im Mittel 80% Sensitivität bei einer 90,75%-igen Spezifität nach der WFS für die Parameterauswahl

Metrik	Decision Tree			SVM mit			Naïve Bayes	k-NN	ANN
	CART	C5.0	RF	linear kernel	RBF kernel	polynomial kernel			
Genauigkeit	0,81	0,82	<b>0,87</b>	0,79	0,77	0,74	0,78	0,84	0,71
Cohen's Kappa	0,68	0,71	<b>0,79</b>	0,67	0,63	0,58	0,65	0,74	0,54
Sensitivität NG	0,73	0,83	0,83	0,81	0,70	0,80	0,82	<b>0,89</b>	0,78
Sensitivität D	0,71	0,83	<b>0,90</b>	0,74	0,73	0,45	0,71	0,83	0,75
Sensitivität MCI	0,69	0,76	0,75	0,61	0,64	0,68	0,65	<b>0,80</b>	0,56
Spezifität NG	<b>0,93</b>	0,84	0,93	0,93	0,88	0,92	0,90	0,87	0,78
Spezifität D	0,98	0,96	0,96	0,94	<b>0,99</b>	0,96	0,96	0,95	0,91
Spezifität MCI	0,96	0,92	<b>0,99</b>	0,92	0,91	0,82	0,86	0,91	0,85

Tabelle 6.20: Durchschnittliche Modell-Leistungen mit WFS erweitert nach Schinle et al., 2021 [471].

Metrik	Decision Tree			SVM mit			Naïve Bayes	k-NN	ANN
	CART	C5.0	random forest	linear kernel	RBF kernel	polynomial kernel			
Genauigkeit	0,64	0,70	<b>0,76</b>	0,63	0,73	0,72	0,76	0,70	0,66
Cohen's Kappa	0,41	0,53	<b>0,62</b>	0,41	0,58	0,55	0,61	0,50	0,47
Sensitivität NG	0,60	0,77	0,77	0,72	0,77	0,80	<b>0,82</b>	0,63	0,77
Sensitivität D	0,48	0,70	0,71	0,73	<b>0,81</b>	0,70	0,69	0,69	0,75
Sensitivität MCI	0,48	0,61	0,61	0,32	0,49	0,60	<b>0,68</b>	0,32	0,40
Spezifität NG	0,83	0,76	0,88	0,77	0,84	0,80	0,84	<b>0,93</b>	0,78
Spezifität D	0,92	0,93	0,95	0,90	0,90	0,95	0,95	<b>0,97</b>	0,89
Spezifität MCI	0,48	0,61	0,61	0,32	0,49	0,60	0,68	0,32	0,40

Tabelle 6.21: Durchschnittliche Modell-Leistungen mit RFE.

p9, p20, p42 und p59 erreicht. Wenn jedoch eine möglichst genaue Diagnose angestrebt wird, ist der RF-Klassifikator aufgrund seiner durchschnittlich höheren Genauigkeit und seines Kappa-Wertes vorzuziehen (Genauigkeit = 86,9%,  $\kappa = 0,79$ ). Mit  $\kappa = 0,87$  erreicht dieser in einer Iteration der 5-fachen Kreuzvalidierung auch den maximalen Wert über alle Modelle hinweg. Der mit dem Algorithmus *CART* trainierte Entscheidungsbaum bietet mit seiner hohen Spezifität die Möglichkeit, die Anzahl der richtig klassifizierten negativen Ergebnisse bei der Unterscheidung G und NG zu maximieren. Hierbei liegt die Spezifität von NG bei 93,33% und einer Sensitivität von 72,93%. Dies bedeutet, dass möglichst wenige Studienteilnehmer fälschlicherweise mit D oder MCI diagnostiziert werden. Wenn die Modelle ohne eine vorhergehende Merkmalsauswahl trainiert werden zeigen die Ergebnisse, dass der Leistungsverlust insgesamt bei etwa 10% hinsichtlich der Genauigkeit liegt (76,88% gegen 86,92%). Die Sensitivität für nicht gesunde Studienteilnehmer sinkt von 89,47% auf 81,73%, während die Spezifität für nicht gesunde Studienteilnehmer mit 93,33% konstant bleibt. Die RFE zeigt gar eine niedrigere Performanz als keine Merkmalsselektion wohingegen durch die Anwendung der statistischen Vorauswahl (FFS) geringfügige Verbesserungen erzielt werden.

Zusammenfassend kann ML dazu beitragen eine höhere Performanz für den angestrebten Zweck einer Früherkennung zu erreichen. Je nach Ausrichtung bezüglich der Genauigkeit mit unterschiedlichen Ansätzen. Dadurch lässt sich die Nützlichkeit einzelner Merkmale und Messmodelle validieren, was die in Unterabschnitt 3.3.3 und Unterabschnitt 3.3.5 postulierten Hypothesen zu den Mehrwerten eines datengetriebenen Demenzdiagnosemodells stützt. Neben dem Ausschluss leitlinienkonformer Parameter durch die damit verbundene Merkmalsselektion kann jedoch je nach eingesetztem Verfahren die Nachvollziehbarkeit für menschliche Entscheider verloren gehen (vgl. Unterabschnitt 2.4.3). Mechanismen mit denen diese Nachvollziehbarkeit in Teilen hergestellt werden kann, werden in Abschnitt 6.4 untersucht.

## 6.4 Plausibilisierung, Risikominimierung und Wissensgenerierung durch erklärbares Maschinelles Lernen

Wie die vorhergehenden Abschnitte zeigen, können zur Automatisierung und Optimierung der Klassifikation von Testergebnissen auf Testergebnis- und Merkmalsebene unterschiedliche ML-Verfahren zum Einsatz kommen. In diesem Abschnitt werden, der grundlegenden Zweckbestimmung von *DemPredict* folgend, Ansätze zur Schaffung von Transparenz für die eingesetzten Verfahren durch Methoden der XAI untersucht (vgl. Unterunterabschnitt 2.4.3). Hierbei stehen vor allem die Fragen im Vordergrund, welche XAI-Ansätze darauf anzuwenden sind und wie man diese auf funktionaler Ebene evaluieren kann. Hierdurch kann einerseits neues Wissen durch die Extraktion der für eine Entscheidung relevanten Merkmale und andererseits eine Grundlage für die Risikobeurteilung durch die Validierung des Systemverhaltens beim Einsatz von intransparenten Modellen entstehen.

### 6.4.1 Auswahl und Anwendung der Verfahren

Während für die in Abschnitt 6.1 angewandten Bildverarbeitungsalgorithmen beispielbasierte Erklärungen möglich sind, setzen die in Abschnitt 6.3 verfolgten ML-Ansätze zusätzliche Schritte voraus. Hier erscheint, je nach Ziel, der Einsatz unterschiedlicher Merkmalsselektions- und entsprechender ML-Verfahren als sinnvoll (vgl. Unterabschnitt 6.3.2). Damit die Eigenschaften unterschiedlicher XAI-Verfahren miteinander vergleichbar werden, ist eine gemeinsame Ausgangssituation zu definieren. Daher werden nachfolgend zwar die Algorithmen mit der höchsten Performanz aus Abschnitt 6.3 untersucht, jedoch nicht mit der individuell besten Merkmalsauswahl. Unter Berücksichtigung der Schnittmenge aus *DemPredict* mit dem CERAD werden als relevante Parameter das Alter, p1, p2, p35, p37, p38, p39, p40, p41, p47, p48, p49 und p50 gewählt. Dabei erreicht der RF einen Kappa-Wert von  $\kappa = 0,58$  bei einer *accuracy* = 0,74 und für den k-NN einen Kappa-Wert von  $\kappa = 0,50$  bei einer *accuracy* = 0,70. Nachfolgend werden dafür sowohl lokale modell-spezifische als auch modell-agnostische Verfahren angewandt und evaluiert, um Auskunft über damit getroffene Prognosen zu bekommen. Hierzu werden die optimierten Modelle aus Absatz 2.4.3 verwendet. Da sich der RF dabei wie ein Black-Box-Modell verhält, bieten sich modell-agnostische Verfahren wie LIME und SHAP an. Für LIME wird eine schnelle Berechnungszeit angegeben sowie Erklärungen, die sich für Ärzte als geeignet gezeigt haben [216, 292]. Auch das SHAP-Verfahren ist für Anwendungen im medizinischen Sektor evaluiert [45, 256, 325]. Dessen solide mathematische Grundlage wird als Voraussetzung zur Beweisbarkeit und Rückverfolgbarkeit im Zuge der Rechtssicherheit gesehen [292]. Für deren Implementierung lässt sich die lime/shapper Bibliothek in R verwenden (vgl. Tabelle 6.23). Auf Grundlage der Testdaten werden Erklärungen generiert, die bei den modell-agnostischen Verfahren durch den mittleren Informationsgehalt der als wichtig identifizierten Merkmale verglichen werden (vgl. Abbildung 6.20).

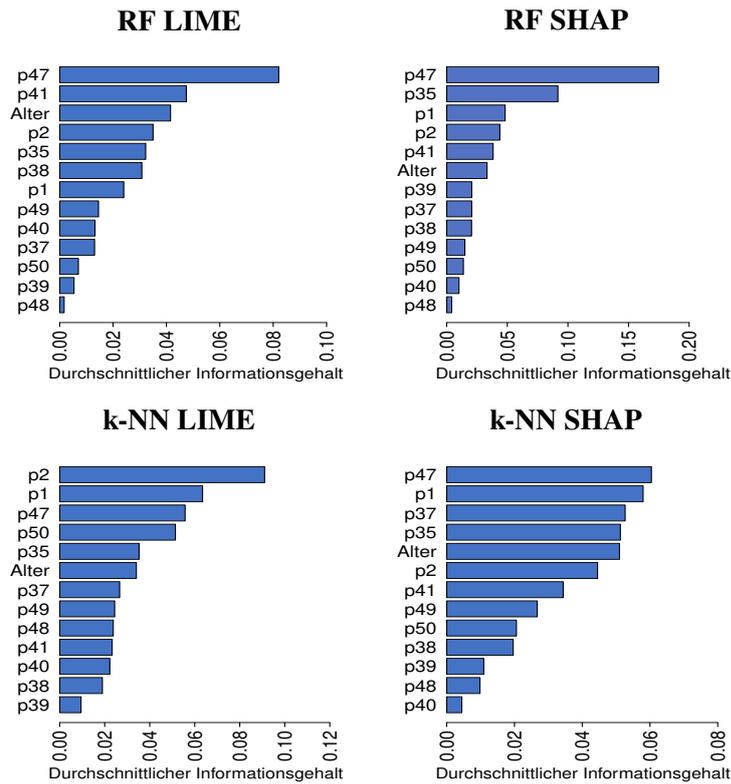


Abbildung 6.20: Gegenüberstellung des mittleren Informationsgehalts der CERAD-Merkmale für die Klassifikationswahrscheinlichkeit.

Die absoluten Werte sind hierbei aufgrund der unterschiedlichen Verfahren zur Berechnung des Informationsgehalts nicht vergleichbar, jedoch zeigt sich welche Merkmale den größten mittleren positiven Beitrag zur Prädiktion leisten. Um Aufschluss über die Stärken und Schwächen der modell-agnostischen Ansätze zu bekommen, werden diese auch mit den modell-spezifischen Ansätzen verglichen, die auf Spezifika der zu erklärenden Algorithmen eingehen, indem sie individuell auf deren Architektur zugeschnitten sind. Da sich der RF aus Decision Trees zusammensetzt, welche als interpretierbare Modelle gelten (vgl. Absatz 2.4.3), wurde CHIRPS als modell-spezifische Möglichkeit entwickelt, um Erklärungen für RFs zu erzeugen. Für das CHIRPS-Verfahren existiert eine öffentlich verfügbare Implementierung in Python [181] (vgl. Tabelle 6.23). CHIRPS extrahiert die benötigten Daten aus dem RF unter Verwendung der *scikit-learn*-Bibliothek. Da die zuvor durchgeführten Optimierungsschritte durch ML-Algorithmen in *R* unter Anwendung des CARET-Pakets [242, 249, 404] entwickelt sind (vgl. Abschnitt 6.3) wird der RF in Python erneut trainiert. Um Vergleichbarkeit zu gewährleisten werden die Hyperparameter des Modells aus *R* übernommen. Als Ergebnis liefert die Implementierung des CHIRPS eine Klassifikationsregel, welche sich als Wenn-Dann-Anweisung interpretieren lässt und dazu gehörende Performanzmaße. Zusätzlich wird die kontrafaktische Erklärung ausgegeben. Die gemittelten Ergebnisse über die elf Test-samples hinweg sind in Tabelle 6.22.

Maß	Coverage	PPW	Sensitivität	F1-Score	Cohen's Kappa
Trainingsdatensatz	0,337	0,871	0,736	0,79	0,635
Testdatensatz	0,336	0,923	0,775	0,801	0,716

Tabelle 6.22: Mittelwerte der Ergebnisse des CHIRPS-Verfahrens.

Als Beispiel zur Erklärung einer Instanz eines gesunden Probanden liefert CHIRPS die Regel:  $p35 \leq 171,841$  Sekunden &  $p47 \leq 66,06$  Sekunden. Als kontrafaktische Erklärung wird durch Inversion der Regel, diese umgekehrt angewandt. Des Weiteren handelt es sich bei k-NN um ein intrinsisch interpretierbares Verfahren, wofür beispielbasierte Erklärungen möglich sind. Hierfür wird ein Explainable k-NN (Xk-NN) Algorithmus umgesetzt (vgl. Algorithmus 1). Für das Modell mit den besten Parametern werden die  $k = 9$  nächsten Nachbarn, die an der Klassifikationsentscheidung beteiligt sind, dem Anwender präsentiert, um nachvollziehen zu können, warum diese Prognose gestellt wird. Die Beispiele der Daten werden nach euklidischer Distanz zur erklärenden Instanz absteigend sortiert angezeigt und entsprechend aus der standardisierten Form rücktransformiert (vgl. Gleichung 2.1). Damit lassen sich Prognosen einordnen und gegebenenfalls weitere Muster durch den Behandelnden feststellen. Lamy et al. schlagen eine Principal Component Analysis (PCA) vor, die dabei hilft Erklärungen grafisch darzustellen (vgl. Abbildung 6.21), indem Dimensionen reduziert werden und die Trainings- und Testdaten zweidimensional in einem Graph dargestellt werden [246].

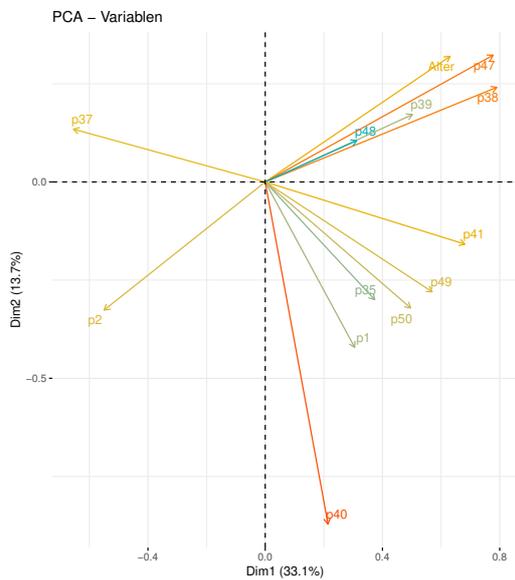


Abbildung 6.21: PCA für Trainingsdaten des Xk-NN.

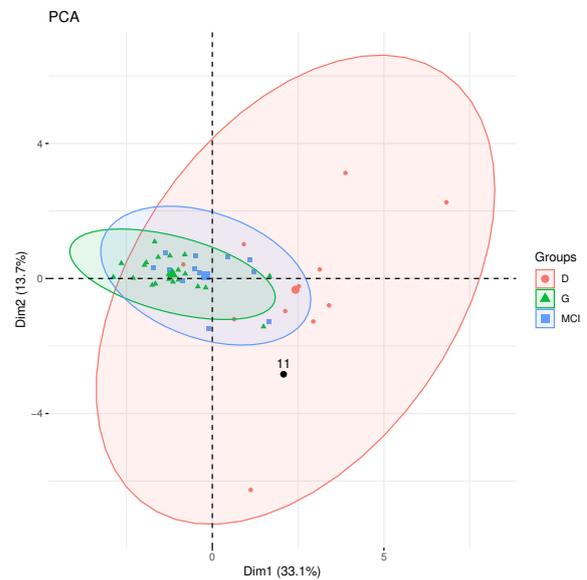


Abbildung 6.22: PCA als Erklärung für Mensch mit D (11).

Zusätzlich kann bei der Visualisierung eine farbliche Abgrenzung zwischen der zu erklärenden Instanz und den unterschiedlichen Klassen (vgl. Abbildung 6.22) vorgenommen werden [246].

Programmiersprache	LIME	SHAP	modell-spezifisch
<i>R</i>	LIME [403]	shapper/DALEX [59, 274]	Xk-NN (vgl. Algorithmus 1)
<i>Python</i>	-	-	CHIRPS [182]

Tabelle 6.23: Verwendete Pakete und selbst geschriebene Funktionen für die Implementierung der XAI-Methoden in *R* und *Python*.

## 6.4.2 Evaluation ausgewählter Verfahren

Die verschiedenen Ebenen der Evaluation werden in Absatz 2.4.3 skizziert. Dabei wird nachfolgend aus der Perspektive von Herstellern auf Funktionsebene evaluiert. Die hierfür erforderlichen Kennzahlen werden auf die Berechnungszeit, die Vertrauenswürdigkeit in Form des Informationswertes eines Merkmals beziehungsweise die Performanz der Erklärungsregeln sowie der Stabilität des Verfahrens hinsichtlich konsistenter Erklärungen für gleiche Instanzen definiert. Tabelle 6.24 fasst die Zusammenhänge zwischen den XAI-Verfahren und der vorgeschlagenen Wahl entsprechender Evaluationsmetriken zusammen. Die Berechnungszeit ist die durchschnittliche Dauer, die ein Verfahren benötigt, um Erklärungen zu den selben zehn Testsamples aus den Trainingsdaten zu generieren. Als weitere zu evaluierende Eigenschaft auf Funktionsebene wird die Stabilität und Konsistenz von Erklärungen untersucht. Hiermit wird sichergestellt, dass unter gleich bleibenden Bedingungen die gleichen Erklärungen generiert werden [413]. Für den  $X_k$ -NN-Algorithmus ist wie für die Vertrauenswürdigkeit per Definition mit einem stabilen Ergebnis der beispielbasierten Erklärungen zu rechnen, da bei konstanten Trainingsdaten die Beispiele einer zu erklärenden Instanz immer deren  $k$ -nächsten Nachbarn entsprechen. Dieser Umstand bezieht sich auch auf das untersuchte modell-spezifische deterministische CHIRPS-Verfahren bei gleich bleibenden Trainingsdaten. Die modell-agnostischen Verfahren können jedoch Instabilitäten aufweisen [413]. Beim LIME-Verfahren liegt das beispielsweise an den zufällig als Beobachtungen generierten synthetischen Daten (vgl. Unterunterabschnitt 3.3.4). Visani et al. und Shankaranarayana et al. schlagen diesbezüglich ein Konzept vor, welches die Stabilität verschiedener modell-agnostischer Verfahren vergleichbar macht [377, 413]. Hierfür wird, mittels eines Samples über mehrere Iterationen hinweg, der Einfluss von Merkmalen auf eine Erklärung aufgezeichnet. Die Streuung der Gewichtung eines Merkmals dient als Maß zur Einordnung der Stabilität, je höher die Streuung desto weniger stabil lässt sich das Verfahren anwenden [413]. Um die Vergleichbarkeit zwischen den Streuungen der Einflüsse eines Merkmals zu gewährleisten werden nachfolgend die Variationskoeffizienten verwendet (vgl. Absatz 2.1.4). Die Ebene der Vertrauenswürdigkeit setzt voraus, dass die von dem eingesetzten XAI-Verfahren gelieferten Erklärungen zutreffend sind [29]. Da als Referenzwerte lediglich die Diagnose-Klassen vorliegen, ist die neurologische und psychologische Ursache pro eingeschränkter Domäne nicht weiter spezifiziert. Vertrauenswürdigkeit wird deshalb dahin gehend evaluiert, indem messbar gemacht wird, wie stark die Komponenten einer Erklärung tatsächlichen Einfluss auf die ML-gestützte Prognose haben. Als Maß der Vertrauenswürdigkeit wird dieser Einfluss quantitativ mittels dem Ansatz Most Relevant First (MoRF) anhand einer Störungskurve dargestellt [360]. Dieses Vorgehen basiert auf dem kontinuierlichen Entzug von Informationen aus dem Testdatensatz, indem die Werte bestimmter Merkmale durch den Mittelwert ersetzt werden und deren Einfluss auf die Prognosewahrscheinlichkeit beobachtet wird. Zum Vergleich mit dem modell-spezifischen CHIRPS-Verfahren, welches eine Klassifikationsregel als Ergebnis liefert, bietet die Störungskurve nicht das richtige Maß. Daher muss anhand der gefundenen Klassifikationsregel evaluiert werden wie vertrauenswürdig eine Erklärung ist. Für diese wird auf die Performanzmaße zur Evaluation von ML-Modellen zurückgegriffen (vgl. Unterunterabschnitt 2.1.4). Dazu werden die Klassifikationsregeln anhand der Trainings- und Testdaten evaluiert und mit den Maßen Präzision, Sensitivität, F1- und Kappa-Wert verglichen. Wenn die gefundenen Klassifikationsregeln ebenso richtig prognostizieren, sind diese als entsprechend vertrauenswürdig einzustufen. Für die Evaluation der beispielbasierten Erklärungen des  $k$ -NN erübrigt sich die Frage nach dem Einfluss einer Komponente auf eine Erklärung und somit der Vertrauenswürdigkeit, da die dazu verwendeten Daten zu denen der Vorhersage äquivalent sind.

	LIME	SHAP	CHIRPS	Xk-NN
Ausgabe	Einfluss der Merkmale	Einfluss der Merkmale	Klassifizierungsregeln	Beispiele
Vertrauenswürdigkeit	MoRF-Störungskurve	MoRF-Störungskurve	Genauigkeitsmaße	optimal
Stabilität	Variationskoeffizient	Variationskoeffizient	optimal	optimal

Tabelle 6.24: Zusammenhang zwischen XAI-Verfahren und Wahl der Evaluationsmetriken.

## Berechnungszeiten

Die Berechnungen wurden unter der Betriebssystemversion Windows 10 Pro auf einem Intel(R) Core(TM) i7 10750H Prozessor mit 2,60 GHz und 31,8 GB DDR4 Arbeitsspeicher durchgeführt. Abbildung 6.23 zeigt die gemittelten Berechnungszeiten pro ML-Modell und XAI-Ansatz.

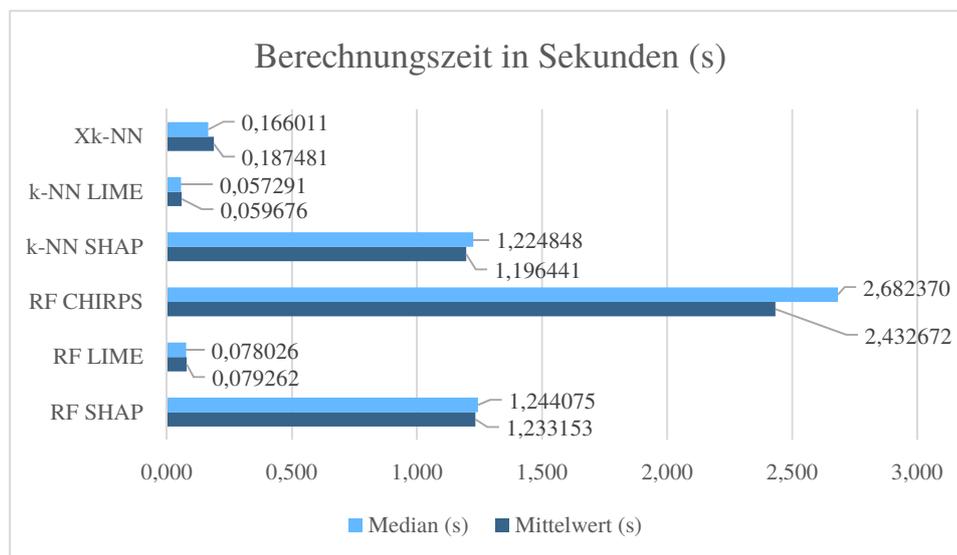


Abbildung 6.23: Berechnungszeit der XAI-Methoden.

Tabelle A.10 listet die Zeiten der einzelnen Samples. Durchweg die kürzesten Berechnungszeiten zur Erklärung einer Instanz liefert das LIME-Verfahren, gefolgt von Xk-NN und SHAP. CHIRPS benötigt um ein Vielfaches länger. Insgesamt liegen die Werte trotzdem mit unter drei Sekunden in einem akzeptablen Bereich bezüglich der Usability-Anforderungen nach Nielsen, der zehn Sekunden Antwortzeit als Akzeptanzobergrenze angibt in der Nutzer sich auf einen Dialog konzentrieren und warten [301]. Dadurch ist der Einsatz dieser Verfahren auch für die Integration in Systemen zur Erklärung einzelner Prognosen wie bspw. in *DemPredict* oder einem PVS denkbar.

## Stabilität und Konsistenz

Die Variationskoeffizienten für die modell-agnostischen Verfahren sind in Abbildung 6.24 auf der Ordinate für die auf der Abszisse aufgetragenen sechs Merkmale mit dem größten berechneten Einfluss absteigend sortiert dargestellt. SHAP zeigt sowohl für den RF als auch für den k-NN die stabilsten Eigenschaften,

indem der Variationskoeffizient im Mittel 0,149% beziehungsweise 0,671% aufweist. LIME hingegen schneidet mit 6,298% für den RF und 2,01% für den k-NN schlechter ab. Auffällig ist, dass die Variabilität steigt, je weniger wichtig der Einfluss der Merkmale wird. LIME zeigt für beide Modelle schlechtere Stabilitätseigenschaften.

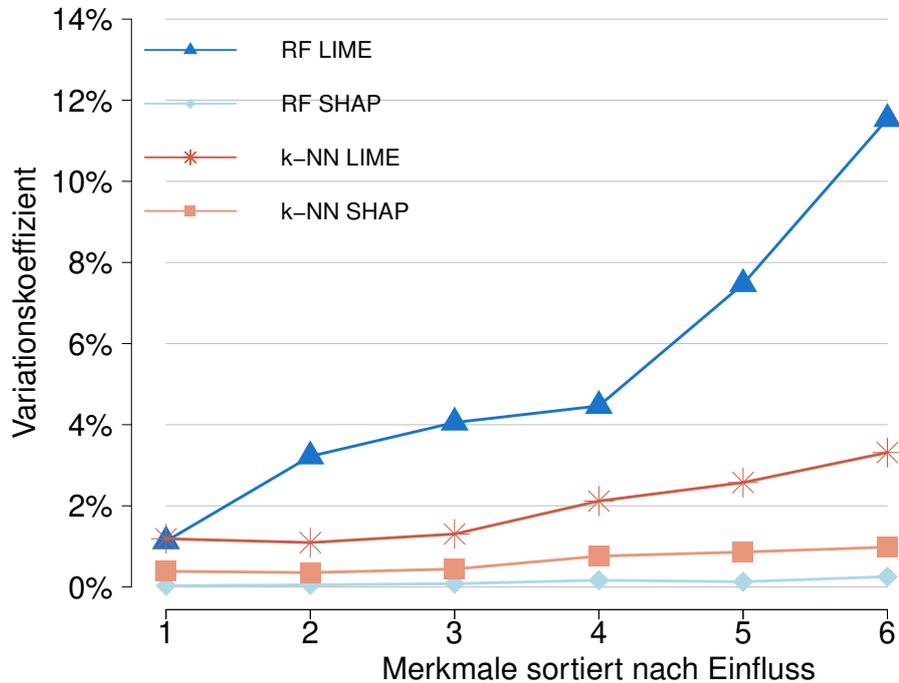


Abbildung 6.24: Variationskoeffizient der wichtigsten Merkmale.

## Vertrauenswürdigkeit

Abbildung 6.25 zeigt die MoRF-Störungskurve des RF. Die Ordinatennachse bildet die mittlere Wahrscheinlichkeit, womit die prognostizierte Klasse durch das entsprechende Verfahren über alle Samples hinweg klassifiziert wird, ab. Die Abszisse trägt die standardisierten Merkmale nach deren Einfluss absteigend sortiert auf. Diese unterscheiden sich in Abhängigkeit der Samples und werden nacheinander auf den Mittelwert Null gesetzt, um deren Einfluss auf das Gesamtergebnis zu betrachten. Die anhand der einzelnen Verfahren bestimmten mittleren positiven Werte des Informationsgehalts werden für den RF in Abbildung 6.20 gegenübergestellt. Dennoch lässt dieser Mittelwert keine Aussage über die Vertrauenswürdigkeit zu, da gegensätzlich zu den MoRF-Kurven die tatsächliche Reihenfolge, in welcher die Merkmale zur Generierung der Störungsstufen verwendet werden, nicht berücksichtigt wird. Nach Berechnung der mittleren AUCs, liegt der RF LIME bei 4,893 und der RF SHAP bei 2,893. Dem gegenüber steht der k-NN LIME mit 4,33 und der k-NN SHAP bei 3,242. Hierbei wird angenommen, je kleiner der AUC-Wert, desto vertrauenswürdiger das Verfahren. Generell zeigt SHAP, dass die als wichtig eingestufte Merkmale einen signifikanten Einfluss auf die Klassifikationswahrscheinlichkeit haben. Die höhere negative Steigung ab Störungsstufe 1 suggeriert mehr Vertrauenswürdigkeit dieses XAI-Verfahrens, da die damit generierten Näherungen an das ursprüngliche Modell dieselben wichtigen Ursachen einer Prognose identifizieren. Der Anstieg von LIME hingegen zeigt, dass das Ersetzen der ersten und damit wichtigsten Merkmale durch den Mittelwert einen positiven Einfluss auf die Klassifikationswahrscheinlichkeit hat. Das liegt zum einen an der geringen Anzahl an Werten, die zur Verfügung stehen, zum anderen daran,

dass diese je nach zu erklärendem Sample nicht gleich viele Merkmale mit einem positiven Einfluss auf die Prognose enthalten (vgl. Abschnitt A.19).

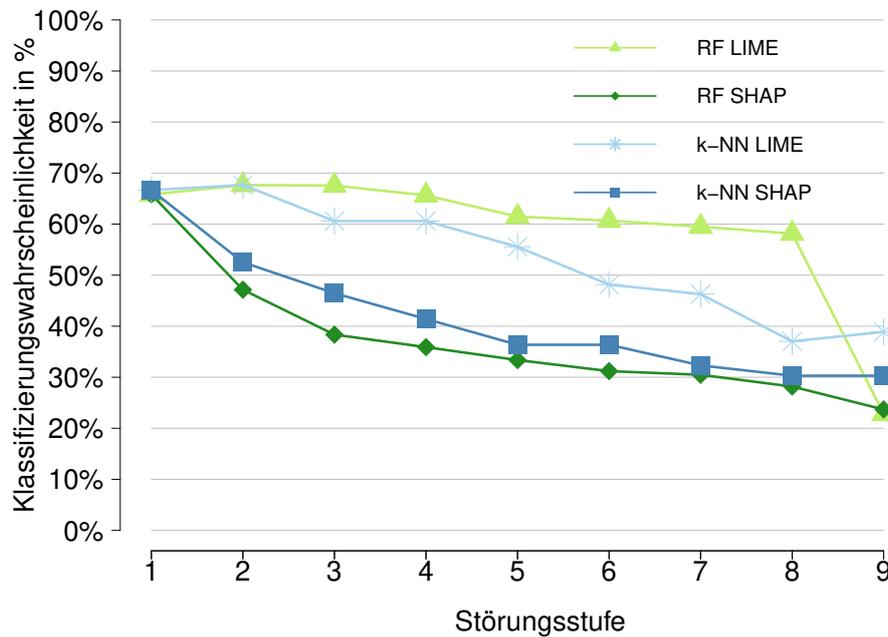


Abbildung 6.25: MoRF-Störungskurve.

Wie in Tabelle 6.22 dargestellt, erreicht das CHIRPS im Mittel für die identifizierten Klassifikationsregeln eine höhere Performanz auf den Testdaten als auf den Trainingsdaten. Der Coverage-Wert zeigt, dass die gefundenen Regeln bei circa einem Drittel der Klassifikationsentscheidungen Anwendung finden können, was ein Indikator für die korrekte Identifikation der für das RF-Modell relevanten Komponenten darstellt. Der F1-Score und Kappa-Wert zeigen ebenfalls, dass sie in einem mit dem RF vergleichbaren Bereich liegen und somit vertrauenswürdige Erklärungen liefern.

Insgesamt zeigen die Ergebnisse dieser Untersuchung ebenfalls den in der Literatur anzutreffenden Trade-Off zwischen Interpretierbarkeit und Genauigkeit von Modellen [312]. Welches Maß an Erklärbarkeit ausreichend ist, hängt neben den anwendungsfallspezifischen Anforderungen, auch mit den regulatorischen Implikationen zusammen, die in Unterabschnitt 7.3.2 näher beleuchtet werden.



# 7 Translation in die Regelversorgung

DIESES Kapitel untersucht gemäß der Phase Translation & Evaluation verschiedene Wege, um deren Anforderungen und Implikationen für den Transfer digitaler Gesundheitsanwendungen aus der Forschung in die Regelversorgung zu bringen (vgl. Unterabschnitt 4.1.2). Hierbei spielen die Perspektiven der unterschiedlichen Interessenvertretergruppen eine Rolle (vgl. Abschnitt 7.1). Dafür wird einerseits Wert auf die Adhärenz der Leistungsempfänger gelegt (vgl. Unterabschnitt 7.1.1). Andererseits werden Akzeptanzkriterien aus Sicht der Leistungserbringenden empirisch erhoben und analysiert unter welchen Bedingungen eine digitale Gesundheitsanwendung wie *DemPredict* in deren Arbeitsalltag Anwendung finden kann (vgl. Unterabschnitt 7.1.2). Hieraus ergibt sich ein Gesamtkonzept, welches weitere Forschungs- und Entwicklungspotentiale hinsichtlich der Digitalisierung strukturierter Präventions- und Interventionsprogramme aufdeckt (vgl. Abschnitt 7.2). Abschließend werden die zum Zeitpunkt der Anfertigung dieser Arbeit gültigen rechtlichen Aspekte und dadurch möglichen Markteintrittsstrategien aus der Perspektive des deutschen Gesundheitswesens beleuchtet (vgl. Unterabschnitt 7.3.1). Hierbei findet eine Einordnung ausgewählter organisatorisch-technischer Implikationen in die zuvor aufgezeigten Gegebenheiten statt (vgl. Unterabschnitt 7.3.2).

## 7.1 Perspektiven unterschiedlicher Akteure

Im deutschen Gesundheitswesen spielen unterschiedliche Akteure bei der Zulassung, Erstattung und Verwendung technischer Lösungen eine Rolle (vgl. Unterabschnitt 2.2.1). Deshalb werden nachfolgend insbesondere die Perspektiven von potentiellen Leistungsempfängern und -erbringern am Beispiel von *DemPredict* untersucht.

### 7.1.1 Adhärenz bei Leistungsempfängenden

Die vorherigen Untersuchungen decken sich mit den in der wissenschaftlichen Literatur vorzufindenden Erkenntnissen, dass die Validität einzelner auch in *DemPredict* enthaltener psychometrischer Tests in digitalen Versionen erhalten bleibt [382]. In Experimenten konnten hierfür die Ergebnisse des analogen Äquivalents reproduziert und weitere digitale Biomarker mittels ML-Methoden gefunden werden [104]. Durch die Computerisierung entstehende Möglichkeiten einer adaptiven Testung können darüber hinaus die Effizienz erhöhen [432]. Dieses Potential ist jedoch nur dann realisierbar, wenn die Akzeptanz digitaler Lösungen und damit verbundene Aufwände bei deren Durchführung durch Nutzende und damit potentielle Leistungsempfänger konstant beispielsweise über einen mit einem Leistungserbringer definierten Zeitraum besteht. Im Kontext der Verordnung einer DiGA kann dieser von einem behandelnden Mediziner festgelegt werden, der einem Patient bei Verdacht eines kognitiven Funktionsverlustes die Durchführung einer solchen Anwendung über einen gewissen Zeitraum verordnet (vgl. Unterabschnitt 7.3.1). Die Adhärenz beschreibt die Bereitschaft der Leistungsempfänger sich an die Verordnung des Arztes zu halten

und wird nachfolgend anhand des Anwendungsfalls von *DemPredict* eruiert, indem die akzeptable Frequenz der Testdurchführung untersucht wird. Hierfür findet eine mehrphasige Beobachtungsstudie statt. Zunächst wird eine Umfrage durchgeführt, welche die demografischen Daten erfasst und vorab eine grobe Einschätzung zur potentiellen Bereitschaft der Studienteilnehmer erfragt, digitale psychometrische Tests mit *DemPredict* durchzuführen, um eine Einschätzung zu verschiedenen kognitiven Leistungsmerkmalen zu erhalten. Die Häufigkeit wird in Form von zeitlichen Angaben über die Testfrequenz erfragt. Hierfür wurden vorab insgesamt 29 Personen online befragt (vgl. Abschnitt A.20). Die Angaben zur vorstellbaren Testfrequenz liegen bei jährlich (31,0%; 9 Teilnehmer), wöchentlich (27,6%; 8), alle 6 Monate (20,7%; 6) und monatlich (10,3%; 3) sowie jeweils bei einem Teilnehmer für alle zwei Wochen, zwei Monate oder quartalsweise. Keiner gibt an die Tests häufiger als wöchentlich durchführen zu wollen. Aus diesen Ergebnissen findet im Anschluss daran eine longitudinale Beobachtungsstudie am Beispiel einer potentiell von einem Behandelnden verordneten Beobachtungsdauer statt. Zur Maximierung der Abtastfrequenz, wird die aus der Umfrage zuvor am häufigsten angegebene Frequenz bei geringstem zeitlichen Abstand zwischen den Tests aufgegriffen. Diese stellt den Wochenrhythmus dar, welcher die Entwicklung der akzeptierten Durchführungsfrequenz über einen Monat hinweg in diesem Turnus neu erfragt. Dabei wird vor Beginn jeder Sitzung mit *DemPredict* die vorstellbare Testfrequenz erfragt, um Wissen über die Änderung entlang des Beobachtungszeitraums zu generieren. Für diese Untersuchung werden insgesamt 18 Teilnehmer nach unterschiedlichen Zeiträumen befragt (vgl. Abschnitt A.20). Zwölf Teilnehmer haben die Studie über vier Wochen absolviert, die restlichen Teilnehmer sind später ein- oder früher ausgestiegen. Insgesamt wurden dadurch 62 Testdurchführungen erzielt. Hiervon beantworteten 48% die Frage nach der geeigneten Testfrequenz mit „wöchentlich“ und 12% mit „alle zwei Wochen“. Die Antwortmöglichkeiten „alle zwei Tage“ und „alle vier Monate“ wurden zu 13% beziehungsweise 10% vergeben. Abbildung 7.1 zeigt die Anzahl aller Antworten sowie deren zeitliche Änderung. Drei

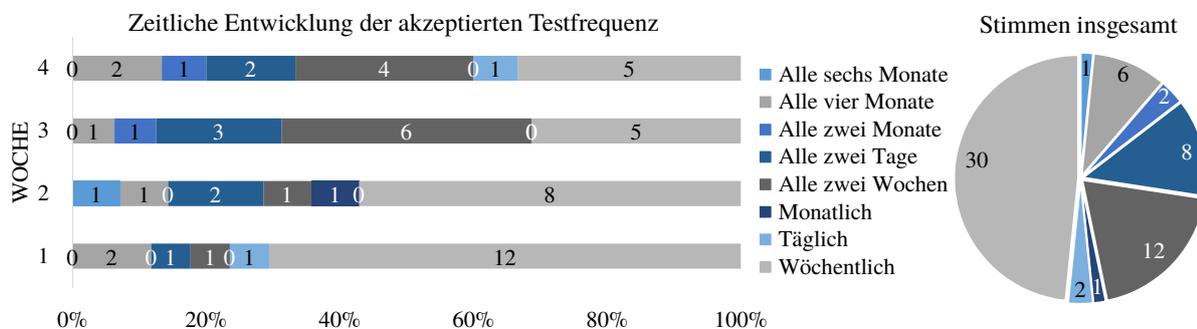


Abbildung 7.1: Umfrageergebnisse bezüglich Durchführungsfrequenz.

der zwölf über den gesamten Zeitraum teilnehmenden Studienteilnehmer halten über die vier Wochen hinweg ihre Antworten konstant, die restlichen neun variieren. Dabei hat sich die initiale Antwort aus der Testsitzung in der ersten Woche im Verlauf vier Mal in Richtung einer niedrigeren Testfrequenz verändert und ein Mal in Richtung einer höheren Testfrequenz. Vier Teilnehmer geben in den folgenden Sitzungen sowohl niedrigere als auch höhere Testfrequenzen an.

Aspekte, die mit der erhöhten Testfrequenz einhergehen sind neben den inhärenten motivationalen Komponenten psychologischer Untersuchungen auch Lerneffekte. Deren Einfluss wird auch bei diesem eingeschränkten Stichprobenumfang sichtbar, muss aber als Teil weiterer Studien zukünftig gesondert untersucht werden. Für diese Stichprobe zeigt sich, dass die Bereitschaft zur Durchführung solcher Testungen über einen definierten Zeitraum von vier Wochen hinweg in einem unterschiedlichen Rhythmus akzeptiert wird. Inwiefern sich diese Bereitschaft über längere Zeiträume entwickelt und wie die Einstellung insbesondere bei Menschen innerhalb der spezifischen Zielgruppe ist, stellt somit auch einen Bestandteil

des Entwicklungsprozesses digitaler Gesundheitsanwendungen dar. Insbesondere für Leistungserbringer könnte dies ein Indikator für die Nützlichkeit solcher Werkzeuge sein, um eine potentielle Behandlung digital zu begleiten. Daher finden nachfolgend Untersuchungen zu weiteren Akzeptanzkriterien aus Perspektive dieser Interessenvertretergruppe statt.

## 7.1.2 Akzeptanz bei Leistungserbringenden

Um herauszufinden, wie Leistungserbringende in der Regelversorgung dem Einsatz einer patientenzentrierten Anwendung gegenüberstehen, erfolgte im Februar 2021 eine Umfrage unter kassenärztlich organisierten Leistungserbringern. Durch die Unterstützung der KV Baden-Württemberg, KV Thüringen und KV Sachsen wurden Fachgruppen innerhalb dieser Bundesländer digital angesprochen. Insgesamt haben 301 Leistungserbringende erfolgreich an der Umfrage teilgenommen. Eine Zusammenfassung über die demografischen Daten der Befragten findet sich in Abschnitt A.22. Der hierfür erarbeitete Fragebogen Digital-Health-Compliance-Questionnaire (DHCQ) konnte als Werkzeug bezüglich der Anwendungsdomäne von *DemPredict* anhand dieser Studie validiert werden. Eine detaillierte Beschreibung der dabei zugrunde gelegten Hypothesen und Messmodelle finden sich in einer vom Autor getätigten Publikation [469]. Durch die Entwicklung dieses neuen Erhebungsinstruments werden diesbezügliche Kriterien zukünftig aus Perspektive der Leistungserbringer erfassbar. Nachfolgend wird auf die Ergebnisse der sekundären Forschungsfragen bezüglich einzelner Akzeptanzkriterien eingegangen. In Abschnitt A.21 ist der DHCQ dargestellt. Er gliedert sich in sieben Abschnitte. Neben demografischen Daten werden die Technikerfahrung sowie die Einstellung zu Technik abgefragt. Gefolgt von Informationen zur Applikation *DemPredict* werden danach die Einstellung zu dieser Anwendung und die Voraussetzungen erfragt, die eine solche Applikation erfüllen muss, um auch in der Regelversorgung zum erfolgreichen Einsatz kommen zu können. Eine detaillierte Darstellung einzelner Antworten findet sich in Abschnitt A.23. Nachfolgend werden die, zur Beantwortung der Fragestellungen dieser Arbeit, relevanten Auszüge aus dieser Umfrage dargestellt. Hierbei wird insbesondere auf folgende übergeordneten Fragen eingegangen:

- F1** Wie hoch ist der Anteil an Leistungserbringern, die bereits Demenzdiagnosen durchführen und welche Werkzeuge kommen zum Einsatz?
- F2** Welche Kriterien sind besonders wichtig damit ein digitales Werkzeug in einer Praxis zum Einsatz kommen könnte?
- F3** Welche Rolle spielen dabei Datenschutzbedenken?
- F4** Wie ist die Einstellung gegenüber automatisierten Auswertungen?
- F5** Würden Mediziner den Test unter Aufsicht / ohne Aufsicht durchführen lassen?
- F6** Werden zu Hause erzielte Testergebnisse akzeptiert?
- F7** Wie sollten die Testergebnisse dargestellt werden?
- F8** Würde ein digitales Behandlungsmanagement entlang von identifizierten Risikofaktoren akzeptiert werden?
- F9** Wie kann die Nutzung abgerechnet werden? (vgl. Unterabschnitt 7.3.1)

Zur Beantwortung von **F1** wird Bezug auf die Angaben zu den Punkten 5. und 6. des Fragebogens genommen (vgl. Abschnitt A.21). Insgesamt geben 78,4% der Befragten an, dass sie demenzspezifische

Parameter in ihren Praxen mit Hilfe von etablierten Werkzeugen erfassen. Dazu zählen 174 von 214 Allgemeinmediziner (Allgemeinmed.), alle 29 befragten Neurologen (Neuro.), 16 von 31 Psychologen/Psychiater (Psych.) und 17 von 27 Fachärzten sonstiger Fachrichtungen (Sonst.). Hierzu kommen primär Kombinationen etablierter psychometrischer Instrumente zum Einsatz (vgl. Unterabschnitt 2.2.5). Am häufigsten wird der MMSE (189 Nennungen), *DemTect* (150), Uhrentest (110), MoCA (29) und CERAD (8) angegeben (vgl. Unterabschnitt 2.2.5). Die Beantwortung der Frage **F2** erfolgt anhand der aggregierten Antworten auf die Fragen 34, 35, 37-39, 42, 44 und 48 (vgl. Abschnitt A.21). Zunächst werden Akzeptanzkriterien bewertet, welche aus Sicht der bei Erstellung des DHCQ eingebundenen Experten die Nutzungswahrscheinlichkeit der Leistungserbringer erhöhen könnten [469]. Hierzu werden Hypothesen aufgestellt, womit die Zustimmung durch die Leistungserbringer über eine fünf-stufige Likert-Skala (trifft nicht zu - trifft zu) erfasst wird. Genannt werden hierbei positive Eigenschaften, die eine Applikation erfüllen müsste, um zu erfahren wie die Leistungserbringer diesen gegenüber stehen. Neben der Vergütung durch Krankenkassen werden solche Aussagen in Bezug auf die Zeitersparnis sowie die Integration der damit erhobenen Daten in das PVS getroffen (Fragen 34, 35, 37 und 38). Des Weiteren wird erfragt, ob eine größere Häufigkeit von Screenings stattfinden würde, wenn die Anwendung einen positiven Einfluss auf die Behandlungszeit hätte sowie, ob die CE-Kennzeichnung für „Medizinprodukte“ einen positiven Einfluss auf die Nutzungswahrscheinlichkeit hat (Frage 39 und 42). Die Ergebnisse in Abbildung 7.2 zeigen in Form von Boxplots, dass für diese Aussagen der Median zwischen trifft eher zu und trifft zu liegt. Vor allem die Integration der Ergebnisse in das PVS und die Zeitersparnis durch die digitale Erfassung und Auswertung werden als wichtig erachtet. Ein damit einhergehender positiver Einfluss auf die Behandlungszeit wird auch als positiv bezüglich der Häufigkeit von Testungen gesehen. Eine angemessene Vergütung durch Krankenkassen und die CE-Zertifizierung spielen zwar relativ zu den anderen Kriterien eine weniger gewichtige Rolle, drücken sich aber dennoch in einer überwiegenden Zustimmung aus. Wie stark Bedenken vor Problemen mit dem Datenschutz existieren, wird auch explizit in **F3** über die Beantwortung der Frage 29 aufgegriffen (vgl. Abbildung 7.2). Demnach sind Bedenken bezüglich des Datenschutzes überwiegend vorhanden. Abbildung 7.2 zeigt ebenfalls die in **F4** erfragte Einstellung gegenüber der Unterstützung einer automatisierten Auswertung der Testergebnisse (Frage 30), welche durch eine solche digitale Lösung ebenfalls als überwiegende Erleichterung des Praxisalltags gesehen wird. Die Fragen **F5** und **F6** zielen darauf ab, die Akzeptanz von Ergebnissen zu prüfen, die nicht unter ärztlicher Aufsicht angefertigt werden. Hierbei wird zunächst gegenübergestellt, wie die Einstellung gegenüber der Anfertigung dieser Tests ohne Aufsicht (Frage 40) oder unter Aufsicht von Personal (Frage 41) vor Ort in der Praxis ausfällt. Abbildung 7.2 zeigt, dass die Akzeptanz der Ergebnisse einer delegierten Testdurchführung unter Aufsicht stärker ausgeprägt ist als durch den potentiell Betroffenen allein durchgeführte Testungen. Auch in Abbildung 7.3 wird durch die Antworten auf Frage 49 deutlich, dass nahezu die Hälfte aller Mediziner keiner diagnostischen Ersteinschätzung trauen, die durch eine Datenerhebung von zu Hause stattfindet. Mehr als ein Drittel der Teilnehmer hat mit vielleicht geantwortet und würde unter bestimmten Bedingungen eine solche Anwendung einsetzen. In den Begründungen dazu wird deutlich, was für Bedenken dabei auftreten und welche Voraussetzungen erfüllt sein müssten, um diese zu adressieren. Nachfolgend werden diese gruppiert dargestellt. Als häufigste Anmerkung wurde genannt, dass die Verwendung der Anwendung von der persönlichen Einschätzung des Leistungserbringers bezüglich der Kontextfaktoren des Patienten abhängt. Hierbei werden der aktuelle Gesundheitszustand des Patients, seine Persönlichkeit, die intellektuelle Ausstattung, Einschränkungen bezüglich der Bedienung, das persönliche Umfeld inklusive der Angehörigen, die Zuverlässigkeit, Kooperationsfähigkeit und Technik-Affinität als Kriterien genannt. Neben den patientenspezifischen Faktoren steht an zweiter Stelle die Sorge um die Verlässlichkeit der Durchführung einer Testung. Hierbei werden Möglichkeiten des „Schummelns“ oder der Beeinflussung von Dritten als Vertrauensprobleme formuliert. Auch wenn

die häusliche Umgebung als stressfreier gesehen wird als die Arztpraxis, muss dies den Möglichkeiten zu mogeln und eventuelle Mithilfe der Angehörigen in Anspruch zu nehmen, gegenübergestellt werden. Auch das richtige Verständnis der Tests vor einer Durchführung und die damit verbundene korrekte Bedienung muss gewährleistet sein. Währenddessen auftretende Probleme sind ebenfalls als Metainformationen relevant. Hier wird zum Beispiel auch die Art wie Antworten gefunden werden oder wie sich Dissimulation und Aggravation auswirken in Betracht gezogen. Zur Testung unter Aufsicht wurde auch aus einer anderen Position eingegangen. Hier wurde die Aufsicht als Voraussetzung genannt, um die Ergebnisse aus dem häuslichen Umfeld zu akzeptieren. Insbesondere vertrauenswürdige Angehörige können nach Meinung der Leistungserbringer bei technischen Fragen unterstützen und versichern, dass der Test durch den Patienten alleine und ohne Hilfsmittel durchgeführt wurde. Eine weitere Voraussetzung für den Einsatz solcher Applikationen im häuslichen Umfeld ist die persönliche Erfahrung durch die Leistungserbringer. So hängt die Akzeptanz auch mit den eigenen Kenntnissen und Erfahrungen bezüglich der Anwendung zusammen, die zunächst durch ausprobieren oder Empfehlungen von Kollegen gewonnen werden muss. Alle weiteren Anmerkungen beziehen sich einerseits auf die Validität der Tests gegenüber Seiteneffekten, wie zum Beispiel Lerneffekten, der Tagesform und dem zeitlichen Versatz zwischen der Testung und dem Praxisbesuch. Vereinzelt wird auch der Missbrauch durch positiven Krankheitsgewinn als Befürchtung geäußert. Abschließend wird der Mehrwert dieser Testergebnisse nur im Kontext darauf basierender Empfehlungen für passende Behandlungsschritte sowie nachfolgende weitere Tests und einem persönlichen Gespräch zur Abklärung gesehen. Um die wichtigsten Voraussetzungen zu identifizieren,

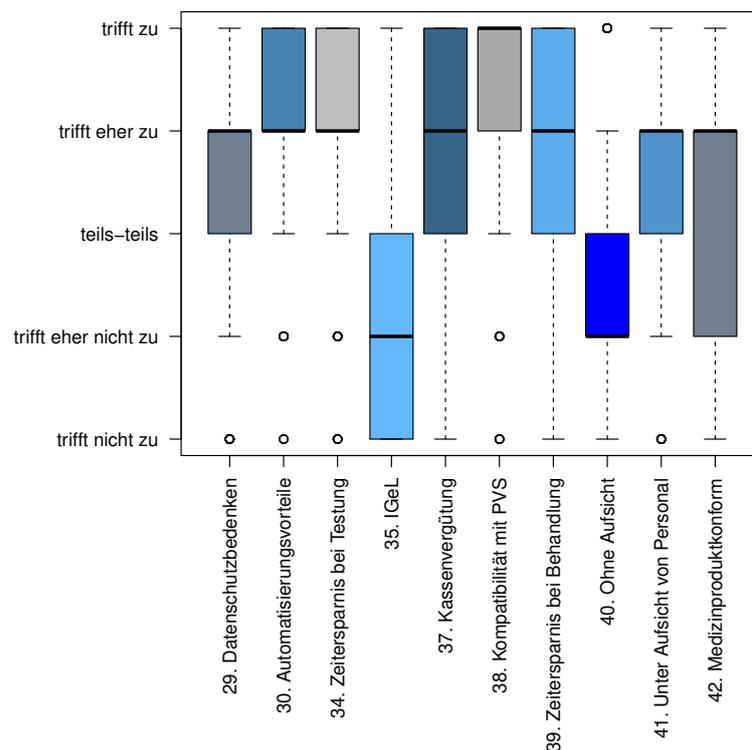


Abbildung 7.2: Umfrageergebnisse der Akzeptanzkriterien zu den Fragen 29, 30, 34, 35, 37, 38, 39, 40, 41 und 42.

die eine solche Anwendung aus Sicht der Leistungserbringer erfüllen müssten, besteht durch Frage 44 die Möglichkeit über ein Freitextfeld die Voraussetzungen für deren Einsatz zu benennen. In Abbildung A.12 werden diese Antworten gruppiert und anhand deren Häufigkeiten in absteigend sortierter Reihenfolge dargestellt. Eindeutig spielt dabei die Praktikabilität der digitalen Lösung sowie die Vergütung gefolgt von potentiellen Effizienzgewinnen und die geprüfte Validität des eingesetzten digitalen Werkzeugs eine

Rolle. Sowohl die laufenden Kosten und die Kompatibilität mit bestehenden Infrastrukturen in den Praxen vor Ort als auch die Erfüllung von Datenschutz- und -sicherheitsaspekten werden als wichtige Themen gesehen. Eine mögliche höhere Aussagekraft der digitalen Testergebnisse ist weniger wichtig als die kostenfreie Bereitstellung der dazu geeigneten Hard- und Software sowie die Akzeptanz bei den Patienten. Frage 48 grenzt die Entscheidung, welche Dinge beim Einsatz der Applikation besonders wichtig sind, ein. Die Verteilung der Antworten in Abbildung A.10c bis Abbildung A.10e zeigt, dass 72% der befragten Mediziner besonderen Wert auf die Messgenauigkeit des diagnostischen Instruments legen. Gefolgt von 63%, welche die Vergütung durch die Krankenkassen als wichtig erachten. Lediglich 42% sehen die Empfehlung der Anwendung durch die KV als wichtig an. In welcher Weise die Testergebnisse für den Arzt angezeigt werden, wird in F7 mit den Möglichkeiten zur Auswahl in Frage 51 adressiert. Hierbei wird im Wesentlichen abgefragt, ob die Klassifikation der Gesamtergebnisse in unterschiedliche Diagnosegruppen (vgl. Abschnitt 6.3) ausreichend ist oder inwiefern ein transparentes Scoring für die Parameter jedes Einzeltests relevant ist (vgl. Abschnitt 6.2). Die Ergebnisse zeigen, dass die Mehrheit der Neurologen keine Klassifikation der Gesamtergebnisse nach Diagnosegruppen benötigt und an einem Scoring zur eigenständigen Einschätzung der Leistungsfähigkeit interessiert ist (vgl. Abbildung 7.3).

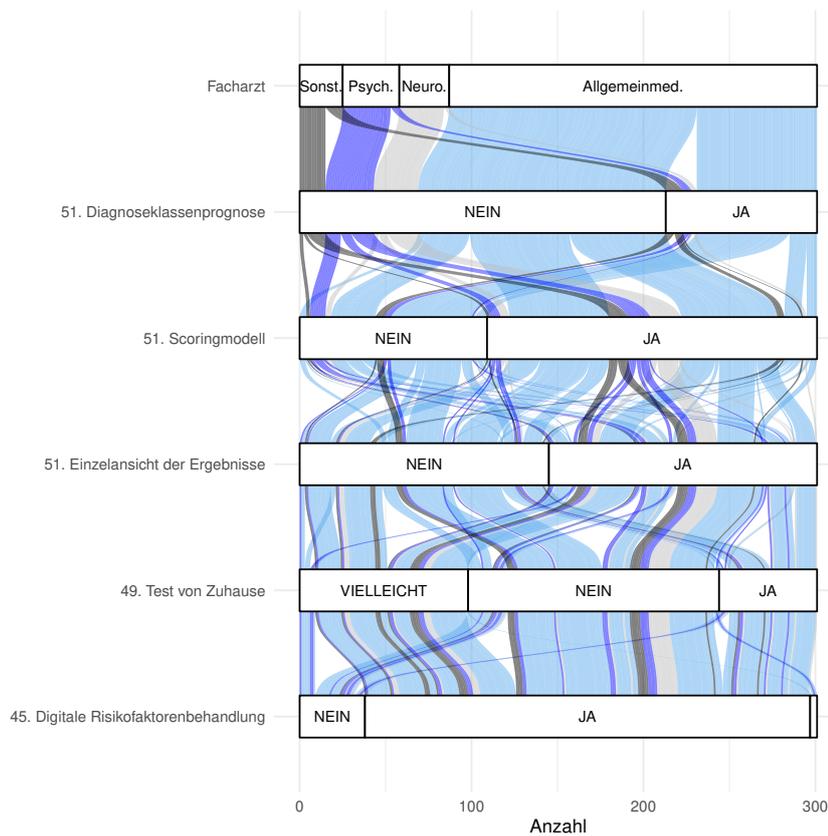


Abbildung 7.3: Umfrageergebnisse zu Frage 45., 49. und 51.

Auch circa zwei Drittel der Allgemeinmediziner zieht ein Scoringverfahren mit unterschiedlichen Detailgraden einer Gesamtklassifikation vor. Als konstruktive weitere Anmerkungen werden der Bedarf einer visuellen Kontrolle der Ergebnisse von Tests und Subtests sowie die Bereitstellung der Rohdaten für eine eigene Auswertung genannt. Darüber hinaus wird die Wichtigkeit der Abgrenzungen zu Erkrankungen mit vergleichbaren Symptomen wie Depressionen genannt sowie nach Möglichkeit eine Einteilung in Schweregrade bei Vorliegen einer Demenz mit Begründung und vergleichenden Werten der entsprechenden Altersgruppe angegeben. In Abhängigkeit des Ergebnisses sollten Empfehlungen zur weiteren

fachärztlichen Abklärung und Vorschläge für ein weiteres Vorgehen gegeben werden. Auf diesen letzten Punkt zielt auch **F8** ab. Studien zufolge stellt die gezielte Adressierung reversibler Risikofaktoren eine Möglichkeit dar in den multifaktoriellen kognitiven Abbauprozess einer Demenz zu intervenieren oder diesem gar vorzubeugen (vgl. Unterunterabschnitt 2.2.4). Um diese adäquat zu behandeln, wird in Frage 45 eine Erweiterung der Anwendung *DemPredict* skizziert, um die Möglichkeit diese Risikofaktoren individuell für einen Patienten in der Praxis einzustellen und das Behandlungsmanagement über die halbjährliche Sprechstunde hinaus zu verfolgen. Diesbezüglich wird die Bereitschaft abgefragt einen solchen Vorgang mittels digitalen Werkzeugen zu unterstützen. Ein Vergleich der Fachrichtungen und deren Neigung diese zu verwenden zeigt sich ebenfalls in Abbildung 7.3. Insgesamt stimmen 88% der antwortenden Allgemeinmediziner, 93% der Neurologen, 70% der Psychologen/Psychiater sowie 93% der niedergelassenen Fachärzte sonstiger Fachrichtungen der Verwendung eines solchen Werkzeugs zu. Wie eine diesbezügliches Systemkonzept ausgestaltet werden kann, ist Teil des Folgeabschnitts.

## 7.2 Der datengetriebene Weg zu personalisierten Präventions- und Behandlungsprogrammen

Demenz ist ein Syndrom, welches aus medizinischer Sicht zum Zeitpunkt der Anfertigung dieser Arbeit noch nicht vollständig verstanden ist (vgl. Unterabschnitt 2.2.3). Der Verlust der Alltagsfähigkeit von Menschen mit Demenz führt zu direkten und indirekten Folgekosten, welche weltweit zu Belastungen der Gesundheitssysteme führen (vgl. Unterunterabschnitt 2.2.3). Daher besteht ein Bedarf an präventiven und interventiven Ansätzen zur Vorbeugung und Behandlung von Demenzen basierend auf der individuellen Adressierung von Risikofaktoren (vgl. Unterunterabschnitt 2.2.4). Die in dieser Thesis erarbeiteten IT-Artefakte stellen Bausteine dar, die zur datengetriebenen Entwicklung diesbezüglicher digitaler Werkzeuge genutzt werden können. Auf den Anwendungskontext der Demenz bezogen, wird nachfolgend ein Weg skizziert, wie strukturierte Präventions- und Behandlungsprogramme dadurch unterstützt werden können. Hierbei spielen unterschiedliche Interessenvertreter eine Rolle, die in das dargestellte Gesamtkonzept in Abbildung 7.4 einbezogen werden.

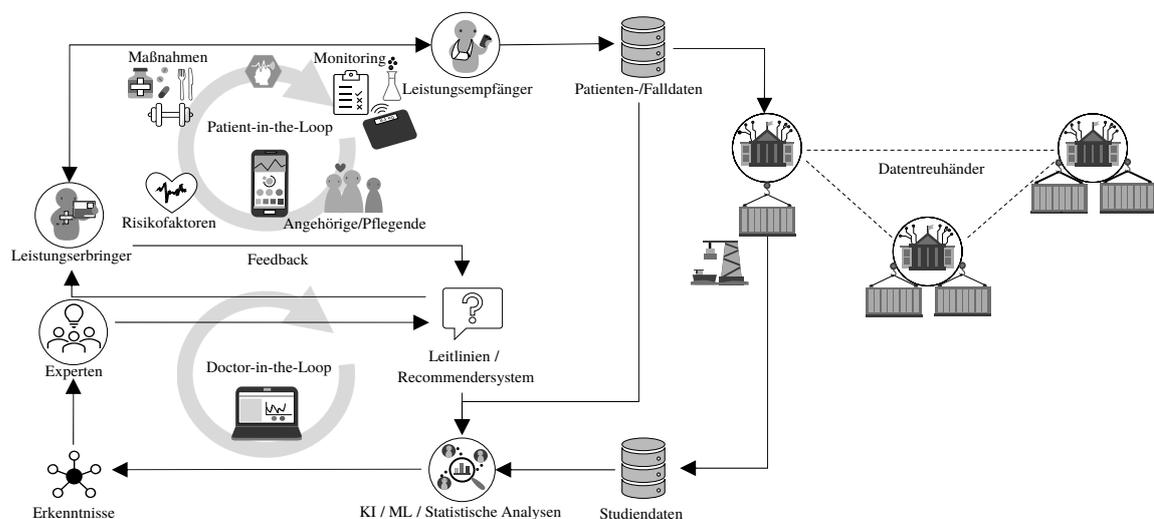


Abbildung 7.4: Entwicklung digitaler Präventions- und Behandlungsprogramme aus der Kontextsicht.

## 7.2.1 Doctor-in-the-Loop

Wie die vorherigen Kapitel zeigen, bestehen für Behandelnde besondere Herausforderungen im Umgang mit Menschen mit Demenz. Was bei der Umsetzung eines digitalen Screenings beachtet werden muss, zeigen die Schritte zur Umsetzung von *DemPredict*. Um die Potenziale digitaler Technologien gewinnbringend für Behandelnde einsetzen zu können, werden auch Lösungen zur Behandlungsunterstützung notwendig. Aufgrund des notwendigen Domänenwissens bezüglich der Erstellung strukturierter Interventionspläne stellt ein Recommendersystem die Möglichkeit dar, wie eine leitlinienkonforme Behandlung durchzuführen ist aber auch darüber hinausgehende individuelle Risikoprofile von Patienten abgeleitet und adressiert werden können. Da das dazu notwendige Wissen kontinuierlich in Studien generiert wird, werden einerseits die aktuellen Leitlinien mit Verzögerungen konsentiert, andererseits trifft das Wissen aus den jeweiligen Studiensettings auf den individuellen Versorgungsalltag. Hierfür ist ein lernendes System denkbar, welches das Erfahrungswissen von Behandelnden und Experten für die adaptive Gestaltung von Präventions- und Behandlungsprogrammen zusammenbringt. Das für die Behandlung und Prävention benötigte Wissen stammt einerseits aus dem Feedback der Experten und Behandler, andererseits lässt es sich mit statistischen Verfahren, KI und ML aus Studiendaten extrahieren und mittels XAI plausibilisieren (vgl. Abschnitt 5.4 und Kapitel 6). Damit lassen sich zusätzlich digitale Lösungen realisieren, die entscheidungsrelevante Merkmale automatisiert bestimmen und die Entscheidungsgrundlage für Behandelnde transparent gestalten und optimieren (vgl. Abschnitt 3.3 und Kapitel 6). Holzinger et al. bezeichnen einen solchen Ansatz, indem Behandelnde in ein lernendes System mit einzubeziehen sind, als „*Doctor-in-the-Loop*“ [193]. Beispiele, wie solche entscheidungsunterstützenden Expertensysteme auf regel- [58, 275], ontologie- [228, 272], fuzzylogikbasierten [28, 100, 135, 271] oder hybriden [53, 310] Inferenzalgorithmen aufgebaut werden, liefert die Literatur. Die Schnittstellen zwischen den Anwendungen werden in einer service-orientierten Architektur mittels Mobile- und Web-Technologien umgesetzt, die aufgrund der Interoperabilitätsanforderungen eine standardisierte Form erfüllen müssen (vgl. Abschnitt 5.3). FHIR liefert hierfür die notwendige Grundlage um zukünftig auch das PVS der Behandelnden und die Versorgungsdaten beispielsweise aus der ePA und eFA der Betroffenen miteinander zu können (vgl. Unterabschnitt 2.4.1). Des Weiteren sind ab dem Jahr 2023 DiGA-Hersteller dazu verpflichtet mit der ePA durch die Verwendung von MIO interoperabel zu sein, sodass der Versicherte die damit erhobenen Daten freiwillig in die ePA übertragen und Dritten freigeben kann. Das eröffnet die Möglichkeit bestehende und damit validierte DiGA zu kombinieren. Insbesondere im Kontext der Sekundärprävention bei Demenzen könnten diese zur Früherkennung, der Verlaufsermittlung und Behandlung individueller Risikofaktoren über eine solche Plattform orchestriert werden. Da außerdem möglichst geringe Aufwände, die Abrechenbarkeit und Delegierbarkeit bei den Leistungserbringern als wichtigste Akzeptanzkriterien erachtet werden (vgl. Abbildung A.12), muss die Einbindung von medizinischem Fachpersonal entsprechend effizient gestaltet werden (vgl. Unterabschnitt 7.1.2). Zumindest initial müssen diese in der Lage sein, Risikofaktoren bei Leistungsempfängern zu diagnostizieren und einen individuellen Behandlungsplan zu erstellen sowie im Verlauf der Behandlung die von Patienten generierten Daten einsehen zu können. Auf eine detaillierte Auflistung von Systemanforderungen muss an dieser Stelle verzichtet werden, da diese Teil von zukünftigen Arbeiten darstellen. In Abbildung 7.5 lässt sich anhand des Prototyps der Anwendung für Leistungserbringende erahnen, in welche Richtung sich ein solches System entwickeln könnte.

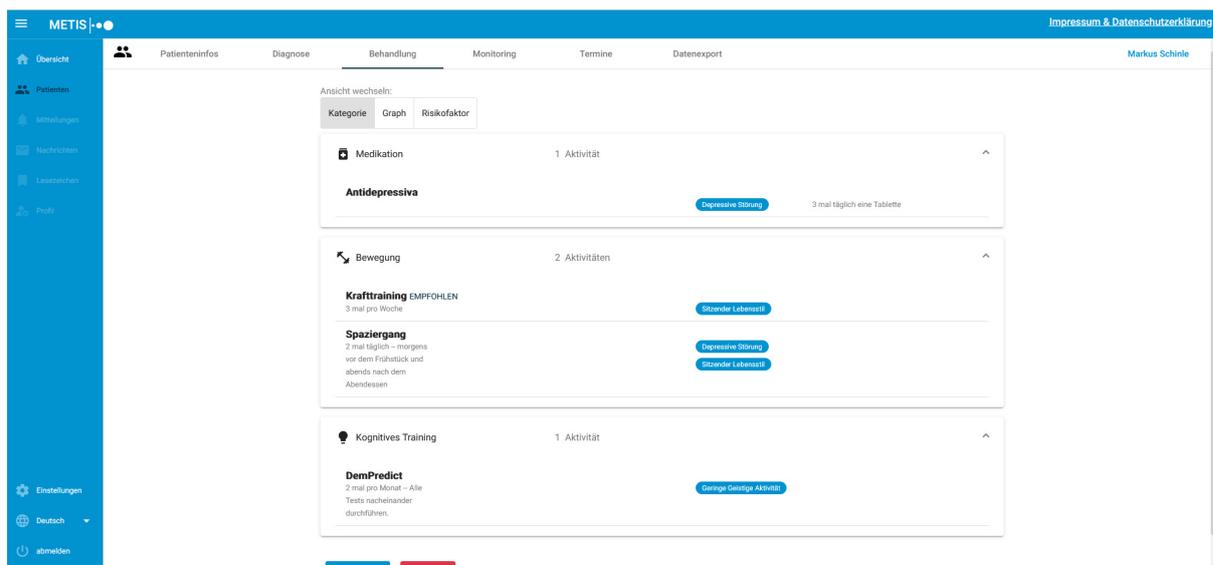


Abbildung 7.5: Prototyp der Anwendung für Leistungserbringer.

## 7.2.2 Patient-in-the-Loop

Zur Erfassung von durch Patienten generierte Daten (PRO) werden digitale Anwendungen notwendig, in denen die Durchführung von Früherkennungs-, Präventions- und Behandlungsmaßnahmen unterstützt werden. Grundsätzlich sind hierbei die Darstellung der Risikofaktoren und Behandlungsmaßnahmen sowie die Erinnerung und das Monitoring der Durchführung notwendig. Die Integration von bestehenden digitalen Anwendungen, die in Form von DiGA oder auch DiPA interoperable Daten generieren, bietet einerseits den Vorteil der Austauschbarkeit bei Unwirksamkeit, andererseits das Risiko von Wechselwirkungen durch unabhängig von einander getroffenen Entscheidungen bezüglich Interventionsmaßnahmen. Ähnlich wie bei einem Medikationsplan werden diesbezüglich Mechanismen notwendig, welche solche Widersprüchlichkeiten für alle an der Behandlung beteiligten Akteure aufzeigen. Daher stellen individuelle Anpassungen und die Erstellung eigener digitaler Interventionskonzepte eine Ergänzung zur reinen Kombination der bestehenden Anwendungen dar. Um damit durchzuführende neue Interventionen zu entwickeln und evaluieren, existieren Konzepte wie das Six Steps in Quality Intervention Development (6SQuID) von Wight et al. [418]. Hierbei wird das Problem und die entsprechenden Ursachen definiert und identifiziert, welche Faktoren in diesem Kontext bestehen beziehungsweise diese beeinflussen. Dabei wird sich darauf fokussiert, welche Faktoren das größte Veränderungspotenzial haben. Anschließend werden Veränderungsmechanismen identifiziert für welche zu klären ist, wie diese umzusetzen sind. Durch kontinuierliche Tests in kleinerem Maßstab findet solange eine iterative Entwicklung des Konzepts und damit der Anwendung statt, bis die Sammlung von Evidenz zur Effektivität in umfangreicher angelegten Interventionsstudien erfolgen kann [418, S. 521]. Solche Testungen könnten zukünftig in Praxen entwickelt und mit anderen Behandlern geteilt und evaluiert werden. Aus Sicht der Leistungserbringer spielt dabei zudem die Adhärenz der Leistungsempfänger eine Rolle, welche durch Anreizmechanismen beispielsweise aus dem Bereich der Gamification (vgl. Absatz 2.4.2) gesteigert werden kann. Die Erfassung behandlungsbegleitender Daten dient daher ebenfalls zum Monitoring. Hier können sich Anwendungen wie *DemPredict* einfügen, welche sich neben der Diagnoseunterstützung für Leistungserbringer auch zum kognitiven Training oder zur punktuellen Ermittlung des kognitiven Leistungsniveaus über den Krankheitsverlauf anbieten. Weitere Funktionen innerhalb mobiler Anwendungen, die beispielsweise von Wearables oder über das Betriebssystem erfasste Vitalparameter auslesen und die

Beteiligten zur Erinnerung an Interventionen benachrichtigen, funktionieren zeit- und ortsungebunden sowie temporär unabhängig von externen Infrastrukturen wie Internetverbindungen (vgl. Abbildung 7.6).

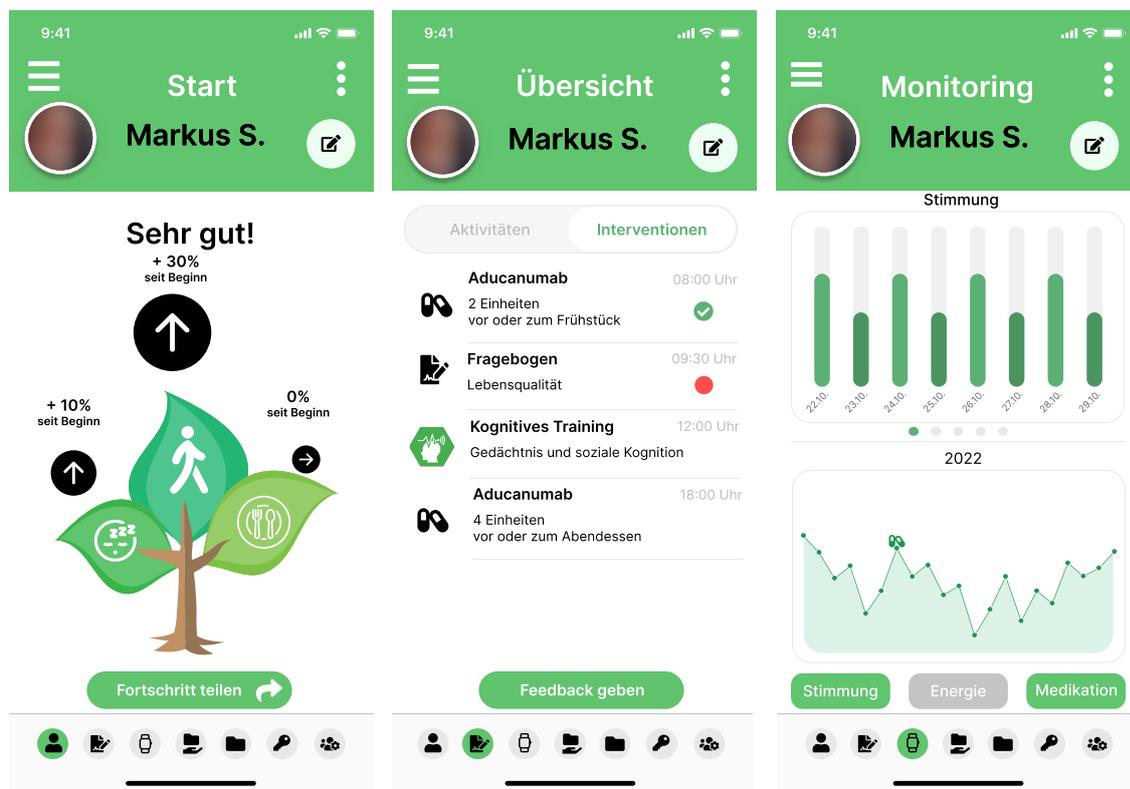
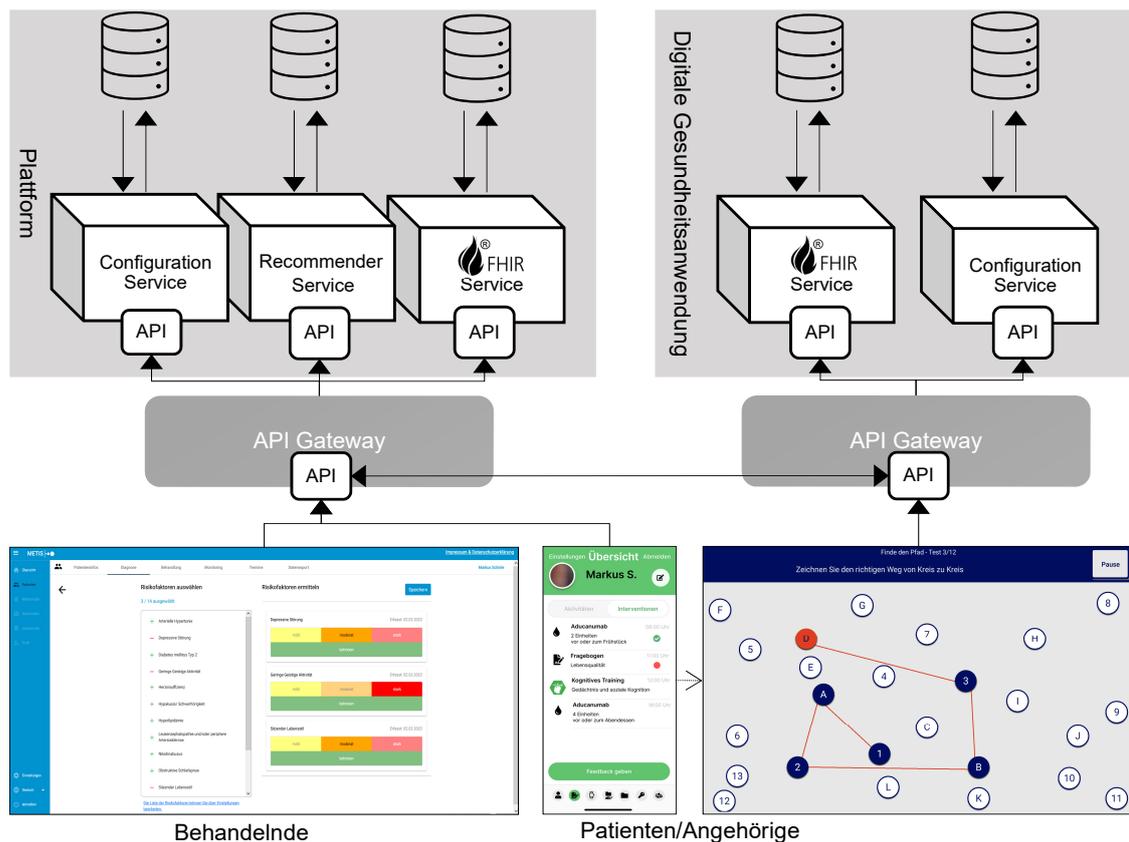


Abbildung 7.6: Mockup der Anwendung für Leistungsempfängende.

Aufgrund des progressiven Verlaufs des Krankheitsbildes ist der Einbezug von Angehörigen oder gar pflegenden Personen ab einem gewissen Zeitpunkt unablässig. Hierbei ist sowohl eine Unterstützung bei der Durchführung interventiver Maßnahmen als auch zur Fremdanamnese von Nöten, um einen „*Patient-in-the-Loop*“-Ansatz zu erreichen. In die Anwendung der Leistungsempfänger wird daher im Verlaufe der Erkrankung die Integration der Angehörigen und Pflegenden notwendig. Hierbei wird das Ziel verfolgt die Unterstützung der Betroffenen bei der Durchführung von Interventionen und perspektivisch die Schnittstelle zwischen Leistungserbringern, Angehörigen und Leistungsempfängern zu digitalisieren, um ein gesamtheitliches Gesundheitsprofil eines Betroffenen zu erlangen. Daraus kann eine effektivere und effizientere Behandlungsplanung, eine optimierte Betreuung sowie verbesserte Lebenssituation für Betroffene und deren Angehörige resultieren.

Das beschriebene Plattformkonzept zeigt, dass Psychosyndrome wie Demenz eine vielfältige Werkzeugpalette voraussetzen, da von individuellen Verlaufs- und Behandlungspfaden auszugehen ist. Wie die Umsetzung dieser Kombination aus digitalen Gesundheitsanwendungen und Plattform zur Konfiguration und Entwicklung von Präventions- und Behandlungsprogrammen aussehen kann, ist in Abbildung 7.7 skizziert.

Abbildung 7.7: Architektur der Präventions- und Behandlungsplattform.<sup>1</sup>

Um diese nachhaltig zu etablieren sind tragfähige Geschäftsmodelle notwendig, deren Rahmenbedingungen nachfolgend erörtert werden.

## 7.3 Markteinführung

Dieser Abschnitt setzt sich mit den Rahmenbedingungen für eine erfolgreiche Markteinführung auseinander. Außerdem werden organisatorisch-technische Implikationen der zuvor beschriebenen Anforderungen bezüglich einer datengetriebenen Entwicklung und damit verbundene offene Fragen skizziert.

### 7.3.1 Rahmenbedingungen für Markteintrittsstrategien

Den Zeitraum der Anfertigung dieser Arbeit (2016-2022) begleiten die Einführung neuer regulatorischer Anforderungen, welche sowohl Einfluss auf die Erforschung als auch die Entwicklung von digitalen Gesundheitsanwendungen und deren möglichen Markteintrittsstrategien haben. Um zu vermitteln, welche Aspekte bei der Translation in die medizinische Versorgung zu beachten sind, werden die Möglichkeiten des Markteintritts und die regulatorischen Anforderungen daran nachfolgend skizziert. Die in Abbildung 2.13 dargestellte Chronologie zeigt den gesetzlichen Einfluss der letzten Jahre auf das digitale Gesundheitswesen in Deutschland.

Auf der einen Seite stehen für die Konformitätsbewertung von Medizinprodukten die am 25. Mai 2017

<sup>1</sup> <https://www.hl7.org/fhir>, zuletzt zugegriffen am 12.05.2022.

in Kraft getretenen und seit Ende Mai 2021 geltende MDR und In-Vitro Diagnostic Medical Devices Regulation (IVDR) (vgl. Unterabschnitt 2.3.1). Dies kann ökonomische Hürden implizieren, die für eine notwendige (Re-)Zertifizierung der Hersteller genommen werden müssen. Für ein Medizinprodukt der Klasse I schätzt die Wirtschaftsprüfungsgesellschaft KPMG bei ihren Kunden dafür im Jahr 2019 allein für Personalkosten eine Spanne von 54.000 bis 67.000 € [237]. Hinzu kommen je nach Risikoklasse weitere Kosten für die klinische Bewertung und Prüfung, technische Dokumentation, Registrierung, die Zertifizierung des Qualitätsmanagementsystems und die Post-Market-Surveillance [237]. Auf der anderen Seite ist insbesondere bei der Verarbeitung von personenbezogenen Daten, den Datensubjekten ein erweiterter Schutz zuzumessen, der sich ebenfalls in technischen und organisatorischen Anforderungen niederschlägt.

Bei der Markteinführung einer digitalen Anwendung wie *DemPredict* ist dieser Aufwand zu berücksichtigen und kann Einfluss auf die Zweckbestimmung und damit Markteintrittsstrategie haben. Mit inhaltlich reduziertem Umfang könnte *DemPredict* als reines Informationsangebot platziert werden, das lediglich Wissen bezüglich Demenz und deren Diagnose- und Behandlungsmöglichkeiten bereitstellt sowie Informationen zu damit einhergehenden psychometrischen Tests anbietet. Dies setzt voraus, dass die Anwendung nicht zu diagnostischen oder therapeutischen Zwecken einzusetzen ist und keine Ableitung medizinisch aussagekräftiger Parameter stattfindet. Während die reine Darstellung von Informationen wie oben benannt, nicht in Anforderungen an ein Medizinprodukt mündet, stellt eine weitere Möglichkeit, um die Erkenntnisse der letzten Kapitel in der Realität zu nutzen, die Entwicklung der Anwendung als Medizinprodukt mit dem Zweck einer diagnostischen Ersteinschätzung dar. Hierfür werden demografische und psychometrische Daten erfasst und analysiert, welche eine Form von personenbezogenen Daten darstellen wodurch Aussagen über die psychische Gesundheit von natürlichen Personen getroffen und somit Auskunft über deren Gesundheitszustand gegeben werden kann (Art. 4 Abs. 15 DSGVO, vgl. Unterabschnitt 2.3.2). Wie zuvor erörtert handelt es sich dabei laut MDR um ein aktives Medizinprodukt und muss mindestens der Klasse IIa (vgl. Abbildung 2.2) zugeordnet werden [78, S. 72 ff.].

Diese zwei alternativen Ansätze der Verwertung lassen unterschiedliche Geschäftsmodelle zu. Wenn durch Anwendungen wie *DemPredict* die Qualität und Wirtschaftlichkeit der Versorgung nachweislich verbessert werden kann sind Kooperationen mit Krankenkassen denkbar. Über Selektivverträge können einzelne Krankenkassen, welche die Anwendung ihren Kassenpatienten anbieten wollen, als Anbieter gewonnen werden. Die Hersteller werden dafür zu einem ausgehandelten Tarif entsprechend entlohnt. Im Jahre 2020 zählten die Techniker Krankenkasse (TK) mit 10,6 Millionen vor der Barmer mit 8,97 Millionen und der DAK Gesundheit mit 5,6 Millionen der insgesamt 72,8 Millionen gesetzlich Krankenversicherten in Deutschland zu den drei größten GKVen [390]. Zusammen mit den Allgemeinen Ortskrankenkassen (AOK) und deren regionalen Ausprägungen decken diese bereits mit über 50 Millionen den größten Teil des deutschen Krankenversichertenmarktes und damit des Marktpotentials ab. Seit dem Jahr 2020 stellt nach dem DVG und der DiGAV die Aufnahme einer sogenannten DiGA ein zusätzliches Geschäftsmodell in der Regelversorgung dar (vgl. Unterabschnitt 2.3.1). Digitale Anwendungen können nach einem Prüfverfahren durch das BfArM als DiGA im DiGAVZ gelistet und auf Rezept erstattungsfähig verschrieben werden [83]. Zum Zeitpunkt der Anfertigung dieser Arbeit sind die Voraussetzungen dafür:

- ein auf digitalen Technologien basierendes
- Medizinprodukt der Klassen I oder IIa,
- welches nachweislich zu einem positiven Versorgungseffekt beiträgt und

- vom Patienten allein oder gemeinsam mit dem Leistungserbringer genutzt wird [78, 82].

Inwiefern eine digitale Screening-Anwendung wie *DemPredict* vor allem den vorletzten Punkt erfüllen kann bleibt offen, solange keine für die Beteiligten positiven Konsequenzen aus den Ergebnissen folgen. Hier könnte die in Abschnitt 7.2 skizzierte Plattformlösung eine Möglichkeit bieten, indem als Kombination daraus eine DiGA zur Sekundärprävention realisiert wird. Eine in diesem Kontext zusätzlich zu eruiende Problemstellung ist die Frage, wie die Anwendung einzuordnen ist sobald Angehörige den Patienten bei der Benutzung unterstützen müssen. Gerade in späten Phasen einer dementiellen Erkrankung und bereits eingeschränkten kognitiven Fähigkeiten kann dies notwendig werden (vgl. Unterunterabschnitt 2.2.3). Sobald durch eine solche Anwendung auch ein Mehrwert für pflegende Angehörige entsteht, sei es durch den Erhalt von Fähigkeiten des Patienten, könnte das Konzept der DiPA eine Rolle spielen.

Unter der Annahme, dass sich *DemPredict* als Medizinprodukt umsetzen lässt, bestehen die zuvor erörterten Möglichkeiten in den Markt einzutreten. Hierfür lässt sich die in Unterabschnitt 7.1.2 gestellte Frage **F9** aufgreifen, die das am zweithäufigsten genannte Akzeptanzkriterium der Vergütung beleuchtet (vgl. Abbildung A.12). Sollte eine Forderung der Übernahme durch die Krankenkassen mittels Selektivverträgen, eine DiGA oder im Rahmen der Gebührenordnungspositionen 03360 (hausärztlich-geriatisches Basisassessment) und 03362 (hausärztlich-geriatischer Betreuungskomplex) nicht erfüllbar sein, können Ärzte auch sogenannte IGeL anbieten und abrechnen. Diese sind freiwillig und die Kosten müssen in der Regel von den Patienten selbst getragen werden. Hierfür hat eine Teilmenge ( $n = 127$ ) der 301 Ärzte auf Frage 36 zahlenmäßige Preisschätzungen zwischen 5 € und 140 € abgegeben, welche als Orientierungshilfe für eine potentiellen Preisbildung dienen können (angegebene Preisspannen wurden arithmetisch gemittelt:  $\bar{x}_{Preis} = 28,49$  €,  $\sigma_{Preis} = 18$ ,  $\tilde{x}_{Preis} = 25$  €). Die breite Streuung lässt sich teilweise dadurch erklären, dass der Preis in Abhängigkeit zu einem damit verbundenen noch unbekanntem Aufwand angesetzt wurde, was sich auch in zusätzlichen Kommentaren zu den Preisen niederschlägt. Hierin werden Schätzungen auf 1-2 € pro Minute Aufwand abgegeben, welche die zeitliche Dimension mit berücksichtigen.

Die in diesem Abschnitt dargelegten Ansätze setzen alle die Bereitschaft und das Vertrauen von Leistungsempfänger und Leistungserbringer voraus auf Grundlage der mittels einer digitalen Lösung erhobenen Daten die Ableitung richtiger Folgeschritte zu treffen. Die Einbindung der Leistungsträger wird insbesondere für Selektivverträge und für den Nachweis eines positiven Versorgungseffekts einer DiGA notwendig. Ob und in welchem Umfang Leistungsträger zur Übernahme solcher Leistungen bereit sind, muss zukünftig untersucht werden.

### 7.3.2 Organisatorisch-technische Implikationen

Die im Rahmen dieser Arbeit gewonnenen Erkenntnisse heben hervor, dass digitale Gesundheitsanwendungen wie *DemPredict* in Abhängigkeit von deren Zweckbestimmung als SaMD einzustufen und daher als Medizinprodukt zu klassifizieren sein können. Bezüglich der Konformitätsbewertungsverfahren für ein SaMD müssen daher die Anforderungen der MDR erfüllt sein (vgl. Unterunterabschnitt 2.3.1). Welche zusätzlichen Normen dafür zu berücksichtigen sind, ist in der existierenden Literatur dokumentiert (vgl. Unterunterabschnitt 2.3.1). Nachfolgend wird deshalb lediglich auf die Besonderheiten bei der Konformitätsbewertung eingegangen, die bezüglich der Sicherheit von ML-/KI-Komponenten auf dem Weg zur SaMD zusätzlich zu antizipieren sind. Zum Zeitpunkt der Fertigstellung dieser Arbeit fehlt diesbezüglich eine harmonisierte Norm, die dafür notwendigen Punkte werden jedoch weltweit von Experten, Regu-

lierungsbehörden und in der wissenschaftlichen Literatur bereits diskutiert [148, 149, 162, 180, 213]. Johner et al. leiten zur Anwendung von KI und ML in Medizinprodukten über 150 Anforderungen ab, an denen sich Hersteller, die den Einsatz dieser Verfahren in ihren Produkten im europäischen und nordamerikanischen Markt vorsehen, orientieren können [213]. Als eine Anforderung für die klinische Bewertung und das Risikomanagement zur Konformitätsbewertung sehen diese, dass der Hersteller begründen kann, weshalb der Einsatz eines Verfahrens des ML einem anderen Verfahren überlegen ist, welche Risiken damit einhergehen und wie damit umgegangen wird [213]. Im Jahr 2018 haben in den USA bereits ML-basierte SaMD eine Marktzulassung durch die FDA erhalten [162]. Gerke et al. berichten von einem autonomen System, das Screenings nach einer Augenkrankheit ermöglicht [162]. Darin integrierte Algorithmen werden als „Locked Algorithms“ bezeichnet, weil daran nach der Zulassung keine Anpassungen durch beispielsweise eine Änderung der Leistungsfähigkeit von eingesetzten Modellen durch zusätzliche Daten über die Zeit erlaubt sind [162]. Eine Veränderung zieht derzeit eine erneute Prüfung durch die FDA nach sich [162]. Lernende Systeme zu entwickeln, die aus deren Anwendung im Feld neu gewonnene Daten integrieren und sich dadurch adaptiv optimieren können (Reinforcement Learning), unterliegen daher aktuell diesen regulatorischen Beschränkungen [162]. Für dieses „Update Problem“ werden kontinuierliche Monitoring-Prozesse gefordert, welche die damit einhergehenden Risiken überwachen [162]. Gerke et al. schlagen diesbezüglich einen „System Ansatz“ vor, da im Gegensatz zu Medikamenten die Varianz der Performanz solcher Systeme zwischen der Test- und Realumgebung von unterschiedlichen Umgebungsfaktoren wie dem Datenqualitätsmanagement oder den Kenntnissen der Anwender im Zuge der Nutzung abhängt, was somit dazu führt, dass das Gesamtsystem in dessen Anwendungskontext betrachtet werden sollte [162]. Auf diese Diskussionen hin veröffentlichte die FDA im Jahr 2021 einen Aktionsplan für den Umgang mit KI- und ML-basierter SaMD [149]. Die zu verfolgenden zukünftigen Ziele dabei sind die Erstellung eines Rahmenwerks inklusive Vorlagen für Änderungskontrollpläne wie dem „Algorithm Change Protocol“ und der Harmonisierung von guten Praktiken für den Umgang mit ML [149]. Außerdem soll ein patientenzentrierter Ansatz entwickelt werden, der eine gewisse Transparenz für die Nutzer beinhaltet [149]. Methoden, die Verzerrungen und die Robustheit von Algorithmen berücksichtigen sowie Prozesse zur realweltlichen Leistungsüberwachung dieser Systeme sollen zukünftig ebenfalls adressiert werden [149]. All diese organisatorisch-technischen Faktoren stellen Hersteller und Regulatoren vor neue Herausforderungen, welche sich auch durch die Frage charakterisieren lassen, ob insbesondere medizinische Entscheidungen auf Basis der Daten dieser Systeme durch Mediziner anders getroffen werden, wenn diese wissen, dass die Entscheidungsgrundlage grundsätzlich erklärbar ist oder wenn sie glauben, dass die Ergebnisse des Algorithmus nicht nachvollziehbar sind [162]. Die daran anknüpfenden Herausforderungen sind Ausgangspunkt zukünftiger Arbeiten. Aus Sicht der Leistungserbringer zeigt sich zumindest eine aktuelle Tendenz hin zu einer transparenten Bewertungsmetrik und keiner autonomen Klassifikation (vgl. Unterabschnitt 7.1.2).

# 8 Diskussion der Ergebnisse

DIESER Abschnitt diskutiert die Ergebnisse der durchgeführten Studien zur abschließenden Evaluation der erarbeiteten Artefakte in Bezug auf die definierten Forschungsfragen. Da der Entwicklungsprozess am Beispiel eines ambulanten Systems zur Demenzfrüherkennung agil erfolgt, werden mehrere Iterationen und Phasen unterschieden. Die Ergebnisse der aktuellsten Inkremente werden demnach interpretiert sowie deren Auswirkungen aufeinander erörtert. Die entsprechenden Teilaspekte sind der bestehenden Kapitelstruktur dieser Arbeit folgend untergliedert.

## 8.1 Digitalisierungsphase

Die Phase der Digitalisierung setzt sich zum Ziel die Datengenerierung und Informationsbereitstellung nutzerzentriert zu gestalten. Hierfür werden Anforderungen erhoben woraus sich ein, den gesamten Entwicklungsprozess begleitendes, Vorgehensmodell ergibt, welches anhand des Anwendungsfalls *DemPredict* evaluiert wird (vgl. Kapitel 4). In der ersten Phase wird hierbei die Digitalisierung und damit eine Zieldefinition sowie die Koordination dafür zu erfassender Parameter modelliert und diskutiert. Gefolgt von einer Referenzarchitektur für digitale Gesundheitsanwendungen findet die Technologieauswahl für eine zu implementierende Lösung statt. Die Schnittstelle zwischen Mensch und Maschine wird als Einstiegspunkt der Entwicklung betrachtet. Die Gebrauchstauglichkeit der digitalen Gesundheitsanwendung in Kombination mit dem vorgesehenen Interaktionsmedium muss deshalb kontinuierlich geprüft und sichergestellt werden (vgl. Absatz 2.4.2). Dafür existierende Evaluationskriterien ergeben sich für einen gebrauchstauglichkeitsorientierten Entwicklungsprozess aus den entsprechenden Normen (vgl. Unterabschnitt 2.3.1). Die Entwicklung einer grafischen Benutzerschnittstelle wird durch formative Bewertungen methodisch von Nutzerstudien mit insgesamt 91 Probanden begleitet (vgl. Unterabschnitt 5.2.2). Da im Kontext dieser Arbeit vier Iterationen durchgeführt wurden, wird nachfolgend dargestellt, inwiefern sich die Usability von *DemPredict* durch die Implementierung der identifizierten zielgruppenspezifischen Anforderungen von Iteration III bis Iteration IV verändert hat. Hierfür wird ein für die Zielgruppe altersstrukturell spezifisches Nutzerkollektiv befragt. Die Einstellung dieser zur Technik und deren -erfahrung wird mittels zuvor eingeführtem Fragebogen erfasst (vgl. Abschnitt A.7). Die abschließende Studie umfasst elf Anwender im Alter von 52 bis 79 Jahren und entspricht somit der Zielgruppe (vgl. Abschnitt 4.2 und Anhang A.25). Die in Unterabschnitt 5.4.3 eingeführten Indices zur persönlichen Einstellung gegenüber Technik (PE,  $\bar{x}_{PE} = 2,47$ ,  $\sigma_{PE} = 0,64$ ) und der Erfahrung im Umgang mit Technik (TE,  $\bar{x}_{TE} = 2,59$ ,  $\sigma_{TE} = 0,8$ ) zeigen, dass die Befragten dem Thema Technik und deren Nutzung neutral gegenüberstehen. Im Allgemeinen werden von den Studienteilnehmern damit Informationsmöglichkeiten, Fortschritt, Kommunikation sowie eine Erleichterung von Arbeitsabläufen und des alltäglichen Lebens verbunden. Erfahrungen im Umgang mit Technik liegen bei jedem Teilnehmer vor, hierbei werden Erfahrungszeiträume von mindestens zwölf Jahren im Umgang mit Computern angegeben. Weitere Details zur Einstellung und Erfahrung bezüglich Technik finden sich in Tabelle A.24. Mit diesem Teilnehmerkollektiv fand auf Grundlage definierter Szenarien in Kombination mit der TAM eine formative Evaluation

statt. Als Maß zur Bewertung wurde der ISO-Norm-Fragebogen (vgl. Abschnitt A.11) herangezogen, womit jeweils eine Bewertung von *DemPredict* auf dem Stand von Iteration III und den Entwurf für die Neugestaltung in Iteration IV durchgeführt wurde. Die Richtlinie „Individualisierbarkeit“ der Norm „Grundsätze der Dialoggestaltung“ wird in der neuen Norm „Interaktionsprinzipien“ durch die Richtlinie „Benutzerbindung“ ersetzt. Da der Fragebogen zum Zeitpunkt der Durchführung dieser Befragung noch nicht aktualisiert und revalidiert wurde, werden diese neutral mit 0 Punkten bewertet, um zu vermeiden, dass die allgemeine Bewertung dadurch beeinflusst wird. Für Iteration III liegen die Bewertungen der einzelnen Fragen demnach zwischen 2,27 und -0,45. Das arithmetische Mittel der Bewertungen aller Teilnehmer beträgt 0,73 ( $\bar{x}_{\text{Bewertungen}} = 0,91$ ) und liegt somit unter dem Indikatorwert von 1, was als Optimierungsbedarf im Hinblick auf die Usability beziehungsweise die Normenkonformität gewertet werden muss. Die durchschnittlichen Bewertungen der Richtlinien für die „Selbstbeschreibungsfähigkeit“ (0,18), „Steuerbarkeit“ (0,27) und „Robustheit gegenüber Benutzerfehlern“ (0,24) liegen ebenfalls unter dem Richtwert von 1. Diese Richtlinien stellen daher besondere Schwachstellen von *DemPredict* in Iteration III dar. Demnach kann eine Umgestaltung in Iteration IV insbesondere in diesen Bereichen Verbesserungspotential heben. Der GUI-Entwurf für Iteration IV implementiert den in dieser Arbeit entwickelten Leitfaden (vgl. Unterabschnitt 5.2.2) und ist auszugsweise in Abschnitt A.26 dargestellt. Die dafür abgegebenen durchschnittlichen Bewertungen der einzelnen Fragen liegen zwischen 2,64 und 0,36. Alle durchschnittlichen Bewertungen der Richtlinien liegen dabei über dem Richtwert von +1 und weisen somit eine durchgehend positive Usability und Konformität mit der Norm auf. Der Leitfaden zeigt sich also als anwendbar und tauglich gegenüber der zielgruppenspezifischen Gestaltung für Anwendungen wie *DemPredict*. Insgesamt konnte damit die Bewertung im Mittel von 0,73 auf 1,58 ( $\bar{x}_{\text{Bewertungen}} = 1,82$ ) erhöht werden. Ein detaillierter Vergleich der Bewertungen ist in Anhang A.25 zu finden. Insbesondere die größeren Bedienelemente, die Vereinheitlichung der Gestaltung und die Übersichtlichkeit der Anwendung werden von den Studienteilnehmern als Verbesserungen genannt und führen zu einer höheren Selbstbeschreibungsfähigkeit, Robustheit gegenüber Benutzerfehlern und Steuerbarkeit. Generell zeigen die Studienteilnehmer Interesse an *DemPredict* und geben an die Anwendung als solche gerne zukünftig verwenden zu wollen. Bezüglich des Umgangs mit der Anwendung werden Tablet-Computer gegenüber Smartphones aufgrund der Bildschirmgröße bevorzugt. Die TAM in Kombination mit dem ISO-Norm-Fragebogen zeigen sich für den Anwendungskontext als valide Werkzeuge zur Gestaltung einer Mensch-Maschine-Schnittstelle. Wie den Entwicklungsprozess begleitend mit digitalen Gesundheitsanwendungen erhobene Daten verwaltet und letztlich in aufkommende Gesundheitsdatenökosysteme integriert werden könnten, zeigt die Gegenüberstellung und beispielhafte Implementierung aktueller Standards in Abschnitt 5.3. Technologische Möglichkeiten wie diese Daten interoperabel beschrieben werden sowie damit einhergehende Datennutzungsbedingungen zu gestalten sind, konnten darin aufgezeigt werden. Durch empirische Untersuchungen wurde deren Notwendigkeit verifiziert sowie weitere Anforderungen insbesondere hinsichtlich Standardisierungsbestrebungen daran abgeleitet und priorisiert. Diese zeigen auch den Reifegrad der damit verbundenen Datensouveränitätstechnologien im Kontext der rechtlichen Anforderungen. Zusätzlich werden in dieser Arbeit Methoden zur Überprüfung der Qualität damit erhobener Daten im Zuge der Vorverarbeitung notwendig. Diese werden in Abschnitt 5.4 durch die Identifikation und Untersuchung von Einflüssen der Digitalisierung mittels statistischer Verfahren adressiert und implizit evaluiert. Hierfür geeignete Maße hängen von der Art und Menge zur Verfügung stehender Daten ab.

## 8.2 Automatisierungs- und Optimierungsphase

Bezüglich der in Abschnitt 6.1 durchgeführten Automatisierungsansätze für Test J zeigt sich, dass auch hierdurch entsprechende Risiken durch Ungenauigkeiten in solche Systeme gebracht werden. Hierbei gilt es eine initial möglichst transparente Bewertungsmetrik zu finden, die eine gewisse Fehlertoleranz voraussetzt. Zur Bewertung der Parameter zeigen sich in Tabelle 6.4 für den maximierten Cohen's Kappa schwache bis deutliche Übereinstimmungen zwischen menschlichen und algorithmischen Entscheidern. Hierbei ist jedoch, wie am Beispiel von PD bei gesunden Probanden (G) ersichtlich wird, der Cohen's Kappa gleich null, da keiner der gesunden Studienteilnehmer die Dreidimensionalitätsdimension nicht reproduzieren konnte. Dieser Umstand zeigt den Mehrwert dieses Maßes zur Optimierung und verdeutlicht die Notwendigkeit heterogener Datenmengen, welche jedoch aufgrund der Beschaffenheit der Tests in Bezug zum Krankheitsbild für gewisse Gruppen nicht zu erwarten sind. Dadurch geht auch hiermit eine Abwägungsentscheidung einher, wodurch die Aufwände der Automatisierung denen einer Auswertung durch menschliche Entscheider sowie deren individuellen Fehleranfälligkeiten gegenübergestellt werden muss.

Die vergleichende Implementierung der verschiedenen ML-Modelle und schwellwertbasierten Verfahren zeigt, dass es in den meisten Fällen möglich ist, die Studienteilnehmer entsprechend ihrer Diagnosegruppe anhand von *DemPredict* korrekt zu klassifizieren. Die nachfolgend bezüglich Abschnitt 6.3 beschriebene Zusammenfassung der Ergebnisse ist in Teilen bereits publiziert [471]. Zur Beurteilung der Screening-Leistung wird die Sensitivität der Klasse NG als Hauptkriterium betrachtet. Um die Leistungsfähigkeit der gelernten Modelle zu beurteilen, bietet sich ein Vergleich von *DemPredict* mit anderen digitalen Demenz-Screeningwerkzeugen an. Die Arbeit von Thabtah et al. stellt die im Jahr 2020 verfügbaren mobilen Screening-Anwendungen gegenüber [402]. In Bezug auf das Ziel von *DemPredict* ist jedoch nur eine Vergleichbarkeit für die Lösungen *Brain Test* und *DST* festzustellen, da diese neuropsychologische Tests anbieten, die ohne Hilfe von zu Hause aus durchgeführt werden können (vgl. Unterunterabschnitt 3.3.2). Sie beinhalten eine erste Einschätzung bezüglich einer Diagnose und für ihre digitalisierten Versionen sind Leistungsdaten verfügbar. Als weitere Referenz wird das MoCA gewählt, das zur Unterstützung von medizinischem Fachpersonal entwickelt wurde und von dessen Entwicklern als das sensitivste kognitive Screeningwerkzeug bezeichnet wird [297]. Da für den *DST* in der Übersichtsarbeit von Thabtah et al. nur die Sensitivität bei 96%, aber keine Spezifität angegeben wird, ist dieser Wert nicht vergleichbar [402]. Angaben auf deren Homepage zufolge erreicht der *DST* eine Spezifität von 91% [373]. Da sich die angegebenen Leistungsdaten auf eine Studie zu einem digitalisierten Test beziehen, welcher im Gegensatz zur mobilen Lösung des *DST* an einem Desktop-PC evaluiert wurde, werden diese Daten daher nicht weiter betrachtet. Der *Brain Test* (Sensitivität 71%, Spezifität 90%) und die MoCA App (Sensitivität 81%, Spezifität 86%) schneiden geringfügig schlechter ab als *DemPredict* unter Verwendung des RF Klassifikators (durchschnittliche Sensitivität 83%, Spezifität 93%). Sensitivität und Spezifität sind jedoch nicht die einzigen Merkmale der Anwendungen, die verglichen werden sollten. Weitere relevante Kriterien sind die Testdauer, die Abdeckung der DSM-V-Kriterien sowie die unterstützten Geräte und Betriebssysteme. Hinsichtlich der Gesamttestdauer sind *Brain Test* und *DST* mit Bearbeitungszeiten zwischen 10 und 15 Minuten sowie das MoCA mit maximal 10 Minuten kürzer als *DemPredict*, jedoch findet beim *Brain Test* und MoCA eine Auswertung durch menschliche Experten statt. Die Daten der *DemPredict*-Studie zeigen bei Berücksichtigung von den nach der WFS relevanten Tests eine durchschnittliche Bearbeitungszeit bei circa 12 Minuten ( $\bar{x}_{\text{Bearbeitungszeit}} = 11$  Minuten) und sind damit vergleichbar mit denen der anderen Screeninglösungen. Bei Abdeckung aller DSM-V-Kriterien wird eine durchschnittliche Bearbeitungszeit von circa 33 Minuten ( $\bar{x}_{\text{Bearbeitungszeit}} = 30$  Minuten) erreicht. Da für die CERAD-Testbatterie bei

weniger Kriterien eine Gesamtbearbeitungszeit von 45 bis 60 Minuten für die gleichen Teilnehmer in dieser Studie benötigt werden, bietet die Digitalisierung einen Effizienzgewinn. Insgesamt kristallisiert sich durch diese Art der datengetriebenen Optimierung eine Abwägungsentscheidung zwischen Genauigkeit, Durchführungsdauer und Leitlinienkonformität der Testung heraus, welche je nach angestrebter Zieldefinition für eine zukünftige Weiterentwicklung getroffen werden muss.

Um die geforderte Transparenz für die verwendeten ML-Algorithmen zu erhalten bieten sich unterschiedliche XAI-Ansätze an. Eine hohe Wiederholungsgenauigkeit erreichen die modell-spezifischen Verfahren sowie SHAP aufgrund der gezeigten stärker ausgeprägten Stabilitätseigenschaften. Sind kurze Berechnungszeiten gefordert, ist das LIME-Verfahren besonders geeignet. Ein Vergleich der Vertrauenswürdigkeit zwischen allen Verfahren ist nur bedingt möglich. Xk-NN erzeugt via Definition vertrauenswürdige Ergebnisse. Bei den modell-agnostischen Verfahren zeigen die MoRF-Kurven, dass SHAP grundsätzlich besser geeignet ist. Zur besseren Vergleichbarkeit findet der SHAP in seiner ursprünglichen Form Anwendung, insbesondere für die Nutzung des RF sollte der TreeSHAP in diesem Kontext zukünftig weiter untersucht werden. Aufgrund der Konstruktion des CHIRPS ist kein direkter Vergleich mit den anderen Verfahren möglich, die Performanzmaße zeigen jedoch, dass eine grundsätzliche Vertrauenswürdigkeit der Erklärungen anzunehmen ist. Zusammengefasst dienen die XAI-Verfahren als Plausibilisierungswerkzeuge für ML-gestützte Ergebnisse, zeigen aber auch die Abwägung auf, dass im Zweifel entsprechend genauere Modelle mit geringer Transparenz weniger genauen transparenten Verfahren weichen müssen.

### **8.3 Herausforderungen und Potentiale in der Translation**

Neben der Darstellung möglicher Verwertungsperspektiven und den daran gekoppelten ökonomischen und regulatorischen Implikationen stehen die Adhärenz und Akzeptanz bei den Leistungsempfängern beziehungsweise -erbringern im Fokus einer Translation. Die Erhebung der Anforderungen und damit verbundenen Akzeptanzkriterien erfolgt, neben den abschließend durchgeführten Umfragen, von Beginn an als Teil der Anforderungsanalyse im Rahmen der Entwicklungszyklen mit Experten und Nutzern. Diese Umstände begleiten den gesamten Entwicklungsprozess bis hin zu einer möglichen CE-Kennzeichnung der digitalen Gesundheitsanwendung. Hiermit verbundene Herausforderungen und Potentiale werden nachfolgend zusammengefasst sowie deren Abhängigkeiten gegenübergestellt.

Wie die vorhergehenden Kapitel andeuten, stellen die auf technischer, medizinischer, ökonomischer und regulatorischer Ebene zu erfüllenden Anforderungen die übergeordneten Herausforderungen bei der Digitalisierung von Gesundheitsanwendungen dar (vgl. Abschnitt 4.1). Neben der Wahl einzusetzender Technologien (vgl. Unterabschnitt 5.1.4) spielen hierbei auch die damit einhergehenden Interaktionskonzepte eine Rolle, um potentielle Nutzer anzusprechen (vgl. Abschnitt 5.2). Aufgrund der kontinuierlichen Weiterentwicklung regulatorischer Rahmenbedingungen entstehen vermehrt Ansprüche an die Verwaltung und Vernetzung damit erfasster digitaler Gesundheitsdaten. Hierbei stehen die individuellen Rechte der Datensubjekte an ihren Daten, deren Nutzen für die Allgemeinheit gegenüber [314]. Im Rahmen dafür entstehender Datenökosysteme ist dieser Umstand ebenfalls für digitale Gesundheitsanwendungen in den entsprechenden Anwendungsdomänen individuell zu berücksichtigen. Zwar zeigt sich der Reifegrad damit verbundener Standards und Technologien zum Zeitpunkt der Anfertigung dieser Arbeit noch als unzureichend für eine rechtssichere Lösung (vgl. Abschnitt 5.3). Dennoch stellt aus Sicht von Herstellern die Anpassung ihrer digitalen Gesundheitsanwendungen an diese Anforderungen eine zukünftige Herausforderung dar. Um zielgenau den Bedarf von Nutzern zu adressieren sehen agile Entwicklungs-

prozesse außerdem möglichst viele Zyklen zur Integration von neu gewonnenem Wissen durch Feedback von Anwendern vor. Damit die Validität der digitalisierten Verfahren geprüft werden kann, müssen daher dafür entwickelte Komponenten auf Systemebene mittels etablierten Methoden unter realen Nutzungsbedingungen regelmäßig auf den Prüfstand gestellt werden können (vgl. Abschnitt 5.4). Die Durchführung hierfür notwendiger Nutzerstudien setzt einen kontinuierlichen Zugang zu Studienteilnehmern aller Interessenvertretergruppen voraus. Insbesondere im Kontext gerontopsychiatrischer Erkrankungen stellt die Studienteilnehmerakquise, wie beispielsweise anhand der geringen Teilnahmebereitschaft und hohen Drop-Out-Quote von circa 26% (33% nach Vorverarbeitung der Daten, vgl. Tabelle A.3) über den Studienzeitraum ersichtlich wird, eine Herausforderung dar [152]. Diese in Unterabschnitt 5.4.1 beschriebene Studie zeigt, dass für dementielle Syndrome entsprechend repräsentative Datenmengen schwer verfügbar zu machen sind. Die Erhebung zusätzlicher Daten ist mit Aufwänden verbunden und hängt von externen Faktoren, wie der Bereitschaft aller Akteure zur Durchführung weiterer Studien zusammen. Zusätzlich kann die Qualität der erhobenen Daten durch die dabei eingesetzten Technologien, von Umgebungsbedingungen der geplanten Einsatzszenarien sowie menschlichen Faktoren der Zielgruppen beeinflusst werden (vgl. Abschnitt 5.4). Um die Potentiale der über digitale Gesundheitsanwendungen gesammelten Daten zu heben, muss daher eine möglichst heterogene Datenbasis geschaffen werden.

Vorteile der Digitalisierung liegen in der Möglichkeit zur Automatisierung und Optimierung bestehender Vorgehensweisen. Die Automatisierung steht insbesondere in Bezug zur Unabhängigkeit von Ort und Zeit der Nutzung. Die Entwicklung hierfür potentiell einzusetzender KI- oder ML-gestützter Produkte impliziert in diesem Kontext zu Beginn jedoch ein Henne-Ei-Problem, da die Nutzung einen Mehrwert für die Dateneigentümer voraussetzt, der zunächst nicht in direkter Verbindung mit den dafür zur Verfügung gestellten Daten stehen kann. Deshalb muss ein diesbezüglicher Katalysator für alle Interessenvertreter durch zusätzliche Anreizmechanismen entstehen, welcher als Bestandteil zukünftiger Forschungsarbeiten vorzusehen ist. Außerdem geht die Wahl des geeigneten Klassifikationsmodells mit regulatorischen Anforderungen einher, welche bezüglich der Transparenz solcher Verfahren sowie der Nachvollziehbarkeit darauf basierender Entscheidungen zu stellen sind und damit beeinflussen, inwieweit diese Verfahren als Bestandteil eines Produkts geeignet sind (vgl. Abschnitt 6.4). Auch Kapitel 7 zeigt, dass aus Sicht der menschlichen Entscheider die Operationalisierung der Bewertungsmetrik transparent und nachvollziehbar sein muss. Letztlich ist die Auswahl des trainierten Modells auch eng mit ethischen und gesundheitspolitischen Fragen verbunden, die ein akzeptables Niveau von Fehlern erster und zweiter Art bei den Ersteinschätzungen definieren. Auch auf sich ändernde Anforderungen, die nicht zuletzt auf Anpassungen in der gesellschaftlichen Wahrnehmung zurückzuführen sein können, lässt sich durch Adaption dieser Leistungsmaße gezielt eingehen. In Abschnitt 6.1 wird gezeigt, welche Herausforderungen und Potenziale bei der automatisierten Erfassung und Auswertung einzelner Leistungsparameter bestehen. Zum einen zeigt der Einsatz von digitalen Technologien deren Grenzen im Kontext von Genauigkeit und Fehlertoleranz bei der Eingabe auf. Andererseits sind damit, wie am Beispiel der Sprachanalysesoftware in Test B deutlich wird, Interaktionsparadigmen verbunden, die bei der entsprechenden Zielgruppe etabliert sein müssen. Die verfügbare Datenmenge, insbesondere von bereits erkrankten Studienteilnehmern, stellt auch hier ein Kriterium dar, welches den konsequenten Einsatz moderner Verfahren des ML und der KI limitieren. Durch die Automatisierung entstehende Ungenauigkeiten liefern schließlich eine zusätzliche potentielle Fehlerquelle, die innerhalb des Gesamtergebnisses berücksichtigt werden muss. Um das Vertrauen in digitale Gesundheitsanwendungen zu stärken, muss ebenso beachtet werden, dass fehlerhafte Systeme weitreichende Folgen haben und damit eine Gefahr für die entsprechende Zielgruppe darstellen können [319]. Des Weiteren stellt dies aus Perspektive der Translation insbesondere auch eine Herausforderung bezüglich der Konformitätsbewertung für sich kontinuierlich weiterentwickelnde Algorithmen dar. Hierfür geeignete Konzepte sind zum Zeitpunkt der Anfertigung dieser Arbeit noch Teil

der Forschung. Zusammenfassend stellen die Herausforderungen bei der Optimierung die Merkmalsselektion für das darauf basierende Modelltraining, die Qualität und Quantität verfügbarer Datenmengen, die Transparenz der Verfahren sowie der Übergang von datensparsamen Algorithmen hin zu Verfahren, deren performante Anwendung ausreichend große Datenmengen voraussetzen, dar. Demgegenüber besteht vor allem die Chance im Einsatz von KI und ML in Medizinprodukten mittels datengetriebenen Verfahren weitere bisher unbekannte Parameter als signifikant in Bezug auf gesundheitsrelevante Ziele zu identifizieren, welche erst durch die digitale Erfassung und Kombination mit weiteren Daten sowie damit verbundenen Auswertemöglichkeiten untersucht werden können (vgl. Abschnitt 3.3). Diese Erweiterbarkeit bestehender Metriken und die Möglichkeit einer datengetriebenen Abbildung von Abhängigkeiten zwischen den Merkmalen sind als Potentiale hervorzuheben. Welche darin zu findenden Merkmale mit in die Bewertung einfließen, stellt am Beispiel von *DemPredict*, eine Abwägung zwischen der Performanz und den DSM-V-Kriterien dar. Ob den medizinischen Leitlinien folgende Prüfkriterien zu berücksichtigen sind, auch wenn diese für die klinische Diagnose lediglich als Empfehlung dienen, muss Teil des weiteren wissenschaftlichen Diskurses sein. Insbesondere wenn die Zweckbestimmung eines Medizinproduktes verfolgt wird, sind jedoch klinische Bewertungen für die Konformitätsbewertung vorgeschrieben sowie der Nachweis positiver Versorgungseffekte, falls eine langfristige Aufnahme in das DiGAVZ angestrebt wird. Leitlinienkonforme Entwicklungen könnten beim diesbezüglichen Studiendesign dabei helfen einen erwartbaren positiven Effekt zu erhalten. Welche weiteren Markteintrittsstrategien für Anwendungen wie *DemPredict* existieren fasst Abschnitt 7.3 zusammen. Insbesondere durch die sich geänderten gesetzlichen Rahmenbedingungen des DVG entstand für Hersteller von patientenzentrierten Gesundheitsanwendungen über die DiGA ein neues Geschäftsmodell, um über den ersten Gesundheitsmarkt potentielle Kunden niederschwelliger zu erreichen. Zum Zeitpunkt der Beendigung dieser Arbeit besteht dieses Modell seit knapp 20 Monaten, daher lässt es noch kein Urteil über dadurch langfristig realisierte Chancen zu. Eine Erweiterung dieses Konzepts durch die DiPA ist ebenfalls bereits beschlossen. Die dadurch entstehende Abrechenbarkeit für Leistungserbringer könnte letztlich auch den Weg zu personalisierten Präventions- und Interventionsprogrammen ebnen, die von zeitaufwändigen und datenintensiven Behandlungsmaßnahmen begleitet, nun aber durch IKT effizienter und effektiver gestaltet werden können (vgl. Abschnitt 7.2). Insgesamt bergen die zuvor skizzierten, primär durch Regulierung entstehenden Hürden jedoch die Gefahr, die Chancen für eine erfolgreiche Etablierung dieser innovativen Konzepte zumindest für junge Unternehmen zu schmälern.

# 9 Abschließende Betrachtung

**D**IE in dieser Dissertation erarbeiteten Erkenntnisse werden nachfolgend resümiert. Neben den, in Form von IT-Artefakten entwickelten Beiträgen, stehen Perspektiven zur Verwertung und Erweiterung dieser sowie potentiell daran anknüpfende Forschungsaktivitäten im Fokus. Welcher Zusammenhang zwischen den Artefakten (A) zur Beantwortung der Forschungsfragen besteht, ist in Abschnitt 1.4 gelistet.

## 9.1 Zusammenfassung

Beschleunigt durch die pandemische Lage hält die Digitalisierung in Gesundheitssystemen weltweit Einzug. Aufbauend auf der Analyse entstehender digitaler Gesundheitsdateninfrastrukturen, werden sowohl Methoden und Werkzeuge identifiziert als auch Anforderungen beschrieben, welche für die damit verbundene Transformation einen Rahmen für ein mögliches Zusammenspiel zwischen einer datengetriebenen und einer wissensbasierten Entwicklung digitaler Gesundheitsanwendungen definiert (A1). Daraus werden einzelne Schritte abgeleitet, welche in den Phasen der Digitalisierung, Automatisierung und Optimierung sowie Translation durch die Beantwortung von Leitfragen zu adressieren sind. Diese werden als Prozess unter Beachtung entsprechender Normen, Standards und Leitlinien verknüpft (A2). Im Kontext der Anwendungsdomäne Demenz wird dieser Prozess, insbesondere bezüglich digitaler Lösungen zur Früherkennung dementieller Syndrome, exemplarisch durchlaufen. Dabei werden zusätzliche Anforderungen an digitale Gesundheitsanwendungen beschrieben, welche explorativ durch ein nutzerzentriertes Digitalisierungskonzept erarbeitet werden. Anhand des daraus resultierenden ambulanten Systems zur Erfassung psychometrischer Daten werden diese validiert sowie durch qualitative und quantitative Methoden evaluiert (A3). Für die Erfassung und Verarbeitung dafür notwendiger Daten werden Algorithmen und Methoden entworfen (A4) sowie Softwareartefakte implementiert (A6), womit daran geknüpfte Herausforderungen und Chancen im Feld eruiert werden. Darauf aufbauend wird die Translation digitaler Gesundheitsanwendungen aus der Perspektive unterschiedlicher Interessenvertreter beleuchtet. Werkzeuge zur Erfassung und Evaluation von Anforderungen aus deren Sicht werden diesbezüglich entwickelt (A7). Ausgehend davon wird der Bedarf an weiteren Konzepten zur Digitalisierung von ambulanten Gesundheitsanwendungen im Bereich dementieller Syndrome erschlossen und konzeptionell in einer Plattformlösung skizziert (A5). Abschließend werden Möglichkeiten und Hürden für einen potentiellen Markteintritt in das deutsche Gesundheitswesen erörtert.

## 9.2 Fazit

Die in dieser Arbeit durchgeführten Untersuchungen zeigen, dass insbesondere im deutschen Gesundheitswesen einer datengetriebenen Entwicklung digitaler Gesundheitsanwendungen neue Herausforderungen und Chancen gegenüberstehen. Einerseits sind durch die Einführung des DVG neue Geschäftsmodell-

möglichkeiten entstanden, die Anbietern digitaler Gesundheitsanwendungen weitere Kanäle eröffnen. Andererseits wurden in den letzten Jahren zusätzliche regulatorische Hürden aufgebaut, welche die Entwicklung von gesundheitsbezogenen Innovationen und den Zugang zu potentiellen Nutzergruppen und deren Daten erschweren. Die zusätzliche stigmatisierende Wirkung im Bereich psychischer und altersbedingter Erkrankungen schränkt ebenfalls die für eine Digitalisierung notwendige Datenverfügbarkeit ein. Dennoch konnten Indikatoren gefunden werden, die aufzeigen, welche Potentiale in computerisierten Methoden und Modellen stecken, um einen Beitrag zur Entlastung des Gesundheitssystems leisten zu können. Aus Patientensicht zeigt sich, welche Priorität der selbstbestimmte Umgang mit den gesammelten Daten besitzt und für welche Zwecke diese eingesetzt werden können. Ein selbstbestimmtes Datenmanagement zum Schutz dieser sticht als wichtiges Merkmal heraus. Technologische Konzepte zu deren Umsetzung befinden sich zum Zeitpunkt der Anfertigung dieser Arbeit noch in der Forschungsphase. Darüber hinaus wurden in Beobachtungsstudien über einen Zeitraum von fünf Jahren hinweg Anforderungen an die Gebrauchstauglichkeit grafischer Benutzerschnittstellen zur ambulanten Bestimmung neuropsychologischer Zustände gesammelt und letztlich eine hohe Normenkonformität erreicht. Da zur Entwicklung hierfür geeigneter Modelle und Algorithmen Referenzdaten von medizinischem Fachpersonal benötigt werden und diese den gängigen medizinischen Standards genügen müssen, zeigt sich, dass die Digitalisierung und Datenerfassung als ressourcenintensive Teile des Entwicklungszyklus mit Bedacht geplant und umgesetzt werden müssen. Um die Plausibilität und Qualität der dabei generierten und verwendeten Daten zu beurteilen, sind statistische Werkzeuge und Methoden notwendig. Die Aufbereitung dieser Daten bedeutet auch, Strategien für den Umgang mit fehlerhaften oder fehlenden Werten zu finden sowie ungeeignete Merkmale und deren Implementierungen zu identifizieren. Die in dieser Arbeit verwendeten Methoden erweisen sich für den Anwendungskontext als passend. Diese bilden auch die Grundlage für die Wahl der Trainings- und Evaluationsverfahren. Als eine Herausforderung bei der Merkmalsselektion stellt sich eine mögliche Inkonsistenz zwischen den von Leitlinien vorgegebenen Diagnosekriterien und algorithmisch gewählten Merkmalen heraus. Innerhalb dieser Modellierungsphase müssen des Weiteren geeignete Optimierungskriterien definiert werden. Im Schritt der Modellevaluation konnte der *Cohen's Kappa* als ein für die Klassifikationsalgorithmen geeignetes Optimierungskriterium validiert werden. Die Genauigkeit der Klassifikation könnte insbesondere durch einen vollständigeren und umfassenderen Datensatz verbessert werden. Auch sollte beachtet werden, dass eine höhere Performanz bei entsprechender Merkmalsselektion erreicht wird. Hierbei bleibt die Frage offen, ob dieser Verlust durch die zuverlässige Erfüllung der DSM-V-Kriterien und der damit einhergehende zusätzliche Zeitaufwand gerechtfertigt ist. In diesem Fall ist die Auswahl auch eng mit der ethischen und gesundheitspolitischen Frage verbunden wie hoch ein akzeptables Leistungsniveau sein muss.

Insbesondere beim Einsatz von Methoden des ML zeigt sich neben den Forderungen der Leistungserbringer auch von Seiten der Regulatoren ein Bedarf nach Methoden zur Risikominimierung durch Schaffung von mehr Transparenz beim Einsatz solcher Modelle. Diese schließen durch Plausibilisierung der Ergebnisse den Kreis zwischen einem datengetriebenen und einem wissensbasierten Ansatz. Daher werden geeignete XAI-Verfahren im Kontext der Anwendungsdomäne identifiziert und funktional evaluiert. Die Auswahl eingesetzter Verfahren steht zumindest zum Zeitpunkt der Anfertigung dieser Arbeit jedoch noch in Abhängigkeit zu den regulatorischen Unklarheiten bezüglich der Konformitätsbewertung von KI- oder ML-gestützter Medizinprodukte. Dafür wird zukünftig eine entsprechende Norm notwendig.

Die Adhärenz von Nutzern, die als potentielle Leistungsempfänger bezüglich einer digitalen Lösung für kognitive Leistungsmessungen auftreten, konnte empirisch erfragt sowie akzeptable Zeitintervalle zwischen den Messungen bestimmt werden. Ob eine langfristige Nutzerbindung an ein digitales Werkzeug möglich ist, muss für den jeweiligen Anwendungsfall untersucht werden. Im Zuge dessen hat sich ebenso gezeigt, dass die Akzeptanz digitaler Gesundheitsanwendungen durch kassenärztlich organisierte

Leistungserbringer in Deutschland primär von der Erstattungsfähigkeit und Delegierbarkeit der damit erbringbaren Leistungen abhängt. Dementsprechend zeigt sich auch für diese Gruppe die zeitökonomische Durchführbarkeit auf Grundlage einfacher Bedienbarkeit von übergeordneter Relevanz bei der Entwicklung digitaler Gesundheitsanwendungen. Daher ist weniger von einer patienten- als von einer abrechnungsorientierten Zieldefinition im deutschen Gesundheitswesen auszugehen. Um die Akzeptanz durch alle Akteure zu steigern stellt sich eine kontinuierliche Kommunikation zwischen den Entwickelnden und Interessenvertretergruppen im Entwicklungsprozess als Notwendigkeit heraus. Hierfür erforderliche Anreizmechanismen sind politisch zu adressieren.

## 9.3 Ausblick

Durch die weltweit steigenden Lebenserwartungen wird der Anteil an gerontopsychiatrischen Erkrankungen in den nächsten Jahren zunehmen. Begleitet von der Digitalisierung unterschiedlicher Lebensbereiche, werden sich Erfahrungen im Umgang mit digitalen Technologien zukünftig auch zunehmend in den damit verbundenen Zielgruppen etablieren. Wie die in dieser Arbeit erhobenen Studiendaten und Charakteristiken bezüglich der Anwendungsdomäne dementieller Syndrome zeigen, werden dafür vor allem langfristig angelegte Studien notwendig. Um Modelle zu entwickeln, die einen sich teilweise über Jahrzehnte hinweg erstreckenden Verlauf einer dementiellen Erkrankung abbilden können, müssen sensitive Testverfahren identifiziert und Längsschnittdatenbanken aufgebaut werden. Hierin sind Aussagen über Lerneffekte, die Adhärenz sowie weitere psychologische und physiologische Marker digital zu sammeln, welche eine effiziente und effektive Kombination der notwendigen medizinischen Parameter im Versorgungsalltag zulassen. Des Weiteren lässt der jüngst veröffentlichte medizinisch-wissenschaftliche Konsens, dass bei bis zu 40% der Betroffenen ein multifaktorieller Behandlungsansatz eine angehende Demenz durch die Identifikation und Korrektur von Risikofaktoren sowohl positiv beeinflussen als auch präventiv verhindern kann, einen steigenden Bedarf an digitalen Gesundheitsanwendungen zur Prävention, Früherkennung und Behandlungsbegleitung erwarten. Hierdurch entstehende Digitalisierungsziele sollten im Zuge dessen von den beteiligten Interessenvertretergruppen konsentiert werden. Dem in dieser Arbeit identifizierten Vorgehen folgend, können diesbezügliche Lösungsinkremente in digitale Gesundheitsanwendungen transferiert werden. Um für diese Verknüpfung möglichst realitätsnahe Bedingungen zu erreichen, sind vorab jedoch die Voraussetzungen für eine normenkonforme Entwicklung von Medizinprodukten zu klären, welche den gesetzlichen Anforderungen entsprechen und Datenerhebungen sowie deren Verwertung auch für kommerzielle Zwecke ermöglichen. Weitere Untersuchungen, die zeigen wie sich die Bereitschaft in der Gesellschaft entwickelt diese Daten langfristig zur Verfügung zu stellen, wie dafür Anreize geschaffen werden können und wie sich dieser Vorgang technologisch abbilden lässt ohne die Datensouveränität einzuschränken, sind ebenfalls Ausgangspunkte für zukünftige Forschungsaktivitäten. Die Verwertung der in dieser Arbeit entwickelten Artefakte findet zum einen durch daran geknüpfte Publikationen sowie insbesondere durch die Verfügbarmachung der mobilen Anwendung *DemPredict* als digitales Informationsangebot statt. Da in Deutschland ein kuratives Gesundheitssystem implementiert ist, werden zum anderen, Teilkomponenten wie die erarbeiteten Anforderungen, der abgeleitete Entwicklungsprozess und das Konzept zur Entwicklung digitaler Präventions- und Interventionsplattformen aktiv in weiterführende Forschungs- und Forschungstransferprojekte integriert. Diese stellen auch die Grundlage zur Erforschung eines präventiven Umgangs mit solchen Erkrankungen dar. Zukünftig ist ebenfalls zu prüfen, inwiefern sich die hierfür identifizierten Technologien, Werkzeuge und Methoden auf weitere Krankheitsbilder übertragen lassen.



# A Anhang

## A.1 Anforderungen an den Umgang mit Gesundheitsdaten

<b>Strukturell</b>	<p><b>SI - Austausch und gemeinsame Nutzung von Gesundheitsdaten:</b> Der Austausch von Gesundheitsdaten über Systemgrenzen hinweg muss ermöglicht werden (z.B. zwischen medizinischen Einrichtungen oder Familienmitgliedern) [19, 47, 169, 174, 206, 215, 260, 269, 308, 313, 321, 330, 333, 338, 341, 342, 374, 396, 421, 425, 438].</p> <p><b>SI.1 - Die Freigabe von Daten an Dritte:</b> Es muss die Möglichkeit bestehen, Daten mit kommerziellen Nutzergruppen (z.B. Universitäts-Spin-offs, Start-ups) zu teilen, um Innovationen z.B. im Bereich der Datenanalyse und KI zu fördern [207, 433].</p> <p><b>SI.2 - Organisatorische Interoperabilität:</b> Ermöglichung von systemübergreifenden Prozessen, Identitäten und Rechten zwischen medizinischen Einrichtungen [174, 185, 324, 374, 421].</p>
	<p><b>S2 - Souveräne Gesundheitsdatenverwaltung:</b> Die Patienten erhalten Zugang und Kontrolle über ihre Gesundheitsdaten, wodurch sie die aktuellen und klinisch korrekten Daten, die über sie gespeichert sind, überwachen können. Der Patient hat die Macht, jeden Datenaustausch zu verwalten, indem er den genauen Inhalt, den Zweck und die Zugriffsrechte definiert (Recht des Patienten auf informationelle Selbstbestimmung) [37, 47, 174, 205, 215, 235, 268, 270, 309, 313, 330, 374, 396, 421, 425, 433].</p> <p><b>S2.1 - Verwaltung von Zustimmungen:</b> Medizinische Daten dürfen nur verarbeitet werden, wenn der Dateneigentümer seine Zustimmung gegeben hat. Der Dateneigentümer hat das Recht, seine Einwilligung jederzeit zu widerrufen [19, 47, 103, 169, 174, 206, 235, 308, 321, 330, 333, 338, 374, 385, 393, 393, 396, 421, 433].</p> <p><b>S2.2 - Vertrauenswürdigkeit:</b> Zwischen Dateneigentümer, Datenverwahrer und -konsument muss Vertrauen herrschen. Einerseits in die technologische Umgebung des Datenaustauschs und der Datenspeicherung, andererseits in die daran beteiligten Akteure selbst, damit nichts mit den Daten geschieht, was dem Willen des Dateneigentümers zuwiderläuft. Im Gegensatz dazu muss für den Datenkonsument gewährleistet werden, dass die Daten vertrauenswürdig und korrekt sind. Dieses Vertrauen muss spürbar sein [18, 169, 174, 206, 207, 235, 253, 308, 313, 316, 330, 338, 396, 421, 425, 433, 438].</p> <p><b>S2.3 - Rückverfolgbarkeit des Datenlebenszyklus:</b> Der Dateneigentümer kann seine Gesundheitsdaten und die Historie aller Aktionen (z.B. Zugriff, Verwendung in Forschung und klinischen Studien) zu diesen Daten sowie die entsprechenden Identitäten und den Zweck hinter diesen Aktionen einsehen, um den gesamten Datenlebenszyklus zu verfolgen [47, 186, 215, 217, 235, 374, 396, 410, 425].</p>

Strukturell	<p><b>S3 - Datenspende für die Forschung:</b> Der Dateneigentümer sollte die eigenen Gesundheitsdaten selektiv freigeben können, damit mögliche Forschungsprojekte und Studien mit ihren Daten arbeiten können. Dabei sollte der Dateneigentümer aktiv durch eine informierte Einwilligung entscheiden können, ob, was und an wen er spendet [19, 34, 103, 126, 169, 174, 206, 207, 215, 253, 308, 330, 333, 385, 393, 396, 425].</p> <p><b>S3.1 - Post-mortem-Datenspende:</b> Patienten sollten die Möglichkeit haben, ihre medizinischen Daten nach ihrem Tod für Forschungszwecke zu spenden, ähnlich wie bei der Organspende [141].</p> <p><b>S3.2 - Gemeinsame Nutzung von Forschungsergebnissen:</b> Rückmeldung von Forschungsergebnissen zurück an Datenbesitzer und Patienten mit ähnlichen Krankheiten und Symptomen [34, 207, 385, 438].</p> <p><b>S3.3 - Integration von patientengenerierten Längsschnittdaten:</b> Integration mobiler Patientenanwendungen und IoT-Geräte zur Erfassung, gemeinsamen Nutzung und Überwachung patientengenerierter Daten, wie z.B. Vitalparameter, körperliche Aktivitäten, Nahrungsaufnahme sowie individuelle Beurteilungen und Beschreibungen des eigenen Wohlbefindens [19, 174, 186, 206, 215, 217, 268, 269, 270, 309, 313, 316, 321, 338, 385, 396, 399, 410, 421, 438].</p>
	<p><b>S4 - Transparenz und Reproduzierbarkeit von Forschungsergebnissen:</b> Die gemeinsame Nutzung der entsprechenden Forschungsdaten (Metadaten, Quellcode, Parameter-Definitionen, deidentifizierte Daten) mit den Kostenträgern der Forschungsprojekte, den Zeitschriften, welche die wissenschaftlichen Publikationen veröffentlichen und mit der Öffentlichkeit über zugängliche Datenrepositorien. Transparente Dokumentation über die Methodik und die Verfahren, die zur Erhebung der Daten verwendet werden, um Aufzeichnungen über den gesamten Prozess und die Möglichkeiten der Wiederverwendung der Forschungsdaten zu erhalten [215, 393].</p> <p><b>S4.1 - Antrag auf Verwendung in der Forschung:</b> Für ein spezifisches Forschungsprojekt können die Gesundheitsdaten eines Patienten dem anfragenden Partner auf der Grundlage der Zustimmung des Patienten zur Verfügung gestellt werden [34, 174, 207, 308, 330, 341, 374, 396]. Bei Datenanfragen müssen Art und Zweck der Datenverarbeitung durch den Datenkonsumenten angegeben werden [235, 425].</p> <p><b>S4.2 - Sekundäre Nutzung von medizinischen Forschungsdaten:</b> Bereitstellung medizinischer Forschungsdaten zur sekundären Verwendung und Analyse der Daten durch weitere Forscher und Studien [141, 207, 253, 308, 316, 330, 393, 410].</p> <p><b>S4.3 - Registrierung von Studien:</b> Registrierung von Forschungsstudien einschließlich der Angabe von beabsichtigten Analysen in einem öffentlichen Register vor der Datenerhebung [393].</p>
Ökonomisch	<p><b>E1 - Wahrnehmbarer Nutzen der Datenspende:</b> Der Datenspender erhält eine Entschädigung in Form von Geld- oder Dienstleistungsvorteilen (z.B. über Krypto-Token), indem er seine persönlichen Gesundheitsdaten oder Forschungsergebnisse an Datenkonsumenten weitergibt [169, 207, 253, 393, 433].</p>
	<p><b>E2 - Finanzierung der Weitergabe von Forschungsdaten:</b> Finanzierung von Forschern und Forschungsprojekten, um den zusätzlichen Aufwand zu kompensieren, der für die Aufbereitung und den Austausch von Forschungsdaten zur weiteren Verwendung erforderlich ist [215].</p>

Medizinisch	<p><b>M1 - Medizinische Relevanz:</b> Die Daten, die über eine Plattform oder eine Spende eingegeben und zur Verfügung gestellt werden können, sind an die Bedürfnisse von medizinischen Fachkräften und klinischen Forschern angepasst. Die Informationen und Dienstleistungen sollten so aufbereitet sein, dass diese Personengruppen entsprechend dem klinischen Anwendungsfall unterstützt werden [18, 37, 174, 309, 396, 437].</p>
	<p><b>M2 - Datengetriebene personalisierte Medizin:</b> Verbesserung der Patientenversorgung, insbesondere der Therapie und Diagnostik aufgrund einer erhöhten Datenmenge. Ziel ist es, Therapien so individuell und präzise wie möglich für die Patienten zu gestalten und mögliche Erkenntnisse an die Patienten zurückzugeben [18, 19, 103, 205, 206, 207, 215, 217, 245, 253, 270, 321, 330, 333, 385, 396, 410, 433, 436].</p>
	<p><b>M3 - Kollaborative Entscheidungsfindung unterstützen:</b> Bereitstellung von historischen Patienten- und Studiendaten sowie des notwendigen Hintergrundwissens in visuell angepasster Form für Patienten, medizinisches Personal und klinische Forschung, um die Entscheidungsfindung zu bestimmten Zeitpunkten bestmöglich zu unterstützen und das Verständnis des Patienten für seinen eigenen Zustand zu stärken [37, 174, 217, 268, 270, 309, 313, 340, 396, 399, 436, 437].</p>
Technisch	<p><b>T1 - Datenschutz:</b> Gewährleistung des Datenschutzes von Patienten und Studienteilnehmern gemäß der gesetzlichen Regelungen (z.B. DSGVO) bei der Speicherung und Übertragung von Daten durch eine Privacy-by-Design-Architektur [18, 19, 34, 47, 103, 141, 169, 174, 186, 206, 207, 215, 217, 218, 235, 253, 257, 260, 268, 269, 308, 313, 316, 321, 330, 333, 338, 341, 342, 356, 396, 399, 410, 421, 425, 433, 438]. Verhindern des Datenverlustes und von Datenlecks [207, 235, 433].</p> <p><b>T1.1 - Datenlöschung:</b> Löschungsmöglichkeit für die gespeicherten Daten durch den Dateneigentümer [47, 207, 399].</p> <p><b>T1.2 - Deidentifizierung:</b> Möglichkeit, direkt identifizierende Informationen aus den medizinischen Daten durch Anonymisierung oder Pseudonymisierung zu entfernen [19, 34, 103, 174, 207, 215, 218, 235, 257, 308, 316, 330, 333, 338, 342, 356, 374, 421, 425, 433, 436].</p> <p><b>T1.3 - Datenschutzwahrende Analysemöglichkeiten:</b> Unterstützung für datenschutzfreundliche Analysealgorithmen (z.B. datenschutzfreundliche Smart Contracts [47], homomorphe Verschlüsselung [316], verteilte Datenverarbeitung oder Containerisierung [207]) so dass die Daten ihren lokalen Speicher nicht verlassen müssen, d.h. die Algorithmen kommen zu den Daten nicht die Daten zu den Algorithmen [174, 207, 235, 253, 260, 308, 316, 330, 333, 356]. Außerdem sollten die Algorithmen funktional sicher und verständlich sowie deren Verarbeitungsschritte transparent nachvollziehbar sein [207].</p> <p><b>T1.4 - Nutzungs- und Zugriffsverwaltung:</b> Feingranulare Definition, Überwachung und Kontrolle des Zugriffs auf Gesundheitsdaten durch den Dateneigentümer. Die Einhaltung sollte kontrollierbar sein. Der Zugang und der Zweck der Nutzung sollten dokumentiert und vom Dateneigentümer für einen bestimmten Zeitraum autorisiert werden [169, 174, 186, 217, 268, 269, 308, 316, 330, 333, 338, 341, 342, 356, 374, 385, 399, 421, 425].</p> <p><b>T1.5 - Benutzerauthentifizierung und -autorisierung:</b> Ermöglichung der Authentifizierung des Benutzers (z.B. mit einem Single-Sign-On-Ansatz, Zwei-Faktor-Authentifizierung) oder der Benutzergeräte, um die Gesundheitsdaten einsehen und Nutzungs- und Zugriffsrechte vergeben und überprüfen zu können [217, 235, 268, 330, 333, 338, 342, 356, 374, 421, 425].</p>

<b>Technisch</b>	<p><b>T2 - Datensicherheit:</b> Anwendung des Security-by-Design-Ansatzes für Systementwurf, Implementierung und Betrieb [19, 47, 103, 169, 174, 186, 206, 207, 215, 217, 235, 253, 268, 269, 308, 313, 316, 321, 330, 333, 338, 341, 342, 356, 374, 399, 421, 425, 433, 438].</p> <p><b>T2.1 - Daten-Verschleierung:</b> Verschleierung von Daten durch Verschlüsselung, digitale Wasserzeichen oder Steganographie zur Gewährleistung der Datensicherheit und Vertraulichkeit [47, 141, 186, 206, 217, 235, 268, 316, 338, 341, 399, 421, 433].</p> <p><b>T2.2 - Datengenauigkeit:</b> Gewährleistung der Richtigkeit der Daten [313].</p> <p><b>T2.3 - Datenunveränderlichkeit:</b> Verifizierung der Unveränderlichkeit der Daten zur Vertrauensbildung [47, 207, 217, 235, 269, 421, 433].</p> <p><b>T2.4 - Verfügbarkeit:</b> Eine (permanente) Verfügbarkeit von Daten für alle autorisierten Akteure sollte gewährleistet werden [19, 169, 186, 206, 215, 268, 269, 356].</p> <p><b>T2.5 - Beweisbarkeit:</b> Die Verknüpfung der Daten mit einer Zustimmung zur Nutzung und ein Nachweis der Existenz dieser Verknüpfung sollte vorhanden sein [235].</p>
	<p><b>T3 - Gebrauchstauglichkeit:</b> Benutzerfreundliche und patientenzentrierte Tools zur Visualisierung und Verwaltung vorhandener Gesundheitsdaten und Zugriffsrechte. Die Einstiegshürde für nicht-technische Anwender sollte niedrig sein, um die Akzeptanz für die Nutzung zu erhöhen [37, 169, 174, 215, 270, 333, 340, 341, 399, 421, 436].</p>
	<p><b>T4 - Datenauffindbarkeit:</b> Bereitstellung einer maschinenverarbeitbaren Beschreibung, eines Datenkatalogs oder von Daten-Tags, die die aggregierten Metadaten der verfügbaren Daten repräsentieren, um die Suche und Datenabfrage nach relevanten Daten zu ermöglichen [235, 316, 333, 374, 396, 425, 433].</p>
	<p><b>T5 - Datenqualität:</b> Messung und Dokumentation der Datenqualität im Originalzustand und entlang des Datenlebenszyklus auch in aggregierter Form [215, 330, 333, 396, 399, 433].</p>

<b>Technisch</b>	<p><b>T6 - Nachhaltigkeit:</b> Daten in Technologie-neutralen und offenen Formaten ausdrücken, so dass Daten und ihre Definitionen effizient von einem Technologie-Stack in einen anderen migriert werden können [174, 333, 393, 425].</p> <p><b>T6.1 - Domänengetriebene Datenmodellierung:</b> Einbeziehung von Domänenexperten (z.B. Forscher, Gesundheitsexperten) in die Datenmodellierung und die Trennung zwischen den technischen Implementierungsdetails und der Domänenmodellierung [174, 330].</p> <p><b>T6.2 - Vermeidung von Lock-in-Effekten:</b> Die Wahl von Standards, Rahmenwerken und Systemkomponenten, so dass Herstellerabhängigkeiten vermieden werden, z.B. durch Bevorzugung von Open-Source-Software sowie offenen Standards, Modellen, Frameworks und Spezifikationen [174, 330, 333, 342, 396].</p> <p><b>T6.3 - Semantische Interoperabilität:</b> Verwendung international anerkannter, bewährter und technologisch offener Terminologien und Standards (z.B. HL7 FHIR, DICOM, IHE, openEHR, SNOMED CT), um die Kommunikation zwischen verschiedenen Systemen zu ermöglichen [18, 19, 126, 174, 206, 330, 338, 341, 342, 374, 425, 436].</p> <p><b>T6.4 - Datenaggregation und -harmonisierung:</b> Aufbau eines gemeinsamen Verständnisses von Informationen und Daten aus verschiedenen Datensilos durch Zusammenführen der verschiedenen Datentypen und Aufbau konsistenter Daten durch berechenbare semantische Modelle (z.B. Ontologien) [18, 19, 126, 174, 215, 260, 330, 333, 341, 342, 421, 436].</p> <p><b>T6.5 - Strukturelle Interoperabilität:</b> Ermöglichung von Datenflüssen zwischen verschiedenen Systemen und Datensilos durch die Definition von Datenstrukturen und -syntax [174, 206, 215, 330, 333, 374, 399, 421, 425].</p> <p><b>T6.6 - Grundlegende Interoperabilität:</b> Ermöglichung von Datenflüssen zwischen verschiedenen Systemen durch Definition von Schnittstellen für den Transport der Daten [185, 324].</p>
	<p><b>T7 - Dezentralität:</b> Dezentralisierte Speicherung der Daten dort, wo sie erzeugt wurden. Im Falle eines notwendigen Datentransfers verlassen die Daten die lokalen Datenspeicher nicht ohne die Zustimmung des Dateneigentümers [174, 330, 425].</p>
	<p><b>T8 - Prädiktive Analysemöglichkeiten:</b> Einsatz von Analyse- und Clustering-Methoden zur Vorhersage oder Identifizierung von Mustern und Zusammenhängen zwischen Patienten, ihren Krankheitsverläufen und Befunden, um frühzeitig Entscheidungen über Diagnose und Behandlung treffen zu können [18, 19, 103, 174, 206, 207, 270, 316, 330, 333, 425, 436].</p>
	<p><b>T9 - Datenanreicherung:</b> Analyse aggregierter Daten (z.B. unter Verwendung von Natural Language Processing (NLP), ontologischen Annotationen oder Techniken des maschinellen Lernens), um aussagekräftiges Wissen in strukturierter Form zu extrahieren [18, 103, 174, 333, 425].</p>
	<p><b>T10 - Skalierbarkeit und Leistung:</b> Möglichkeit zur Handhabung und Verarbeitung einer großen Menge komplexer und sich ständig ändernder Daten [19, 103, 174, 330, 333, 341, 399, 438].</p>
	<p><b>T11 - Flexibilität:</b> Eine Datenanalyse-Infrastruktur sollte ein hohes Maß an Flexibilität gewährleisten, um je nach Anwendung unterschiedliche Konfigurations- und Weiterverarbeitungsmöglichkeiten der Daten zu unterstützen (z.B. Unterstützung verschiedener Dateiformate für den Export) [174, 338].</p>
	<p><b>T12 - Erweiterbarkeit:</b> Die IT-Architektur sollte modular und erweiterbar sein, so dass eine Vielzahl von Anwendungsfällen bedient werden kann [174, 330, 333, 399, 425].</p>

Tabelle A.1: Detaillierte Anforderungen an den Umgang mit Gesundheitsdaten in Gesundheitsdatenökosystemen nach Schinle et al., 2021 [473].

## A.2 Beispiel für Usage Policy

```
1 EquivalentClasses(  
2   spl:Authorization  
3   ObjectIntersectionOf(  
4     ObjectSomeValuesFrom(spl:hasData ObjectUnionOf(dpr:  
       Psychometric dpr:Pseudonymized svd:Demographic svd:  
       Location svd:Statistical svd:Profile))  
5     ObjectSomeValuesFrom(spl:hasProcessing ObjectUnionOf(dpr:  
       Profiling dpr:Pseudonymize svpr:Aggregate svpr:Analyse  
       svpr:Anonymize svpr:Derive svpr:Query))  
6     ObjectSomeValuesFrom(spl:hasPurpose ObjectUnionOf(dpr:  
       NeuropsychologicalState dpr:ScreeningApplication))  
       ObjectSomeValuesFrom(spl:hasRecipient ObjectUnionOf(dpr:  
         :Developer svr:Ours)) ObjectSomeValuesFrom(spl:  
         hasStorage  
7     ObjectIntersectionOf(  
8     ObjectUnionOf(svl:EU ObjectSomeValuesFrom(spl:hasLocation  
       svl:OurServers))  
9     DataMinCardinality(5 spl:durationInDays xsd:  
       positiveInteger)  
10    ))  
11  )  
12 )
```

**Listing A.1** Beispielhafte Nutzungsbedingungen für *DemPredict* in OWL2 Functional Syntax

## A.3 Fragebogen zur Umfrage Datennutzung

Fragebogen zur Erfassung der Einstellung gegenüber Gesundheitsdatenspenden und der Gestaltung diesbezüglicher Nutzungsbestimmungen.

### Umfrage zum Thema Datenspende im Gesundheitswesen

#### Einstellung zu Gesundheitsdatenverwaltung

Kreuzen Sie bitte an, inwieweit Sie den folgenden Aussagen zustimmen (1 - Stimme überhaupt nicht zu | 2 - Stimme nicht zu | 3 - Stimme weder zu noch nicht zu | 4 - Stimme zu | 5 - Stimme völlig zu):

1. "Ich möchte einen besseren Überblick über meine Gesundheitsdaten bekommen, indem ich diese und den jeweiligen Speicherort einsehen kann."\*

**1**
**2**
**3**
**4**
**5**

Stimme überhaupt nicht zu      Stimme völlig zu

2. "Wenn es ein System gäbe, in dem man einen Einblick in die Gesamtheit seiner Gesundheitsdaten erhalten würde. Dann würde ich das System benutzen, wenn Datenschutz und Datensicherheit garantiert werden würden."\*

**1**
**2**
**3**
**4**
**5**

Stimme überhaupt nicht zu      Stimme völlig zu

3. Kreuzen Sie an, welchem Akteur im Gesundheitswesen Sie es zutrauen würden, eine Plattform zu betreiben, die Ihre Gesundheitsdaten verwaltet?

- Eigene Krankenkasse  
 Kassenärztliche Vereinigung  
 Bundesministerium für Gesundheit  
 Nationale Koordinierungsstelle E-Health-Deutschland  
 Keinem der genannten  
 Sonstiges: \_\_\_\_\_

#### Selbstgenerierte Gesundheitsdaten

4. Benutzen Sie einen Fitnesstracker oder eine Smart Watch mit ähnlichen Funktionen? \*

- Smart Watch  
 Fitness Tracker  
 Keine Angabe  
 Nein  
 Sonstiges: \_\_\_\_\_

5. Welche der folgenden Gesundheitsdaten zeichnen Sie aktuell auf? \*

- |  |   |
|--|---|
| <input type="radio"/> Gewicht              | <input type="radio"/> Periode             |
| <input type="radio"/> Trainingseinheiten   | <input type="radio"/> Müdigkeit           |
| <input type="radio"/> Nahrungsaufnahme     | <input type="radio"/> Schlaf              |
| <input type="radio"/> Wasseraufnahme       | <input type="radio"/> keine der genannten |
| <input type="radio"/> Blutdruck            | <input type="radio"/> Sonstiges: _____    |
| <input type="radio"/> Puls                 |   |
| <input type="radio"/> Blutzucker           |   |
| <input type="radio"/> Kopfschmerzen        |   |
| <input type="radio"/> Schrittzahl          |   |
| <input type="radio"/> Medikamenteneinnahme |   |

Kreuzen Sie bitte eine der folgenden Antworten auf die Fragen an (1 - Keinesfalls | 2 - wahrscheinlich nicht | 3 - ich weiß nicht | 4 - sehr wahrscheinlich | 5 - Ganz sicher):

6. Würden Sie Ihre selbstgenerierten Gesundheitsdaten bzw. einen Teil davon anonymisiert an eine Forschungseinrichtung zu Forschungszwecken spenden? \*

	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	
Keinesfalls	<input type="radio"/>	Ganz sicher				

7. Würden Sie Ihre selbstgenerierten Gesundheitsdaten an Ihre Krankenversicherung senden, um den Krankenkassenbeitrag zu senken? \*

	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	
Keinesfalls	<input type="radio"/>	Ganz sicher				

8. Inwiefern glauben Sie, dass Sie in der Zukunft von medizinischen Datenspenden profitieren könnten? \*

---



---

9. Wie wichtig ist es in Ihrem persönlichen Empfinden auf einer Skala von 1 (nicht wichtig) bis 5 (äußerst wichtig), nachvollziehen zu können, wo, von wem und zu welchem Zweck Ihre anonymisierten Gesundheitsdaten verarbeitet werden? (Bsp: Uniklinik Heidelberg zur Demenzforschung) \*

	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	
Nicht wichtig	<input type="radio"/>	Äußerst wichtig				

10. Angenommen sie stehen unter einem Demenzverdacht und machen einen Test zur Diagnose per App. Würden Sie Ihre Daten der Demenzforschung spenden? \*

- Ja
- Nein
- Keine Angabe

**Nutzungsbestimmungen für die Daten**

Durch die Nutzungsbedingungen möchte man den Nutzer bestimmen lassen zu welchem Zweck er die Daten spenden möchte. So werden die Daten vor Zweckentfremdung geschützt. Das folgende Bild zeigt wie ein solches Auswahlformular aussehen könnte.

Art der Daten	
<input type="checkbox"/> Anonymisiert	oder <input type="checkbox"/> Pseudoanonymisiert
<input type="checkbox"/> Personenbezogene (z.B. Alter, Größe) <input type="checkbox"/> Medizinisch <input type="checkbox"/> Selbstgenerierte (z.B. Lebensstil, Ernährung, Beruf, Schlaf, Gesundheitsapps)	
Zweck	
<input type="checkbox"/> Forschung <input type="checkbox"/> Bildung <input type="checkbox"/> Regierung (z.B. Paul-Ehrlich-Institut oder Bundesinstitut für Arzneimittel und Medizinprodukte)	
Datenverarbeitung	
<input type="checkbox"/> Aggregieren <small>① mehrere Daten zu übergeordneten Einheiten zusammenfassen, um dadurch eine bessere Beziehung zw. Datenebene und Konstruktebene zu erhalten</small> <input type="checkbox"/> Analysieren <input type="checkbox"/> Ableitungen <small>① Daten werden durch eine mathematische, logische oder andere Art der Transformation zu neuen Datenelementen</small> <input type="checkbox"/> Einzelne Abfragen	
Ort der Datenspeicherung	
<b>Server:</b> <input type="checkbox"/> Des Datenverwalters <input type="checkbox"/> Der Datenverarbeiter	
<b>Server in:</b> <input type="checkbox"/> Deutschland <input type="checkbox"/> EU <input type="checkbox"/> Länder mit ähnlichen Datenschutzgesetzen <input type="checkbox"/> Andere Länder	
Dauer der Datenspeicherung	
<input type="checkbox"/> Für Zeit der Forschung <input type="checkbox"/> Zeit der Datenaggregation/-aufbereitung Angegebener Zeitraum: <input type="checkbox"/> 5 <input type="checkbox"/> 10 <input type="checkbox"/> 20 Jahre <input type="checkbox"/> Unlimitiert	

11. Ist die obige Darstellung der Nutzungsbestimmungen für Sie verständlich? \*

- Ja  
 Nein  
 Sonstiges: \_\_\_\_\_

12. Falls nein, wieso?

\_\_\_\_\_

\_\_\_\_\_

13. Mithilfe der Nutzungsbedingungen können Sie festlegen, was mit Ihren Daten nach der Datenspende passieren darf. Welche Wichtigkeit besitzt welche der aufgeführten Kategorien für Sie? \*

Bitte kreuzen Sie die für Sie zutreffende Aussage an	Verstehe ich nicht	Nicht wichtig	Mittelmäßig wichtig	Sehr wichtig
Art der Daten (Anonymisiert/Pseudoanonymisiert);				

personenbezogene Daten, medizinische Daten, selbstgenerierte Daten)				
Zweck (Forschung, Bildung, Regierung)				
Datenverarbeitung				
Datenempfänger (Forschungsinstitutionen)				
Ort der Datenspeicherung				
Länge der Datenaufbewahrung				

14. Welche Art der Nutzungsbedingung fehlt Ihnen unter den oben aufgeführten?

\_\_\_\_\_

\_\_\_\_\_

Kreuzen Sie bitte eine der folgenden Antworten auf die Frage an (1 - Keinesfalls | 2 - wahrscheinlich nicht | 3 - ich weiß nicht | 4 - sehr wahrscheinlich | 5 - Ganz sicher):

15. Wenn Sie feststellen könnten, wer wann zu welchem Zweck auf ihre Daten zugegriffen hat, könnten Sie sich vorstellen, Ihre Daten zu spenden? (Bsp: Verhandlungsverlauf einer Krankheit) \*

- |                            |                       |                       |                       |                       |                            |
|----------------------------|-----------------------|-----------------------|-----------------------|-----------------------|----------------------------|
| <b>1</b>                   | <b>2</b>              | <b>3</b>              | <b>4</b>              | <b>5</b>              |                            |
| <input type="radio"/>      | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |                            |
| <small>Keinesfalls</small> |                       |                       |                       |                       | <small>Ganz sicher</small> |

**Allgemeine Fragen**

16. Alter: \* \_\_\_\_\_

17. Geschlecht: \*

- Männlich
- Weiblich
- Anderes
- Möchte ich nicht angeben

18. Bundesland Ihres ersten Wohnsitzes \*

- |   |  |
|---|--|
| <ul style="list-style-type: none"> <li><input type="radio"/> Ausland</li> <li><input type="radio"/> Baden-Württemberg</li> <li><input type="radio"/> Rheinland-Pfalz</li> <li><input type="radio"/> Bayern</li> <li><input type="radio"/> Nordrhein-Westfalen</li> <li><input type="radio"/> Sachsen</li> <li><input type="radio"/> Sachsen-Anhalt</li> <li><input type="radio"/> Berlin</li> <li><input type="radio"/> Bremen</li> </ul> | <ul style="list-style-type: none"> <li><input type="radio"/> Hessen</li> <li><input type="radio"/> Mecklenburg-Vorpommern</li> <li><input type="radio"/> Hamburg</li> <li><input type="radio"/> Niedersachsen</li> <li><input type="radio"/> Saarland</li> <li><input type="radio"/> Brandenburg</li> <li><input type="radio"/> Thüringen</li> <li><input type="radio"/> Schleswig-Holstein</li> </ul> |
|---|--|

19. Höchster Bildungsabschluss \*

- Abitur (allgemeine oder fachgebundene Hochschulreife)
- Fachhochschulreife (allgemeine oder fachgebundene Fachhochschulreife)
- Mittlerer Schulabschluss (Realschulabschluss und vergleichbare Schulabschlüsse)
- Hauptschulabschluss und qualifizierender Hauptschulabschluss
- Studium (Bachelor, Master/Diplom/Magister, Promotion)
- Keine Angabe

20. Leiden Sie unter einer chronischen Krankheit? (Diabetes, Herzkrankheit, Krebs, ...) \*

- Ja
- Nein
- Keine Angabe



# A.4 Demografische Daten zu Datennutzung

				Chronisch krank			gesamt
				Ja	Nein	keine Angabe	
männlich	Anzahl	Höchster Bildungsabschluss	Studium	5	64	1	70
			Abitur	3	42	3	48
			Fachhochschulreife	0	8	1	9
			Mittlere Reife	1	3	1	5
			Sonstiges	2	0	0	2
	Herkunft	Baden-Württemberg	5	68	1	74	
		Rheinland-Pfalz	1	29	1	31	
		Bayern	1	6	1	8	
		Hessen	1	8	0	9	
		Niedersachsen	1	1	1	3	
NRW	0	0	0	0			
Sonstiges	2	5	2	9			
Studienteilnehmer				11	117	6	134
Alter	Maximum			67,00	79,00	61,00	79,00
	Minimum			22,00	20,00	22,00	20,00
	Mittelwert			43,91	29,69	44,67	31,53
	Median			25,00	25,00	25,00	25,00
	Standardabweichung			17,63	11,68	17,73	13,33
weiblich	Anzahl	Höchster Bildungsabschluss	Studium	6	38	1	45
			Abitur	7	32	0	39
			Fachhochschulreife	1	6	0	7
			Mittlere Reife	2	4	0	6
			Sonstiges	0	2	0	2
	Herkunft	Baden-Württemberg	8	38	0	46	
		Rheinland-Pfalz	1	22	1	24	
		Bayern	2	4	0	6	
		Hessen	0	5	0	5	
		Niedersachsen	3	8	0	11	
NRW	0	4	0	4			
Sonstiges	2	1	0	3			
Studienteilnehmer				16	82	1	99
Alter	Maximum			57,00	63,00	32,00	63,00
	Minimum			23,00	19,00	32,00	19,00
	Mittelwert			36,56	31,27	32,00	32,13
	Median			25,00	25,00	25,00	25,00
	Standardabweichung			11,41	13,30	0,00	13,03
keine Angabe	Anzahl	Höchster Bildungsabschluss	Studium	0	0	1	1
			Abitur	1	0	0	1
			Fachhochschulreife	0	0	0	0
			Mittlere Reife	0	1	0	1
			Sonstiges	0	0	2	2
	Herkunft	Baden-Württemberg	0	0	0	0	
		Rheinland-Pfalz	1	0	0	1	
		Bayern	0	1	1	2	
		Hessen	0	0	0	0	
		Niedersachsen	0	0	0	0	
NRW	0	0	0	0			
Sonstiges	0	0	2	2			
Studienteilnehmer				1	1	3	5
Alter	Maximum			22,00	18,00	33,00	33,00
	Minimum			22,00	18,00	24,00	18,00
	Mittelwert			22,00	18,00	27,33	24,40
	Median			25,00	25,00	25,00	25,00
	Standardabweichung			0,00	0,00	4,93	5,50
gemischt	Anzahl	Höchster Bildungsabschluss	Studium	11	102	3	116
			Abitur	11	74	3	88
			Fachhochschulreife	1	14	1	16
			Mittlere Reife	3	8	1	12
			Sonstiges	2	2	2	6
	Herkunft	Baden-Württemberg	13	106	1	120	
		Rheinland-Pfalz	3	51	2	56	
		Bayern	3	11	2	16	
		Hessen	1	13	0	14	
		Niedersachsen	4	9	1	14	
NRW	0	4	0	4			
Sonstiges	4	6	4	14			
Studienteilnehmer				28	200	10	238
Alter	Maximum			67,00	79,00	61,00	79,00
	Minimum			22,00	18,00	22,00	18,00
	Mittelwert			38,93	30,28	38,20	31,63
	Median			25,00	25,00	25,00	25,00
	Standardabweichung			14,54	12,36	15,86	13,10

Tabelle A.2: Studienteilnehmerdaten der Umfrage unter Leistungserbringern.

# A.5 Umfrageergebnisse zu Datennutzung

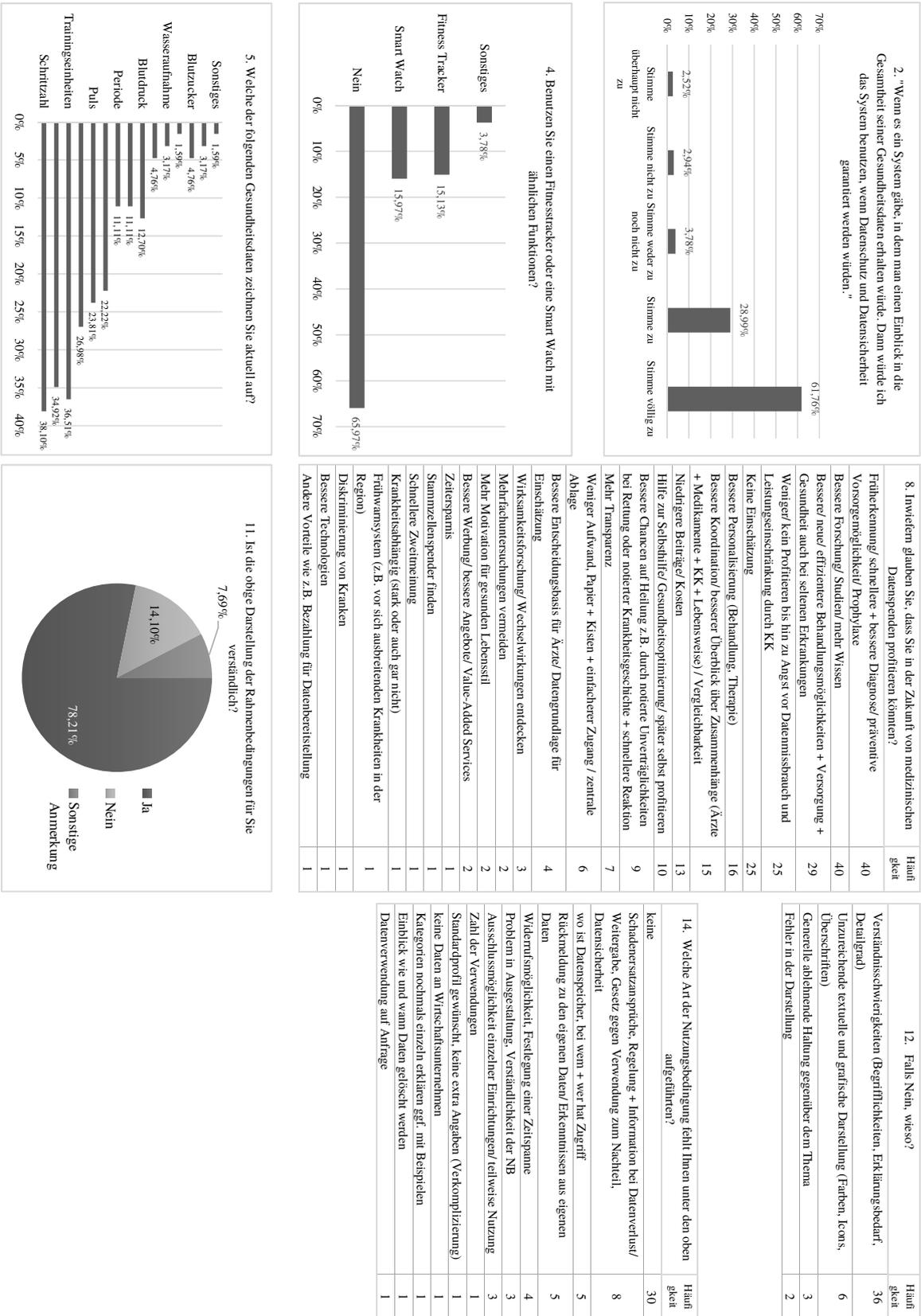


Abbildung A.1: Antworten auf Umfrage aus Abschnitt A.3.

## A.6 Beispielhafte FHIR-Ressource

```
1 {
2   "resourceType": "DiagnosticReport",
3   "id": "2535064",
4   "identifier": [
5     {
6       "system": "https://dempredict.fzi.de:8888/parse/TestResult",
7       "value": "3UeiH3GFOV"
8     }
9   ],
10  "status": "final",
11  "code": {
12    "coding": [
13      {
14        "system": "http://snomed.info/sct",
15        "code": "165320004",
16        "display": "Demenz Test"
17      }
18    ]
19  },
20  "subject": {
21    "reference": "Patient/2535022",
22    "display": "schinle"
23  },
24  "effectiveDateTime": "2021-08-27T08:30:00+11:00",
25  "result": [
26    {
27      "reference": "Observation/2535062",
28      "display": "DemPredict"
29    }
30  ],
31  "conclusion": "450",
32  "conclusionCode": [
33    {
34      "coding": [
35        {
36          "system": "http://snomed.info/sct",
37          "code": "52448006",
38          "display": "Alzheimer Demenz"
39        }
40      ]
41    }
42  ]
43 }
```

**Listing A.2** Beispielhafte Abbildung einer Testsitzung aus *DemPredict* auf eine FHIR DiagnosticReport Ressource im JSON-Format

## **A.7 Fragebogen zur Technikerfahrung**

Die Einstellung gegenüber Technik und Nutzungshäufigkeit von technischen Geräten wird mit Hilfe von Fragen bestimmt, deren Einsatz in Forschungsprojekten zur Gestaltung altersgerechter Technik validiert wurde [152].

**Technik-Fragebogen**

Sehr geehrte Teilnehmerin,

Sehr geehrter Teilnehmer,

zur besseren Auswertung der Tests sind wir an Ihren Einstellungen bezüglich technischer Geräte und Entwicklungen im Allgemeinen interessiert. Zudem sind Ihre bisherigen Erfahrungen im Umgang mit (Multimedia-)Technik (= Computer und ähnliches) für uns von Interesse.

Bitte beantworten Sie dazu folgende Fragen.

**Herzlichen Dank für Ihre Mitarbeit!**

**Einstellung zur Technik**

Bitte kreuzen Sie die für Sie zutreffende Aussage an	Trifft überhaupt nicht zu	Trifft kaum zu	Teils/teils	Trifft überwiegend zu	Trifft voll zu
Der technische Fortschritt hat den Menschen überwiegend Gutes gebracht (PE <sub>1</sub> ).					
Wenn man unseren gegenwärtigen Lebensstandard aufrechterhalten will, muss man bei der technologischen Entwicklung mithalten (PE <sub>2</sub> ).					
Technik unterstützt die Erhaltung meiner Selbstständigkeit (PE <sub>3</sub> ).					
Technik ist aus meinem Alltag nicht mehr wegzudenken (PE <sub>4</sub> ).					
Ich war stets daran interessiert, die neuesten technischen Geräte zu besitzen (PE <sub>5</sub> ).					
Mich frustriert oder entmutigt die Nutzung von Technik (PE <sub>6</sub> ).					

Was verbinden Sie ganz allgemein mit Technik?

---



---

Alter: \_\_\_\_\_

männlich

weiblich

**Technik-Erfahrungen**

Bitte beantworten Sie folgende Fragen und kreuzen in den Tabellen die Aussagen an, die auf Sie zutreffen.

1. Haben Sie bereits früher (z.B. in Ihrem Berufsleben) Erfahrungen mit dem Computer gesammelt?      Ja            Nein

Wenn Ja: Vor wie vielen Jahr etwa? \_\_\_\_\_

	Täglich	3-4 Mal pro Woche	1-2 Mal pro Woche	1-2 Mal pro Monat	Seltener	Nie
Wie häufig genutzt? (HG <sub>1</sub> )						

2. Welche technischen Geräte haben Sie im vergangenen Jahr genutzt? Wie häufig?

	Täglich	3-4 Mal pro Woche	1-2 Mal pro Woche	1-2 Mal pro Monat	Seltener	Nie
Fester Computer (HG <sub>2</sub> )						
Tragbarer Computer (HG <sub>3</sub> )						
Tablet (=ohne Tastatur) (HG <sub>4</sub> )						
Mobiletelefon (HG <sub>5</sub> )						
Smartphone (HG <sub>6</sub> )						
Fernseher (HG <sub>7</sub> )						

3. Für welche Zwecke nutzen Sie im vergangenen Jahr Ihr Gerät? Wie häufig?

	Täglich	3-4 Mal pro Woche	1-2 Mal pro Woche	1-2 Mal pro Monat	Seltener	Nie
E-Mail- Kontakte pflegen (HZ <sub>1</sub> )						
Briefe o.ä. schreiben (HZ <sub>2</sub> )						
Internet-Recherche (HZ <sub>3</sub> )						
Spielen (HZ <sub>4</sub> )						
Programmieren (HZ <sub>5</sub> )						
Filme/Sendungen schauen (HZ <sub>6</sub> )						
Online-Banking (HZ <sub>7</sub> )						



## A.8 Studienteilnehmer der Beobachtungsstudie

			G	MCI	D	gesamt	
männlich	Anzahl	Verfügbare Referenz	GDS	16	2	8	26
			CERAD	7	1	4	12
			Technik	14	2	6	22
	DemPredict Sitzungen		T1	9	4	4	17
			T2	7	0	4	11
			Studienteilnehmer	10	4	4	18
	Alter		Maximum	86	79	84	86
			Minimum	62	64	60	60
			Mittelwert	71,20	71,75	72,00	71,50
			Median	69,50	72,00	72,00	69,50
Standardabweichung			8,09	6,60	12,25	8,30	
weiblich	Anzahl	Verfügbare Referenz	GDS	27	16	14	57
			CERAD	16	16	8	40
			Technik	27	16	14	57
	DemPredict Sitzungen		T1	16	11	10	37
			T2	12	7	6	25
			Studienteilnehmer	16	11	10	37
	Alter		Maximum	81	88	96	96
			Minimum	55	55	70	55
			Mittelwert	71,31	77,18	85,20	76,81
			Median	73,50	80,00	86,00	79,00
Standardabweichung			8,62	9,92	8,69	10,50	
gemischt	Anzahl	Verfügbare Referenz	GDS	43	18	22	83
			CERAD	23	17	12	52
			Technik	41	18	20	79
	DemPredict Sitzungen		T1	25	15	14	54
			T2	19	7	10	36
			Studienteilnehmer	26	15	14	55
	Alter		Maximum	86,00	88,00	96,00	96,00
			Minimum	55,00	55,00	60,00	55,00
			Mittelwert	71,27	75,73	81,43	75,07
			Median	84,50	77,00	70,00	77,00
Standardabweichung			11,19	9,26	8,26	10,08	

Tabelle A.3: Studienteilnehmerdaten der Beobachtungsstudie nach Iteration II.

# A.9 DemPredict nach Iteration II

**AUFGABE ÜBERSPRINGEN** **PAUSIEREN** **ABBRECHEN**

1. Welches Datum haben wir heute? Geben Sie das heutige Datum ein.

1999 2000 2001

↑ ↓

12 31

↑ ↓

Dezember  
Januar  
Februar

↑ ↓

OK

(a) Test A.

**PAUSIEREN** **ABBRECHEN**

Berühren Sie bitte das „Obst“ und sagen Sie mir, welches das ist.

Ergebnis: 0/3

(b) Test B.

**PAUSIEREN** **ABBRECHEN**

(c) Test C.

**PAUSIEREN** **ABBRECHEN**

(d) Test D.

**ZURÜCK** **WEITER**

**Aufgabe**

Finden Sie den unsichtbaren Weg von Start (S) zu Ziel (Z) wie soeben in der Beispielaufgabe.

(e) Test E.

**PAUSIEREN** **ABBRECHEN**

**Aufgabe**

Durchgang 1  
Wiederholen Sie den gefundenen Weg so gut Sie können.

(f) Test F.

**PAUSIEREN** **ABBRECHEN**

**Aufgabe**

Bitte schreiben Sie folgende Zahl in Worten: 243

Q W E R T Z U I O P Ü  
A S D F G H J K L Ö Ä  
Y X C V B N M

Löschen OK

(g) Test G.

**PAUSIEREN** **ABBRECHEN**

Tippen Sie die gezeigte Zahlenfolge in umgekehrter Reihenfolge ein.

Ihre Eingabe:

Bitte nur die Zahlenfolge eingeben!

4	5	6
7	8	9
0		

Löschen OK

(h) Test H.

**ZURÜCK** **WEITER**

**Beispielaufgabe**

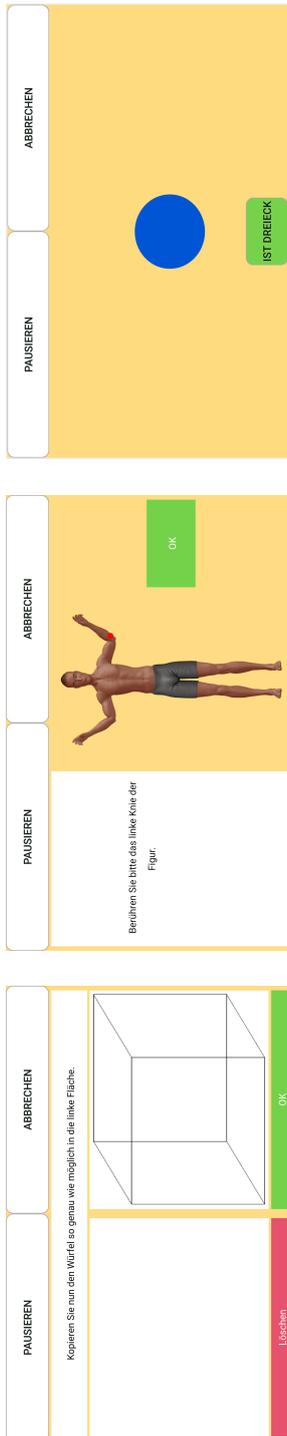
Ziehen Sie kontinuierlich die Zahl 7 ab. Sie beginnen bei 30.

1. Ergebnis:

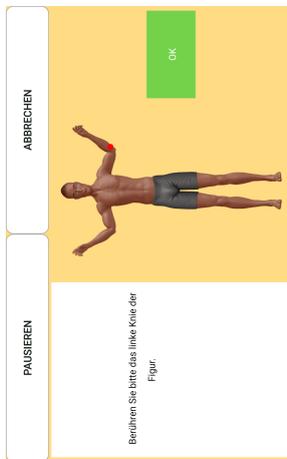
1	2	3
4	5	6
7	8	9
0		

Löschen OK

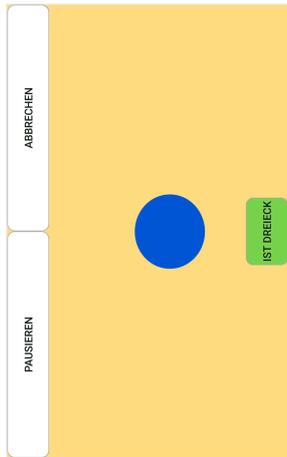
(i) Test I.



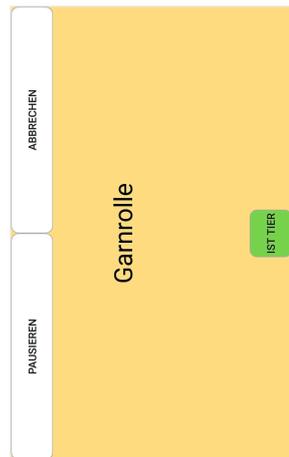
(a) Test J.



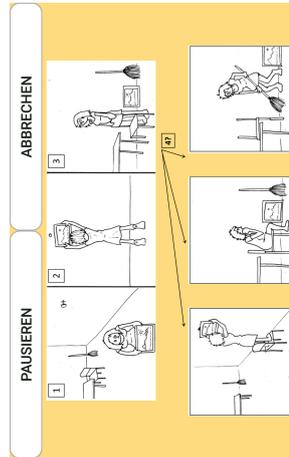
(b) Test K.



(c) Test L.



(d) Test M.



(e) Test N.

## A.10 Mögliche digitale Parameter

Die Vorverarbeitung zeigt, dass sich durch den Digitalisierungsschritt zwar nicht alle domänenspezifischen Parameter nutzen lassen, jedoch zusätzliche Parameter und -Kombinationen abgeleitet werden können. Tabelle A.4 zeigt eine Übersicht, über die in *DemPredict* verwendbaren Parameter.

Parameter	Test	Variable
<i>p1</i>	J	Bearbeitungsdauer
<i>p2</i>	J	Würfel-Punktzahl = $PP+PD+PF+PI$
<i>p3</i>	N	Bearbeitungsdauer
<i>p4</i>	N	Anzahl Fehler
<i>p5</i>	F	Bearbeitungsdauer
<i>p6</i>	F	Anzahl Fehler
<i>p7</i>	F	Anzahl ER
<i>p8</i>	F	Anzahl EL
<i>p9</i>	F	Anzahl EP
<i>p10</i>	E	Bearbeitungsdauer
<i>p11</i>	E	Anzahl Fehler
<i>p12</i>	E	Anzahl ER
<i>p13</i>	E	Anzahl EL
<i>p14</i>	E	Anzahl EP
<i>p15</i>	E	Bearbeitungsdauer Durchlauf 1 (D1)
<i>p16</i>	E	Anzahl Fehler (D1)
<i>p17</i>	E	Anzahl ER (D1)
<i>p18</i>	E	Anzahl EL (D1)
<i>p19</i>	E	Anzahl EP (D1)
<i>p20</i>	E	Bearbeitungsdauer Durchlauf 2 (D2)
<i>p21</i>	E	Anzahl Fehler (D2)
<i>p22</i>	E	Anzahl ER (D2)
<i>p23</i>	E	Anzahl EL (D2)
<i>p24</i>	E	Anzahl EP (D2)
<i>p25</i>	E	Bearbeitungsdauer-Differenz D2-D1 (Durchschnitt T1 & T2)
<i>p26</i>	E	Anzahl Fehler-Differenz D2-D1 (Durchschnitt T1 & T2)
<i>p27</i>	E	ER-Differenz D2-D1 (Durchschnitt T1 & T2)
<i>p28</i>	E	EL-Differenz D2-D1 (Durchschnitt T1 & T2)
<i>p29</i>	E	EP-Differenz D2-D1 (Durchschnitt T1 & T2)
<i>p30</i>	E+F	Bearbeitungsdauer
<i>p31</i>	E+F	Anzahl Fehler
<i>p32</i>	E+F	Anzahl ER
<i>p33</i>	E+F	Anzahl EL
<i>p34</i>	E+F	Anzahl EP
<i>p35</i>	A	Bearbeitungsdauer
<i>p36</i>	A	Anzahl Fehler

---

<i>p37</i>	A	Summe richtiger Antworten
<i>p38</i>	A	Differenz Tag
<i>p39</i>	A	Differenz Monat
<i>p40</i>	A	Differenz Jahr
<i>p41</i>	A	Datum-Differenz-Punktzahl
<i>p42</i>	I	Bearbeitungsdauer
<i>p43</i>	I	Anzahl Fehler
<i>p44</i>	L	Bearbeitungsdauer
<i>p45</i>	L	Fehler-Quote
<i>p46</i>	L	Mittlere Reaktionszeit
<i>p47</i>	C	Bearbeitungsdauer
<i>p48</i>	C	Anzahl Fehler
<i>p49</i>	D	Bearbeitungsdauer
<i>p50</i>	D	Anzahl Fehler
<i>p51</i>	C-D	Bearbeitungsdauer-Differenz
<i>p52</i>	C-D	Bearbeitungsdauer-Quotient
<i>p53</i>	M	Bearbeitungsdauer
<i>p54</i>	M	Fehler-Quote
<i>p55</i>	M	Mittlere Reaktionszeit
<i>p56</i>	H	Bearbeitungsdauer
<i>p57</i>	H	Maximale Länge gemerkter Zahlenfolge
<i>p58</i>	G	Bearbeitungsdauer L2N
<i>p59</i>	G	Bearbeitungsdauer N2L
<i>p60</i>	G	Bearbeitungsdauer
<i>p61</i>	G	Anzahl richtiger Antworten L2N
<i>p62</i>	G	Anzahl richtiger Antworten N2L
<i>p63</i>	G	Anzahl Fehler L2N
<i>p64</i>	G	Anzahl Fehler N2L
<i>p65</i>	G	Anzahl Fehler
<i>p66</i>	K	Bearbeitungsdauer
<i>p67</i>	K	Anzahl Fehler
<i>p68</i>	B	Anzahl freier Abruf Durchlauf 1 (D1)
<i>p69</i>	B	Anzahl Abruf nach Hinweis (D1)
<i>p70</i>	B	Anzahl Abruf nach Hilfestellung (D1)
<i>p71</i>	B	Anzahl freier Abruf Durchlauf 2 (D2)
<i>p72</i>	B	Anzahl Abruf nach Hinweis (D2)
<i>p73</i>	B	Anzahl Abruf nach Hilfestellung (D2)

---

<i>p74</i>	B	Anzahl freier Abruf Durchlauf 3 (D3)
<i>p75</i>	B	Anzahl Abruf nach Hinweis (D3)
<i>p76</i>	B	Anzahl Abruf nach Hilfestellung (D3)
<i>p77</i>	B	Anzahl verzögerter freier Abruf
<i>p78</i>	B	Anzahl verzögerter Abruf nach Hinweis
<i>p79</i>	B	Bearbeitungsdauer

Tabelle A.4: Übersicht über die möglichen digitalen Parameter.

## A.11 Normalverteilungsannahme für mit DemPredict erhobene Parameter

Tabelle A.5 stellt die erzielten Ergebnisse des SWTs zur Überprüfung der Normalverteilungsannahme ( $H_0$ ) für mit *DemPredict* erhobene Parameter dar. Bei  $p \leq 0,05$  ist die Normalverteilungsannahme abzulehnen. Bei  $p > 0,05$  wird die Normalverteilung angenommen.

Test	Parameter	<i>p</i> -Wert	Test	Parameter	<i>p</i> -Wert
A	p35	<b>0,143</b>	G	p9 (EP)	$5,35 * 10^{-12}$
	p37	0,0261		p60	0,039
	p38	$8,89 * 10^{-8}$		p65	0,005
	p39	$6,60 * 10^{-9}$	H	p56	0,036
	p40	$6,78 * 10^{-8}$		p57	0,001
	p41	$1 * 10^{-6}$	I	p42	$4,32 * 10^{-5}$
C	p47	0,001		p43	$4,90 * 10^{-6}$
	p48	0,002	J	p1	$4,54 * 10^{-5}$
D	p49	0,026		p2	0,0003
	p50	0,002	K	p66	<b>0,098</b>
E	p10	$3,39 * 10^{-8}$		p67	$6,32 * 10^{-5}$
	p11	$1,25 * 10^{-7}$	L	p46	$4,30 * 10^{-9}$
	p12 (ER)	$3,30 * 10^{-10}$		p45	$4,63 * 10^{-7}$
	p13 (EL)	<b>0,035</b>		p46	<b>0,721</b>
	p14 (EP)	$8,10 * 10^{-13}$	M	p53	$1,08 * 10^{-9}$
F	p5	$3,13 * 10^{-6}$		p54	$3,96 * 10^{-8}$
	p6	0,001		p55	<b>0,179</b>
	p7 (ER)	$8,56 * 10^{-7}$	N	p3	$1,80 * 10^{-5}$
	p8 (EL)	<b>0,057</b>		p4	<b>0,136</b>

---

Test Parameter  $p$ -Wert

---

Tabelle A.5: Ergebnisse des SWTs für mit *DemPredict* erhobene Parameter.

## A.12 Temporale Effekte bezüglich der Testzeitpunkte

Mit Hilfe des zweiseitigen WVRT wird die Nullhypothese  $H_0$  untersucht: Zwei gepaarte Stichproben folgen der gleichen zentralen Tendenz. In diesem Fall passiert dieser Vergleich auf Grundlage der *Dem-Predict*-Daten aus den zwei Testzeitpunkten T1 und T2. Lässt sich  $H_0$  ablehnen, ist anzunehmen dass es zwischen T1 und T2 signifikante Veränderungen in den Rängen gegeben hat und dadurch eine Verschiebung der Tendenz erfolgte. Die Ablehnung erfolgt, wenn  $p \leq p_{\text{lokal}}$  (vgl. Unterabschnitt 5.4.1). Da dies für keinen der Werte in Tabelle A.6 zutrifft, ist die Nullhypothese für alle Parameter anzunehmen.

Parameter	$p$	$p_{\text{lokal}}$	Parameter	$p$	$p_{\text{lokal}}$
p3	0,00080	0,00076	p16	0,29478	0,00128
p45	0,00457	0,00077	p62	0,33074	0,00132
p15	0,02559	0,00078	p64	0,33074	0,00135
p4	0,05005	0,00079	p1	0,35739	0,00139
p9	0,05783	0,00081	p39	0,37109	0,00143
p19	0,08898	0,00082	p8	0,37211	0,00147
p50	0,09086	0,00083	p46	0,38065	0,00152
p14	0,09824	0,00085	p23	0,39347	0,00156
p29	0,09824	0,00086	p57	0,41241	0,00161
p38	0,10035	0,00088	p49	0,42163	0,00167
p41	0,10035	0,00089	p67	0,42570	0,00172
p25	0,11088	0,00091	p34	0,43925	0,00179
p10	0,11091	0,00091	p51	0,45369	0,00185
p6	0,11463	0,00093	p22	0,45411	0,00192
p61	0,12943	0,00094	p59	0,46889	0,00200
p63	0,12943	0,00096	p42	0,48037	0,00208
p18	0,12981	0,00098	p17	0,48080	0,00217
p40	0,17357	0,00100	p56	0,53102	0,00227
p58	0,18709	0,00102	p32	0,53278	0,00238
p13	0,18824	0,00104	p66	0,54336	0,00250
p28	0,18824	0,00106	p48	0,55430	0,00263
p53	0,19728	0,00109	p35	0,57493	0,00278
p37	0,20164	0,00111	p52	0,58222	0,00294
p30	0,20213	0,00114	p36	0,58869	0,00313
p60	0,23120	0,00116	p43	0,59196	0,00333
p20	0,26222	0,00119	p31	0,62618	0,00357
p5	0,26384	0,00122	p44	0,63856	0,00385
p7	0,27538	0,00125	p47	0,64429	0,00417

Parameter	$p$	$p_{lokal}$
p54	0,64674	0,00455
p11	0,67308	0,00500
p26	0,67308	0,00556
p21	0,68253	0,00625
p55	0,80917	0,00714
p2	0,82110	0,00833
p65	0,95147	0,01000
p12	1,00000	0,01250
p24	1,00000	0,01667
p27	1,00000	0,02500
p33	NaN	0,05000

Tabelle A.6: Ergebnisse des WRTs für die mit *DemPredict* domänenspezifischen Parameter in T1 und T2.

# A.13 Aggregierter Vergleich zwischen CERAD und DemPredict

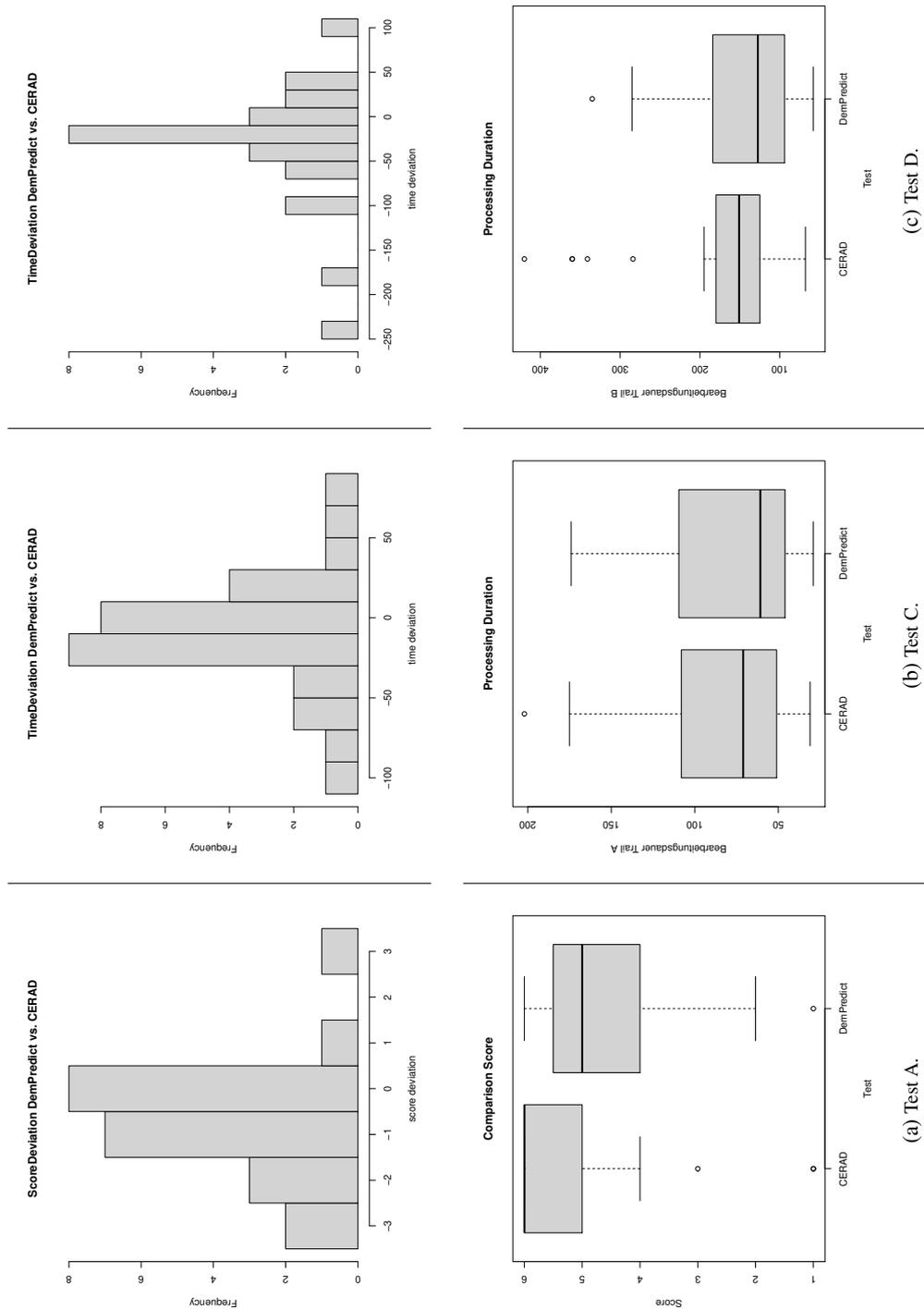


Abbildung A.5: Vergleich zwischen aggregierten Testergebnissen der CERAD- und DemPredict-Version.

## A.14 Individueller Vergleich zwischen CERAD und DemPredict

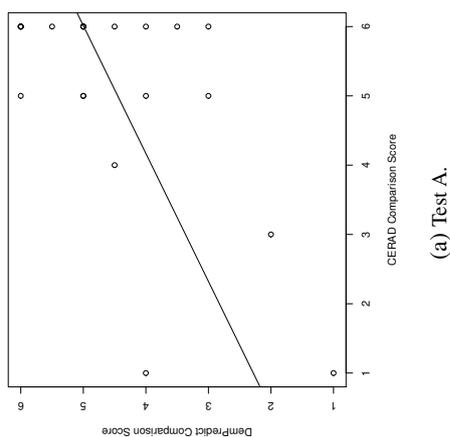
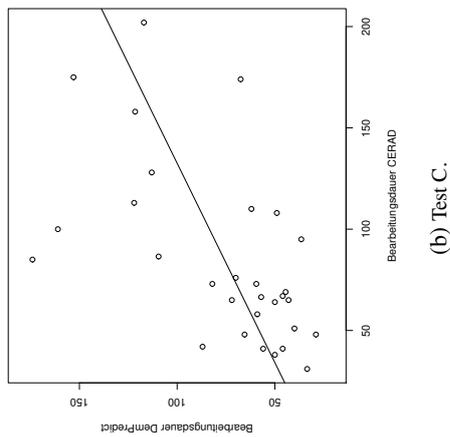
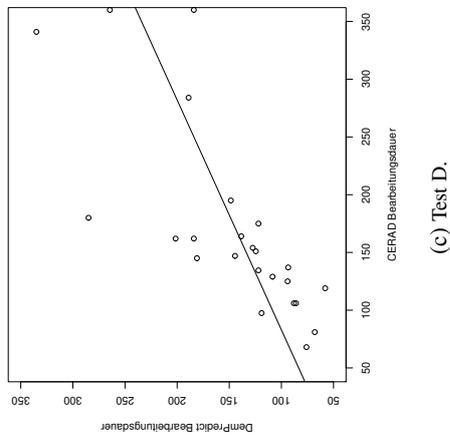


Abbildung A.7: Korrelationsanalyse zwischen individuellen Testergebnissen der CERAD- und *DemPredict*-Version nach Pearson.

## A.15 Korrelation zwischen Technologie-Erfahrung und den domänenspezifischen Parametern

Tabelle A.7 stellt die erzielten Ergebnisse der Spearman-Korrelationsanalyse zwischen der Technologie-Erfahrung (*TE*) und den digital (*DemPredict*) gesammelten Parametern dar. Bei  $p < 0,05$  besteht ein statistisch signifikanter Zusammenhang. Bei  $p \geq 0,05$  wird kein statistisch signifikanter Zusammenhang angenommen.

Test	Parameter	$\rho$	p-Wert
A	p35	-0,095	0,758
	<i>DemPredict</i> p37	0,368	0,07
	CERAD p37	<b>0,661</b>	<b>0,014</b>
	Differenz <i>DemPredict</i> -CERAD p37	-0,392	0,185
	<i>DemPredict</i> p38	-0,518	0,07
	<i>DemPredict</i> p39	-0,352	0,238
	<i>DemPredict</i> p40	-0,232	0,445
	<i>DemPredict</i> p41	-0,442	0,13
C	<i>DemPredict</i> p47	<b>-0,766</b>	<b>0,0001</b>
	CERAD p47	<b>-0,664</b>	<b>0,002</b>
	Differenz <i>DemPredict</i> -CERAD p47	-0,114	0,642
	<i>DemPredict</i> p48	-0,437	0,061
	CERAD p48	0,02	0,934
	Differenz <i>DemPredict</i> -CERAD p48	-0,329	0,17
D	<i>DemPredict</i> p49	<b>-0,625</b>	<b>0,007</b>
	CERAD p49	<b>-0,758</b>	<b>0,0004</b>
	Differenz <i>DemPredict</i> -CERAD p49	0,092	0,725
	<i>DemPredict</i> p50	-0,196	0,451
	CERAD p50	-0,296	0,249
	Differenz <i>DemPredict</i> -CERAD p50	0,113	0,665
E	p10	-0,258	0,135
	p11	-0,222	0,2
	p12 ( <i>ER</i> )	-0,149	0,393
	p13 ( <i>EL</i> )	-0,151	0,387
	p14 ( <i>EP</i> )	-0,21	0,226
	F	p5	-0,104
p6		0,227	0,191
p7 ( <i>ER</i> )		0,207	0,232
p8 ( <i>EL</i> )		0,234	0,177
p9 ( <i>EP</i> )		0,075	0,669

<i>G</i>	p60	<b>-0,537</b>	<b>0,002</b>
	p58	-0,313	0,081
	p59	<b>-0,534</b>	<b>0,002</b>
	p65	-0,053	0,774
	p63	-0,05	0,787
	p64	-0,048	0,794
<i>H</i>	p56	0,105	0,531
	p57	0,311	0,06
<i>I</i>	p42	-0,135	0,418
	p43	-0,161	0,333
<i>J</i>	p1	-0,332	0,165
	<i>DemPredict</i> p2	<b>0,6</b>	<b>0,011</b>
	CERAD p2	<b>0,571</b>	<b>0,011</b>
	Differenz <i>DemPredict</i> -CERAD p2	0,155	0,552
<i>K</i>	p66	<b>-0,327</b>	<b>0,042</b>
	p67	-0,109	0,509
<i>L</i>	p46	-0,151	0,394
	p45	-0,08	0,653
	p46	-0,192	0,277
<i>M</i>	p53	0,0589	0,734
	p54	-0,156	0,362
	p55	0,068	0,695
<i>N</i>	p3	-0,302	0,065
	p4	0,109	0,515

Tabelle A.7: Korrelation zwischen Technologie-Erfahrung und den domänenspezifischen Parametern.

## A.16 Diskriminationsfähigkeit zwischen Diagnose und den domänenspezifischen Parametern

Tabelle A.8 stellt die erzielten Ergebnisse des KWTs und WRTs für die digital (*DemPredict*) sowie gegebenenfalls analog (CERAD) gesammelten Parameter dar. Bei  $p < 0,05$  besteht ein statistisch signifikanter Zusammenhang. Bei  $p \geq 0,05$  wird kein statistisch signifikanter Zusammenhang angenommen.

Test	Parameter	KWT $p$ -Wert	WRT	WRT	WRT
			$p$ -Wert (G-MCI)	$p$ -Wert (MCI-D)	$p$ -Wert (G-D)
A	p35	<b>0,014</b>	<b>0,011</b>	1	0,074
	p37	0,079	0,59	0,13	0,2

	p38	<b>0,008</b>	1	0,22	0,22
	p39	0,261	0,7	0,7	0,43
	p40	0,411	0,8	0,79	0,79
	p41	<b>0,01</b>	0,804	<b>0,041</b>	<b>0,041</b>
<i>C</i>	p47	<b>0,001</b>	<b>0,003</b>	<b>0,044</b>	<b>0,015</b>
	p48	0,227	0,98	0,98	0,14
<i>D</i>	p49	<b>0,009</b>	<b>0,012</b>	0,8	0,166
	p50	0,38	0,79	0,79	0,88
<i>E</i>	p10	<b>0,002</b>	<b>0,028</b>	0,096	<b>0,008</b>
	p11	<b>0,002</b>	0,206	0,206	<b>0,002</b>
	p12 ( <i>ER</i> )	<b>0,003</b>	0,115	0,218	<b>0,003</b>
	p13 ( <i>EL</i> )	<b>0,002</b>	0,23	<b>0,025</b>	<b>0,002</b>
	p14 ( <i>EP</i> )	0,063	0,89	0,14	0,11
	p15	<b>0,003</b>	<b>0,036</b>	0,185	<b>0,014</b>
	p16	<b>0,005</b>	0,175	0,175	<b>0,006</b>
	<i>ER<sub>E1</sub></i> (p17)	<b>0,004</b>	0,061	0,180	<b>0,009</b>
	<i>EL<sub>E1</sub></i> (p18)	<b>0,005</b>	0,160	0,075	<b>0,006</b>
	<i>EP<sub>E1</sub></i> (p19)	0,539	0,98	0,98	0,98
	p20	<b>0,000</b>	<b>0,005</b>	0,328	<b>0,005</b>
	p21	<b>0,001</b>	0,110	0,110	<b>0,001</b>
	<i>ER<sub>E2</sub></i> (p22)	<b>0,018</b>	0,232	0,364	<b>0,017</b>
	<i>EL<sub>E2</sub></i> (p23)	<b>0,004</b>	0,338	0,071	<b>0,002</b>
	<i>EP<sub>E2</sub></i> (p24)	<b>0,006</b>	0,633	0,117	<b>0,009</b>
<i>F</i>	p5	<b>0</b>	<b>0,027</b>	<b>0,027</b>	<b>0,001</b>
	p6	<b>0,01</b>	0,117	0,117	<b>0,022</b>
	p7 ( <i>ER</i> )	<b>0,015</b>	0,271	0,162	<b>0,014</b>
	p8 ( <i>EL</i> )	<b>0,007</b>	0,126	0,126	<b>0,012</b>
	p9 ( <i>EP</i> )	<b>0</b>	0,517	<b>0,001</b>	<b>0</b>
<i>G</i>	p60	<b>0,018</b>	0,062	0,775	0,071
	p58	0,302	0,65	0,75	0,65
	p59	<b>0,05</b>	0,425	0,425	0,073
	p65	0,052	0,219	0,948	0,083
	p63	<b>0,012</b>	0,196	0,489	<b>0,015</b>
	p64	0,583	1	1	1
<i>H</i>	p56	0,594	1	1	0,9
	p57	<b>0,002</b>	0,107	0,068	<b>0,003</b>
<i>I</i>	p42	<b>0</b>	0,139	<b>0,004</b>	<b>0</b>

	p43	0,162	0,79	0,26	0,26
<i>J</i>	p1	<b>0,029</b>	<b>0,028</b>	0,412	0,412
	p2	<b>0,002</b>	<b>0,033</b>	0,133	<b>0,006</b>
<i>K</i>	p66	<b>0,038</b>	0,052	0,164	0,215
	p67	<b>0,004</b>	0,231	0,231	<b>0,004</b>
<i>L</i>	p46	0,165	0,32	0,67	0,32
	p45	<b>0,048</b>	0,552	0,089	0,089
	p46	0,263	0,32	0,95	0,71
<i>M</i>	p53	<b>0,018</b>	0,062	0,589	0,062
	p54	<b>0,002</b>	0,073	0,101	<b>0,004</b>
	p55	0,059	0,059	0,491	0,491
<i>N</i>	p3	0,663	1	1	1
	p4	<b>0,019</b>	0,31	0,31	<b>0,02</b>

Tabelle A.8: Diskriminationsfähigkeit zwischen Diagnosegruppen anhand von KWT und WRT.

## A.17 Mögliche Schwellwerte auf Merkmalsebene

In Tabelle A.9 ist eine Übersicht über die individuellen Schwellwerte der digitalen Parameter. Hierbei gilt für die meisten Merkmale je geringer der Wert, desto höher die kognitive Leistungsfähigkeit und somit die Entscheidung in Richtung gesünderer Gruppe. Ausgenommen sind umgekehrt skalierten Merkmale wie die Punktzahlen in p2, p37, p57, p61 und p62, bei denen die entsprechend umgekehrte Logik gilt.

Parameter	Abzugrenzende Klassenpaare			
	G-NG	G-MCI	MCI-D	G-D
p1	95,5	95,5	103	103
p2	3	3	2,5	3
p3	248	248	270	270
p4	3,62	3,62	4,5	4
p5	55,5	55,5	80	80
p6	9,38	9,38	16	11
p7	3	3,15	6,556	3
p8	6	6	8	8
p9	0,5	Inf	0,5	0,5
p10	132	132	172,5	134
p11	21	21	39,5	29
p12	10	14	16	10
p13	14,192	14,192	20	18,5
p14	0,5	1,077	0,5	0,5
p15	70	70	105	86
p16	12,5	13	17,5	12,5
p17	3,5	3,5	9,19	3,5
p18	8,192	8,19	11,44	10
p19	0,308	0,308	0,438	0,438
p20	62	62	85,5	69
p21	12	12	16,5	16
p22	6	2,5	9,375	6
p23	6	6	8	7
p24	0,5	0,769	0,5	0,5
p25	-8,375	-12,654	-8,375	-8,375
p26	-1,9375	-2	-1,938	-1,938
p27	0	0	0,188	0,188
p28	0	0	-2,188	-2,188
p29	0,063	0,462	0,063	0,063
p30	-10,229	-10,23	-9,33	-9,33
p31	-1,56	-1,56	-5,5	6

p32	0	0	0	0
p33	-0,146	-0,5	1	1
p33	0,25	-0,083	0,25	0,25
p35	179	179	210,5	162
p37	4,5	4,5	4,5	4,5
p38	2	0,7	3	3
p39	0,05	0,05	1,917	1,917
p40	8	0,9	1	8
p41	3	1,25	3	3
p42	54,615	49,5	70,5	70,5
p43	0,538	0,538	1	1,136
p44	62	61,5	60,409	62
p45	6,92	6,92	7,08	13,46
p46	1129,58	1131,21	1170,27	1129,58
p47	62	62	87	65,5
p48	1,5	1	1,5	1,5
p49	144,5	138,5	207	148,5
p50	2,5	2,5	2,5	2,5
p51	80	72	117,46	117,46
p52	1,99	2,046	2,33	1,99
p53	61,7	63	64,5	61,7
p54	4,25	4,245	9,899	5,634
p55	1224,423	1290	1358,222	1224,423
p56	81	113	81	81
p57	2,5	3	2	2,5
p58	67,313	56	68	68
p59	152,5	190,438	152,5	152,5
p60	292	295	336,25	292
p61	2	1,813	2	2
p62	1,063	1,063	1	1
p63	1	1,188	1	1
p64	1,938	1,938	2	2
p65	3	3	3,417	3,417
p66	335,5	335,5	393	303
p67	1,5	1,5	2	

Tabelle A.9: Übersicht über die individuellen Schwellwerte der digitalen Parameter.

## A.18 Algorithmus für Explainable k-Nearest Neighbors

---

**Algorithmus 1** : Algorithmus für Explainable k-Nearest Neighbors

---

**Input** : Eingabe instanzen, trainingsdaten

**Output** : Ergebnisvektor e

```

1 datenPCA ← PCA(trainingsdaten) // PCA auf Trainingsdaten ausführen
2 nachbarn ← vector()
3 for i in instanzen do
4   neueDatenPCA ← PCAhinzu(datenPCA, i) // Instanz als Punkt zu PCA hinzufügen
5   plot(neueDatenPCA)
6   for t in Trainingsdaten do
7     neueDistanz ← distanz(i, t) // euklidische Distanz zwischen i und t
8     distanzen ← c(distanzen, neueDistanz)
9   sort(distanzen) // Sortieren des Vektors in absteigender Reihenfolge
10  indices ← index(distanzen) // Indizes der Entfernungen ermitteln
11  nachbarn[i] ← findSamples(indices) // Trainingsdaten mit gleichen Indizes finden
12  inverseStandardization(nachbarn) // Standardisierung rückgängig machen
13 e ← c(e, nachbarn)
14 return: e

```

---

## A.19 Evaluationsdaten der XAI-Ansätze

Verfahren \ Sample	RF			k-NN		
	SHAP	LIME	CHIRPS	SHAP	LIME	Xk-NN
1	1,114801	0,086768	2,947434	1,052275	<b>0,059840</b>	0,400724
2	1,181745	0,076671	1,981789	1,221874	<b>0,058819</b>	0,154525
3	1,302874	0,076504	1,215481	1,097692	<b>0,065305</b>	0,177620
4	1,115923	0,083625	2,722929	1,179938	<b>0,072128</b>	0,166986
5	1,163006	0,077756	2,670389	1,102564	<b>0,056956</b>	0,163857
6	1,265648	0,077082	2,645522	1,260992	<b>0,056550</b>	0,161849
7	1,294876	0,079418	1,922528	1,227822	<b>0,057625</b>	0,166140
8	1,343950	0,078795	2,718031	1,258102	<b>0,056706</b>	0,150632
9	1,222501	0,078295	2,808266	1,242083	<b>0,056863</b>	0,166590
10	1,326201	0,077707	2,694350	1,321071	<b>0,055966</b>	0,165882

Tabelle A.10: Berechnungszeiten in Sekunden.

Iteration \ Sample	Sample											
	1	2	3	4	5	6	7	8	9	10	11	12
0	0,874	0,868	0,392	0,938	0,98	0,498	0,482	0,486	0,502	0,836	0,38	0,662
1	0,814	0,846	0,636	0,77	0,958	0,452	0,552	0,72	0,494	0,944	0,302	0,63
2	0,742	0,86	0,63	0,684	0,958	0,412	0,662	0,714	0,662	0,884	0,302	0,596
3	0,742	0,806	0,658	0,596	0,894	0,412	0,63	0,722	0,664	0,884	0,272	0,596
4	NA	0,818	0,618	0,508	NA	0,41	0,644	0,686	0,636	0,89	0,322	NA
5	NA	0,854	0,642	0,398	NA	NA	0,692	0,674	0,68	NA	0,308	NA
6	NA	0,852	0,644	0,34	NA	NA	0,732	0,678	0,672	NA	0,246	NA
7	NA	0,846	0,636	0,314	NA	NA	0,732	0,652	0,678	NA	0,212	NA
8	NA	NA	NA	0,282	NA	NA	NA	NA	NA	NA	0,174	NA
9	NA	NA	NA	0,274	NA	NA	NA	NA	NA	NA	0,162	NA
10	NA	NA	NA	0,228	NA	NA	NA	NA	NA	NA	0,162	NA
11	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
12	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
13	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Tabelle A.11: MoRF-Störungswerte des RF LIME.

Iteration \ Sample	Sample											
	1	2	3	4	5	6	7	8	9	10	11	12
0	0,874	0,868	0,392	0,938	0,98	0,498	0,482	0,486	0,502	0,836	0,38	0,662
1	0,536	0,846	0,376	0,77	0,422	0,296	0,552	0,474	0,494	0,298	0,302	0,292
2	0,214	0,78	0,38	0,684	0,23	0,258	0,512	0,48	0,476	0,196	0,272	0,12
3	0,198	0,806	0,35	0,584	0,206	0,264	0,492	0,426	0,454	0,174	0,238	0,116
4	0,178	0,78	0,34	0,494	0,186	0,244	0,458	0,394	0,438	0,174	0,202	0,116
5	0,186	0,73	0,32	0,398	0,186	0,238	0,462	0,38	0,434	0,148	0,172	0,09
6	0,168	0,696	0,314	0,39	0,158	0,208	0,572	0,354	0,432	0,124	0,152	0,09
7	0,142	0,694	0,308	0,356	0,16	0,208	0,57	0,346	NA	0,114	0,126	0,076
8	0,142	0,704	NA	0,318	0,154	0,172	NA	0,338	NA	0,104	0,126	0,074
9	0,136	0,7	NA	0,26	0,118	0,162	NA	0,33	NA	0,102	NA	0,062
10	0,104	0,684	NA	0,242	0,112	0,158	NA	NA	NA	0,082	NA	0,06
11	0,102	0,678	NA	0,196	0,112	0,16	NA	NA	NA	0,082	NA	NA
12	0,1	0,72	NA	0,184	0,11	NA						
13	NA	0,696	NA									

Tabelle A.12: MoRF-Störungswerte des RF SHAP.

Iteration \ Sample	Sample										
	1	2	3	4	5	6	7	8	9	10	11
0	0,556	0,667	0,556	0,556	0,778	0,778	0,556	0,778	0,556	0,889	0,667
1	0,556	0,778	0,556	0,667	0,778	0,778	0,556	0,667	0,333	0,889	0,667
2	0,556	0,667	0,333	0,556	0,778	0,667	0,556	0,667	0,556	0,667	0,667
3	0,556	0,556	0,333	0,556	0,778	0,667	0,556	0,667	0,444	0,778	0,778
4	0,556	0,667	0,556	0,444	0,778	NA	NA	NA	0,444	0,667	0,667
5	0,444	0,667	0,556	0,333	NA	NA	NA	NA	0,444	NA	0,444
6	0,444	0,667	0,556	0,333	NA	NA	NA	NA	0,444	NA	0,333
7	NA	0,444	NA	0,333							
8	NA	0,333	NA	0,444							
9	NA	0,111	NA	0,444							
10	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
11	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
12	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
13	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Tabelle A.13: MoRF-Störungswerte des k-NN LIME.

Iteration \ Sample	Sample										
	1	2	3	4	5	6	7	8	9	10	11
0	0,556	0,667	0,556	0,556	0,778	0,778	0,556	0,778	0,556	0,889	0,667
1	0,556	0,444	0,333	0,667	0,667	0,667	0,556	0,667	0,333	0,556	0,333
2	0,556	0,333	0,333	0,556	0,556	0,667	0,444	0,556	0,333	0,556	0,333
3	0,556	0,222	0,333	0,556	0,556	0,556	0,333	0,556	0,111	0,556	0,333
4	0,556	0,222	0,333	0,556	0,444	0,556	0,333	0,444	0,111	0,333	0,111
5	0,556	0,111	0,333	0,556	0,444	0,444	0,333	0,444	0,222	0,444	0,111
6	0,556	0	0,333	0,444	0,444	0,444	0,333	0,444	0,111	0,444	0
7	0,444	0	0,333	0,333	0,444	0,444	0,333	0,444	0,111	0,444	0
8	0,444	0	0,333	0,333	0,444	0,444	0,333	0,444	0,111	0,444	0
9	0,444	0	0,333	0,333	0,444	0,444	0,333	0,444	0,333	0,444	NA
10	0,444	0	0,333	0,333	0,444	0,444	NA	0,556	0,333	0,444	NA
11	NA	NA	NA	0,333	0,444	0,444	NA	NA	0,333	NA	NA
12	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
13	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Tabelle A.14: MoRF-Störungswerte des k-NN SHAP.

Iteration \ Merkmal	1	2	3	4	5	6	7	8	9	10
p47	-0,125245318	-0,125685795	-0,124419599	-0,123326655	-0,123717722	-0,124598751	-0,126937928	-0,125672514	-0,124503196	-0,127478765
p35	-0,066183515	-0,06554982	-0,067664878	-0,066110081	-0,064331028	-0,067861506	-0,063695066	-0,065613627	-0,067555416	-0,063831882
p1	-0,045683551	-0,041604378	-0,043698015	-0,042905227	-0,04036824	-0,044138434	-0,043194573	-0,044822136	-0,04071134	-0,044465901
p2	0,032822261	0,035526085	0,034369812	0,037367354	0,03831463	0,032450327	0,03764045	0,034905664	0,036969963	0,037515328
Alter	-0,029185434	-0,032992942	-0,029784034	-0,029675748	-0,027669125	-0,027317777	-0,026461913	-0,028858838	-0,028337834	-0,028217296
p50	-0,012283448	-0,014185144	-0,011837387	-0,013787781	-0,016433393	-0,016948902	-0,011423043	-0,014084075	-0,011169431	-0,018714142

Tabelle A.15: Gewichtung der Merkmale für RF LIME.

Iteration \ Merkmal	1	2	3	4	5	6	7	8	9	10
p47	0,18580854	0,185819258	0,185866386	0,185720377	0,185789223	0,18574141	0,185740461	0,185789368	0,185797298	0,185758322
p35	0,129604034	0,129478391	0,129600183	0,129562015	0,129540581	0,129485873	0,129513463	0,129575236	0,129448035	0,129477255
p1	0,058002482	0,057954043	0,057959064	0,057970843	0,05805207	0,057982461	0,05796822	0,058044484	0,058043712	0,057996104
p2	0,040744445	0,04090695	0,040926836	0,041014296	0,04085031	0,040954407	0,040864007	0,04096535	0,040916311	0,040961582
Alter	0,02609801	0,026116858	0,026097461	0,026056028	0,026016268	0,026139764	0,026144268	0,026099674	0,026188493	0,026193432
p37	0,015650419	0,015639591	0,015581309	0,015716971	0,015698672	0,015669683	0,015672571	0,015668372	0,015584853	0,015770858

Tabelle A.16: Gewichtung der Merkmale für RF SHAP.

Iteration \ Merkmal	1	2	3	4	5	6	7	8	9	10
p47	-0,100103531	-0,101385197	-0,099214912	-0,099503031	-0,100066723	-0,100750565	-0,098695745	-0,097924093	-0,099004748	-0,098842324
p2	0,090457161	0,090336376	0,090048173	0,091848157	0,09042477	0,091402397	0,090940738	0,091540308	0,091606706	0,090825255
Alter	-0,070511462	-0,070388453	-0,072832348	-0,073936004	-0,071086804	-0,073470667	-0,072277013	-0,074824605	-0,07315929	-0,072182682
p35	-0,062983271	-0,061946734	-0,061548814	-0,061812315	-0,061341771	-0,063818539	-0,063077991	-0,062168068	-0,066207124	-0,061290991
p1	-0,059786329	-0,061012813	-0,060726115	-0,058656546	-0,061099292	-0,060634944	-0,063439619	-0,061917846	-0,064675418	-0,060054485
p49	-0,057356734	-0,055864271	-0,05700997	-0,055064942	-0,052665431	-0,053315267	-0,056941788	-0,056720885	-0,056015152	-0,054062857

Tabelle A.17: Gewichtung der Merkmale für k-NN LIME.

Iteration \ Merkmal	1	2	3	4	5	6	7	8	9	10
p2	0,083770127	0,083478404	0,083189988	0,083431864	0,083679831	0,083449272	0,083645209	0,083727381	0,083250479	0,083378466
p1	0,063958969	0,064339181	0,064382478	0,064102149	0,064451535	0,064426695	0,064084209	0,064775932	0,064522459	0,064281122
Alter	0,06352014	0,063820961	0,063510336	0,063448521	0,063849346	0,06315365	0,063810867	0,063781281	0,063789646	0,063651535
p35	0,037730059	0,037002758	0,037590045	0,037370959	0,037592535	0,037409812	0,037576983	0,037616475	0,037592193	0,037307746
p47	0,037392759	0,037908235	0,037523121	0,037463678	0,036838345	0,037796157	0,037208542	0,037589139	0,037195583	0,037661047
p50	0,022774372	0,023016012	0,023001741	0,022961523	0,021812061	0,022575248	0,022819065	0,022703361	0,022333108	0,022425059

Tabelle A.18: Gewichtung der Merkmale für k-NN SHAP.

## A.20 Demografische Daten zur Umfrage der Adhärenz

Studie	Geschlecht (Anzahl)		Alter in Jahren			
	<i>männlich</i>	<i>weiblich</i>	$\bar{x}$	$\sigma$	<i>max.</i>	<i>min.</i>
vorab	24	5	26,8	12,18	60	18
longitudinal	14	4	26	11,95	60	19

Tabelle A.19: Demografische Daten der Studie zur Adhärenz.

## A.21 Umfrage zur Akzeptanz bei Leistungserbringern

Die Akzeptanz für digitale Unterstützungslösungen wie *DemPredict* bei niedergelassenen Ärzten wurde durch die Konzeption des DHCQ im Rahmen einer Umfrage in Kooperation mit der Universität Koblenz-Landau unter Einbezug der KV Baden-Württemberg, Thüringen und Sachsen erhoben.



---

nutzungswahrscheinlichkeit App → qnr1

11.03.2021, 07:48

**Seite 01**

EE

**Die Richtlinien guter ethischer Forschung sehen vor, dass sich die Teilnehmer/innen an empirischen Studien explizit und nachvollziehbar mit der Teilnahme einverstanden erklären.**

**Freiwilligkeit.** Ihre Teilnahme an dieser Untersuchung ist freiwillig. Es steht Ihnen zu jedem Zeitpunkt dieser Studie frei, Ihre Teilnahme abubrechen, ohne dass Ihnen daraus Nachteile entstehen.

**Anonymität.** Ihre Daten sind selbstverständlich vertraulich, werden nur in anonymisierter Form ausgewertet und nicht an Dritte weitergegeben. Demographische Angaben wie Alter oder Geschlecht lassen keinen eindeutigen Schluss auf Ihre Person zu.

**Fragen.** Falls Sie noch Fragen zu dieser Studie haben sollten, finden Sie im Anschluss ein Impressum mit Kontaktdaten der Studienleiter.

- Hiermit bestätige ich, dass ich mindestens 18 Jahre alt bin sowie die Einverständniserklärung gelesen und verstanden habe

## Datenschutz

Die Ergebnisse und Daten dieser Studie können als wissenschaftliche Publikation veröffentlicht werden. Dies geschieht in anonymisierter\* Form, d. h. ohne dass die Daten einer spezifischen Person zugeordnet werden können. Die vollständig anonymisierten Daten dieser Studie können im Internet in einem öffentlichen Datenarchiv zugänglich gemacht werden. Damit folgt diese Studie den Empfehlungen der Deutschen Forschungsgemeinschaft (DFG) und der Deutschen Gesellschaft für Psychologie (DGPs) zur Qualitätssicherung in der Forschung.

\*Anonymisieren ist das Verändern personenbezogener Daten in der Weise, dass Einzelangaben über persönliche und sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbarer natürlichen Person zugeordnet werden können

Sie können Ihre Einwilligung zur Speicherung der Daten bis zum Ende der Datenerhebung widerrufen, ohne dass Ihnen daraus Nachteile entstehen. Nach Abschluss der Studie können Sie die Daten nicht mehr löschen lassen, da der Datensatz anonymisiert wurde und somit nicht mehr Ihrer Person zugeordnet werden kann.

Während der Studie wird als einzige persönliche Angabe ihre Mailadresse zwecks Kontaktaufnahme während der Studierhebung separat von den Studien-bezogenen Daten gespeichert, sodass kein Bezug zwischen Mailadresse und Ihren anonymisierten Studiendaten hergestellt werden kann. Eine nachträgliche Zuordnung der Daten zu einer bestimmten Person ist aufgrund der Anonymisierung nicht möglich.

Die für die Studie wichtigen Daten werden ausschließlich in anonymisierter Form gespeichert, ausgewertet und gegebenenfalls weitergegeben. Die Daten sind gegen unbefugten Zugriff gesichert. Die personenbezogenen Daten werden bereits zum Zeitpunkt der Datenerhebung anonymisiert.

Dritte erhalten keinen Einblick in Originalunterlagen. Die in der Studie erhobenen Daten werden für einen Zeitraum von mind. 10 Jahren aufbewahrt. Studienteilnehmende können sich nach Abschluss der Studie unter den unten angegebenen Kontaktdaten über die allgemeinen Ergebnisse der Studie informieren. Überdies werden die Daten spätestens fünf Jahre nach Abschluss der Studie anonym veröffentlicht.

Verantwortung für die sachgemäße Datenverarbeitung trägt der Studienleiter (Herrn Markus Schinle – [schinle@fzi.de](mailto:schinle@fzi.de)).

Darüber hinaus weisen wir Sie auf das Beschwerderecht bei einer Datenschutz-Aufsichtsbehörde hin:

Kontaktdaten der zuständigen Aufsichtsbehörde:  
Der Landesbeauftragte für den Datenschutz Rhein-land-Pfalz, Hintere Bleiche 34, 55116 Mainz  
Postanschrift: Postfach 30 40 55020 Mainz  
Telefon +49 6131 2082449  
Telefax +49 6131 2082497  
E-Mail: [poststelle@datenschutz.rlp.de](mailto:poststelle@datenschutz.rlp.de)

Kontaktdaten des zuständigen Datenschutzbeauftragten, Campus Landau:  
Universität Koblenz-Landau,  
Datenschutzbeauftragte: Dr. Susanne Weis,  
Campus Landau, Fortstr. 7, 76829 Landau,  
E-Mail: [datenschutz@uni-landau.de](mailto:datenschutz@uni-landau.de).

Ich habe die Datenschutz Richtlinien gelesen und bin mit dem Vorgehen einverstanden

**1. Alter**

Alter

<30

45-60

30-45

>60

**2. Geschlecht**

Geschlecht

- männlich
- weiblich
- neutral

**3. Fachrichtung**

Allgemeinmedizin (Hausarzt)

Facharzt Neurologie

Sonstiges

**4. Kassenzulassung**

- ja
- nein

**5. Führen Sie Demenzdiagnosen durch?**

- ja
- nein

**6. Wenn ja, mit welchem Test?**

Name des Tests

**7. Besitzen Sie ein Smartphone oder Tablet?**

- ja
- nein

**8. Kennen Sie Applikationen, die die Arbeit von Ärzten erleichtern sollen?**

- ja
- nein

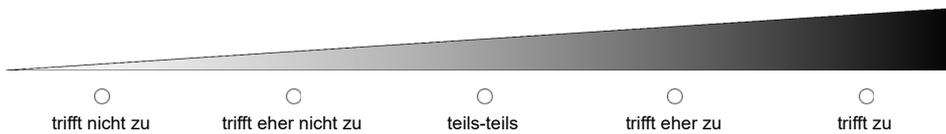
**9. Wenn ja, welche?**

Name der Applikation/en

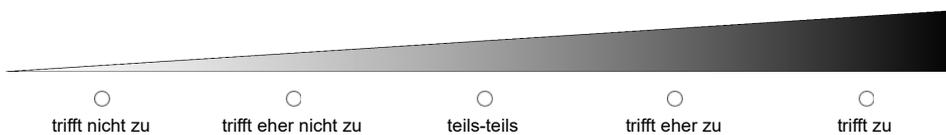
**10. Sind ihre Patientenakten digital oder in Papierform abgelegt?**

- digital
- papierform
- gemischt

**11. Im Umgang mit Technik bin ich sicherer als der Durchschnitt**



**12. Technik ist aus meinem Alltag nicht mehr wegzudenken**



**13. Ich war stets daran interessiert, die neuesten technischen Geräte zu besitzen**

trifft nicht zu     trifft eher nicht zu     teils-teils     trifft eher zu     trifft zu

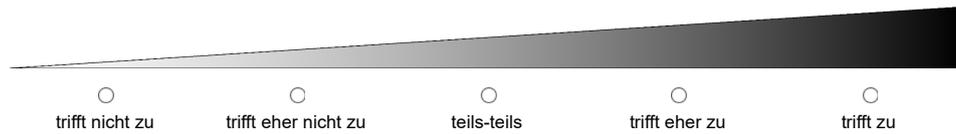
**14. Es fällt mir leicht, die Bedienung eines elektronischen Geräts zu lernen**

trifft nicht zu     trifft eher nicht zu     teils-teils     trifft eher zu     trifft zu

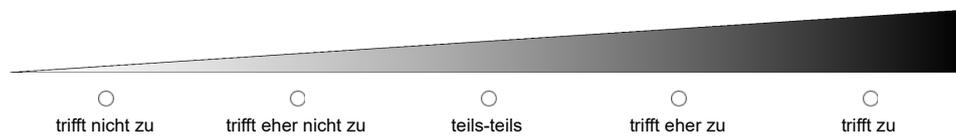
**15. Ich kenne mich im Bereich elektronischer Geräte aus**

trifft nicht zu     trifft eher nicht zu     teils-teils     trifft eher zu     trifft zu

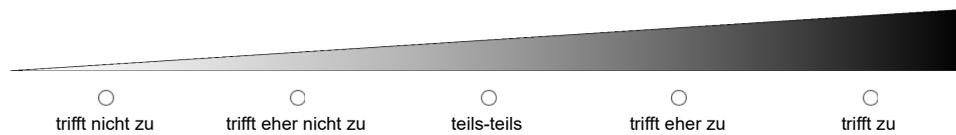
**16. Der technische Fortschritt hat den Menschen überwiegend Gutes gebracht**



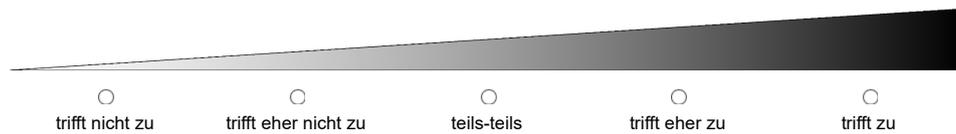
**17. Wenn man unseren gegenwärtigen Lebensstandard aufrechterhalten will, muss man bei der technologischen Entwicklung mithalten**



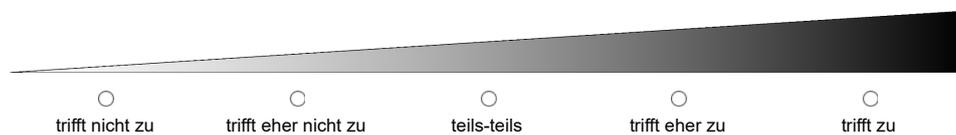
**18. Ich vermute, dass sich die Arbeit in meiner Praxis in den nächsten Jahren durch die Technik ändern wird**



**19. Der Technikfortschritt in der Medizinbranche bereitet mir Unbehagen**



**20. Elektronische Geräte erleichtern mir den Alltag**



**21. Elektronische Geräte verursachen bei mir Stress**

trifft nicht zu     trifft eher nicht zu     teils-teils     trifft eher zu     trifft zu

**22. Elektronische Geräte machen vieles Umständlicher**

trifft nicht zu     trifft eher nicht zu     teils-teils     trifft eher zu     trifft zu

**23. Im Umgang mit Technik habe ich Angst etwas kaputt zu machen**

trifft nicht zu     trifft eher nicht zu     teils-teils     trifft eher zu     trifft zu

**24. Technik ist für mich nur ein Mittel zum Zweck**

trifft nicht zu     trifft eher nicht zu     teils-teils     trifft eher zu     trifft zu

**25. Ich habe keine Scheu davor technische Geräte auszuprobieren**

trifft nicht zu     trifft eher nicht zu     teils-teils     trifft eher zu     trifft zu

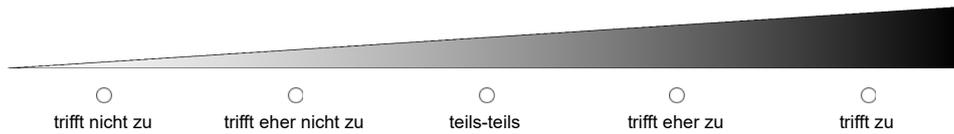
**26. Potenzielle Demenzpatienten sollen direkt beim Hausarzt durch ein digitales Screening identifiziert werden. Die durch die Patienten selbstständige Verwendung dieser App bspw. im Wartezimmer, ermöglicht die Einschätzung der kognitiven Leistungsfähigkeit und somit die Abgrenzung zwischen gesunden Menschen von Menschen mit milden kognitiven Einschränkungen sowie Demenz. Diese Vorgehensweise soll als Aufgreifkriterium für Menschen im ambulanten Umfeld dienen und den niederschweligen Einstieg in das Versorgungssystem – besonders durch eine enorme Zeitersparnis auf Seiten der Ärzte – ermöglichen.**

**Die Applikation DemPredict wurde in anderen Studien bereits auf ihre Bedienbarkeit (an einer Stichprobe älterer Patienten) und auf ihre Validität und Reliabilität (an einer Stichprobe demenzerkrankter Personen) getestet.**

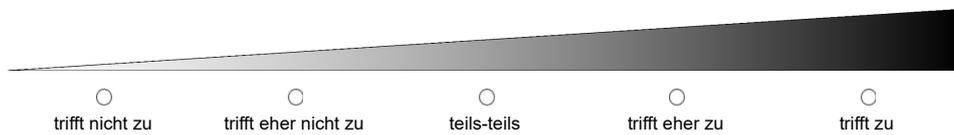
Um die Auswertbarkeit der Studie zu gewährleisten, beschreiben Sie bitte kurz in eigenen Worten, was die Applikation DemPredict tun soll.

Beschreibung:

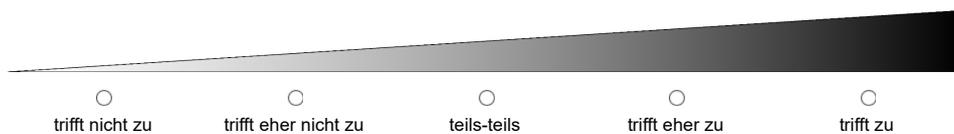
**27. Eine sinnvoll entworfene Applikation kann eine Anamnese genauso gut unterstützen, wie ein Papiertest**



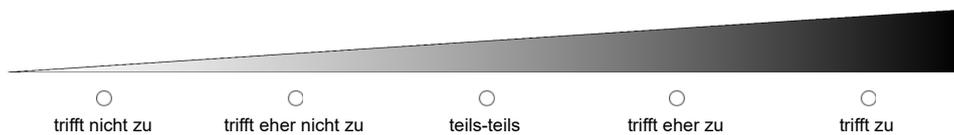
**28. Ich denke, dass eine Applikation mehr Aufwand macht als ein Papiertest**



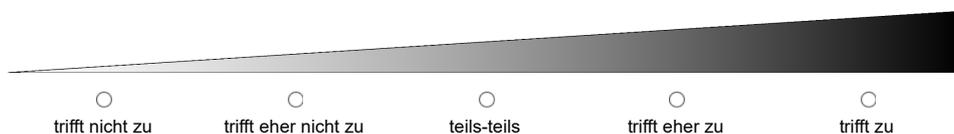
**29. Ich befürchte Probleme mit dem Datenschutz**

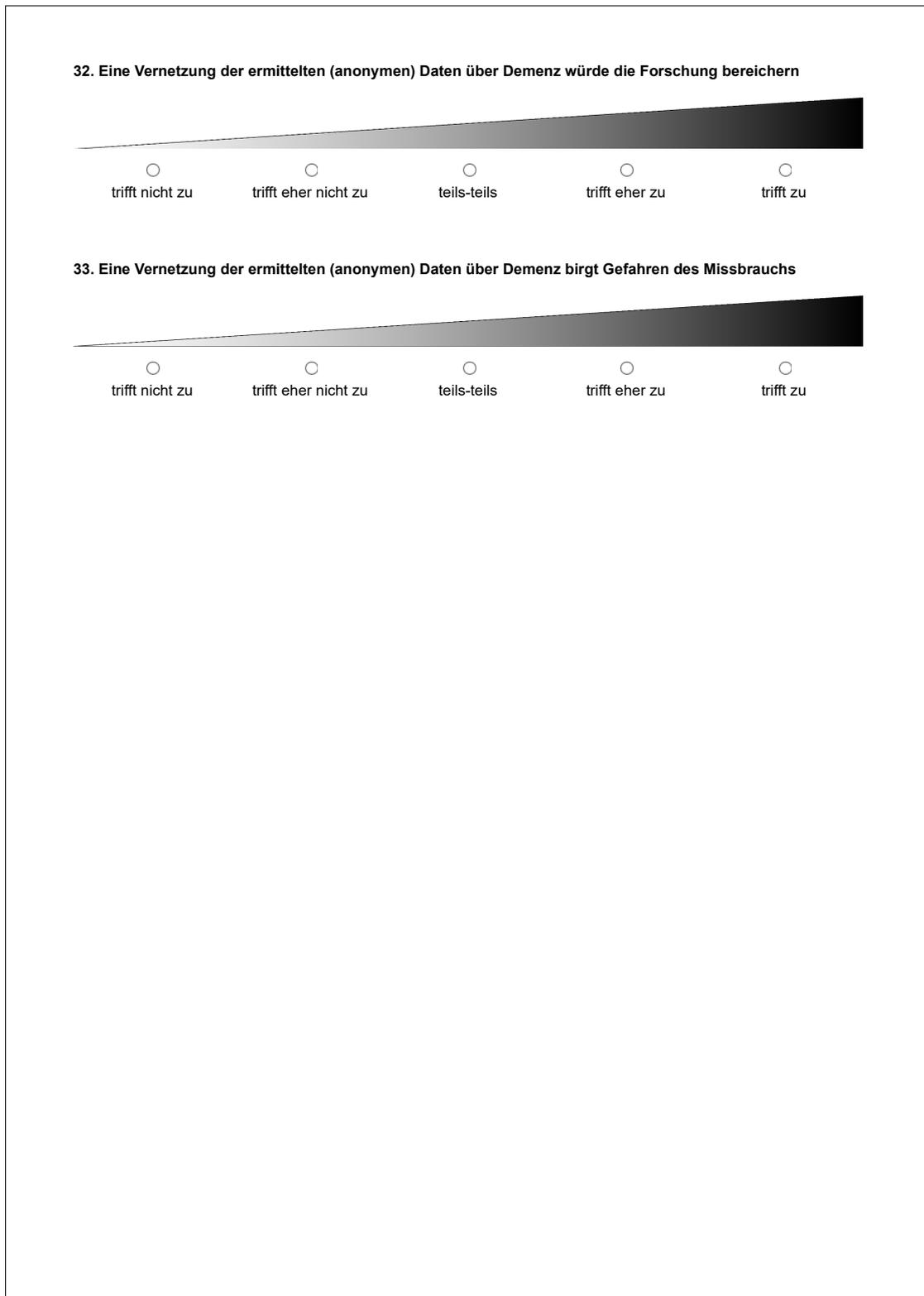


**30. Eine automatische Auswertung erleichtert mir den Praxisalltag**

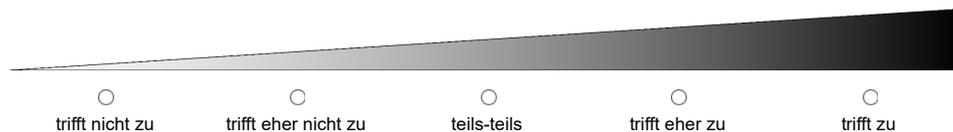


**31. Ich halte es für zeitaufwändiger mit älteren Menschen eine App zu nutzen, als einen Papiertest**

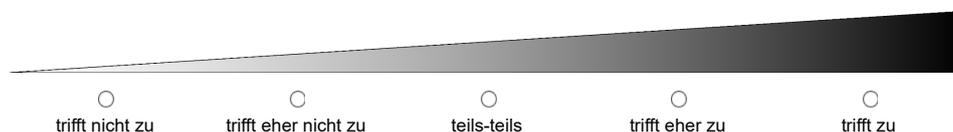




34. Ich würde die Applikation dem Papiertest vorziehen, wenn es mir eine Zeitersparnis bringt



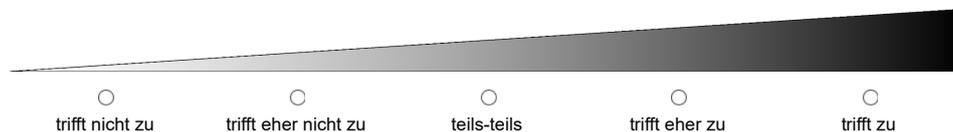
35. Ich würde das Demenz-Screening über eine IGEL Leistung in meiner Praxis anbieten



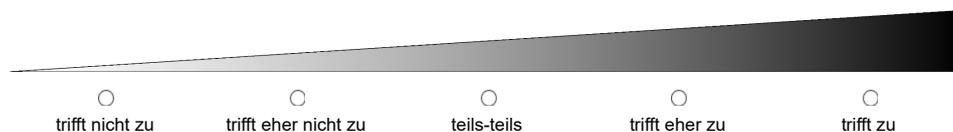
36. Ich würde für eine IGEL Leistung „Demenz-Screening“ über die Applikation folgendes Verlangen

€ (Euro)

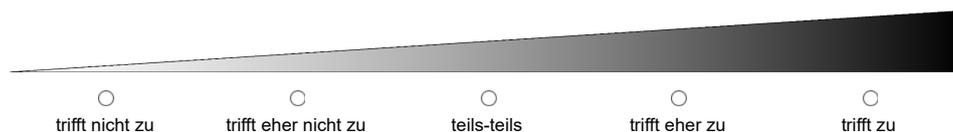
37. Ich würde die Applikation dem Papiertest vorziehen, wenn das Screening von der Krankenkasse vergütet wird



38. Ich würde die Applikation dem Papiertest vorziehen, wenn sie mir die Ergebnisse direkt im Praxisverwaltungssystem zur Verfügung stellt und ich mir damit eine Auswertung von Hand sparen kann



39. Ich würde mehr Demenz-Screenings durchführen wenn die App die Behandlungszeit verkürzt



**40. Ich würde die Patienten ohne Aufsicht den Test im Wartezimmer bearbeiten lassen**



trifft nicht zu    
  trifft eher nicht zu    
  teils-teils    
  trifft eher zu    
  trifft zu

**41. Ich würde die Patienten den Test unter Aufsicht einer Arzthelferin bearbeiten lassen**



trifft nicht zu    
  trifft eher nicht zu    
  teils-teils    
  trifft eher zu    
  trifft zu

**42. Wenn die Applikation als „Medizinprodukt“ zertifiziert ist steigert das die Wahrscheinlichkeit einer Nutzung meinerseits**



trifft nicht zu    
  trifft eher nicht zu    
  teils-teils    
  trifft eher zu    
  trifft zu

**43. Nach dem derzeitigen Wissensstand würde ich die Applikation DemPredict in meiner Praxis für Demenz-Screenings einsetzen**



trifft nicht zu    
  trifft eher nicht zu    
  teils-teils    
  trifft eher zu    
  trifft zu

**44. Ich würde die Applikation DemPredict unter folgenden Voraussetzungen zum Demenz-Screening einsetzen**

Voraussetzungen

**45. Neusten wissenschaftlichen Studien zufolge, lassen sich Demenzen durch eine möglichst frühzeitige Behandlung von Risikofaktoren verhindern oder zumindest deren Verlauf positiv beeinflussen (Weitere Informationen: Livingston et al., Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. doi: 10.1016/S0140-6736(20)30367-6) **

Sobald es eine Möglichkeit gibt mittels digitaler Unterstützungslösungen die Risikofaktoren einer Demenz bei meinen Patienten frühzeitig zu erkennen und zu überwachen, würde ich diese Unterstützung in meiner Praxis anbieten

ja  
 nein

**46. Brauchen Sie zusätzliche Informationen von uns, um die App anzubieten?**

- Ich würde zu einer Vorstellung der Applikation erscheinen, um mich zu informieren
- Ich hätte gerne einen Besuch in meiner Praxis, in dem mir die Applikation und ihr Nutzen erläutert wird
- Ich würde in einem Forum meine Fragen stellen

sonstiges

**47. Was würde eine alternative Abrechnungsmethode für Sie darstellen? (Alternativ zu DiGA oder IGeL)**

Antwort

**48. Welche Dinge sind Ihnen bei einem Einsatz der Applikation besonders wichtig?**

- Eine Sensitivität/ Spezifität zwischen 80% und 95%
- Dass die Nutzung der Applikation mit der Krankenkasse abgerechnet werden kann
- Dass eine Empfehlung der kassenärztlichen Vereinigung vorliegt

Sonstiges

**49. Würden Sie auf ein Testergebnis vertrauen, wenn der Patient den Test im Vorfeld zu Hause durchgeführt hat?**

- ja
- nein
- vielleicht

**50. Bei Auswahl „vielleicht“**

bitte erläutern Sie ihre Auswahl

**51. Wie genau wünschen Sie sich eine Anzeige der Testergebnisse?**

- Ich brauche nur eine Diagnose: Demenz, Mild Cognitive Impairment, Gesund
- Ich möchte eine detailliertere Information in Form eines Scorings
- Ich möchte eine Einzelansicht der Subtests bekommen

Sonstiges

**52. Weitere Anmerkungen**

Sollten Sie weitere Anmerkungen, Kritik oder Bedenken haben, lassen Sie es uns bitte wissen! Nur mit Ihrer Hilfe können wir ein Produkt entwickeln, von dem Sie als Ärzteschaft profitieren können.

Anmerkungen

---

**Letzte Seite**

**Vielen Dank für Ihre Teilnahme!**

Wir möchten uns ganz herzlich für Ihre Mithilfe bedanken.

Falls Sie Rückfragen haben, wenden Sie sich bitte an: Markus Schinle, [schinle@fzi.de](mailto:schinle@fzi.de) oder Mayumi Kaliciak, [kali5768@uni-landau.de](mailto:kali5768@uni-landau.de).

Ihre Antworten wurden gespeichert, Sie können das Browser-Fenster nun schließen.

## A.22 Demografische Daten zur Umfrage bei Leistungserbringern

In Tabelle A.20 werden die demografischen Daten der mittels des in Abschnitt A.21 dargestellten Fragebogens befragten Teilnehmer dargestellt.

Geschlecht	Profession	Alter	Anzahl	Geschlecht	Profession	Alter	Anzahl	Geschlecht	Profession	Alter	Anzahl		
Männlich	Allgemeinmed.	< 30	0	Neutral	Allgemeinmed.	< 30	0	Weiblich	Allgemeinmed.	< 30	0	214	
		30-45	18			30-45	0			30-45	21		
		45-60	51			45-60	0			45-60	71		Summe
		> 60	35			> 60	0			> 60	18		
	Neurologie	< 30	0		Neurologie	< 30	0		29				
		30-45	1			30-45	0			30-45	3		
		45-60	8			45-60	1			45-60	11		
		> 60	4			> 60	0			> 60	1		
	Psychologie/ Psychatrie	< 30	0		Psychologie/ Psychatrie	< 30	0		31				
		30-45	2			30-45	0			30-45	3		
		45-60	6			45-60	0			45-60	10		
		> 60	2			> 60	0			> 60	8		
	Sonstige	< 30	0		Sonstige	< 30	0		27				
		30-45	4			30-45	1			30-45	3		
		45-60	5			45-60	0			45-60	8		
		> 60	2			> 60	0			> 60	4		
	Insgesamt	< 30	0		Insgesamt	< 30	0		301				
		30-45	25			30-45	1			30-45	30		
		45-60	70			45-60	1			45-60	100		
		> 60	43			> 60	0			> 60	31		
<b>Summe</b>		<b>138</b>	<b>Summe</b>		<b>2</b>	<b>Summe</b>		<b>161</b>					

Tabelle A.20: Studienteilnehmerdaten der Umfrage unter Leistungserbringern.

# A.23 Umfrageergebnisse zur Akzeptanz bei Leistungserbringern

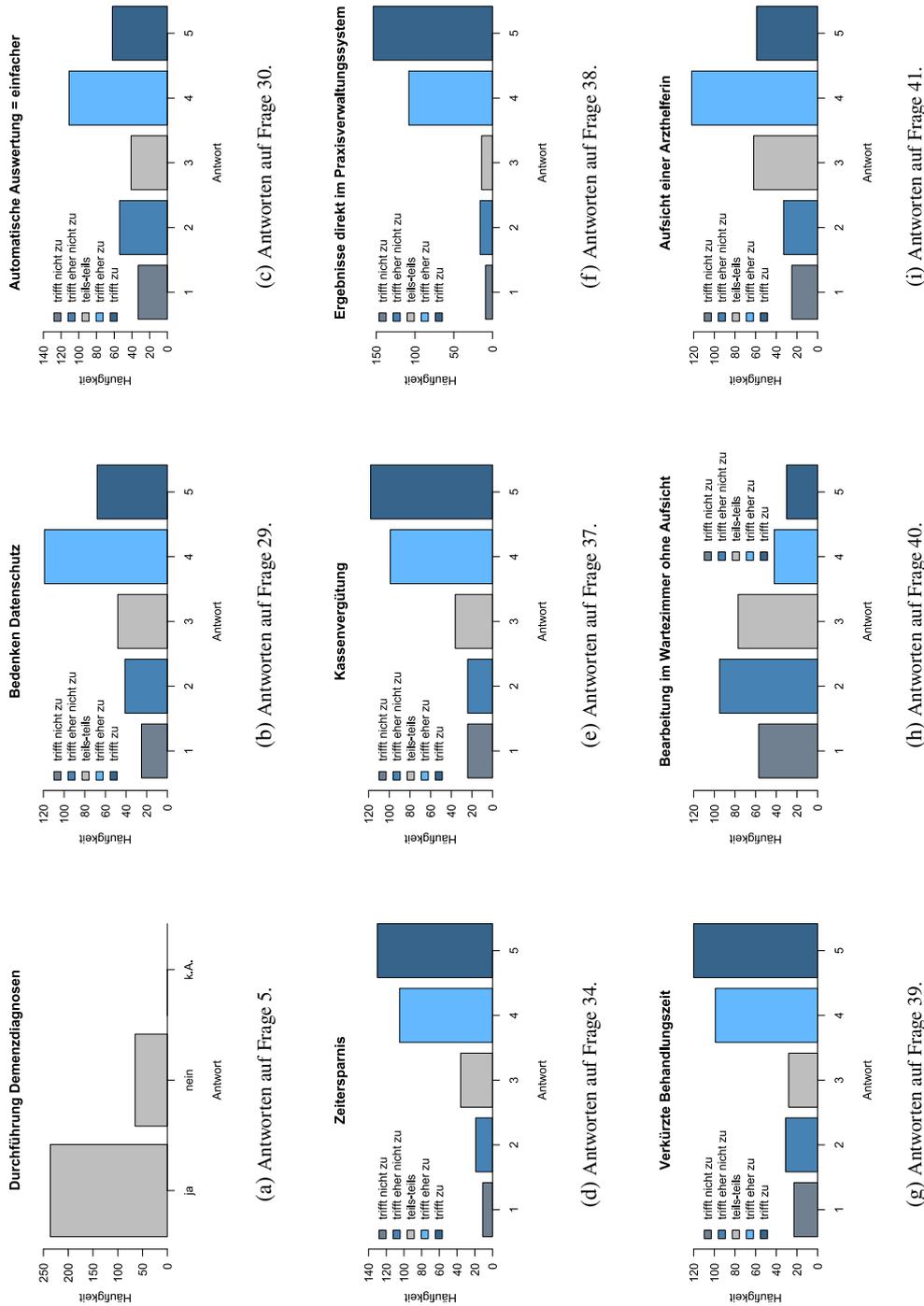


Abbildung A.9: Antworten auf die Fragen zur Beantwortung der übergeordneten Fragen F1-F4 und F5.

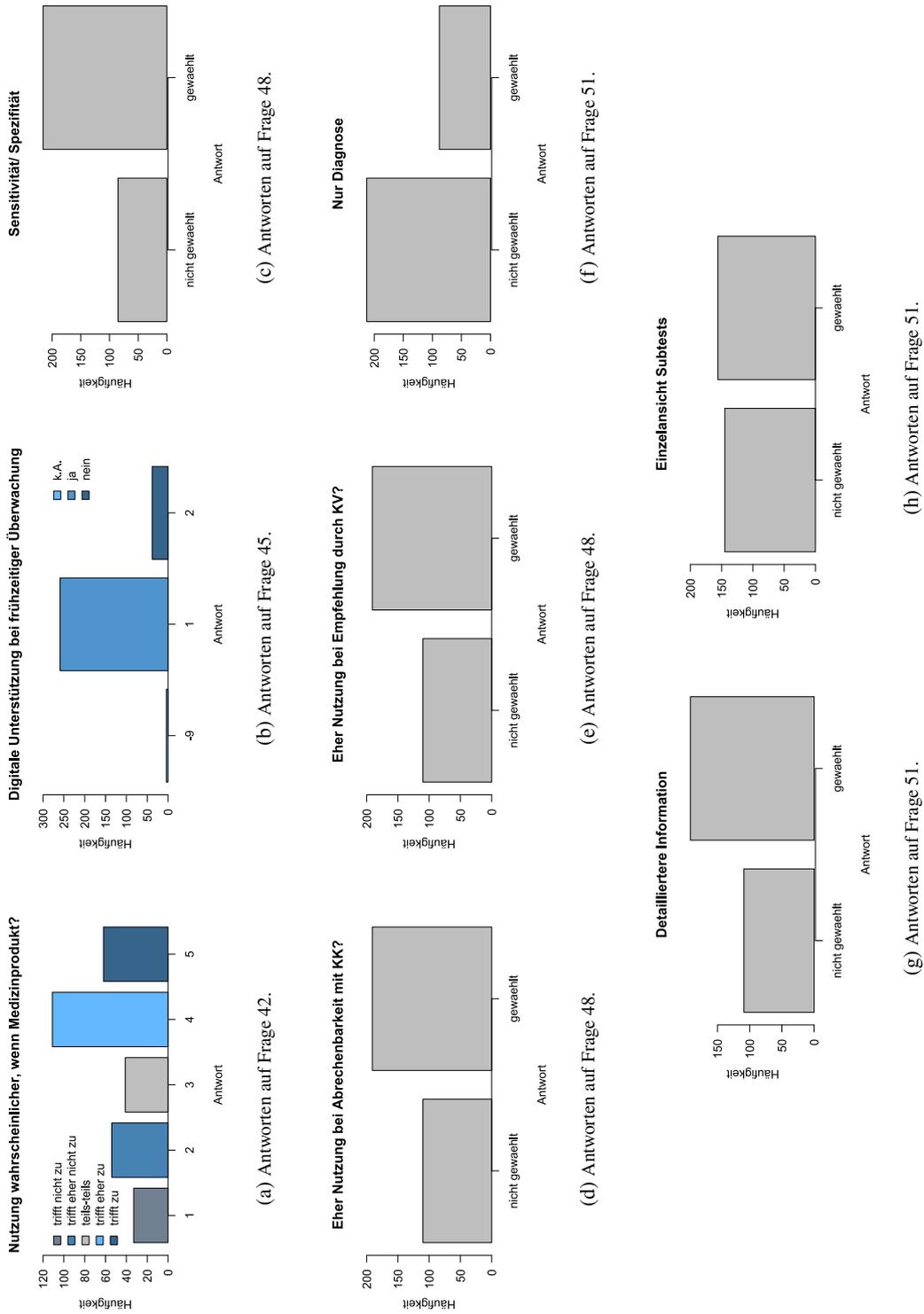


Abbildung A.11: Antworten auf die Fragen zur Beantwortung der übergeordneten Fragen **F2**, **F7** und **F8**.

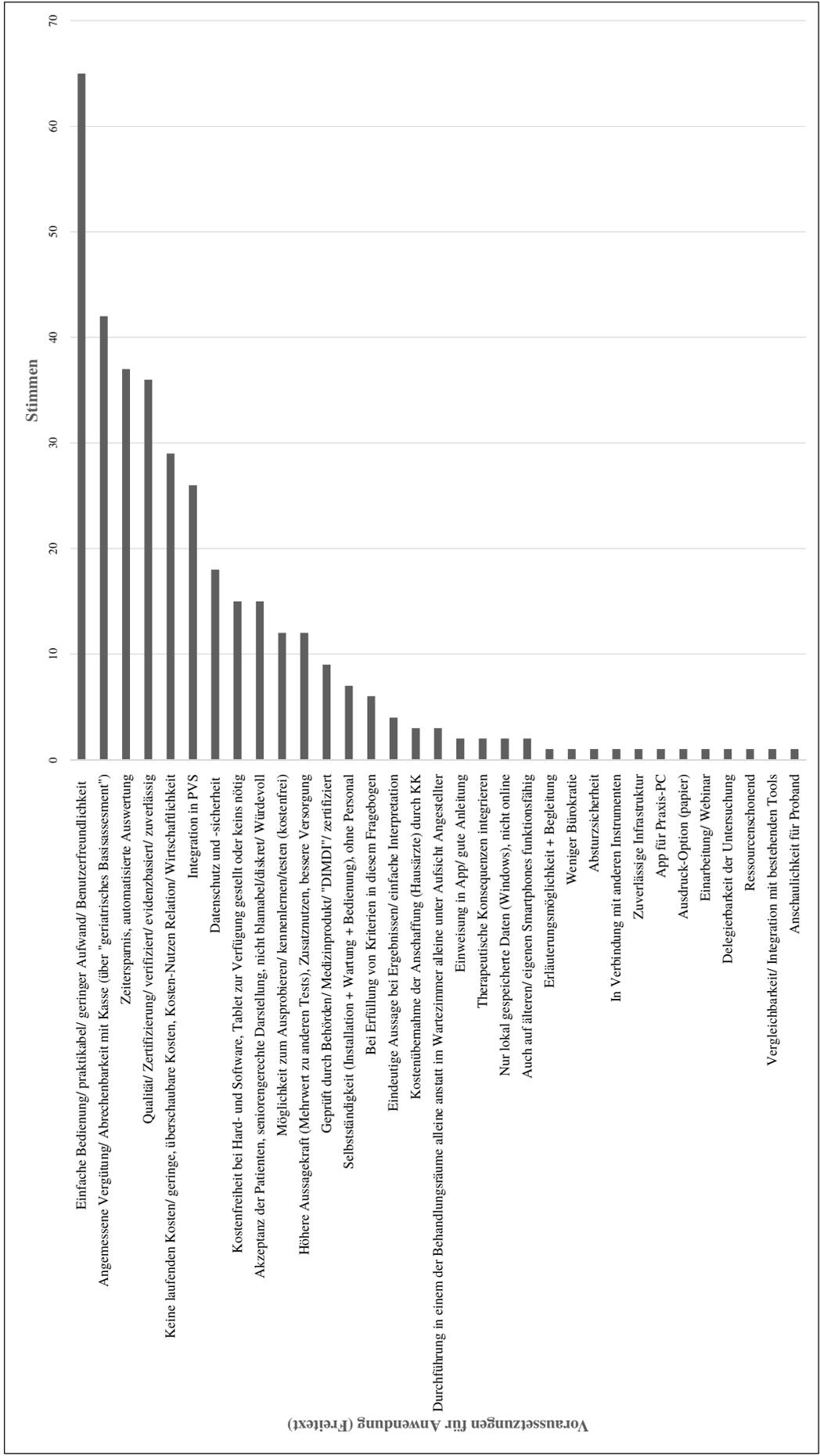


Abbildung A.12: Umfrageergebnisse zu den wichtigsten Akzeptanzkriterien (Frage 44).

Antwort auf die Frage 44	Antwort auf die Frage 44
Qualität,Dauer	ich kenne die App nicht
Wenn DemPredict die Fragen 37-42 (41) erfüllt. s. Ausführungen oben!	einfach anwendbar, Tablett wird zur Verfügung gestellt
Zeit- und Effizienzsteigerung	Geringe Kosten
Wenn es mir nicht mehr Arbeit macht	Kostenübernahme Krankenkasse
sicherer Datenschutz, d.h. keine direkte Vernetzung	Kosten Faktor/Zeitaufwand / unterschriebene Datenschutzerklärung
nicht absehbar	Einfache Anwendung, hinreichend sicheres Ergebnis
Geringer technischer Aufwand, von mir selbst zu installieren und ins Praxissystem einbindbar.	kassenleistung, zeitersparnis
Nutzung ohne laufende Kosten und Notwendigkeit von EDV Support	gute Validierung der Ergebnisse, GKV Leistung
Gute selbsterklärende Bediembbarkeit	Kein zusätzlicher Aufwand
Vergütung durch die Krankenkasse, valide Testergebnisse, Erkenntniszugewinn	Validierung der App
Zertifiziert sollte sie sein	kostenfrei
mal ausprobieren...	Sicherung Datenschutz + kein zusätzlicher Zeitaufwand notwendig
muss praktikabel sein	wenn die Kosten auch für Hausärzte von den Krankenkassen übernommen werden
Wenn die Aussagekraft einen höhere ist als bei den vorhandenen	weniger Bürokratie!!!!!!!!!!!!!!
Verifiziert	Ich kenne die App noch nicht
angemessene Vergütung, Verbesserung der Versorgung	wenn sie ganz einfach zu bedienen ist
eingehende Erläuterungsmöglichkeit für Patienten , Begleitung durch Angehörige	absturzsicher, datenschutzkonform, senioren-gerechte darstellung
Einfache Anwendung	Siege oben
hohe Datensicherheit, ausreichend evidenzbasiert, unkomplizierte Anwendung ohne Notwendigkeit dafür Personal zur Verfügung stellen zu müssen.	Zusätzliche Vergütungen/ Zeitersparnis/ problemlose Einbindung in Unterschiedliche Arztverwaltungssysteme
Positive Studien	Sicherheit
Leichte Durchführbarkeit, läuft auf i pad, Abrechenbar mindestens als IGEL besser noch KV und HZV	einfache handhabung
Wenn es funktioniert	Einweisung in App, einfaches Handling und Auswertung
ich kenne sie nicht	?
gute Handhabbarkeit, schnelle Auswertung und Übertragung ins Praxisinformationssystem	unkomplizierte Anbindung an Praxissoftware, Datenschutzrechtliche Problematik gelöst
	leichte und schnelle Bediembbarkeit. Klare und zuverlässige Ergebnisse. Niedrige Kosten.
	siehe oben
	Wenn pat mitmachen und es verlangen
	leicht ausführbar

<b>Antwort auf die Frage 44</b>
Einbindung ins Verwaltungssystem, leichte Bedienbarkeit
wenn sie sicher ist bzgl. Datenschutz
einfaches Handling, in die Praxissoftware integrierbar
die App soll für Pemenz-Positive im Test leicht verständlich sein und einen hohen prädiktiven Wert haben
Sie muss leicht benutzbar sein.
Vergütung durch die Krankenkasse
kostengünstig, zuverlässig, etabliert
Einfache Bedienung.
Kasse, Implemmentiert in Praxissoftware
nähere Kenntnis DemPredict, Datenlage Nutzen/Zeitaufwand/Vergütung
evaluiert und einfach anwendbar
keine zusätzliche Kosten,
Einfache Handhabung.
s.oben
Einfache Nutzung
Kosten werde von den Kassen übernommen
Patiententablet und Anbindung an praxissoftware müssen vorhanden sein
mehr fällt mir nicht ein
leichte Anwendbarkeit für Patient und für Praxis, überschaubare Kosten
Wenn die Durchführung vergütet wird
nicht als einziges Instrument
Datenschutz muss gewährleistet sein
Keine
Vergütung
erst ausprobieren, dann beurteilen
einfache Handhabung
In meiner Praxis schreckt die Anwendung eines elektronischen Gerätes viele über 65 jährige ab - dies wird sich evtl in 5 Jahren bereits ändern

<b>Antwort auf die Frage 44</b>
Günstiger Preis, gut evaluiert, Akzeptanz der Patienten
kostenlos Bereitstellung der Hard- und Software; angemessene Vergütung; überzeugende Sinnhaftigkeit
Datenschutz und Ausschluss Missbrauch muss gewährleistet sein. Zulassung durch die entsprechenden Behörden.
Kostenneutealität, Test validiert
Gute Infrastruktur
Kosten/Nutzenrelation, v.a. fragliche therapeutische Konsequenzen.
s.o.
wenn ich Hausarzt, oder Neurologe wäre
Abrechenbarkeit für geriatrisches Basisassessment, Übernahme in die praxissoftware, Bedingung allein möglich, allerdings nicht im Wartezimmer, sondern in Ruhe in einem unserer Behandlungsräume
NUR Windows-PC-Version mit NUR lokal gespeicherten Daten bei mir
gegebene zusätzliche Honorierbarkeit, Zeiterparnis
Ich finde sie gut und sie muss günstig sein
Kostenlos zur Verfügung
unkompliziert verfü- und bedienbar, keine EDV-Zusatzkosten
Kosten halten sich begrenzt. Nutzen auch verwertbar.
Kostenlose Bereitstellung der Hard- und Software, Abrechnungsmöglichkeit gegeben
einfache Bedienung, möglichst auch Installierbarkeit auf Praxis-PC
Wenn die Nutzung im Praxisalltag zu integrieren und effizienter ist als die App
Hardwareanforderungen
Kostenfrei
zeitlich sinnvoll machbar, Demenz und Vitaminmangelsyndrom wird zu wenig beachtet

<b>Antwort auf die Frage 44</b>
Vereinfachung, Vergütung
Wird von meinen Patienten gut angenommen, wird von meinen Patienten beherrscht
Zertifizierung, Einbindung in PVS
Einbindung Praxissoftware
einfache technische Umsetzung, ansprechend, nicht abschreckend für Patienten, sozusagen \spielerisch\
Wirklich einfache Bedienbarkeit
einfach zu benutzen auf auch älteren Smartphones
Einfache und nicht sehr zeitaufwändige Durchführbarkeit mit gut nachvollziehbar Ergebnis
Datensicherheit. Bereitstellung Tablet durch die GKV (viele Senioren besitzen kein Smartphone)
Zeitersparnis, automatisierte Auswertung
erst mal ausprobieren
Qualität
App ist mir nicht bekannt. Ich würde es aber testen unter Aufsicht Personal
Wenn sie anwendungsfreundlich ist
Ergebnis hat therapeutische oder praktische Konsequenzen für den Patienten
unkomplizierte Technik, Zeitersparnis, leichte Handhabbarkeit
Einfache Bearbeitung durch die Patienten
habe die Fragen nach Aufforderung d. KVBA-Wü beantwortet, die DemPredict APP habe ich nicht getestet
Wenn Sie durch die Pat. angenommen wird, valide Ergebnisse liefert u. das Ergebnis leicht in die elektronische Akte übernommen werden kann
wenn es eine vernünftige Therapie vs. Demenz gäbe!!!!!!
Einfaches und schnelles handling
kann ich erst beurteilen wenn ich die APP kenne

<b>Antwort auf die Frage 44</b>
wenn der Test eine therapeutische Konsequenz hätte
wenn: zugelassen, vergütet, praktisch
wissenschaftliche Evidenz muss vorliegen
einfache, rasche, verständliche Anwendung
wenn sie mir Daten liefert, wie ich die aktuelle Situation und das Leben der Patientin oder des Patienten im häuslichen Umfeld einschätzen kann
Einfache Handhabung
gut ohne Hilfe durchführbar, valides Ergebnis, einfache Einbindung in PVS
Einfache Durchführung, eindeutige Aussage, schnelles Ergebnis
Geeigneter Patient
Weniger zeit , keine Mitarbeiter erforderlich , bessere Vergütung
Zeitersparnis
ich müsste sie zuerst einmal kennenlernen. Ich glaube ich habe etwas verpasst. Traurig ist nur, dass Pat. mit eindeutiger Demenz keine adäquate Therapie durch den Neurologen erhalten. Ich hoffe Sie befragen auch Neurologen.
Kassenleistung,
Günstig einfach und schnell
Kostenfreie 30 Tage Testversion
Sicherer Umgang Zeitersparnis und Zulassung
Aushändigung digital und in Papierform zum Ausdrucken
Dass sie sich bei Kollegen bewährt hat
anerkanntes Medizinprodukt
einfach, sicher, transparent, nicht blamabel für Pat., reproduzierbare Ergebnisse
Kostengünstig oder finanziert durch Krankenkassen, Einfache Anwendung
Einfache Haltbarkeit, kein Igelleistung
Zugelassen/Validiert/adäquat vergütet
Einfache Handhabung
ausreichend Zeit

<b>Antwort auf die Frage 44</b>
gute Anbindung an Praxis-EDV
valide, reliabel, kostenfrei
Erstattungsfähigkeit, Kostenübernahme durch GKV
Erschwinglichkeit gegeben, am besten über GKV vergütet, Teil des geriatrischen Basisassessments, wirkkl. Arbeitserleichterung, gute Handhabbarkeit
Gesicherter Datenschutz
kostenloses Tablet, Einarbeitung, gute Vergütung
Kosten?
leichte Bedienbarkeit
gute Vergütung, Zeitersparnis, Nutzen
-
preisgünstige Bereitstellung, Datensicherheit, Abrechnungsfähigkeit
Zeitersparnis gegenüber Papiertest
leicht durchführbar und aussagekräftig
s.o.
Einfach zu verstehen und einfach zu handhaben
einfach, Zeitersparnis, diskret, ohne emotionale Belastung für den Patienten
Zulassung durch DIMDI und Bezahlung durch KK
Ich halte dieses Projekt für albsolut falsch, denn es verstößt gegen die Würde des Patienten und die Ergebnisse sind irrelevant.
keine Risiken
Adäquate Honorierung durch KV, Regelmäßige Anwendbarkeit und Zuverlässigkeit der Ergebnisse. Über Bisheriges hinausgehender Nutzen.
zunächst müsste ich nähere informationen erhalten und die sache kennenlernen
wenn er mich überzeugt
keinesfalls online
Anrechenbar bzw. kostenneutral und in pvs (T2Med) angebunden

<b>Antwort auf die Frage 44</b>
Sicherheit, ZUverlässigkeit
kann ich nicht beantworten da ich die App nicht kenne
Leichte und schnelle Bedienung
Datensicherheit
Wenn ältere Patienten mit der Technik zurecht kommen würden
Zeitersparnis, Aussagekräftigkeit,
Gute Handhabbarkeit bei Menschen, die nicht gewohnt sind mit Technik umzugehen
Zeitersparnis, Verbesserung der Behandlung, für mich kostenneutral
einfache Handhabung v.a. für Patient
Zuverlässige Ergebnisse
nicht zu teuer, einfach durchzuführen, Abrechnungsmöglichkeit wäre natürlich toll
Wirtschaftlichkeit, einfache Anwendung
wenn ich dafür keine zusätzlichen Geräte wie ein Tablet anschaffen müsste
die APP kann mit dem Verwaltungssystem der Praxis verbunden werden und wird durch die Kassen vergütet
Zeitersparnis, keine neuen Kosten, Praktikabilität
Technische Möglichkeit fehlt in der Praxis
muss mir erst ein Patientengeeignetes Tablet kaufen
einfache Anwendung, Akzeptanz der Patienten
Einbindung i System, kostenfrei
Kosten für Anschaffung Tablet/ Smartphone werden von KV/ Krankenkasse übernommen.
freiwillige TEilnahme von Klienten ab einem bestimmten Alter
Gesicherte Zeitersparnis, keine Zusatzkosten und -anschaffungen
kostenlos verfügbar
Tablet Apps werden von Demenzpatienten eher verhalten akzeptiert
wenn eine Erstattung durch die Krankenkasse möglich wäre.

**Antwort auf die Frage 44**

vernünftige Kosten-Nutzen-Relation (Abrechenbarkeit), Verknüpfbarkeit Praxis-EDV

Delegierbarkeit der Untersuchung und nahtlose Einbindung in die Dokumentation der Praxissoftware

Kassenleistung

positive Zulassungsstudie

wissenschaftl fundiert, in Praxissystem integrierbar, geringe/keine Kosten

Kostengünstig, einfaches Handling

Kostenneutral

muß mehr über Validierung wissen

so oder so wenn ich nicht nur Zeit für den Patienten hätte sondern auch eine Verdacht

kostenfreier Bezug

gute Anleitung

Zulassung und Vergütung durch KV

einfache Handhabung; Vergütung

Wenn meine Patienten und deren Angehörigen dies wünschen

leichte Anwendbarkeit

Zuverlässiger prädiktiver Wert, einfache, kurze Handhabung

die nachfolgende Studie stimmt in Ihrer Logik nicht, für Risikofaktoren brauche ich keine App!!!

leichte Handhabung, geringe Kosten, über Krankenkasse anrechenbar

kann ich momentan noch nicht beurteilen

Datenschutz muss erfüllt sein, einfach zu handhaben, schnelle Auswertung

Zeitersparnis

Wenn mir die Anwendung genau gezeigt wird und ich sie dann besser beurteilen kann.

niederschwellige Verfügbarkeit, Kompatibilität mit Praxissoftware, leichte Bedienbarkeit, Auswertung ins Praxissystem

kein zusätzlicher Zeitaufwand durch Installation und Erklärung bei jedem Patienten

**Antwort auf die Frage 44**

Testung primär durch mich auf Praktikabilität

wenn valide, aber wenig zeit- und ressourcen-aufwändig

günstig für den Patienten

mein Klientel ist sicher zu Jung bis 18

kein Abo mit laufenden Kosten

Abrechnung über KV

Wenn sie zertifiziert und abrechnungsfähig ist und ich sowieso ein Tablet dafür hätte.

Zeitersparnis, Erfahrbarkeit der Ergebnisse

Wenn Validität und Reliabilität hoch sind.

Datensicherheit sollte gewährleistet sein, es sollte zu Zeitersparnis führen und die anschließende Versorgung der Patienten gesichert sein

geringer zeitlicher Aufwand

Benutzerfreundlichkeit gegeben

kostenlose oder mindestens kostengünstige/-kostendeckende Nutzung

wenn sie wirklich einfach bei älteren Menschen, auch mit sprachlichen Problemen, anzuwenden ist

ich muss es selbst getestet haben

geht nur unter Aufsicht einer MfA o.Ä.

Abrechnungsfähig über KV und wenn >Verdachtsfall auf Demenz in meine Praxis käme

praktikabel

Anwendung nur in ruhigem Sprechzimmer mit geschulter Assistentin, keinesfalls alleine im Wartezimmer und ohne geschulte Unterstützung

gute Validität

kostenneutral, anwenderfreundlich, in Praxis-EDV ohne großen Aufwand integrierbar

Wenn ich mich vollumfänglich über das Produkt informiert hätte, die Infos aus dieser Umfrage reiche mir für eine Entscheidung nicht aus.

---

**Antwort auf die Frage 44**


---

geprüfte Datenlage

---

ich lerne sie kennen und sie erfüllt die o.g. Erwartungen (Qualität und Zeit)

---

Nur Vielleicht: ... Wenn sie integraler Bestandteil des Patientenverwaltungssystems wäre und die höchsten Sicherheitszertifikate vorlägen,

---

Vergütung über die kv

---

Muss einen organisatorischen Nutzen haben

---

wenn der Patient dies selbstständig auf sein Smartphone oder Smartphone der Angehörigen bearbeiten kann

---

Einfache Bedienbarkeit, Ressourcenschonende Hardware-Voraussetzung, einfache Integration ins System, angemessene Kosten, Integration in Leitlinien/Vergleichbarkeit mit bestehenden Tools

---

Zulassung als Medizinprodukt + Zeitersparnis in meinem Praxisablauf

---

gutes Handling (auch Anschaulichkeit für Proband), Auswertbarkeit, Sicherheit, Zeitersparnis, adäquate Abrechenbarkeit

---

einfache Handhabung, eindeutig vergleichbare Screening-Ergebnisse zu anderen bestehenden Tools

---

wenn die Interpretation so einfach wie beim MMT / Uhrentest wäre

---

kein erhöhter Zeitaufwand

---

Praktikabilität, Sensitivität

---

ich kenne diesen Test nicht

---

Tabelle A.21: Freitext-Antworten auf die Frage 44.

Antwort auf die Frage 48.	Antwort auf die Frage 48.
Dass die o.g. Kriterien erfüllt und den Pat. nicht überfordert	Zeitersparnis
Technischer Aufwand	einfache Handhabung
Nutzen und Praktikabilität muss gewährleistet sein	wissenschaftliche unabhängige Evaluation
Verwendung ohne Aufsicht	Empfehlung der DEGAM
auf iPad Durchführbar	dass die Anwendung einfach und ohne Stress für den Probanden erfolgt
einfache Bedienbarkeit und Deligierbarkeit	Dass es \"Ärzttemuster\"
Einfache Bedienbarkeit	für Patienten akzeptabel und gut zu benutzen
Dass sie funktioniert und weiterhilft	Praktikabilität
Erstmal ausprobieren	Einfache Installation, gute Anwendbarkeit, hohe Praktikabilität
handelbar sein	Alltagstauglichkeit
einfache, unkomplizierte Anwendung	Datensicherheit
Zeit / Nutzen Verhältnis	dass die Patienten es akzeptieren
Das sie so einfach ist, dass auch ein völlig technikfremder sie bedienen kann	das sie durchführbar ist
dass ich von Sinn und Zweck überzeugt bin	das es einen nutzen gibt und keine Datenmengen unnütz gespeichert werden für werbe zwecke
eine reine Windows-App mit ausschließlicher Speicherung der Daten bei mir	einfache Handhabbarkeit, techn. Zuverlässigkeit
Sonst Igel	Spezifität mindestens 90%
Leichtigkeit der Bedienung	nur unter Aufsicht
adäquate Vergütung, analog zu Betriebswirtschaftlern, Juristen und IT ler	praktikabel im Alltag einsetzbar, auch von Seiten der Patienten
wird von den Patienten beherrscht	Rahmenbedingungen in der Praxis müssen gegeben sein
Akzeptanz/Interesse von Seiten der Patienten	Praktikabilität
einfache Nutzbarkeit, nicht verführend zu unnützer Medikation	Einfache und schnelle Nutzbarkeit, Integration in PVS, gute Evaluation/Evidenz, Vergleichbarkeit mit anderen Tests/ Nutzbarkeit zur Diagnose
wissenschaftlich evaluiert	
Dass sinnvolle Hilfe möglich ist	
sollte was taugen und Ausreißer im Testergebnis (durch Unsicherheit oder Depression) erkennen	
nichts	
Sehr hohe Sensitivität mind. 95%	
einfache Handhabung	
Hohe Sensivität	
Einfache Handhabung vor allem für die älteren Patienten und Verlässlichkeit der Ergebnisse	

Tabelle A.22: Freitext-Antworten auf die Frage 48.

# A.24 Fragebogen zur ISO NORM 9241-110

Fragebogen ISONORM 9241/110-S

Seite 2 | 6

Im Folgenden geht es um die Beurteilung von Softwaresystemen auf Grundlage der Internationalen Norm DIN EN ISO 9241-110.

**Bitte beachten Sie:**

- Das Ziel dieser Beurteilung ist es, Schwachstellen bei Softwaresystemen aufzudecken und konkrete Verbesserungsvorschläge zu entwickeln.
- Um dies zu bewerkstelligen, ist Ihr Urteil als Kenner des Softwaresystems von entscheidender Bedeutung! Grundlage Ihrer Bewertung sind Ihre individuellen Erfahrungen mit dem Software-Programm, das Sie beurteilen möchten.
- Dabei geht es nicht um eine Beurteilung Ihrer Person, sondern um Ihre persönliche Bewertung der Software mit der Sie arbeiten.

Bitte machen Sie im folgenden Kasten zunächst einige Angaben zu der Software, auf die sich Ihre Beurteilung im Folgenden beziehen wird.

Auf welches Software-Programm bezieht sich Ihre Beurteilung? (Beurteilen Sie bitte lediglich ein Software-Programm!)							
Name der Software							
Versionsnummer							
gegebenenfalls Teilanwendung / Modul							
Wie gut beherrschen Sie die von Ihnen beurteilte Software? (- - - = sehr schlecht; +++ = sehr gut)	---	--	-	- / +	+	++	+++
	<input type="radio"/>						

Ein Fragebogen von [www.seikumu.de](http://www.seikumu.de)



**Noch ein Hinweis zur Beantwortung des Beurteilungsbogens:**

Im folgenden Fragebogen werden die Anforderungen der Norm über Beschreibungen konkretisiert. Diese Beschreibungen weisen immer folgende Form auf:

**Beispiel 1:**

Die Software ...	---	--	-	- / +	+	++	+++	Die Software ...
ist schlecht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	ist gut.

Im ersten Beispiel wird danach gefragt, wie gut bzw. wie schlecht die Software ist. Die Benutzerin oder der Benutzer beurteilt in diesem Fall die Software zwar als gut, sieht jedoch noch Verbesserungsmöglichkeiten.

- Am besten bearbeiten Sie den Beurteilungsbogen, während Sie das zu bewertende Softwaresystem vor sich am Bildschirm haben. Dadurch haben Sie die Möglichkeit, bei der Beantwortung der einzelnen Fragen die eine oder andere Sache noch einmal zu überprüfen.
- Füllen Sie bitte den Beurteilungsbogen äußerst sorgfältig aus und lassen Sie keine der Fragen aus!

	Die Software ...	--- -- - -/+ + ++ +++	Die Software ...
aa1	bietet nicht alle Funktionen, um die anfallenden Aufgaben effizient zu bewältigen.	<input type="radio"/>	bietet alle Funktionen, um die anfallenden Aufgaben effizient zu bewältigen.
aa2	erfordert überflüssige Eingaben.	<input type="radio"/>	erfordert keine überflüssigen Eingaben.
aa3	ist schlecht auf die Anforderungen der Arbeit zugeschnitten.	<input type="radio"/>	ist gut auf die Anforderungen der Arbeit zugeschnitten.
sb1	liefert in unzureichendem Maße Informationen darüber, welche Eingaben zulässig oder nötig sind.	<input type="radio"/>	liefert in zureichendem Maße Informationen darüber, welche Eingaben zulässig oder nötig sind.
sb2	bietet auf Verlangen keine situationsspezifischen Erklärungen, die konkret weiterhelfen.	<input type="radio"/>	bietet auf Verlangen situationsspezifische Erklärungen, die konkret weiterhelfen.
sb3	bietet von sich aus keine situationsspezifischen Erklärungen, die konkret weiterhelfen.	<input type="radio"/>	bietet von sich aus situationsspezifische Erklärungen, die konkret weiterhelfen.
ek1	erschwert die Orientierung durch eine uneinheitliche Gestaltung.	<input type="radio"/>	erleichtert die Orientierung durch eine einheitliche Gestaltung.

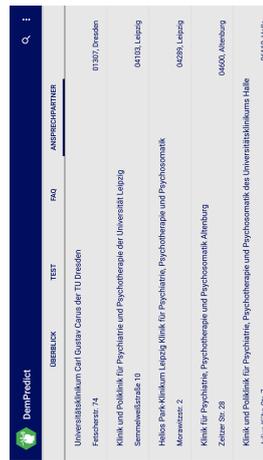
	Die Software ...	--- -- - -/+ + ++ +++	Die Software ...
ek2	informiert in unzureichendem Maße über das, was es gerade macht.	<input type="radio"/>	informiert in ausreichendem Maße über das, was es gerade macht.
ek3	lässt sich nicht durchgehend nach einem einheitlichen Prinzip bedienen.	<input type="radio"/>	lässt sich durchgehend nach einem einheitlichen Prinzip bedienen.
lf1	erfordert viel Zeit zum Erlernen.	<input type="radio"/>	erfordert wenig Zeit zum Erlernen.
lf2	erfordert, dass man sich viele Details merken muss.	<input type="radio"/>	erfordert nicht, dass man sich viele Details merken muss.
lf3	ist schlecht ohne fremde Hilfe oder Handbuch erlernbar.	<input type="radio"/>	ist gut ohne fremde Hilfe oder Handbuch erlernbar.
sk1	erzwingt eine unnötig starre Einhaltung von Bearbeitungsschritten.	<input type="radio"/>	erzwingt keine unnötig starre Einhaltung von Bearbeitungsschritten.
sk2	ermöglicht keinen leichten Wechsel zwischen einzelnen Menüs oder Masken.	<input type="radio"/>	ermöglicht einen leichten Wechsel zwischen einzelnen Menüs oder Masken.

	Die Software ...	--- -- - -/+ + ++ +++	Die Software ...
sk3	erzwingt unnötige Unterbrechungen der Arbeit.	<input type="radio"/>	erzwingt keine unnötigen Unterbrechungen der Arbeit.
ft1	liefert schlecht verständliche Fehlermeldungen.	<input type="radio"/>	liefert gut verständliche Fehlermeldungen.
ft2	erfordert bei Fehlern im Großen und Ganzen einen hohen Korrekturaufwand.	<input type="radio"/>	erfordert bei Fehlern im Großen und Ganzen einen geringen Korrekturaufwand.
ft3	gibt keine konkreten Hinweise zur Fehlerbehebung.	<input type="radio"/>	gibt konkrete Hinweise zur Fehlerbehebung.
lk1	lässt sich schwer erweitern, wenn für mich neue Aufgaben entstehen.	<input type="radio"/>	lässt sich leicht erweitern, wenn für mich neue Aufgaben entstehen.
lk2	lässt sich schlecht an meine persönliche, individuelle Art der Arbeitserledigung anpassen.	<input type="radio"/>	lässt sich gut an meine persönliche, individuelle Art der Arbeitserledigung anpassen.
lk3	lässt sich - im Rahmen ihres Leistungsumfangs - von mir schlecht für unterschiedliche Aufgaben passend einrichten.	<input type="radio"/>	lässt sich - im Rahmen ihres Leistungsumfangs - von mir gut für unterschiedliche Aufgaben passend einrichten.

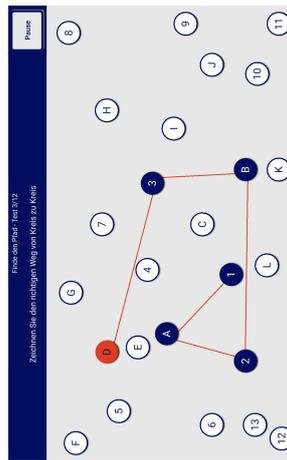
# A.26 DemPredict nach Iteration IV



(a) Startbildschirm.



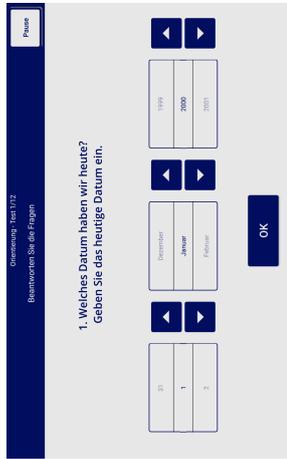
(d) Ansprechpartner.



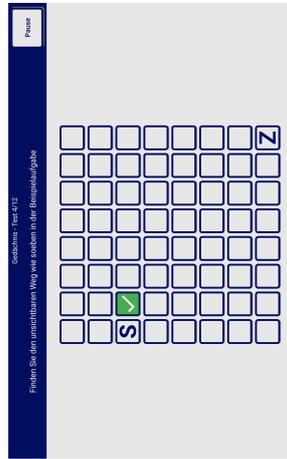
(g) Test D.



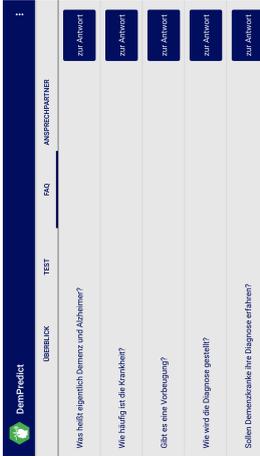
(b) Testung.



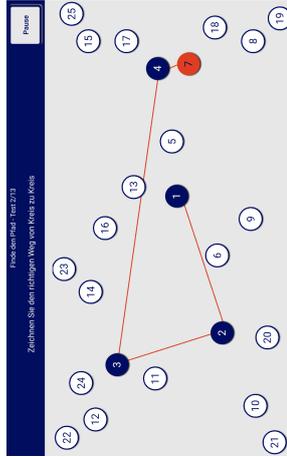
(e) Test A.



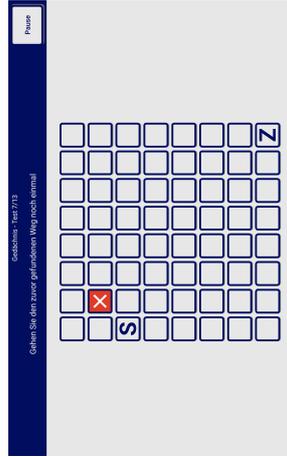
(h) Test E.



(c) FAQ.



(f) Test C.



(i) Test F.

Navigation von Würfen - Test 6/13  
Zeichnen Sie den Würfel so genau wie möglich in der linken Fläche ab.

Löschen Bestätigen

(a) Test G.

Rechnen - Test 5/13  
Ziehen Sie kontinuierlich die Zahl 7 ab. Sie beginnen bei 100.

1. Ergebnis  
0

0 1 2 3 4 5 6  
Eingabe korrigieren 7 8 9 Eingabe bestätigen

(b) Test H.

Zahlenfolge raten - Test 1/13  
Tippen Sie die gezeigte Zahlenfolge in umgekehrter Reihenfolge ein.

Ende der Zahlenfolge!  
38

0 1 2 3 4 5 6  
Eingabe korrigieren 7 8 9 Eingabe bestätigen

(c) Test I.

Aufmerksamkeit - Test 9/13  
Schaltfläche drücken wenn das rote Dreieck erscheint.

ist Dreieck

(d) Test J.

Aufmerksamkeit - Test 10/13  
Schaltfläche drücken wenn das rote Tier im Tier erscheint.

Garnrolle

ist Tier

(e) Test K.

Umsatz von Zahlen - Test 12/13  
Geben Sie die angezeigte Zahl als Wert ein.

1478

0 W E R T Z U J I O P  
A S D F G H J K L O A  
Eingabe korrigieren Y X C V B N M Eingabe bestätigen

(f) Test L.

Navigation von Zahlenreihen - Test 1/13  
Geben Sie die angezeigte Zahl ein.

SECHSZEHTAUSENDVIERHUNDERTDREIUNDSEBZIG

0 1 2 3 4 5 6  
Eingabe korrigieren 7 8 9 Eingabe bestätigen

(g) Test M.

Körperbau - Test 13/13  
Berühren Sie die angegebenen Körperteile.

den rechten Ellenbogen

Eingabe bestätigen

(h) Test N.

Erreichte Punktzahl: 85

277

Gesamtpunktzahl

Fortfahren

(i) Ergebniszwischenstand.



# Literaturverzeichnis

- [17] *S3-Leitlinie Demenzen*. Interdisziplinäre S3-Praxisleitlinien. Springer Berlin Heidelberg, Berlin, Heidelberg, 2017. ISBN: 9783662538746. DOI: 10.1007/978-3-662-53875-3.
- [18] A. Krithara et al. iASiS: Towards Heterogeneous Big Data Analysis for Personalized Medicine. In *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, pages 106–111, 2019. DOI: 10.1109/CBMS.2019.00032.
- [19] A. Suinesiaputra, P. Medrano-Gracia, B. R. Cowan, und A. A. Young. Big Heart Data: Advancing Health Informatics Through Data Sharing in Cardiovascular Imaging. *IEEE Journal of Biomedical and Health Informatics*, 19(4):1283–1290, 2015. DOI: 10.1109/JBHI.2014.2370952.
- [20] Clark C. Abt. *Serious games*. Viking Press, New York, 5. print edition, 1975. ISBN: 0670634905.
- [21] Chantal Aebi. *Validierung der neuropsychologischen Testbatterie CERAD-NP: Eine Multi-Center Studie*. Dissertation, Universität Basel, Basel, 2002. URL: [https://www.memoryclinic.ch/fileadmin/user\\_upload/Memory\\_Clinic/Literatur/2002/Aebi\\_2002.pdf](https://www.memoryclinic.ch/fileadmin/user_upload/Memory_Clinic/Literatur/2002/Aebi_2002.pdf) [zuletzt abgerufen am 10.12.2021].
- [22] Dana AL-Dlaeen und Abdallah Alashqur. Using decision tree classification to assist in the prediction of Alzheimer’s disease. In *2014 6th International Conference on Computer Science and Information Technology (CSIT)*, pages 122–126. ISBN: 978-1-4799-3999-2. IEEE, 3/26/2014 - 3/27/2014. DOI: 10.1109/CSIT.2014.6805989.
- [23] Muna S. Al-Razgan, Hend S. Al-Khalifa, Mona D. Al-Shahrani, und Hessah H. AlAjmi. Touch-Based Mobile Phone Interface Guidelines and Design Recommendations for Elderly People: A Survey of the Literature. In David Hutchison, Takeo Kanade, Josef Kittler, et al., editors, *Neural Information Processing*, volume 7666 of *Lecture notes in computer science*, pages 568–574. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. ISBN: 978-3-642-34477-0. DOI: 10.1007/978-3-642-34478-7 $\backslash$ backslash\$textunderscore.
- [24] Urs-Vito Albrecht. Chancen und Risiken von Gesundheits-Apps. In Marion Albers und Ioannis Katsivelas, editors, *Recht & Netz*, pages 417–430. Nomos Verlagsgesellschaft mbH & Co. KG, 2018. ISBN: 9783845293288. DOI: 10.5771/9783845293288-417.
- [25] Urs-Vito Albrecht, Volker E. Amelung, Ines Aumann, et al. Chancen und Risiken von Gesundheits-Apps (CHARISMHA). 2016. URL: <http://www.digibib.tu-bs.de/?docid=00060000> [zuletzt abgerufen am 10.12.2021], DOI: 10.24355/dbbs.084-201210110913-53.
- [26] Alexandra zu Bentheim. Demenz, 2011. URL: <https://www.bundestag.de/resource/blob/192028/e1a6df8f6c0122d29742ff259adb3e5b/demenz-data.pdf> [zuletzt abgerufen am 10.12.2021].

- [27] Muhammad Salek Ali, Koustabh Dolui, und Fabio Antonelli. IoT Data Privacy via Blockchains and IPFS. In *Proceedings of the Seventh International Conference on the Internet of Things*, IoT '17, pages 14:1–14:7. ACM, 2017.
- [28] Novruz Allahverdi, Serhat Torun, und Ismail Saritas. Design of a fuzzy expert system for determination of coronary heart disease risk. In Plamen Vatchkov, Angel Smrikarov, und Bozhan Zhechev, editors, *Proceedings of the 2007 international conference on Computer systems and technologies - CompSysTech '07*, page 1. ISBN: 9789549641509, New York, New York, USA, 2007. ACM Press. DOI:10.1145/1330598.1330638.
- [29] David Alvarez-Melis und Tommi S. Jaakkola. Towards Robust Interpretability with Self-Explaining Neural Networks. URL: <https://arxiv.org/pdf/1806.07538>.
- [30] A. Alzheimer. Über eigenartige Krankheitsfälle des späteren Alters. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, 4(1):356–385, 1911. DOI:10.1007/BF02866241.
- [31] Alzheimer Forschung Initiative e.V. Amyloide Plaques und Fibrillen, 2021. URL: <https://www.alzheimer-forschung.de/alzheimer/wasistalzheimer/veraenderungen-im-gehirn/amyloide-plaques-und-fibrillen/> [zuletzt abgerufen am 10.12.2021].
- [32] Alzheimer Forschung Initiative e.V. Diagnose-Verfahren, 29.04.2021. URL: <https://www.alzheimer-forschung.de/alzheimer/diagnose>.
- [33] Alzheimer's Association. 2017 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 13(4):325–373, 2017. DOI:10.1016/j.jalz.2017.02.001.
- [34] An-Wen Chan et al. Increasing value and reducing waste: addressing inaccessible research. *The Lancet*, 383(9913):257–266, 2014. URL: <http://www.sciencedirect.com/science/article/pii/S0140673613622965>, DOI:10.1016/S0140-6736(13)62296-5.
- [35] Ana Azevedo und M. Santos. KDD, SEMMA and CRISP-DM: a parallel overview. *undefined*, 2008. URL: <https://www.semanticscholar.org/paper/KDD%2C-SEMMA-and-CRISP-DM%3A-a-parallel-overview-Azevedo-Santos/6bc30ac3f23d43ffc2254b0be24ec4217cf8c845>.
- [36] Ting F. A. Ang, Ning An, Huitong Ding, et al. Using data science to diagnose and characterize heterogeneity of Alzheimer's disease. *Alzheimer's & dementia (New York, N. Y.)*, 5:264–271, 2019. DOI:10.1016/j.trci.2019.05.002.
- [37] Anobel Y. Odisho und John L. Gore. Patient-centered approaches to creating understandable health information. *Urologic Oncology: Seminars and Original Investigations*, 35(9):559–563, 2017. URL: <http://www.sciencedirect.com/science/article/pii/S1078143917303393>, DOI:10.1016/j.urolonc.2017.06.055.
- [38] Manon Ansart, Stéphane Epelbaum, Giulia Bassignana, et al. Predicting the Progression of Mild Cognitive Impairment Using Machine Learning: A Systematic, Quantitative and Critical Review. *Medical Image Analysis*, page 101848, 2020. DOI:10.1016/j.media.2020.101848.
- [39] Australian Government. My Health Record, 2020. URL: <https://www.myhealthrecord.gov.au/> [zuletzt abgerufen am 04.02.2022].

- [40] Jens Baas und Volker E. Amelung. *Zukunft der Gesundheit: Vernetzt, digital, menschlich*. 2019. ISBN: 978-3954664672.
- [41] Klaus Backhaus, Bernd Erichson, Wulff Plinke, und Rolf Weiber. *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Springer Gabler, Berlin, 15., vollständig überarbeitete auflage edition, 2018. ISBN: 9783662566558. DOI: 10.1007/978-3-662-56655-8.
- [42] Alan D. Baddeley und Robert H. Logie. Working Memory: The Multiple-Component Model. In Akira Miyake und Priti Shah, editors, *Models of working memory: Mechanisms of active maintenance and executive control*, pages 28–61. Cambridge Univ. Press, Cambridge, 2003. ISBN: 9780521587211. DOI: 10.1017/CB09781139174909.005.
- [43] Sebastian R. Bader und Maria Maleshkova. Towards Enforceable Usage Policies for Industry 4.0. In *LASCAR@ ESWC*, pages 75–84, 2019. URL: [ceur-ws.org/Vol-2489/paper8.pdf](http://ceur-ws.org/Vol-2489/paper8.pdf).
- [44] Jan Balata, Zdenek Mikovec, und Tomas Slavicek. KoalaPhone: touchscreen mobile phone UI for active seniors. *Journal on Multimodal User Interfaces*, 9(4):263–273, 2015. DOI: 10.1007/s12193-015-0188-1.
- [45] Amie J. Barda, Christopher M. Horvat, und Harry Hochheiser. A qualitative research framework for the design of user-centered displays of explanations for machine learning model predictions in healthcare. *BMC Medical Informatics & Decision Making*, 20(1):257, 2020. DOI: 10.1186/s12911-020-01276-x.
- [46] Roger G. Barker. The Stepping-Stone Maze: A Directly Visible Space-Problem Apparatus. *The Journal of General Psychology*, 5(2):280–285, 1931. DOI: 10.1080/00221309.1931.9918402.
- [47] Aurelie Bayle, Mirko Koscina, David Manset, und Octavio Perez-Kempner. When Blockchain Meets the Right to Be Forgotten: Technology versus Law in the Healthcare Industry. pages 788–792, 2018. DOI: 10.1109/WI.2018.00133.
- [48] Thomas G. Beach, Sarah E. Monsell, Leslie E. Phillips, und Walter Kukull. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. *Journal of neuropathology and experimental neurology*, 71(4):266–273, 2012. DOI: 10.1097/NEN.0b013e31824b211b.
- [49] Beatriz Esteves und Víctor Rodríguez-Doncel. Analysis of Ontologies and Policy Languages to Represent Information Flows in GDPR. *Semantic Web – Interoperability, Usability, Applicability*, 2021. URL: [www.semantic-web-journal.net/content/analysis-ontologies-and-policy-languages-represent-information-flows-gdpr-0](http://www.semantic-web-journal.net/content/analysis-ontologies-and-policy-languages-represent-information-flows-gdpr-0) [zuletzt abgerufen am 10.12.2021].
- [50] Sylvie Belleville, Isabelle Peretz, und Dominique Malenfant. Examination of the working memory components in normal aging and in dementia of the Alzheimer type. *Neuropsychologia*, 34(3):195–207, 1996. DOI: 10.1016/0028-3932(95)00097-6.
- [51] Klaus Bengler und Martin Schmauder. Digitalisierung. *Zeitschrift für Arbeitswissenschaft*, 70(2):75–76, 2016. DOI: 10.1007/s41449-016-0021-z.
- [52] Cara L. Berkowitz, Lisa Mosconi, Olivia Scheyer, et al. Precision Medicine for Alzheimer’s Disease Prevention. *Healthcare (Basel, Switzerland)*, 6(3), 2018. DOI: 10.3390/healthcare6030082.

- [53] Adam Bermingham, Julia O'Rourke, Cathal Gurrin, et al. Automatically recommending multimedia content for use in group reminiscence therap. In Alexia Briasouli, Jenny Benois-Pineau, und Alexander Hauptmann, editors, *Proceedings of the 1st ACM international workshop on Multimedia indexing and information retrieval for healthcare - MIIRH '13*, pages 49–58. ISBN: 9781450323987, New York, New York, USA, 2013. ACM Press. DOI:10.1145/2505323.2505333.
- [54] Tim Berners-Lee. An RDF language for the Semantic Web: Notation 3 Logic, 2005. URL: <https://www.w3.org/DesignIssues/Notation3> [zuletzt abgerufen am 10.12.2021].
- [55] Tim Berners-Lee. Linked Data - Design Issues, 2006. URL: <https://www.w3.org/DesignIssues/LinkedData.html> [zuletzt abgerufen am 10.12.2021].
- [56] Tim Berners-Lee, James Hendler, und Ora Lassila. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, 2001. URL: [https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American\\_%20Feature%20Article\\_%20The%20Semantic%20Web\\_%20May%202001.pdf](https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf) [zuletzt abgerufen am 10.12.2021].
- [57] Daniel M. Bernstein, Wendy Loken Thornton, und Jessica A. Sommerville. Theory of mind through the ages: older and middle-aged adults exhibit more errors than do younger adults on a continuous false belief task. *Experimental aging research*, 37(5):481–502, 2011. DOI:10.1080/0361073X.2011.619466.
- [58] Jinbo Bi und Arun Abraham. An intelligent web-based decision support tool for enhancing asthma guideline adherence. In *Proceedings of the 2nd ACM SIGHT International Health Informatics Symposium, IHI '12*, page 51. ISBN: 9781450307819, New York, NY, USA, 2012. Association for Computing Machinery. DOI:10.1145/2110363.2110373.
- [59] Przemyslaw Biecek und Marcin Kosinski. DALEX: Explainers for Complex Predictive Models in R. *Journal of Statistical Software*, 82(11), 2018. URL: <https://www.jmlr.org/papers/volume19/18-416/18-416.pdf> [zuletzt abgerufen am 24.01.2022], DOI:10.18637/jss.v082.i11.
- [60] Christophe Bintener, Owen Miller, und Jean Georges. Dementia in Europe Yearbook 2019: Estimating the prevalence of dementia in Europe, 2019. URL: <https://www.alzheimer-europe.org/Publications/Dementia-in-Europe-Yearbooks> [zuletzt abgerufen am 10.12.2021].
- [61] Bitkom e.V. Große Offenheit für Spende von Patientendaten. *Bitkom e.V.*, 03.07.2020. URL: <https://www.bitkom.org/Presse/Presseinformation/Grosse-Offenheit-fuer-Spende-von-Patientendaten> [zuletzt abgerufen am 10.12.2021].
- [62] Kaj Blennow, Harald Hampel, Michael Weiner, und Henrik Zetterberg. Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nature reviews. Neurology*, 6(3):131–144, 2010. DOI:10.1038/nrneuro1.2010.4.
- [63] Piero Bonatti, Bert Bos, Stefan Decker, et al. Data Privacy Vocabularies and Controls: Semantic Web for Transparency and Privacy. 2018. URL: <https://www.w3.org/Privacy/permissions-ws-2018/papers/axel-polleres.pdf> [zuletzt abgerufen am 06.12.2021].

- [64] Piero Bonatti, Sabrina Kirrane, Illiana Mineva Petrova, et al. The SPECIAL Usage Policy Language, 2019. URL: <https://ai.wu.ac.at/policies/policylanguage/> [zuletzt abgerufen am 10.12.2021].
- [65] Jürgen Bortz und Christof Schuster. *Statistik für Human- und Sozialwissenschaftler: Extras online*. Springer-Lehrbuch. Springer, Berlin and Heidelberg, limitierte sonderausgabe, 7., vollständig überarbeitete und erweiterte auflage edition, 2016. ISBN: 3642127703.
- [66] Brain Test Inc. BrainTest - Online Dementia, MCI and Alzheimer's Screening Test, 2016. URL: <https://braintest.com/> [zuletzt abgerufen am 10.12.2021].
- [67] Brain Test Inc. BrainTest - Apps on Google Play, 2021. URL: <https://play.google.com/store/apps/details?id=com.braintest.app> [zuletzt abgerufen am 10.12.2021].
- [68] Mark L. Braunstein. Health Care in the Age of Interoperability: The Potential and Challenges. *IEEE Pulse*, 9(5):34–36, 2018. DOI:10.1109/MPUL.2018.2856941.
- [69] Mark L. Braunstein. Healthcare in the Age of Interoperability: The Promise of Fast Healthcare Interoperability Resources. *IEEE Pulse*, 9(6):24–27, 2018. DOI:10.1109/MPUL.2018.2869317.
- [70] Mark L. Braunstein. Health Care in the Age of Interoperability Part 5: The Personal Health Record. *IEEE Pulse*, 10(3):19–23, 2019. DOI:10.1109/MPULS.2019.2911804.
- [71] Mark L. Braunstein. Healthcare in the Age of Interoperability: Part 3. *IEEE Pulse*, 10(1):26–29, 2019. DOI:10.1109/MPULS.2018.2885831.
- [72] Dale E. Bredesen. Reversal of cognitive decline: a novel therapeutic program. *Aging*, 6(9):707–717, 2014. DOI:10.18632/aging.100690.
- [73] Dale E. Bredesen, Edwin C. Amos, Jonathan Canick, et al. Reversal of cognitive decline in Alzheimer's disease. *Aging*, 8(6):1250–1258, 2016. DOI:10.18632/aging.100981.
- [74] Dale E. Bredesen, Kenneth Sharlin, David Jenkins, et al. Reversal of Cognitive Decline: 100 Patients. *Journal of Alzheimer's Disease & Parkinsonism*, 08(05), 2018. DOI:10.4172/2161-0460.1000450.
- [75] Leo Breiman. *Classification and regression trees*. [Routledge], [Abingdon], 1984. ISBN: 9781315139470. DOI:10.1201/9781315139470.
- [76] Leo Breiman. Random Forests. *Machine learning*, 45(1):5–32, 2001. DOI:10.1023/A:1010933404324.
- [77] Dan Brickley, R. V. Guha, und Brian McBride. RDF Schema 1.1, 2014. URL: <https://www.w3.org/TR/rdf-schema/> [zuletzt abgerufen am 10.12.2021].
- [78] Jan B. Brönneke, Jörg F. Debatin, Julia Hagen, et al. *DiGA VADEMECUM: Was man zu Digitalen Gesundheitsanwendungen wissen muss*. Medizinisch Wissenschaftliche Verlagsgesellschaft and ProQuest, Berlin and Ann Arbor, Michigan, 1st ed. edition, 2020. ISBN: 978-3-95466-569-3.
- [79] Isabelle Budin-Ljøsne, Harriet J. A. Teare, Jane Kaye, et al. Dynamic Consent: a potential solution to some of the challenges of modern biomedical research. *BMC medical ethics*, 18(1):4, 2017. DOI:10.1186/s12910-016-0162-9.

- [80] Bundesgesundheitsministerium. Spitzenverband Bund der Krankenkassen (GKV-Spitzenverband), 2021. URL: <https://www.bundesgesundheitsministerium.de/themen/gesundheitswesen/selbstverwaltung/spitzenverband-bund-der-krankenkassen-gkv-spitzenverband.html> [zuletzt abgerufen am 10.12.2021].
- [81] Bundesinstitut für Arzneimittel und Medizinprodukte. BfArM - Digitale Gesundheitsanwendungen, 05.10.2020. URL: [https://www.bfarm.de/DE/Medizinprodukte/DVG/\\_node.html](https://www.bfarm.de/DE/Medizinprodukte/DVG/_node.html) [zuletzt abgerufen am 10.12.2021].
- [82] Bundesinstitut für Arzneimittel und Medizinprodukte. Das Fast Track Verfahren für digitale Gesundheitsanwendungen (DiGA) nach § 139e SGB V: Ein Leitfaden für Hersteller, Leistungserbringer und Anwender, 2020. URL: [https://www.bfarm.de/SharedDocs/Downloads/DE/Service/Beratungsverfahren/DiGA-Leitfaden.pdf;jsessionid=9D9D67741329073A14EA93597A07B908.1\\_cid319?\\_\\_blob=publicationFile&v=10](https://www.bfarm.de/SharedDocs/Downloads/DE/Service/Beratungsverfahren/DiGA-Leitfaden.pdf;jsessionid=9D9D67741329073A14EA93597A07B908.1_cid319?__blob=publicationFile&v=10) [zuletzt abgerufen am 10.12.2021].
- [83] Bundesinstitut für Arzneimittel und Medizinprodukte. DiGA-Verzeichnis, 2020. URL: <https://diga.bfarm.de/de/verzeichnis> [zuletzt abgerufen am 04.02.2022].
- [84] Bundesinstitut für Arzneimittel und Medizinprodukte. Orientierungshilfe Medical Apps, 2021. URL: [https://www.bfarm.de/DE/Medizinprodukte/Abgrenzung/MedicalApps/\\_node.html](https://www.bfarm.de/DE/Medizinprodukte/Abgrenzung/MedicalApps/_node.html).
- [85] Bundesministerium des Innern, für Bau und Heimat. Bundesdatenschutzgesetz, 25.09.2017. URL: <https://www.bmi.bund.de/DE/themen/it-und-digitalpolitik/datenpolitik/bundesdatenschutzgesetz/bundesdatenschutzgesetz-node.html>.
- [86] Bundesministerium für Familie, Senioren, Frauen und Jugend. Nationale Demenzstrategie. 2020. URL: [https://www.nationale-demenzstrategie.de/fileadmin/nds/pdf/2020-07-01\\_Nationale\\_Demenzstrategie.pdf](https://www.nationale-demenzstrategie.de/fileadmin/nds/pdf/2020-07-01_Nationale_Demenzstrategie.pdf) [zuletzt abgerufen am 28.02.2022].
- [87] Bundesministerium für Gesundheit. Gesetz für schnellere Termine und bessere Versorgung: Terminservice- und Versorgungsgesetz TSVG, 06.05.2019.
- [88] Bundesministerium für Gesundheit. Ärzte sollen Apps verschreiben können, 2019. URL: <https://www.bundesgesundheitsministerium.de/digitale-versorgung-gesetz.html> [zuletzt abgerufen am 04.02.2022].
- [89] Bundesministerium für Gesundheit. Gesetz für eine bessere Versorgung durch Digitalisierung und Innovation (Digitale-Versorgung-Gesetz): DVG, 2019. URL: [https://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger\\_BGBl&jumpTo=bgbl119s2562.pdf#\\_\\_bgbl\\_\\_%2F%2F\\*%5B%40attr\\_id%3D%27bgbl119s2562.pdf%27%5D\\_\\_1594022757147](https://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl119s2562.pdf#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl119s2562.pdf%27%5D__1594022757147) [zuletzt abgerufen am 10.12.2021].
- [90] Bundesregierung. Gesetz zur Modernisierung der gesetzlichen Krankenversicherung: (GKV-Modernisierungsgesetz - GMG), 2003.
- [91] Neil Burgess, Eleanor A. Maguire, und John O'Keefe. The Human Hippocampus and Spatial and Episodic Memory. *Neuron*, 35(4):625–641, 2002. URL: <https://www.sciencedirect.com/science/article/pii/S0896627302008309>, DOI:10.1016/S0896-6273(02)00830-9.

- [92] H. Buschke. Cued recall in amnesia. *Journal of clinical neuropsychology*, 6(4):433–440, 1984. DOI: 10.1080/01688638408401233.
- [93] Haewon Byeon. Developing a random forest classifier for predicting the depression and managing the health of caregivers supporting patients with Alzheimer’s Disease. *Technology and health care : official journal of the European Society for Engineering and Medicine*, 27(5):531–544, 2019. DOI: 10.3233/THC-191738.
- [94] Douglas W. Challener, Larry J. Prokop, und Omar Abu-Saleh. The Proliferation of Reports on Clinical Scoring Systems: Issues About Uptake and Clinical Utility. *JAMA*, 321(24):2405–2406, 2019. DOI: 10.1001/jama.2019.5284.
- [95] A. Mark Clarfield. The decreasing prevalence of reversible dementias: an updated meta-analysis. *Archives of internal medicine*, 163(18):2219–2229, 2003. DOI: 10.1001/archinte.163.18.2219.
- [96] Jacob Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960. DOI: 10.1177/001316446002000104.
- [97] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Taylor and Francis, Hoboken, 2nd ed. edition, 2013. ISBN: 9780805802832.
- [98] Randall J. Cohrs, Tyler Martin, Parviz Ghahramani, et al. Translational Medicine definition by the European Society for Translational Medicine. *European Journal of Molecular & Clinical Medicine*, 2(3):86, 2017. URL: [https://www.researchgate.net/publication/274142125\\_Translational\\_Medicine\\_definition\\_by\\_the\\_European\\_Society\\_for\\_Translational\\_Medicine](https://www.researchgate.net/publication/274142125_Translational_Medicine_definition_by_the_European_Society_for_Translational_Medicine) [zuletzt abgerufen am 10.12.2021], DOI: 10.1016/j.nhtm.2014.12.002.
- [99] Fabienne Collette, H el ene Amieva, St ephane Adam, et al. Comparison of Inhibitory Functioning in Mild Alzheimer’s Disease and Frontotemporal Dementia. *Cortex*, 43(7):866–874, 2007. DOI: 10.1016/S0010-9452(08)70686-5.
- [100] Mark Jade S. Corpuz, Junnel Avestro, Fredilyn B. Calanda, und Roselia C. Morco. eRheumatologist: Mobile-Based Expert System for Rheumatology Utilizing Fuzzy Logic Algorithm. In Unknown, editor, *Proceedings of the 2017 International Conference on Computer Science and Artificial Intelligence - CSAI 2017*. ISBN: 9781450353922, New York, New York, USA, 2017. ACM Press. DOI: 10.1145/3168390.
- [101] Corinna Cortes und Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995. DOI: 10.1007/BF00994018.
- [102] Lorrie Cranor, Brooks Dobbs, Serge Egelman, et al. The Platform for Privacy Preferences 1.1 (P3P1.1) Specification, 2006. URL: <https://www.w3.org/TR/P3P11/> [zuletzt abgerufen am 10.12.2021].
- [103] D. Seth, N. Biswas, und D. Ghosh. Big health data: Cardiac remodelling and functional interactions of big brain based implications in body sensor networks. In *2017 7th International Conference on Communication Systems and Network Technologies (CSNT)*, pages 339–344, 2017. DOI: 10.1109/CSNT.2017.8418563.

- [104] Jessamyn Dahmen, Diane Cook, Robert Fellows, und Maureen Schmitter-Edgecombe. An analysis of a digital variant of the Trail Making Test using machine learning techniques. *Technology and health care : official journal of the European Society for Engineering and Medicine*, 25(2):251–264, 2017. DOI: 10.3233/THC-161274.
- [105] Alireza Darvishy, Hans-Peter Hutter, und Alexander Seifert. Altersgerechte mobile Applikationen: Grundlagen und Empfehlungen. 2016. URL: [https://www.zfg.uzh.ch/dam/jcr:7304ebb1-0d7c-47d7-ae9-0c07dece8db2/AppBrochure\\_20160822\\_WEB.pdf](https://www.zfg.uzh.ch/dam/jcr:7304ebb1-0d7c-47d7-ae9-0c07dece8db2/AppBrochure_20160822_WEB.pdf) [zuletzt abgerufen am 10.12.2021].
- [106] Matthew Davis, Thomas O Connell, Scott Johnson, et al. Estimating Alzheimer’s Disease Progression Rates from Normal Cognition Through Mild Cognitive Impairment and Stages of Dementia. *Current Alzheimer research*, 15(8):777–788, 2018. DOI: 10.2174/1567205015666180119092427.
- [107] Ana Correia de Barros, Roxanne Leitão, und Jorge Ribeiro. Design and Evaluation of a Mobile User Interface for Older Adults: Navigation, Interaction and Visual Design Recommendations. *Procedia Computer Science*, 27:369–378, 2014. DOI: 10.1016/j.procs.2014.02.041.
- [108] DEPARTMENT OF HEALTH AND HUMAN SERVICES. 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program, 2020. URL: <https://www.govinfo.gov/content/pkg/FR-2020-05-01/pdf/2020-07419.pdf> [zuletzt abgerufen am 10.12.2021].
- [109] Deutsche Alzheimer Gesellschaft. Die Häufigkeit von Demenzerkrankungen. 2018. URL: [http://www.deutsche-alzheimer.de/fileadmin/alz/pdf/factsheets/infoblatt1\\_haeufigkeit\\_demenzerkrankungen\\_dalzg.pdf](http://www.deutsche-alzheimer.de/fileadmin/alz/pdf/factsheets/infoblatt1_haeufigkeit_demenzerkrankungen_dalzg.pdf) [zuletzt abgerufen am 10.12.2021].
- [110] Deutscher Ärzteverlag GmbH, Redaktion Deutsches Ärzteblatt. RKI: Mehr als eine halbe Million Teilnehmer bei „Datenspende“-App, 2020. URL: <https://www.aerzteblatt.de/nachrichten/112636/RKI-Mehr-als-eine-halbe-Million-Teilnehmer-bei-Datenspende-App> [zuletzt abgerufen am 04.02.2022].
- [111] Deutsches Parlament. Sozialgesetzbuch (SGB) Erstes Buch (I) - Allgemeiner Teil - (Artikel des Gesetzes vom 11. Dezember 1975, BGBl. I S. 3015): SGB I, 11.12 / 1975.
- [112] Deutsches Parlament. Bürgerliches Gesetzbuch: BGB, 18.08.1896. URL: <https://www.gesetze-im-internet.de/bgb/BJNR001950896.html#BJNR001950896BJNG002202377> [zuletzt abgerufen am 10.12.2021].
- [113] Deutsches Parlament. Gesetz über Medizinprodukte - Medizinproduktegesetz: MPG, 20.11.2019. URL: <http://www.gesetze-im-internet.de/mpg/> [zuletzt abgerufen am 10.12.2021].
- [114] Deutsches Parlament. Gesetz für sichere digitale Kommunikation und Anwendungen im Gesundheitswesen sowie zur Änderung weiterer Gesetze: E-Health-Gesetz, 28.12.2015.
- [115] DICOM. DICOM is the international standard to transmit, store, retrieve, print, process, and display medical imaging information, 2021. URL: <https://www.dicomstandard.org/> [zuletzt abgerufen am 10.12.2021].
- [116] DIN. Ergonomie der Mensch-System-Interaktion: Teil 110: Grundsätze der Dialoggestaltung, 2008-09. DOI: 10.31030/1464024.

- [117] DIN. Ergonomie der Mensch-System-Interaktion: Teil 112: Grundsätze der Informationsdarstellung, 2017-08. DOI: 10.31030/2580439.
- [118] DIN. Ergonomie der Mensch-System-Interaktion: Teil 125: : Empfehlungen zur visuellen Informationsdarstellung, 2018-05. DOI: 10.31030/2674631.
- [119] DIN. Ergonomie der Mensch-System-Interaktion: Teil 110: Interaktionsprinzipien, 2019-07-26. DOI: 10.31030/3071981.
- [120] DIN. Anwendung des Risikomanagements auf Medizinprodukte, 2020-07. DOI: 10.31030/3062541.
- [121] Alessandra Dodich, Chiara Cerami, Nicola Canessa, et al. A novel task assessing intention and emotion attribution: Italian standardization and normative data of the Story-based Empathy Task. *Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, 36(10):1907–1912, 2015. DOI: 10.1007/s10072-015-2281-3.
- [122] Guntram Doelfs. TSVG Verabschiedet: Minister Spahn entmachtet Selbstverwaltung in der gematik. *kma - Klinik Management aktuell*, 24(04):8, 2019. DOI: 10.1055/s-0036-1595749.
- [123] Asuman Dogac, Gokce B. Laleci, Thomas Aden, und Marco Eichelberg. Enhancing IHE XDS for federated clinical affinity domain support. *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, 11(2):213–221, 2007. DOI: 10.1109/titb.2006.874928.
- [124] Finale Doshi-Velez und Been Kim. Towards A Rigorous Science of Interpretable Machine Learning, 2017. URL: <https://arxiv.org/pdf/1702.08608> [zuletzt abgerufen am 10.12.2021].
- [125] John H. Dougherty, Rex L. Cannon, Christopher R. Nicholas, et al. The computerized self test (CST): an interactive, internet accessible cognitive screening test for dementia. *Journal of Alzheimer's disease : JAD*, 20(1):185–195, 2010. DOI: 10.3233/JAD-2010-1354.
- [126] Martin Dugas, Philipp Neuhaus, Alexandra Meidt, et al. Portal of medical data models: information infrastructure for medical research and healthcare. *Database: The Journal of Biological Databases & Curation*, 2016:1–9, 2016. URL: <http://www.redi-bw.de/db/ebsco.php/search.ebscohost.com/login.aspx%3fdirect%3dtrue%26db%3daph%26AN%3d119919977%26site%3dehost-live>.
- [127] E. Kaplan, H. Goodglass, S. Weintraub. *The Boston Naming Test*. Lea & Febiger, Philadelphia, 1983.
- [128] Howard Eichenbaum. The hippocampus and mechanisms of declarative memory. *Behavioural brain research*, 103(2):123–133, 1999. URL: <https://pubmed.ncbi.nlm.nih.gov/10513581/> [zuletzt abgerufen am 10.12.2021], DOI: 10.1016/s0166-4328(99)00044-3.
- [129] Nabil El Ioini und Claus Pahl. A Review of Distributed Ledger Technologies. *OTM Confederated International Conferences*, pages 277–288, 2018. DOI: 10.1007/978-3-030-02671-4{\textunderscore}16.
- [130] Shaker El-Sappagh, Jose M. Alonso, S. M. Riazul Islam, et al. A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease. *Scientific reports*, 11(1):2660, 2021. DOI: 10.1038/s41598-021-82098-3.

- [131] Stéphane Epelbaum, Rémy Genthon, Enrica Cavedo, et al. Preclinical Alzheimer's disease: A systematic review of the cohorts underlying the concept. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 13(4):454–467, 2017. DOI: [10.1016/j.jalz.2016.12.003](https://doi.org/10.1016/j.jalz.2016.12.003).
- [132] Dorothea Erharter. SeniorInnen und Smartphones bzw. Tablets: Anforderungen an Usability, Funktionalitäten, Bildungsangebote, Kaufberatung & Support. In Susanna Finker, Karin Kicker-Frisinghelli, und Claudia Stöckl, editors, *Die Gesellschaft des langen Lebens: Soziale und individuelle Herausforderungen*, Gesellschaft der Unterschiede, pages 93–108. transcript-Verlag, 2016. ISBN: 9783839434260. DOI: [10.14361/9783839434260-007](https://doi.org/10.14361/9783839434260-007).
- [133] Dorothea Erharter, Bernhard Jungwirth, Bente Knoll, et al. Smartphones, Tablets, App für Seniorinnen und Senioren. In Guido Kempfer und Walter Ritter, editors, *Assistenztechnik für betreutes Wohnen: Beiträge zum Usability Day XII, 16. Mai 2014*. Pabst Science Publ, Lengerich, 2014. ISBN: 978-3-89967-943-4.
- [134] Dorothea Erharter und Elka Xharo. Developer-Guidelines Usability von Apps für Seniorinnen und Senioren. 2016.
- [135] Besime Erin und Rahib H. Abiyev. Diagnosis of Common Diseases Using Type-2 Fuzzy System. In Unknown, editor, *Proceedings of the 3rd International Conference on Machine Learning and Soft Computing - ICMLSC 2019*, pages 239–243. ISBN: 9781450366120, New York, New York, USA, 2019. ACM Press. DOI: [10.1145/3310986.3311028](https://doi.org/10.1145/3310986.3311028).
- [136] Gerhard W. Eschweiler, Thomas Leyhe, Stefan Klöppel, und Michael Hüll. New developments in the diagnosis of dementia. *Deutsches Arzteblatt international*, 107(39):677–683, 2010. DOI: [10.3238/arztebl.2010.0677](https://doi.org/10.3238/arztebl.2010.0677).
- [137] Deborah Estrin. Small data, where n = me. *Communications of the ACM*, 57(4):32–34, 2014. DOI: [10.1145/2580944](https://doi.org/10.1145/2580944).
- [138] Europäisches Parlament. Verordnung (EU) 2017/745 des Europäischen Parlaments und des Rates über Medizinprodukte: MDR, 25. Mai 2017.
- [139] Europäisches Parlament. VERORDNUNG (EU) 2016/679 DES EUROPÄISCHEN PARLAMENTES UND DES RATES vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung): DSGVO, 27.04.2016. URL: <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32016R0679&from=DE> [zuletzt abgerufen am 10.12.2021].
- [140] G. Eysenbach. What is e-health? *Journal of medical Internet research*, 3(2):E20, 2001. DOI: [10.2196/jmir.3.2.e20](https://doi.org/10.2196/jmir.3.2.e20).
- [141] F. Holl, L. Reschke, S. Müller-Miellitz, et al. Secondary Use of Research Data: A Review of Availability and Utilization of Public Use Files and Initial Steps for the Development of a Process to Facilitate Medical Research Data Donation. In *2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA)*, pages 1–5, 2019. DOI: [10.1109/AICCSA47632.2019.9035331](https://doi.org/10.1109/AICCSA47632.2019.9035331).
- [142] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006. URL: <https://www.math.ucdavis.edu/~protect\unhbox\voidb@x\protect\penalty>

- \@M\saito/data/roc/fawcett-roc.pdf [zuletzt abgerufen am 18.05.2021], DOI: 10.1016/j.patrec.2005.10.010.
- [143] U. Fiedler, J. Wiltfang, N. Peters, und J. Benninghoff. Fortschritte in der Diagnostik der Alzheimer-Demenz. *Der Nervenarzt*, 83(5):661–673, 2012. DOI: 10.1007/s00115-012-3486-9.
- [144] Kathrin Figl. Deutschsprachige Fragebögen zur Usability-Evaluation im Vergleich. In *Zeitschrift für Arbeitswissenschaft*, volume 4, pages 321–337. 2010. URL: <https://epub.wu.ac.at/5648/> [zuletzt abgerufen am 10.12.2021].
- [145] Gerda G. Fillenbaum, Gerald van Belle, John C. Morris, et al. CERAD (Consortium to Establish a Registry for Alzheimer’s Disease) The first 20 years. *Alzheimer’s & dementia : the journal of the Alzheimer’s Association*, 4(2):96–109, 2008. DOI: 10.1016/j.jalz.2007.08.005.
- [146] Findata. Finnish Social and Health Data Permit Authority, 2020. URL: <https://findata.fi/en/> [zuletzt abgerufen am 04.02.2022].
- [147] Marshal F. Folstein, Susan E. Folstein, und Paul R. McHugh. “Mini-mental state”. *Journal of Psychiatric Research*, 12(3):189–198, 1975. DOI: 10.1016/0022-3956(75)90026-6.
- [148] Food and Drug Administration. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML) - Based Software as a Medical Device (SaMD): Discussion Paper and Request for Feedback. 2019. URL: <https://www.fda.gov/media/122535/download> [zuletzt abgerufen am 10.12.2021].
- [149] Food and Drug Administration. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. 2021. URL: <https://www.fda.gov/media/145022/download> [zuletzt abgerufen am 10.12.2021].
- [150] Forschungszentrum Jülich GmbH. Personalisierte Medizin im Gesundheitssystem der Zukunft, 2009. URL: [https://www.isi.fraunhofer.de/content/dam/isi/dokumente/cc\\_t/2009/personalisierte\\_medizin.pdf](https://www.isi.fraunhofer.de/content/dam/isi/dokumente/cc_t/2009/personalisierte_medizin.pdf) [zuletzt abgerufen am 10.12.2021].
- [151] Hans Förstl. *Demenzatlant spezial*. Thieme, Stuttgart, 2012. ISBN: 978-3131693914.
- [152] Frank Schwärzler, Frank Peter Roland. PflegeCoDe Abschlussbericht (Zeitraum 2015 - 2018), 2018. URL: <https://www.tib.eu/suchen/id/TIBKAT:1668453436/> [zuletzt abgerufen am 10.12.2021].
- [153] Helen Frankish und Richard Horton. Prevention and management of dementia: a priority for public health. *The Lancet*, 390(10113):2614–2615, 2017. DOI: 10.1016/S0140-6736(17)31756-7.
- [154] Mara Friedrich und Christophe Kunze. Usability-Aspekte bei der Gestaltung mobiler Nutzeroberflächen für technikdistanzierte ältere Nutzer. In Pabst Science Publishers, editor, *Umgebungsunterstütztes Leben: Beiträge zum Usability Day XV*. 2017. ISBN: 978-3958533165.
- [155] Teresa Froböse und Alexander Kurz. So wichtig ist frühe Demenzdiagnose: Alzheimer, Lewy-Körper-Krankheit oder Durchblutungsstörung? Zertifizierte Fortbildung. *DNP – Der Neurologe & Psychiater*, page 13, 2012. URL: <https://link.springer.com/content/pdf/10.1007/s15202-012-0077-3.pdf> [zuletzt abgerufen am 10.12.2021].

- [156] gematik. Elektronische Patientenakte: Spezifikationen und Zulassungsverfahren veröffentlicht, Informationsveranstaltung im Januar für Zulassungsnehmer, 2018. URL: <https://www.gematik.de/news/news/elektronische-patientenakte-spezifikationen-und-zulassungsverfahren-veroeffentlicht-informationsver/>.
- [157] gematik. Gemeinsam für eine große Aufgabe, 2021. URL: <https://www.gematik.de/ueber-uns/>.
- [158] gematik. Unternehmensstruktur, 2021. URL: <https://www.gematik.de/ueber-uns/unternehmensstruktur/>.
- [159] gematik, Kassenärztliche Bundesvereinigung, Spitzenverband Bund der Krankenkassen, und Kassenzahnärztliche Bundesvereinigung. Letter of Intent - Elektronische Patientenakte, 2018. URL: [https://www.kbv.de/media/sp/LoI\\_ePA\\_final.pdf](https://www.kbv.de/media/sp/LoI_ePA_final.pdf) [zuletzt abgerufen am 10.12.2021].
- [160] Gemeinsamer Bundesausschuss. Mitglieder - Gemeinsamer Bundesausschuss, 2021. URL: <https://www.g-ba.de/ueber-den-gba/wer-wir-sind/mitglieder/> [zuletzt abgerufen am 10.12.2021].
- [161] Gerd Gigerenzer, Kirsten Schlegel-Matthies und Gert G. Wagner. eHealth und mHealth: Chancen und Risiken der Digitalisierung im Gesundheitsbereich. 2016. URL: <http://www.svr-verbraucherfragen.de/wp-content/uploads/Digitale-Welt-und-Gesundheit.pdf> [zuletzt abgerufen am 10.12.2021].
- [162] Sara Gerke, Boris Babic, Theodoros Evgeniou, und I. Glenn Cohen. The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. *NPJ digital medicine*, 3:53, 2020. URL: <https://www.nature.com/articles/s41746-020-0262-2.pdf> [zuletzt abgerufen am 10.12.2021], DOI: [10.1038/s41746-020-0262-2](https://doi.org/10.1038/s41746-020-0262-2).
- [163] GKV-Spitzenverband. Aufgaben und Ziele - GKV-Spitzenverband, 2021. URL: [https://www.gkv-spitzenverband.de/gkv\\_spitzenverband/der\\_verband/aufgaben\\_und\\_ziele/aufgaben\\_und\\_ziele.jsp](https://www.gkv-spitzenverband.de/gkv_spitzenverband/der_verband/aufgaben_und_ziele/aufgaben_und_ziele.jsp) [zuletzt abgerufen am 10.12.2021].
- [164] Richard E. Gliklich, Michelle B. Leavy, und Nancy A. Dreyer. Tools and Technologies for Registry Interoperability, 2nd addendum of Registries for Evaluating Patient Outcomes, A User's Guide, 3rd ed, 2019. URL: [https://www.ncbi.nlm.nih.gov/books/NBK551879/pdf/Bookshelf\\_NBK551879.pdf](https://www.ncbi.nlm.nih.gov/books/NBK551879/pdf/Bookshelf_NBK551879.pdf) [zuletzt abgerufen am 10.12.2021], DOI: [10.23970/AHRQPCREGISTRIES3ADDENDUM2](https://doi.org/10.23970/AHRQPCREGISTRIES3ADDENDUM2).
- [165] Pablo Gomez, Roger Ratcliff, und Manuel Perea. A model of the go/no-go task. *Journal of experimental psychology. General*, 136(3):389–413, 2007. DOI: [10.1037/0096-3445.136.3.389](https://doi.org/10.1037/0096-3445.136.3.389).
- [166] Barry Gordon. Lexical access and lexical decision: mechanisms of frequency sensitivity. *Journal of Verbal Learning and Verbal Behavior*, 22(1):24–44, 1983. DOI: [10.1016/S0022-5371\(83\)80004-8](https://doi.org/10.1016/S0022-5371(83)80004-8).
- [167] Barry Gordon und Alfonso Caramazza. Lexical decision for open- and closed-class words: Failure to replicate differential frequency sensitivity. *Brain and Language*, 15(1):143–160, 1982. DOI: [10.1016/0093-934X\(82\)90053-0](https://doi.org/10.1016/0093-934X(82)90053-0).

- [168] Adela Grando, Julia Ivanova, Megan Hiestand, et al. Mental health professional perspectives on health data sharing: Mixed methods study. *Health informatics journal*, page 1460458219893848, 2020. DOI:10.1177/1460458219893848.
- [169] Maria Adela Grando, Anita Murcko, Srividya Mahankali, et al. A Study to Elicit Behavioral Health Patients' and Providers' Opinions on Health Records Consent. *Journal of Law, Medicine & Ethics*, 45(2):238–259, 2017. URL: <http://www.redibw.de/db/ebsco.php/search.ebscohost.com/login.aspx?fdirect=true&d%3daph%26AN%3d124159301%26site%3dehost-live>.
- [170] Ellen Grober und Herman Buschke. Genuine memory deficits in dementia. *Developmental neuropsychology*, 3(1):13–36, 1987. DOI:10.1080/87565648709540361.
- [171] Ellen Grober, Amy E. Sanders, Charles Hall, und Richard B. Lipton. Free and cued selective reminding identifies very mild dementia in primary care. *Alzheimer disease and associated disorders*, 24(3):284–290, 2010. DOI:10.1097/WAD.0b013e3181cfc78b.
- [172] Swati Gupta, Karl Fua, David Pautler, und Ilya Farber. Designing Serious Games for Elders, 2013. DOI:10.13140/RG.2.1.5166.9848.
- [173] Isabelle Guyon, Jason Weston, Stephen Barnhill, und Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1/3):389–422, 2002. DOI:10.1023/A:1012487302797.
- [174] Birger Haarbrandt, Björn Schreiweis, Sabine Rey, et al. HiGHmed - An Open Platform Approach to Enhance Care and Research across Institutional Boundaries. *Methods of Information in Medicine*, 57(S 01):e66–e81, 2018. DOI:10.3414/ME18-02-0002.
- [175] Juho Hamari, Jonna Koivisto, und Harri Sarsa. Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In *IEEE 8th International Symposium on Service-Oriented System Engineering (SOSE), 2014: 7 - 11 April 2014, Oxford, United Kingdom ; [including workshop/symposium papers]*, pages 3025–3034. ISBN: 978-1-4799-2504-9, Piscataway, NJ, 2014. IEEE. DOI:10.1109/HICSS.2014.377.
- [176] H. Hampel, S. E. O'Bryant, J. I. Castrillo, et al. PRECISION MEDICINE - The Golden Gate for Detection, Treatment and Prevention of Alzheimer's Disease. *The journal of prevention of Alzheimer's disease*, 3(4):243–259, 2016. DOI:10.14283/jpad.2016.112.
- [177] Jiawei Han, Micheline Kamber, und Jian Pei. *Data mining: Concepts and techniques*. The Morgan Kaufmann series in data management systems. Elsevier/Morgan Kaufmann, Amsterdam, 3. ed. edition, 2012. ISBN: 9781283171175. URL: <https://ebookcentral.proquest.com/lib/subhh/detail.action?docID=729031> [zuletzt abgerufen am 10.12.2021].
- [178] Robert M. Haralick und Linda G. Shapiro. *Computer and robot vision*. Addison-Wesley, Reading, Mass., 1993. ISBN: 978-0-201-56943-8.
- [179] Christopher G. Harris, Mike Stephens, et al. A combined corner and edge detector. In *Alvey vision conference*, volume 15, pages 10–5244, 1988. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.434.4816&rep=rep1&type=pdf> [zuletzt abgerufen am 10.12.2021].

- [180] Mark Hastenteufel und Sina Renaud. *Software als Medizinprodukt: Entwicklung und Zulassung von Software in der Medizintechnik*. Lehrbuch. Springer Fachmedien Wiesbaden, Wiesbaden, 2019. ISBN: 978-3-658-26487-1. DOI: 10.1007/978-3-658-26488-8.
- [181] Julian Hatwell. CHIRPS: Explanation System for Tree Ensembles, 2020. URL: [https://github.com/julianhatwell/explain\\_te](https://github.com/julianhatwell/explain_te) [zuletzt abgerufen am 19.01.2022].
- [182] Julian Hatwell, Mohamed Medhat Gaber, und R. Muhammad Atif Azad. Ada-WHIPS: explaining AdaBoost classification with applications in the health sciences. *BMC Medical Informatics & Decision Making*, 20(1):250, 2020. DOI: 10.1186/s12911-020-01201-2.
- [183] Julian Hatwell, Mohamed Medhat Gaber, und R. Muhammad Atif Azad. CHIRPS: Explaining random forest classification. *Artificial Intelligence Review*, 53(8):5747–5788, 2020. DOI: 10.1007/s10462-020-09833-6.
- [184] Health Level Seven. HL FHIR Standard, 2021. URL: [http://hl7.org/fhir/?utm\\_referrer=http%3A%2F%2Fwww.hl7.org%2F](http://hl7.org/fhir/?utm_referrer=http%3A%2F%2Fwww.hl7.org%2F) [zuletzt abgerufen am 10.12.2021].
- [185] Healthcare Information Management Systems Society. Interoperability in Healthcare. *HIMSS*, 2020, 04.08.2020. URL: <https://www.himss.org/resources/interoperability-health-care> [zuletzt abgerufen am 10.12.2021].
- [186] Hedieh Sajedi und Shabnam Rahbar Yaghobi. Information hiding methods for E-Healthcare. *Smart Health*, 15:100104, 2020. URL: <http://www.sciencedirect.com/science/article/pii/S2352648317300016>, DOI: 10.1016/j.smhl.2019.100104.
- [187] Eric B. Hekler, Predrag Klasnja, Guillaume Chevaance, et al. Why we need a small data paradigm. *BMC Medicine*, 17(1):133, 2019. DOI: 10.1186/s12916-019-1366-x.
- [188] Alan R. Hevner, Salvatore T. March, Jinsoo Park, und Sudha Ram. Design science in information systems research. *MIS quarterly*, 28(1):75–105, 2004.
- [189] Nils B. Heyen. Digitale Selbstvermessung und Quantified Self: Potenziale, Risiken und Handlungsoptionen, 2016. URL: [https://www.isi.fraunhofer.de/content/dam/isi/dokumente/cct/2016/Policy-Paper-Quantified-Self\\_Fraunhofer-ISI.pdf](https://www.isi.fraunhofer.de/content/dam/isi/dokumente/cct/2016/Policy-Paper-Quantified-Self_Fraunhofer-ISI.pdf) [zuletzt abgerufen am 10.12.2021].
- [190] Pascal Hitzler. *Semantic Web: Grundlagen*. eXamen.press. Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg, 2008. ISBN: 978-3-540-33993-9. DOI: 10.1007/978-3-540-33994-6.
- [191] Stephan Hofmeister. Elektronische Arztakte und elektronische Patientenakte: Vorstellungen der KBV, 20.06.2018. URL: [https://www.kbv.de/media/sp/2018\\_06\\_20\\_Diskussionsforum\\_KBV\\_Hofmeister.pdf](https://www.kbv.de/media/sp/2018_06_20_Diskussionsforum_KBV_Hofmeister.pdf).
- [192] Andreas Holzinger, Chris Biemann, Constantinos S. Pattichis, und Douglas B. Kell. What do we need to build explainable AI systems for the medical domain? URL: <https://arxiv.org/pdf/1712.09923>.
- [193] Andreas Holzinger, André Calero Valdez, und Martina Zieffle. Towards Interactive Recommender Systems with the Doctor-in-the-Loop. DOI: 10.18420/muc2016-ws11-0001.

- [194] C. P. Hughes, L. Berg, W. L. Danziger, et al. A new clinical scale for the staging of dementia. *The British journal of psychiatry : the journal of mental science*, 140:566–572, 1982. DOI: 10.1192/bjp.140.6.566.
- [195] Julie Hugo und Mary Ganguli. Dementia and cognitive impairment: epidemiology, diagnosis, and treatment. *Clinics in Geriatric Medicine*, 30(3):421–442, 2014. DOI: 10.1016/j.cger.2014.04.001.
- [196] Hyperledger Fabric. A Blockchain Platform for the Enterprise, 2021. URL: <https://hyperledger-fabric.readthedocs.io/en/release-2.2/> [zuletzt abgerufen am 10.12.2021].
- [197] Renato Iannella. The Open Digital Rights Language: XML for Digital Rights Management. *Information Security Technical Report*, 9(3):47–55, 2004. URL: <http://www.sciencedirect.com/science/article/pii/S1363412704000317>, DOI: 10.1016/S1363-4127(04)00031-7.
- [198] Renato Iannella, Susanne Guth, Daniel Paehler, und Andreas Kasten. ODRL Version 2.1 Core Model, 2015. URL: <https://www.w3.org/community/odrl/2-1-2/> [zuletzt abgerufen am 10.12.2021].
- [199] Marcello Ienca, Effy Vayena, und Alessandro Blasimme. Big Data and Dementia: Charting the Route Ahead for Research, Ethics, and Policy. *Frontiers in medicine*, 5:13, 2018. DOI: 10.3389/fmed.2018.00013.
- [200] R. Ihl und B. Grass-Kapanke. Neuropsychologisches Test-Screening bei Demenzpatienten. *Neuro-Ger*, (2):21–24, 2005. URL: [https://www.hippocampus.de/media/316/cms\\_4a94ea54c216.pdf](https://www.hippocampus.de/media/316/cms_4a94ea54c216.pdf) [zuletzt abgerufen am 10.12.2021].
- [201] Steve Iliffe, Jill Manthorpe, und Alison Eden. Sooner or later? Issues in the early diagnosis of dementia in general practice: a qualitative study. *Family practice*, 20(4):376–381, 2003. DOI: 10.1093/fampra/cm407.
- [202] Ingo Weber, Qinghua Lu, An Binh Tran, et al. A Platform Architecture for Multi-Tenant Blockchain-Based Systems. *CoRR*, 2019.
- [203] Richard S. Isaacson, Christine A. Ganzer, Hollie Hristov, et al. The clinical practice of risk reduction for Alzheimer’s disease: A precision medicine approach. *Alzheimer’s & dementia : the journal of the Alzheimer’s Association*, 14(12):1663–1673, 2018. DOI: 10.1016/j.jalz.2018.08.004.
- [204] R. J. Ivnik, G. E. Smith, J. A. Lucas, et al. Free and cued selective reminding test: MOANS norms. *Journal of clinical and experimental neuropsychology*, 19(5):676–691, 1997. DOI: 10.1080/01688639708403753.
- [205] J. Lewis und T. Siaw-Liaw. Using information management systems and processes to support shared care for colorectal cancer survivors. In *2017 IEEE International Symposium on Technology and Society (ISTAS)*, pages 1–5, 2017. DOI: 10.1109/ISTAS.2017.8319083.
- [206] J. Ma, C. Peng, und Q. Chen. Health Information Exchange for Home-Based Chronic Disease Self-Management – A Hybrid Cloud Approach. In *2014 5th International Conference on Digital Home*, pages 246–251, 2014. DOI: 10.1109/ICDH.2014.54.

- [207] Jacob L. Jaremko, Marleine Azar, Rebecca Bromwich, et al. Canadian Association of Radiologists White Paper on Ethical and Legal Issues Related to Artificial Intelligence in Radiology. *Canadian Association of Radiologists Journal*, 70(2):107–118, 2019. URL: <http://www.sciencedirect.com/science/article/pii/S0846537119300063>, DOI:10.1016/j.carj.2019.03.001.
- [208] Paul T. Jaeger, Jimmy Lin, und Justin M. Grimes. Cloud Computing and Information Policy: Computing in a Policy Cloud? *Journal of Information Technology & Politics*, 5(3):269–283, 2008. DOI:10.1080/19331680802425479.
- [209] Vishal Jain und Jyotir Moy Chatterjee. *Machine learning with health care perspective: Machine learning and healthcare*, volume volume 13 of *Learning and analytics in intelligent systems*. Springer, Cham, Switzerland, 2020. ISBN: 978-3-030-40849-7. DOI:10.1007/978-3-030-40850-3.
- [210] Gareth James, Daniela Witten, Trevor Hastie, und Robert Tibshirani. *An introduction to statistical learning: With applications in R*, volume 103 of *Springer texts in statistics*. Springer, New York and Heidelberg and Dordrecht and London, corrected at 8th printing edition, 2017. ISBN: 978-1-4614-7137-0. DOI:10.1007/978-1-4614-7138-7.
- [211] Laszlo A. Jeni, Jeffrey F. Cohn, und Fernando de La Torre. Facing Imbalanced Data—Recommendations for the Use of Performance Metrics. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013: 2-5 September 2013, Geneva, Switzerland ; proceedings*. ISBN: 9780769550480, Los Alamitos, 2013. IEEE. DOI:10.1109/acii.2013.47.
- [212] Fei Jiang, Yong Jiang, Hui Zhi, et al. Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology*, 2(4):230–243, 2017. DOI:10.1136/svn-2017-000101.
- [213] Christian Johner, Christoph Molnar, Andreas Purde, et al. Leitfaden zur KI bei Medizinprodukten, 2020. URL: [https://github.com/johner-institut/ai-guideline/blob/master/Guideline-AI-Medical-Devices\\_DE.md](https://github.com/johner-institut/ai-guideline/blob/master/Guideline-AI-Medical-Devices_DE.md) [zuletzt abgerufen am 10.12.2021].
- [214] Kerina H. Jones, Graeme Laurie, Leslie Stevens, et al. The other side of the coin: Harm due to the non-use of health-related data. *International Journal of Medical Informatics*, 97:43–51, 2017. URL: <https://pubmed.ncbi.nlm.nih.gov/27919394/> [zuletzt abgerufen am 10.12.2021], DOI:10.1016/j.ijmedinf.2016.09.010.
- [215] Jerome Jourquin, Stephanie Birkey Reffey, Cheryl Jernigan, et al. Susan G. Komen Big Data for Breast Cancer Initiative: How Patient Advocacy Organizations Can Facilitate Using Big Data to Improve Patient Outcomes. *JCO Precision Oncology*, 3:1–9, 2019. URL: <http://www.redi-bw.de/db/ebSCO.php/search.ebscohost.com/login.aspx?fdirect%3dtrue%26db%3daph%26AN%3d138593254%26site%3dehost-live> [zuletzt abgerufen am 10.12.2021].
- [216] Jürgen Dieber und Sabrina Kirrane. Why model why? Assessing the strengths and limitations of LIME. *CoRR*, abs/2012.00093, 2020.
- [217] K. Azbeg, O. Ouchetto, S. J. Andaloussi, et al. Blockchain and IoT for Security and Privacy: A Platform for Diabetes Self-management. In *2018 4th International Conference on Cloud Computing Technologies and Applications (Cloudtech)*, pages 1–5, 2018. DOI:10.1109/CloudTech.2018.8713343.

- [218] K. T. Pickard. Exploring Markets of Data for Personal Health Information. In *2014 IEEE International Conference on Data Mining Workshop*, pages 477–480, 2014. DOI:10.1109/ICDMW.2014.120.
- [219] Jeffrey P. Kahn, Effy Vayena, und Anna C. Mastroianni. Opinion: Learning as we go: lessons from the publication of Facebook’s social-computing research. *Proceedings of the National Academy of Sciences of the United States of America*, 111(38):13677–13679, 2014. DOI:10.1073/pnas.1416405111.
- [220] E. Kalbe, J. Kessler, P. Calabrese, et al. DemTect: a new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *International journal of geriatric psychiatry*, 19(2):136–143, 2004. DOI:10.1002/gps.1042.
- [221] E. Kalbe, J. Kessler, R. Smith, et al. The DemTect®: a very sensitive screening instrument for mild dementia. *European Psychiatry*, 17:131, 2002. DOI:10.1016/S0924-9338(02)80579-X.
- [222] Dawid Kasprowicz und Stefan Rieger. *Handbuch Virtualität*. Springer reference. Springer VS, Springer Fachmedien Wiesbaden, Wiesbaden, 2020. ISBN: 978-3-658-16341-9. DOI:10.1007/978-3-658-16342-6.
- [223] Kassenärztlichen Bundesvereinigung. Patient - MIO Impfpass, 05.10.2020. URL: <https://mio.kbv.de/display/IM1x0/Patient> [zuletzt abgerufen am 10.12.2021].
- [224] Kassenärztlichen Bundesvereinigung. Der Computer-Führer für Ärzte: Die Schnittstellen der KBV: xDT – Synonym für elektronischen Datenaustausch in der Arztpraxis, 2006. URL: [http://www.medizin-edv.de/ARCHIV/die\\_schnittstellen\\_der\\_kbv\\_o6.pdf](http://www.medizin-edv.de/ARCHIV/die_schnittstellen_der_kbv_o6.pdf) [zuletzt abgerufen am 10.12.2021].
- [225] Gajendra Jung Katuwal und Robert Chen. Machine Learning Model Interpretability for Precision Medicine. URL: <https://arxiv.org/pdf/1610.09045>.
- [226] Reem Kayyali, Aliko Peletidi, Muhammad Ismail, et al. Awareness and Use of mHealth Apps: A Study from England. *Pharmacy (Basel, Switzerland)*, 5(2), 2017. DOI:10.3390/pharmacy5020033.
- [227] J. Kessler, M. Bley, R. Mielke, und E. Kalbe. Strategies and structures in verbal fluency tasks in patients with Alzheimer’s disease. *Behavioural neurology*, 10(4):133–135, 1997. DOI:10.3233/BEN-1997-10406.
- [228] Atif Khan, John A. Doucette, und Robin Cohen. Validation of an ontological medical decision support system for patient treatment using a repository of patient data. *ACM Transactions on Intelligent Systems and Technology*, 4(4):1–31, 2013. DOI:10.1145/2508037.2508049.
- [229] Sabrina Kirrane, Javier D. Fernández, Piero Bonatti, et al. The SPECIAL-K Personal Data Processing Transparency and Compliance Platform, 2020. URL: <https://www.specialprivacy.eu/images/documents/report-D4-1.pdf> [zuletzt abgerufen am 10.12.2021].
- [230] Sabrina Kirrane, Javier D. Fernández, Wouter Dullaert, et al. A Scalable Consent, Transparency and Compliance Architecture. In Aldo Gangemi, Anna Lisa Gentile, Andrea Giovanni Nuzzolese, et al., editors, *The Semantic Web: ESWC 2018 Satellite Events*, volume 11155 of *Lecture notes in computer science*, pages 131–136. Springer International Publishing, Cham, 2018. ISBN: 978-3-319-98191-8. DOI:10.1007/978-3-319-98192-5\backslashtextunderscore.

- [231] Sabrina Kirrane, Uroš Milošević, Javier Fernández, et al. Scalable Policy-aware Linked Data Architecture for privacy, transparency and compliance: Transparency Framework V2: Deliverable D2.7. 2018, 2018. URL: [https://specialprivacy.ercim.eu/images/documents/SPECIAL\\_D27\\_M23\\_V10.pdf](https://specialprivacy.ercim.eu/images/documents/SPECIAL_D27_M23_V10.pdf) [zuletzt abgerufen am 06.12.2021].
- [232] Miia Kivipelto, Tiia Ngandu, Tiina Laatikainen, et al. Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. *The Lancet Neurology*, 5(9):735–741, 2006. DOI: [10.1016/S1474-4422\(06\)70537-3](https://doi.org/10.1016/S1474-4422(06)70537-3).
- [233] C. Klein, J. Hagenah, B. Landwehrmeyer, et al. Das präsymptomatische Stadium neurodegenerativer Erkrankungen. *Der Nervenarzt*, 82(8):994–1001, 2011. DOI: [10.1007/s00115-011-3258-y](https://doi.org/10.1007/s00115-011-3258-y).
- [234] Wolfgang Kohn. *Statistik: Datenanalyse und Wahrscheinlichkeitsrechnung*. SpringerLink Bücher. Springer-Verlag, Berlin/Heidelberg, 2005. ISBN: 9783540267683. URL: <http://swbplus.bsz-bw.de/bsz264347048cov.htm>, DOI: [10.1007/b138042](https://doi.org/10.1007/b138042).
- [235] Mirko Koscina, David Manset, Claudia Negri, und Octavio Perez. Enabling trust in healthcare data exchange with a federated blockchain-based architecture. In Payam Barnaghi, Georg Gottlob, Dimitrios Katsaros, et al., editors, *IEEE/WIC/ACM International Conference on Web Intelligence on - WI '19 Companion*, pages 231–237. ISBN: 9781450369886, New York, New York, USA, 2019. ACM Press. DOI: [10.1145/3358695.3360897](https://doi.org/10.1145/3358695.3360897).
- [236] Lampros C. Kourtis, Oliver B. Regele, Justin M. Wright, und Graham B. Jones. Digital biomarkers for Alzheimer’s disease: the mobile/ wearable devices opportunity. *NPJ digital medicine*, 2, 2019. DOI: [10.1038/s41746-019-0084-2](https://doi.org/10.1038/s41746-019-0084-2).
- [237] KPMG AG. Zulassungsverfahren für Medizinprodukte: Kostentreiber MDR. 2019. URL: <https://assets.kpmg/content/dam/kpmg/de/pdf/Themen/2019/12/2019-mdr-in-mtd-meyer-schmidt-janssen.pdf> [zuletzt abgerufen am 10.12.2021].
- [238] Helmut Krcmar. *Informationsmanagement*. Springer Gabler, Wiesbaden, 6., überarb. aufl. edition, 2015. ISBN: 978-3-662-45862-4. DOI: [10.1007/978-3-662-45863-1](https://doi.org/10.1007/978-3-662-45863-1).
- [239] Jenny Krutzinna und Luciano Floridi, editors. *The Ethics of Medical Data Donation*, volume 137 of *Philosophical Studies Series*. Springer International Publishing, Cham, 2019. ISBN: 9783030043643. DOI: [10.1007/978-3-030-04363-6](https://doi.org/10.1007/978-3-030-04363-6).
- [240] Miroslav Kubat. *An introduction to machine learning*. Springer, Cham, 2015. ISBN: 9783319200095. DOI: [10.1007/978-3-319-20010-1](https://doi.org/10.1007/978-3-319-20010-1).
- [241] Bertolt Kuhn und Volker Amelung. Kapitel 4 - Gesundheits-Apps und besondere Herausforderungen: Chancen und Risiken von Gesundheits-Apps (CHARISMHA) / Chancen und Risiken von Gesundheits-Apps (CHARISMHA). 2016. DOI: [10.24355/DBBS.084-201210110913-58](https://doi.org/10.24355/DBBS.084-201210110913-58).
- [242] Max Kuhn und Kjell Johnson. *Applied predictive modeling*. Springer, New York, corrected 5th printing edition, 2016. ISBN: 9781461468493.
- [243] Thomas Kunz, Benjamin Lange, und Annika Selzer. Datenschutz und Datensicherheit in Digital Public Health. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, 63(2):206–214, 2020. DOI: [10.1007/s00103-019-03083-w](https://doi.org/10.1007/s00103-019-03083-w).

- [244] LUKASZ A. KURGAN und PETR MUSILEK. A survey of Knowledge Discovery and Data Mining process models. *The Knowledge Engineering Review*, 21(1):1–24, 2006. DOI: [10.1017/S0269888906000737](https://doi.org/10.1017/S0269888906000737).
- [245] L. Koumakis, C. Chatzaki, E. Kazantzaki, et al. Dementia Care Frameworks and Assistive Technologies for Their Implementation: A Review. *IEEE Reviews in Biomedical Engineering*, 12:4–18, 2019. DOI: [10.1109/RBME.2019.2892614](https://doi.org/10.1109/RBME.2019.2892614).
- [246] Jean-Baptiste Lamy, Boomadevi Sekar, Gilles Guezennec, et al. Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach. *Artificial Intelligence in Medicine*, 94:42–53, 2019. DOI: [10.1016/j.artmed.2019.01.001](https://doi.org/10.1016/j.artmed.2019.01.001).
- [247] Claire Lancaster, Ivan Koychev, Jasmine Blane, et al. Gallery Game: Smartphone-based assessment of long-term memory in adults at risk of Alzheimer’s disease. *Journal of clinical and experimental neuropsychology*, 42(4):329–343, 2020. DOI: [10.1080/13803395.2020.1714551](https://doi.org/10.1080/13803395.2020.1714551).
- [248] J. R. Landis und G. G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.
- [249] Brett Lantz. *Machine learning with R: Discover how to build machine learning algorithms, prepare data, and dig deep into data prediction techniques with R*. Community experience distilled. Packt Publishing, Birmingham, UK, second edition edition, 2015. ISBN: 9781784393908. URL: <http://proquest.tech.safaribooksonline.de/9781784393908> [zuletzt abgerufen am 10.12.2021].
- [250] Brett Lantz. *Machine learning with R: Expert techniques for predictive modeling*. Expert insight. Packt, Birmingham, third edition edition, 2019. ISBN: 9781788295864.
- [251] Ora Lassila und Ralph Swick. Resource Description Framework (RDF) Model and Syntax Specification, 1999. URL: <https://www.w3.org/TR/PR-rdf-syntax/Overview.html> [zuletzt abgerufen am 10.12.2021].
- [252] A. T. Latal, T. Hinze, N. Roeder, und D. Franz. Aufbau des selbstverwalteten Gesundheitswesens in Deutschland. *Zeitschrift für Herz-,Thorax- und Gefäßchirurgie*, 31(1):54–59, 2017. DOI: [10.1007/s00398-016-0101-6](https://doi.org/10.1007/s00398-016-0101-6).
- [253] Mark Lawler und Tim Maughan. From Rosalind Franklin to Barack Obama: Data Sharing Challenges and Solutions in Genomics and Personalised Medicine. *New Bioethics*, 23(1):64–73, 2017. URL: <http://www.redi-bw.de/db/ebSCO.php/search.ebscohost.com/login.aspx%3fdirect%3dtrue%26db%3daph%26AN%3d123112438%26site%3dehost-live> [zuletzt abgerufen am 10.12.2021].
- [254] Timothy Lebo, Satya Sahoo, und Deborah McGuinness. PROV-O: The PROV Ontology, 2013. URL: <https://www.w3.org/TR/prov-o/> [zuletzt abgerufen am 10.12.2021].
- [255] Mickael Leclercq, Benjamin Vittrant, Marie Laure Martin-Magniette, et al. Large-Scale Automatic Feature Selection for Biomarker Discovery in High-Dimensional OMICs Data. *Frontiers in genetics*, 10:452, 2019. DOI: [10.3389/fgene.2019.00452](https://doi.org/10.3389/fgene.2019.00452).
- [256] Leonoor E.M. Tideman, Lukasz G. Migas, Katerina V. Djambazova, et al. Automated biomarker candidate discovery in imaging mass spectrometry data through spatially localized Shapley additive

- explanations. *Analytica Chimica Acta*, 1177:338522, 2021. URL: <https://www.sciencedirect.com/science/article/pii/S0003267021003482>, DOI:10.1016/j.aca.2021.338522.
- [257] Charley S. Lewis, Jordyn Gunville, Ryan T. Goeckner, et al. Prescription Medication Misuse Among American Indians in the Midwestern US. *Journal of Health Disparities Research & Practice*, 12(1):151–158, 2019. URL: <http://www.redibw.de/db/ebsco.php/search.ebscohost.com/login.aspx%3fdirect%3dtrue%26db%3daph%26AN%3d136144450%26site%3dehost-live>.
- [258] Thomas Leyhe, Charles F. Reynolds, Tobias Melcher, et al. A common challenge in older adults: Classification, overlap, and therapy of depression and dementia. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 13(1):59–71, 2017. DOI:10.1016/j.jalz.2016.08.007.
- [259] Q. Li. *User Modelling for Older Adults' Mobile Interaction Behaviour: Evaluation of User Characteristics, Task Demands and Interface Design*. School of Design, The Hong Kong Polytechnic University, 2019. URL: <https://books.google.de/books?id=yy0CzQEACAAJ> [zuletzt abgerufen am 10.12.2021].
- [260] Yan Li, Changxin Bai, und Chandan K. Reddy. A Distributed Ensemble Approach for Mining Healthcare Data under Privacy Constraints. *Information Sciences*, 330:245–259, 2016. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4677334/pdf/nihms-731200.pdf> [zuletzt abgerufen am 10.12.2021], DOI:10.1016/j.ins.2015.10.011.
- [261] Li Lin, Wei Liang, Chao-Feng Li, et al. Development and implementation of a dynamically updated big data intelligence platform from electronic health records for nasopharyngeal carcinoma research. *British Journal of Radiology*, 92(1102):N.PAG, 2019. URL: <http://www.redibw.de/db/ebsco.php/search.ebscohost.com/login.aspx%3fdirect%3dtrue%26db%3daph%26AN%3d138674258%26site%3dehost-live> [zuletzt abgerufen am 10.12.2021].
- [262] Gill Livingston, Jonathan Huntley, Andrew Sommerlad, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet*, 396(10248):413–446, 2020. DOI:10.1016/S0140-6736(20)30367-6.
- [263] Victoria López, Alberto Fernández, Salvador García, et al. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences*, 250:113–141, 2013. DOI:10.1016/j.ins.2013.07.007.
- [264] Bruno Loureiro und R. Rodrigues. Design Guidelines and Design Recommendations of Multi-Touch Interfaces for Elders. In *ACHI 2014*, 2014.
- [265] Francesca Lunardini, Matteo Luperto, Katia Daniele, et al. Validity of digital Trail Making Test and Bells Test in elderlies. In *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pages 1–4. ISBN: 978-1-7281-0848-3. IEEE, 2019. DOI:10.1109/BHI.2019.8834513.
- [266] Scott M. Lundberg, Gabriel G. Erion, und Su-In Lee. Consistent Individualized Feature Attribution for Tree Ensembles. URL: <https://arxiv.org/pdf/1802.03888.pdf> [zuletzt abgerufen am 12.01.2022].

- [267] Jeantine E. Lunshof, Ruth Chadwick, Daniel B. Vorhaus, und George M. Church. From genetic privacy to open consent. *Nature reviews. Genetics*, 9(5):406–411, 2008. DOI: 10.1038/nrg2360.
- [268] M. A. Al-Tae, W. Al-Nuaimy, Z. J. Muhsin, et al. Mapping Security Requirements of Mobile Health Systems into Software Development Lifecycle. In *2016 9th International Conference on Developments in eSystems Engineering (DeSE)*, pages 87–93, 2016. DOI: 10.1109/DeSE.2016.16.
- [269] M. A. Rahman, M. Rashid, S. Barnes, et al. An IoT and Blockchain-Based Multi-Sensory In-Home Quality of Life Framework for Cancer Patients. In *2019 15th International Wireless Communications Mobile Computing Conference (IWCMC)*, pages 2116–2121, 2019. DOI: 10.1109/IWCMC.2019.8766496.
- [270] M. T. Sqalli und D. Al-Thani. AI-supported Health Coaching Model for Patients with Chronic Diseases. In *2019 16th International Symposium on Wireless Communication Systems (ISWCS)*, pages 452–456, 2019. DOI: 10.1109/ISWCS.2019.8877113.
- [271] Maede Maftouni, I. B. Turksen, M. FazelH. Zarandi, und Faezeh Roshani. Type-2 fuzzy rule-based expert system for Ankylosing spondylitis diagnosis. In *2015 Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS) held jointly with 2015 5th World Conference on Soft Computing (WConSC)*, pages 1–5. ISBN: 978-1-4673-7248-0. IEEE, 17.08.2015 - 19.08.2015. DOI: 10.1109/NAFIPS-WConSC.2015.7284195.
- [272] Nesma Mahmoud und Heba Elbeh. IRS-T2D. In Unknown, editor, *Proceedings of the 10th International Conference on Informatics and Systems - INFOS '16*, pages 203–209. ISBN: 9781450340625, New York, New York, USA, 2016. ACM Press. DOI: 10.1145/2908446.2908495.
- [273] W. Maier und U. B. Barnikol. Neurokognitive Störungen im DSM-5 : Durchgreifende Änderungen in der Demenzdiagnostik. *Der Nervenarzt*, 85(5):564–570, 2014. DOI: 10.1007/s00115-013-3984-4.
- [274] Szymon Maksymiuk, Alicja Gosiewska, und Przemyslaw Biecek. An R wrapper of SHAP python library, 2020. URL: <https://github.com/Modeloriented/shapper> [zuletzt abgerufen am 24.01.2022].
- [275] Milena Soriano Marcolino, João Antonio Queiroz Oliveira, Christiane Corrêa Rodrigues Cimini, et al. Development and Implementation of a Decision Support System to Improve Control of Hypertension and Diabetes in a Resource-Constrained Area in Brazil: Mixed Methods Study. *Journal of medical Internet research*, 23(1):e18872, 2021. DOI: 10.2196/18872.
- [276] Mark Brown und John Brocklebank. Data Mining. 2000. URL: <https://support.sas.com/resources/papers/proceedings/proceedings/sugi22/DATAWARE/PAPER128.PDF> [zuletzt abgerufen am 10.12.2021].
- [277] João Maroco, Dina Silva, Ana Rodrigues, et al. Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC research notes*, 4:299, 2011. DOI: 10.1186/1756-0500-4-299.

- [278] I. Martenstein und A. Wienke. Das neue E-Health-Gesetz : Was kommt auf Kliniken und niedergelassene Ärzte zu? *HNO*, 64(7):515–516, 2016. DOI: [10.1007/s00106-015-0118-2](https://doi.org/10.1007/s00106-015-0118-2).
- [279] Paul Maruff, Yen Ying Lim, David Darby, et al. Clinical utility of the cogstate brief battery in identifying cognitive impairment in mild cognitive impairment and Alzheimer’s disease. *BMC psychology*, 1(1):30, 2013. DOI: [10.1186/2050-7283-1-30](https://doi.org/10.1186/2050-7283-1-30).
- [280] Stephan Matzka. *Künstliche Intelligenz in den Ingenieurwissenschaften*. Springer Fachmedien Wiesbaden, Wiesbaden, 2021. ISBN: 978-3-658-34640-9. DOI: [10.1007/978-3-658-34641-6](https://doi.org/10.1007/978-3-658-34641-6).
- [281] Jennifer A. McDicken, Emma Elliott, Gareth Blayney, et al. Accuracy of the short-form Montreal Cognitive Assessment: Systematic review and validation. *International journal of geriatric psychiatry*, 34(10):1515–1525, 2019. DOI: [10.1002/gps.5162](https://doi.org/10.1002/gps.5162).
- [282] McKinsey & Company. Digitalisierung im Gesundheitswesen: die Chancen für Deutschland. 2018. URL: [https://www.mckinsey.de/~media/mckinsey/locations/europe%20and%20middle%20east/deutschland/news/presse/2018/2018-09-25-digitalisierung%20im%20gesundheitswesen/langfassung%20digitalisierung%20im%20gesundheitswesen\\_\\_neu.ashx](https://www.mckinsey.de/~media/mckinsey/locations/europe%20and%20middle%20east/deutschland/news/presse/2018/2018-09-25-digitalisierung%20im%20gesundheitswesen/langfassung%20digitalisierung%20im%20gesundheitswesen__neu.ashx) [zuletzt abgerufen am 10.12.2021].
- [283] Medical Device Coordination Group. MDCG 2019-11: Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR. 2019. URL: [https://ec.europa.eu/health/sites/health/files/md\\_topics-interest/docs/md\\_mdcg\\_2019\\_11\\_guidance\\_en.pdf](https://ec.europa.eu/health/sites/health/files/md_topics-interest/docs/md_mdcg_2019_11_guidance_en.pdf) [zuletzt abgerufen am 10.12.2021].
- [284] Ursula Meidert, Mandy Scheermesser, Yvonne Prieur, et al. *Quantified Self - Schnittstelle zwischen Lifestyle und Medizin*, volume 67 of *TA-SWISS*. vdf, Zürich, 2018. ISBN: 9783728138927. URL: <https://vdf.ch/quantified-self-schnittstelle-zwischen-lifestyle-und-medizin-e-book.html> [zuletzt abgerufen am 10.12.2021], DOI: [10.3218/3892-7](https://doi.org/10.3218/3892-7).
- [285] Dirk Merkel. Docker: lightweight linux containers for consistent development and deployment. *Linux journal*, 2014(239):2, 2014.
- [286] Charles E. Metz. Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8(4):283–298, 1978. DOI: [10.1016/S0001-2998\(78\)80014-2](https://doi.org/10.1016/S0001-2998(78)80014-2).
- [287] Michelle M. Mielke, Mary M. Machulda, Clinton E. Hagen, et al. Performance of the CogState computerized battery in the Mayo Clinic Study on Aging. *Alzheimer’s & dementia : the journal of the Alzheimer’s Association*, 11(11):1367–1376, 2015. DOI: [10.1016/j.jalz.2015.01.008](https://doi.org/10.1016/j.jalz.2015.01.008).
- [288] Walter Miles. The High Relief Finger Maze for Human Learning. *The Journal of General Psychology*, 1(1):3–14, 1928. DOI: [10.1080/00221309.1928.9923408](https://doi.org/10.1080/00221309.1928.9923408).
- [289] Richard Milne und Carol Brayne. We need to think about data governance for dementia research in a digital era. *Alzheimer’s Research & Therapy*, 12(1):17, 2020. DOI: [10.1186/s13195-020-0584-y](https://doi.org/10.1186/s13195-020-0584-y).
- [290] Panagiota Mistridis, Sabine Krumm, Andreas U. Monsch, et al. The 12 Years Preceding Mild Cognitive Impairment Due to Alzheimer’s Disease: The Temporal Emergence of Cognitive Decline. *Journal of Alzheimer’s disease : JAD*, 48(4):1095–1107, 2015. DOI: [10.3233/JAD-150137](https://doi.org/10.3233/JAD-150137).

- [291] Alex J. Mitchell. A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. *Journal of Psychiatric Research*, 43(4):411–431, 2009. URL: <https://pubmed.ncbi.nlm.nih.gov/18579155/> [zuletzt abgerufen am 10.12.2021], DOI: 10.1016/j.jpsychires.2008.04.014.
- [292] Christoph Molnar. *Interpretable machine learning: A guide for making Black Box Models interpretable*. Lulu, Morisville, North Carolina, 2019. ISBN: 9780244768522.
- [293] Manuel Montero-Odasso, Zahinoor Ismail, und Gill Livingston. One third of dementia cases can be prevented within the next 25 years by tackling risk factors. The case "for" and "against". *Alzheimer's Research & Therapy*, 12(1):81, 2020. DOI: 10.1186/s13195-020-00646-x.
- [294] Gary P. Morris, Ian A. Clark, und Bryce Vissel. Questions concerning the role of amyloid- $\beta$  in the definition, aetiology and diagnosis of Alzheimer's disease. *Acta neuropathologica*, 136(5):663–689, 2018. DOI: 10.1007/s00401-018-1918-8.
- [295] J. C. Morris, A. Heyman, R. C. Mohs, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, 39(9):1159–1165, 1989. URL: <https://pubmed.ncbi.nlm.nih.gov/2771064/> [zuletzt abgerufen am 10.12.2021], DOI: 10.1212/wnl.39.9.1159.
- [296] Kevin Murphy. *Machine Learning - A Probabilistic Perspective*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, 2014. ISBN: 9780262305242. URL: <http://gbv.eblib.com/patron/FullRecord.aspx?p=3339490>.
- [297] Ziad S. Nasreddine, Natalie A. Phillips, Valérie Bédirian, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4):695–699, 2005. DOI: 10.1111/j.1532-5415.2005.53221.x.
- [298] NCBI Bookshelf. Data Standards, 05.10.2020. URL: <https://www.ncbi.nlm.nih.gov/books/NBK551886/> [zuletzt abgerufen am 10.12.2021].
- [299] Craig G. J. Newman, John Hodges, Stephen Pearson, und Rupert F. Noad. The design and implementation of a computer supported assessment of Dementia - ACEmobile. *International Journal of Integrated Care (IJIC)*, 14(8), 2014. DOI: 10.5334/ijic.1784.
- [300] Tiia Ngandu, Jenni Lehtisalo, Alina Solomon, et al. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. *The Lancet*, 385(9984):2255–2263, 2015. DOI: 10.1016/S0140-6736(15)60461-5.
- [301] J. Nielsen. *Usability Engineering*. Interactive Technologies. Elsevier Science, 1994. ISBN: 9780080520292. URL: <https://books.google.de/books?id=DB0owF7LqIQc> [zuletzt abgerufen am 10.12.2021].
- [302] J. Nielsen und R. Molich. Teaching user interface design based on usability engineering. *ACM SIGCHI Bulletin*, 21(1):45–48, 1989. DOI: 10.1145/67880.67885.
- [303] Shahriar Nirjon, Ifat Afrin Emi, Md Abu Sayeed Mondol, et al. MOBI-COG: A Mobile Application for Instant Screening of Dementia Using the Mini-Cog Test. In Wendy Nilsen, Julien Penders, Misha Pavel, und Andrew Rajj, editors, *Proceedings of the Wireless Health 2014 on National*

- Institutes of Health*, pages 1–7. ISBN: 9781450331609, New York, NY, 2014. ACM. DOI: 10.1145/2668883.2668889.
- [304] Svenja Noichl, Nadine Bergner, und Ulrik Schroeder. Apps für Seniorinnen und Senioren anpassen mittels spezifischer Tastaturen und Menüeinstellungen, 2017. DOI: 10.18420/muc2017-ws17-0412.
- [305] Svenja Noichl und Ulrik Schroeder. Icons für Seniorinnen und Senioren – Universell vs. Adaptiv, 2018. DOI: 10.18420/MUC2018-WS11-0359.
- [306] Dorothy Bird Nyswander. A Comparison of the High Relief Finger Maze and the Stylus Maze. *The Journal of General Psychology*, 2(2-3):273–289, 1929. DOI: 10.1080/00221309.1929.9918063.
- [307] Konrad Obermann, Peter Müller, und Stefanie Woerns. Ärzte im Zukunftsmarkt Gesundheit 2017: Transsektorale Zusammenarbeit: Eine deutschlandweite Befragung niedergelassener Ärztinnen und Ärzte. 2017, 2017. URL: [https://www.stiftung-gesundheit.de/pdf/studien/aerzte-im-zukunftsmarkt-gesundheit\\_2017.pdf](https://www.stiftung-gesundheit.de/pdf/studien/aerzte-im-zukunftsmarkt-gesundheit_2017.pdf) [zuletzt abgerufen am 10.12.2021].
- [308] Aisling A. O’Kane, Helena M. Mentis, und Eno Thereska. Non-Static Nature of Patient Consent: Shifting Privacy Perspectives in Health Information Sharing. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work, CSCW ’13*, pages 553–562. ISBN: 9781450313315, New York, NY, USA, 2013. Association for Computing Machinery. DOI: 10.1145/2441776.2441838.
- [309] Aisling Ann O’Kane und Helena Mentis. Sharing Medical Data vs. Health Knowledge in Chronic Illness Care. In *CHI ’12 Extended Abstracts on Human Factors in Computing Systems, CHI EA ’12*, pages 2417–2422. ISBN: 9781450310161, New York, NY, USA, 2012. Association for Computing Machinery. DOI: 10.1145/2212776.2223812.
- [310] Luis Oliva-Felipe, Cristian Barrué, Atia Cortés, et al. Health Recommender System design in the context of CAREGIVERSPRO-MMD Project. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference, PETRA ’18*, pages 462–469. ISBN: 9781450363907, New York, NY, USA, 2018. Association for Computing Machinery. DOI: 10.1145/3197768.3201558.
- [311] Niall O’Mahony, Sean Campbell, Anderson Carvalho, et al. Deep Learning vs. Traditional Computer Vision. 2019. URL: <https://arxiv.org/ftp/arxiv/papers/1910/1910.13796.pdf> [zuletzt abgerufen am 19.02.2021].
- [312] Graziella Orrù, Merylin Monaro, Ciro Conversano, et al. Machine Learning in Psychometrics and Psychological Research. *Frontiers in Psychology*, 10:2970, 2019. DOI: 10.3389/fpsyg.2019.02970.
- [313] Dominik Ose, Ines Baudendistel, Sabrina Pohlmann, et al. Persönliche Patientenakten im Internet. Ein narrativer Review zu Einstellungen, Erwartungen, Nutzung und Effekten. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 122:9–21, 2017. DOI: 10.1016/j.zefq.2017.04.006.
- [314] Boris Otto und Anja Burmann. Europäische Dateninfrastrukturen. *Informatik-Spektrum*, 44(4):283–291, 2021. DOI: 10.1007/s00287-021-01386-4.

- [315] OwnYourData. Semantic Container, 2019. URL: [https://www.ownyourdata.eu/wp-content/uploads/2018/12/SEMCON\\_WhitePaper\\_Jan19.pdf](https://www.ownyourdata.eu/wp-content/uploads/2018/12/SEMCON_WhitePaper_Jan19.pdf) [zuletzt abgerufen am 10.12.2021].
- [316] Silvia Paddock, Hamed Abedtash, Jacqueline Zummo, und Samuel Thomas. Proof-of-concept study: Homomorphically encrypted data can support real-time learning in personalized cancer medicine. *BMC Medical Informatics & Decision Making*, 19(1):N.PAG, 2019. URL: <http://www.redi-bw.de/db/ebsco.php/search.ebscohost.com/login.aspx%3fdirect%3dtrue%26db%3daph%26AN%3d140156062%26site%3dehost-live> [zuletzt abgerufen am 10.12.2021].
- [317] Iam Palatnik de Sousa, Marley Maria Bernardes Rebuzzi Vellasco, und Eduardo Da Costa Silva. Local Interpretable Model-Agnostic Explanations for Classification of Lymph Node Metastases. *Sensors (14248220)*, 19(13), 2019. DOI: [10.3390/s19132969](https://doi.org/10.3390/s19132969).
- [318] Harshvardhan J. Pandit, Axel Polleres, Bert Bos, et al. Creating a Vocabulary for Data Privacy. In Hervé Panetto, Christophe Debruyne, Martin Hepp, et al., editors, *On the Move to Meaningful Internet Systems: OTM 2019 Conferences*, volume 11877 of *Lecture notes in computer science*, pages 714–730. Springer International Publishing, Cham, 2019. ISBN: 978-3-030-33245-7. DOI: [10.1007/978-3-030-33246-4%5Cbackslash%5Ctextunderscore](https://doi.org/10.1007/978-3-030-33246-4%5Cbackslash%5Ctextunderscore).
- [319] Peter K. Panegyres, Renee Berry, und Jennifer Burchell. Early Dementia Screening. *Diagnostics (Basel, Switzerland)*, 6(1), 2016. DOI: [10.3390/diagnostics6010006](https://doi.org/10.3390/diagnostics6010006).
- [320] Kathryn V. Papp, Peter J. Snyder, Paul Maruff, et al. Detecting subtle changes in visuospatial executive function and learning in the amnesic variant of mild cognitive impairment. *PLoS ONE*, 6(7):e21688, 2011. DOI: [10.1371/journal.pone.0021688](https://doi.org/10.1371/journal.pone.0021688).
- [321] Sunil Patil, Lu Hui, Catherine L. Saunders, et al. Public preferences for electronic health data storage, access, and sharing - evidence from a pan-European survey. *Journal of the American Medical Informatics Association*, 23(6):1096–1106, 2016. URL: <http://www.redi-bw.de/db/ebsco.php/search.ebscohost.com/login.aspx%3fdirect%3dtrue%26db%3dlxh%26AN%3d118960390%26site%3dehost-live>.
- [322] Alma B. Pedersen, Ellen M. Mikkelsen, Deirdre Cronin-Fenton, et al. Missing data and multiple imputation in clinical epidemiological research. *Clinical epidemiology*, 9:157–166, 2017. DOI: [10.2147/CLEP.S129785](https://doi.org/10.2147/CLEP.S129785).
- [323] Ken Peffers, Tuure Tuunanen, Marcus A. Rothenberger, und Samir Chatterjee. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3):45–77, 2007. DOI: [10.2753/MIS0742-1222240302](https://doi.org/10.2753/MIS0742-1222240302).
- [324] Cong Peng, Prashant Goswami, und Guohua Bai. A literature review of current technologies on health data integration for patient-centered health management. *Health informatics journal*, 26(3):1926–1951, 2020. DOI: [10.1177/1460458219892387](https://doi.org/10.1177/1460458219892387).
- [325] Junfeng Peng, Kaiqiang Zou, Mi Zhou, et al. An Explainable Artificial Intelligence Framework for the Deterioration Risk Prediction of Hepatitis Patients. *Journal of medical systems*, 45(5):61, 2021. DOI: [10.1007/s10916-021-01736-5](https://doi.org/10.1007/s10916-021-01736-5).
- [326] Robert H. Pietrzak, Paul Maruff, Linda C. Mayes, et al. An examination of the construct validity and factor structure of the Groton Maze Learning Test, a new measure of spatial working memory,

- learning efficiency, and error monitoring. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*, 23(4):433–445, 2008. DOI: 10.1016/j.acn.2008.03.002.
- [327] Thomas Ploug und Søren Holm. Routinisation of informed consent in online health care systems. *International Journal of Medical Informatics*, 84(4):229–236, 2015. URL: <https://pubmed.ncbi.nlm.nih.gov/25636823/> [zuletzt abgerufen am 10.12.2021], DOI: 10.1016/j.ijmedinf.2015.01.003.
- [328] Thomas Ploug und Søren Holm. Meta Consent - A Flexible Solution to the Problem of Secondary Use of Health Data. *Bioethics*, 30(9):721–732, 2016. DOI: 10.1111/bioe.12286.
- [329] Sylvia Polgar, Christian Mychajliw, Franz Wortha, et al. Towards the development of a tablet-based screening for cognitive symptoms of neuropsychiatric disorders. In Interaction, A. Special Interest Group on Computer-HumanC.M., editor, *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 253–260. ISBN: 9781450359412, [Place of publication not identified], 2018. ACM. DOI: 10.1145/3236112.3236147.
- [330] Fabian Prasser, Oliver Kohlbacher, Ulrich Mansmann, et al. Data Integration for Future Medicine (DIFUTURE). *Methods of Information in Medicine*, 57(S 01):e57–e65, 2018. DOI: 10.3414/ME17-02-0022.
- [331] Martin Prince, Adelina Comas-Herrera, Martin Knapp, et al. World Alzheimer Report 2016 - Improving healthcare for people living with dementia: Coverage, quality and costs now and in the future. 2016.
- [332] Martin Prince, Anders Wimo, Maëlen Guerchet, et al. World Alzheimer Report 2015 - The Global Impact of Dementia: An Analysis of Prevalence, Incidence, Cost and Trends. 2015.
- [333] Hans-Ulrich Prokosch, Till Acker, Johannes Bernarding, et al. MIRACUM: Medical Informatics in Research and Care in University Medicine. *Methods of Information in Medicine*, 57(S 01):e82–e91, 2018. DOI: 10.3414/ME17-02-0025.
- [334] Jochen Prümper. Der Benutzungsfragebogen ISONORM 9241/10: Ergebnisse zur Reliabilität und Validität. In Rüdiger Liskowsky, Boris M. Velichkovsky, und Wolfgang Wünschmann, editors, *Software-Ergonomie '97*, volume 49 of *Berichte des German Chapter of the ACM*, pages 253–262. Vieweg+Teubner Verlag, Wiesbaden, 1997. ISBN: 978-3-519-02690-7. DOI: 10.1007/978-3-322-86782-7\backslashbackslash\$textunderline.
- [335] Jochen Prümper und Anft Michael. Die Evaluation von Software auf Grundlage des Entwurfs zur internationalen Ergonomie-Norm ISO 9241 Teil 10 als Beitrag zur partizipativen Systemgestaltung - ein Fallbeispiel. In Karl-Heinz Rödiger, editor, *Software-Ergonomie '93: Von der Benutzungsoberfläche zur Arbeitsgestaltung*, pages 154–156, Stuttgart, 1993. B.G.Teubner.
- [336] Shangran Qiu, Prajakta S. Joshi, Matthew I. Miller, et al. Development and validation of an interpretable deep learning framework for Alzheimer’s disease classification. *Brain : a journal of neurology*, 143(6):1920–1933, 2020. DOI: 10.1093/brain/awaa137.
- [337] John Ross Quinlan. *C4.5 - programs for machine learning*. The Morgan Kaufmann series in machine learning. Kaufmann, San Mateo, Calif., 1993. ISBN: 1-55860-238-0.

- [338] R. Sánchez-Guerrero, F. A. Mendoza, D. Díaz-Sánchez, et al. Collaborative eHealth Meets Security: Privacy-Enhancing Patient Profile Management. *IEEE Journal of Biomedical and Health Informatics*, 21(6):1741–1749, 2017. DOI: [10.1109/JBHI.2017.2655419](https://doi.org/10.1109/JBHI.2017.2655419).
- [339] R. Wirth und J. Hipp. Crisp-dm: towards a standard process modell for data mining. *undefined*, 2000. URL: <https://www.semanticscholar.org/paper/Crisp-dm%3A-towards-a-standard-process-modell-for-Wirth-Hipp/48b9293cfd4297f855867ca278f7069abc6a9c24>.
- [340] Yair G. Rajwan und George R. Kim. Medical Information Visualization Conceptual Model for Patient-Physician Health Communication. In *Proceedings of the 1st ACM International Health Informatics Symposium, IHI '10*, pages 512–516. ISBN: 9781450300308, New York, NY, USA, 2010. Association for Computing Machinery. DOI: [10.1145/1882992.1883074](https://doi.org/10.1145/1882992.1883074).
- [341] Praveen R. Rao, Stanley A. Edlavitch, Jeffrey L. Hackman, et al. Towards Large-Scale Sharing of Electronic Health Records of Cancer Patients. In *Proceedings of the 1st ACM International Health Informatics Symposium, IHI '10*, pages 545–549. ISBN: 9781450300308, New York, NY, USA, 2010. Association for Computing Machinery. DOI: [10.1145/1882992.1883081](https://doi.org/10.1145/1882992.1883081).
- [342] Praveen R. Rao, Tivakar Komara Swami, Deepthi S. Rao, et al. A Software Tool for Large-Scale Sharing and Querying of Clinical Documents Modeled Using HL7 Version 3 Standard. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, IHI '12*, pages 473–482. ISBN: 9781450307819, New York, NY, USA, 2012. Association for Computing Machinery. DOI: [10.1145/2110363.2110417](https://doi.org/10.1145/2110363.2110417).
- [343] Khansa Rasheed, Adnan Qayyum, Mohammed Ghaly, et al. *Explainable, Trustworthy, and Ethical Machine Learning for Healthcare: A Survey*. 2021. DOI: [10.36227/techrxiv.14376179](https://doi.org/10.36227/techrxiv.14376179).
- [344] Christiane Reitz. Toward precision medicine in Alzheimer’s disease. *Annals of translational medicine*, 4(6):107, 2016. DOI: [10.21037/atm.2016.03.05](https://doi.org/10.21037/atm.2016.03.05).
- [345] Marco Tulio Ribeiro, Sameer Singh, und Carlos Guestrin. Model-Agnostic Interpretability of Machine Learning, 2016. URL: <https://arxiv.org/pdf/1606.05386> [zuletzt abgerufen am 10.12.2021].
- [346] Rachel L. Richesson und Jeffrey Krischer. Data standards in clinical research: gaps, overlaps, challenges and future directions. *Journal of the American Medical Informatics Association*, 14(6):687–696, 2007. DOI: [10.1197/jamia.M2470](https://doi.org/10.1197/jamia.M2470).
- [347] Craig W. Ritchie, Tom C. Russ, Sube Banerjee, et al. The Edinburgh Consensus: preparing for the advent of disease-modifying therapies for Alzheimer’s disease. *Alzheimer’s Research & Therapy*, 9(1):85, 2017. DOI: [10.1186/s13195-017-0312-4](https://doi.org/10.1186/s13195-017-0312-4).
- [348] José Matías Rivero, Gustavo Rossi, Julián Grigera, et al. From Mockups to User Interface Models: An Extensible Model Driven Approach. In Florian Daniel und Federico Michele Facca, editors, *Current Trends in Web Engineering*, pages 13–24. ISBN: 978-3-642-16985-4, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.
- [349] Rober Koch-Institut. RKI - Themenschwerpunkt Chronische Erkrankungen, 18.05.2020. URL: [https://www.rki.de/DE/Content/Gesundheitsmonitoring/Themen/Chronische\\_Erkrankungen/Chronische\\_Erkrankungen\\_node.html](https://www.rki.de/DE/Content/Gesundheitsmonitoring/Themen/Chronische_Erkrankungen/Chronische_Erkrankungen_node.html) [zuletzt abgerufen am 10.12.2021].

- [350] Rober Koch-Institut. Corona-Datenspende, 2020. URL: <https://corona-datenspende.de/> [zuletzt abgerufen am 04.02.2022].
- [351] Robert Koch Institut. RKI - Institut, 2021. URL: [https://www.rki.de/DE/Content/Institut/institut\\_node.html](https://www.rki.de/DE/Content/Institut/institut_node.html) [zuletzt abgerufen am 10.12.2021].
- [352] Jessica Robin, John E. Harrison, Liam D. Kaufman, et al. Evaluation of Speech-Based Digital Biomarkers: Review and Recommendations. *Digital biomarkers*, 4(3):99–108, 2020. DOI:10.1159/000510820.
- [353] A. Rocchi, S. Pellegrini, G. Siciliano, und L. Murri. Causative and susceptibility genes for Alzheimer’s disease: a review. *Brain research bulletin*, 61(1):1–24, 2003. DOI:10.1016/s0361-9230(03)00067-4.
- [354] Isabelle Rouleau, David P. Salmon, Nelson Butters, et al. Quantitative and qualitative analyses of clock drawings in Alzheimer’s and Huntington’s disease. *Brain and Cognition*, 18(1):70–87, 1992. DOI:10.1016/0278-2626(92)90112-y.
- [355] David E. Rumelhart, Geoffrey E. Hinton, und Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986. URL: <https://www.nature.com/articles/323533a0> [zuletzt abgerufen am 10.12.2021], DOI:10.1038/323533a0.
- [356] S. Shahbaz, A. Mahmood, und Z. Anwar. SOAD: Securing Oncology EMR by Anonymizing DICOM Images. In *2013 11th International Conference on Frontiers of Information Technology*, pages 125–130, 2013. DOI:10.1109/FIT.2013.30.
- [357] Perminder S. Sachdev, Deborah Blacker, Dan G. Blazer, et al. Classifying neurocognitive disorders: the DSM-5 approach. *Nature reviews. Neurology*, 10(11):634–642, 2014. DOI:10.1038/nrneuro.2014.181.
- [358] Lothar Sachs und Jürgen Hedderich. *Angewandte Statistik: Methodensammlung mit R*. Springer, Dordrecht, 13., aktualisierte und erw. aufl. edition, 2009. ISBN: 9783540889014.
- [359] Yvan Saeys, Iñaki Inza, und Pedro Larrañaga. A review of feature selection techniques in bioinformatics. *Bioinformatics (Oxford, England)*, 23(19):2507–2517, 2007. DOI:10.1093/bioinformatics/btm344.
- [360] Wojciech Samek, Alexander Binder, Grégoire Montavon, et al. Evaluating the visualization of what a Deep Neural Network has learned. URL: <http://arxiv.org/pdf/1509.06321v1> [zuletzt abgerufen am 17.01.2022].
- [361] Ai Sano. Touchscreen Smartphone User Interfaces for Older Adults. 2017. URL: <https://digitalcommons.morris.umn.edu/cgi/viewcontent.cgi?article=1048&context=horizons> [zuletzt abgerufen am 10.12.2021].
- [362] Emily Nina Satinsky, Corine Driessens, David Crepez-Keay, und Antonis Kousoulis. Mental health service users’ perceptions of data sharing and data protection: a qualitative report. *Journal of innovation in health informatics*, 25(4):239–242, 2018. DOI:10.14236/jhi.v25i4.1033.
- [363] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2008.

- [364] W. Satzger, H. Hampel, F. Padberg, et al. Zur praktischen Anwendung der CERAD-Testbatterie als neuropsychologisches Demenzscreening. *Der Nervenarzt*, 72(3):196–203, 2001. DOI: 10.1007/s001150050739.
- [365] Matthieu-P. Schapranow, Janos Brauer, und Hasso Plattner. The Data Donation Pass: Enabling Sovereign Control of Personal Healthcare Data. In Hamid R. Arabnia, Leonidas Deligiannidis, und Michael B. O’Hara, editors, *HIMS 2017: Proceedings of the 2017 International Conference on Health Informatics and Medical Systems : CSCE ’17, July 17-20, 2017, Las Vegas, Nevada, USA*. CSREA Press, United States of America, 2017. ISBN: 9781601324597.
- [366] Douglas W. Scharre, Shu-Ing Chang, Robert A. Murden, et al. Self-administered Gerocognitive Examination (SAGE): a brief cognitive assessment Instrument for mild cognitive impairment (MCI) and early dementia. *Alzheimer disease and associated disorders*, 24(1):64–71, 2010. DOI: 10.1097/WAD.0b013e3181b03277.
- [367] Douglas W. Scharre, Shu Ing Chang, Haikady N. Nagaraja, et al. Digitally translated Self-Administered Gerocognitive Examination (eSAGE): relationship with its validated paper version, neuropsychological evaluations, and clinical assessments. *Alzheimer’s Research & Therapy*, 9(1):44, 2017. DOI: 10.1186/s13195-017-0269-3.
- [368] Thijs Scheepers. Virtualization and Containerization of ApplicationInfrastructure: A Comparison. 2014. URL: <https://thijs.ai/papers/scheepers-virtualization-containerization.pdf> [zuletzt abgerufen am 10.12.2021].
- [369] Lon Schneider. A resurrection of aducanumab for Alzheimer’s disease. *The Lancet Neurology*, 19(2):111–112, 2020. DOI: 10.1016/S1474-4422(19)30480-6.
- [370] Michael L. Scott. *Programming Language Pragmatics*. Elsevier textbooks, s.l., 2. Aufl. edition, 2005. ISBN: 978-0-12-633951-2. URL: <http://gbv.ebib.com/patron/FullRecord.aspx?p=649018> [zuletzt abgerufen am 10.12.2021].
- [371] Scott M. Lundberg und Su-In Lee. A Unified Approach to Interpreting Model Predictions. 2017. URL: <https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf> [zuletzt abgerufen am 09.01.2022].
- [372] Sebastian Horn. DST - Dementia Screening Test, Alzheimer Test - Apps on Google Play, 2021. URL: <https://play.google.com/store/apps/details?id=com.dementiascreeningtest>.
- [373] Sebastian Horn. DST - Demenz Screening Test, 2021. URL: <https://www.demenz-test.com/> [zuletzt abgerufen am 10.12.2021].
- [374] Sebastian C. Semler, Frank Wissing, und Ralf Heyder. German Medical Informatics Initiative. *Methods of Information in Medicine*, 57(S 01):e50–e56, 2018. DOI: 10.3414/ME18-03-0003.
- [375] Jeff Sevigny, Ping Chiao, Thierry Bussière, et al. The antibody aducanumab reduces A $\beta$  plaques in Alzheimer’s disease. *Nature*, 537(7618):50–56, 2016. DOI: 10.1038/nature19323.
- [376] N. Shadbolt, T. Berners-Lee, und W. Hall. The Semantic Web Revisited. *IEEE Intelligent Systems*, 21(3):96–101, 2006. DOI: 10.1109/MIS.2006.62.

- [377] Sharath M. Shankaranarayana und Davor Runje. ALIME: Autoencoder Based Approach for Local Interpretability. URL: <https://arxiv.org/pdf/1909.02437v1.pdf> [zuletzt abgerufen am 17.01.2022].
- [378] Lloyd S. Shapley. *A Value for N-Person Games*. RAND Corporation, 1952. DOI: 10.7249/P0295.
- [379] Armin Sharafi. *Knowledge Discovery in Databases: Eine Analyse des Änderungsmanagements in der Produktentwicklung: Zugl.: München, Techn. Univ., Diss., 2012*. Springer Gabler Research. Springer Gabler, Wiesbaden, 2013. ISBN: 3658020016.
- [380] Colin Shearer. The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*, 5(4), 2000.
- [381] I. Sim, P. Gorman, R. A. Greenes, et al. Clinical decision support systems for the practice of evidence-based medicine. *Journal of the American Medical Informatics Association*, 8(6):527–534, 2001. DOI: 10.1136/jamia.2001.0080527.
- [382] Chanda Simfukwe, Young Chul Youn, Sang Yun Kim, und Seong Soo An. Digital trail making test-black and white: Normal vs MCI. *Applied neuropsychology. Adult*, pages 1–7, 2021. DOI: 10.1080/23279095.2021.1871615.
- [383] Shireen Sindi, Elisabeth Calov, Jasmine Fokkens, et al. The CAIDE Dementia Risk Score App: The development of an evidence-based mobile application to predict the risk of dementia. *Alzheimer's & dementia (Amsterdam, Netherlands)*, 1(3):328–333, 2015. DOI: 10.1016/j.dadm.2015.06.005.
- [384] Amitojdeep Singh, Sourya Sengupta, und Vasudevan Lakshminarayanan. Explainable deep learning models in medical image analysis. URL: <https://arxiv.org/pdf/2005.13799>.
- [385] T. G. Smith, M. E. Dunn, K. Y. Levin, et al. Cancer survivor perspectives on sharing patient-generated health data with central cancer registries. *Quality of Life Research*, 28(11):2957–2967, 2019. URL: <http://www.redi-bw.de/db/ebsco.php/search.ebscohost.com/login.aspx?direct=true&db%3daph%26AN%3d139255538%26site%3dehost-live>.
- [386] Peter J. Snyder, Martin M. Bednar, Jennifer R. Cromer, und Paul Maruff. Reversal of scopolamine-induced deficits with a single dose of donepezil, an acetylcholinesterase inhibitor. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 1(2):126–135, 2005. DOI: 10.1016/j.jalz.2005.09.004.
- [387] William Souillard-Mandar, Randall Davis, Cynthia Rudin, et al. Learning Classification Models of Cognitive Conditions from Subtle Behaviors in the Digital Clock Drawing Test. *Machine learning*, 102(3):393–441, 2016. DOI: 10.1007/s10994-015-5529-5.
- [388] William Souillard-Mandar, Randall Davis, Cynthia Rudin, et al. Interpretable Machine Learning Models for the Digital Clock Drawing Test, 2016.
- [389] Ton Spil und Richard Klein. Personal Health Records Success: Why Google Health Failed and What Does that Mean for Microsoft HealthVault? In *IEEE 8th International Symposium on Service-Oriented System Engineering (SOSE), 2014: 7 - 11 April 2014, Oxford, United Kingdom ; [including workshop/symposium papers*, pages 2818–2827. ISBN: 978-1-4799-2504-9, Piscataway, NJ, 2014. IEEE. DOI: 10.1109/HICSS.2014.353.

- [390] Statista. Gesetzliche Krankenkassen: Größte Kassen nach Anzahl der Versicherten 2020, 2020. URL: <https://de.statista.com/statistik/daten/studie/856392/umfrage/groesste-gesetzliche-krankenkassen-in-deutschland-nach-der-versichertenzahl/> [zuletzt abgerufen am 10.12.2021].
- [391] Statistisches Bundesamt. Herz-Kreislauf-Erkrankungen verursachen die höchsten Kosten, 2017. URL: [https://www.destatis.de/DE/Presse/Pressemitteilungen/2017/09/PD17\\_347\\_236.html](https://www.destatis.de/DE/Presse/Pressemitteilungen/2017/09/PD17_347_236.html) [zuletzt abgerufen am 04.02.2022].
- [392] Statistisches Bundesamt Deutschland. Statistisches Bundesamt Deutschland - GENESIS-Online: Ergebnis 23631-0003, 2015. URL: <https://www-genesis.destatis.de/genesis/online?operation=previous&levelindex=1&step=1&titel=Ergebnis&levelid=1600682522701&acceptscookies=false#abreadcrumb> [zuletzt abgerufen am 10.12.2021].
- [393] Stephanie Wykstra. Funder Data-Sharing Policies: Overview and Recommendations, 2017.
- [394] Henri-Corto Stoeklé, Marie-France Mamzer-Bruneel, Guillaume Vogt, und Christian Hervé. 23andMe: a new two-sided data-banking market model. *BMC medical ethics*, 17:19, 2016. DOI: 10.1186/s12910-016-0101-9.
- [395] Joella E. Storey, Jeffrey T. J. Rowland, David Basic, et al. The Rowland Universal Dementia Assessment Scale (RUDAS): a multicultural cognitive assessment scale. *International psychogeriatrics*, 16(1):13–31, 2004. DOI: 10.1017/s1041610204000043.
- [396] Veronika Strotbaum, Monika Pobiruchin, Björn Schreiweis, et al. Your data is gold – Data donation for better healthcare? *it - Information Technology*, 61, 2019. DOI: 10.1515/itit-2019-0024.
- [397] David Stück, Alessio Signorini, Tuka Al Hanai, et al. P2-267: Novel Digital Voice Biomarkers of Dementia from the Framingham Study. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 14(7S\_Part\_14):P778–P779, 2018. DOI: 10.1016/j.jalz.2018.06.956.
- [398] YANMIN SUN, ANDREW K. C. WONG, und MOHAMED S. KAMEL. Classification of imbalanced Data: A REVIEW. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(04):687–719, 2009. DOI: 10.1142/S0218001409007326.
- [399] Ali Sunyaev und Dmitry Chorny. Supporting chronic disease care quality: Design and implementation of a health service and its integration with electronic health records. *Journal of Data and Information Quality*, 3, 2012. DOI: 10.1145/2184442.2184443.
- [400] Alisa Surkis und Kevin Read. Research data management. *Journal of the Medical Library Association : JMLA*, 103(3):154–156, 2015. DOI: 10.3163/1536-5050.103.3.011.
- [401] Fadi Thabtah, Ella Mampusti, David Peebles, et al. A Mobile-Based Screening System for Data Analyses of Early Dementia Traits Detection. *Journal of medical systems*, 44(1):24, 2019. DOI: 10.1007/s10916-019-1469-0.
- [402] Fadi Thabtah, Robinson Spencer, und Yongsheng Ye. The correlation of everyday cognition test scores and the progression of Alzheimer's disease: a data analytics study. *Health information science and systems*, 8(1):24, 2020. DOI: 10.1007/s13755-020-00114-8.

- [403] Thomas Lin Pedersen. *Local Interpretable Model-Agnostic Explanations: [R package lime version 0.5.2]*. Comprehensive R Archive Network (CRAN), 2021. URL: <https://cran.r-project.org/web/packages/lime/index.html>.
- [404] Sylvia Tippmann. Programming tools: Adventures with R. *Nature*, 517(7532):109–110, 2015. URL: <https://www.nature.com/news/programming-tools-adventures-with-r-1.16609> [zuletzt abgerufen am 10.12.2021], DOI: 10.1038/517109a.
- [405] T. Tombaugh. Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, 19(2):203–214, 2004. DOI: 10.1016/S0887-6177(03)00039-8.
- [406] Kat Toups, Ann Hathaway, Deborah Gordon, et al. *Precision Medicine Approach to Alzheimer’s Disease: Successful Proof-of-Concept Trial*. 2021. DOI: 10.1101/2021.05.10.21256982.
- [407] Po-Yu Tseng, Yi-Ting Chen, Chuen-Heng Wang, et al. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Critical Care*, 24(1):478, 2020. URL: <https://ccforum.biomedcentral.com/articles/10.1186/s13054-020-03179-9>, DOI: 10.1186/s13054-020-03179-9.
- [408] Usama Fayyad, Gregory Piatetsky-Shapiro, und Padhraic Smyth. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3):37, 1996. URL: <https://ojs.aaai.org//index.php/aimagazine/article/view/1230> [zuletzt abgerufen am 10.12.2021], DOI: 10.1609/aimag.v17i3.1230.
- [409] Jaideep Vaidya, Murat Kantarcioglu, und Chris Clifton. Privacy-preserving Naïve Bayes classification. *The VLDB Journal*, 17(4):879–898, 2008. DOI: 10.1007/s00778-006-0041-y.
- [410] Lucy van Kleunen und Stephen Volda. Challenges in Supporting Social Practices around Personal Data for Long-Term Mental Health Management. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers, UbiComp/ISWC ’19 Adjunct*, pages 944–948. ISBN: 9781450368698, New York, NY, USA, 2019. Association for Computing Machinery. DOI: 10.1145/3341162.3346273.
- [411] Maarten W. van Someren, Yvonne F. Barnard, und Jacobijn A. C. Sandberg. *The think aloud method: A practical guide to modelling cognitive processes*. Academic Press, London, 1994. ISBN: 0127142703.
- [412] José Viña und Jorge Sanz-Ros. Alzheimer’s disease: Only prevention makes sense. *European journal of clinical investigation*, 48(10):e13005, 2018. DOI: 10.1111/eci.13005.
- [413] Giorgio Visani, Enrico Bagli, Federico Chesani, et al. Statistical stability indices for LIME: obtaining reliable explanations for Machine Learning models. URL: <https://arxiv.org/pdf/2001.11757v2.pdf> [zuletzt abgerufen am 17.01.2022].
- [414] Silia Vitoratou und Andrew Pickles. A note on contemporary psychometrics. *Journal of mental health (Abingdon, England)*, 26(6):486–488, 2017. DOI: 10.1080/09638237.2017.1392008.
- [415] W3C OWL Working Group. OWL 2 Web Ontology Language Document Overview: (Second Edition), 2012. URL: <https://www.w3.org/TR/owl2-overview/> [zuletzt abgerufen am 10.12.2021].

- [416] Guanjin Wang, Zhaohong Deng, und Kup-Sze Choi. Detection of epilepsy with Electroencephalogram using rule-based classifiers. *Neurocomputing*, 228:283–290, 2017. DOI: 10.1016/j.neucom.2016.09.080.
- [417] David Wendler. Broad versus blanket consent for research with human biological samples. *The Hastings Center report*, 43(5):3–4, 2013. DOI: 10.1002/hast.200.
- [418] Daniel Wight, Erica Wimbush, Ruth Jepson, und Lawrence Doi. Six steps in quality intervention development (6SQuID). *Journal of epidemiology and community health*, 70(5):520–525, 2016. DOI: 10.1136/jech-2015-205952.
- [419] G. A. Wildenbos, L. W. Peute, und M. W. M. Jaspers. A framework for evaluating mHealth tools for Older Patients on Usability. *Studies in health technology and informatics*, 210:783–787, 2015.
- [420] Mark D. Wilkinson, Michel Dumontier, I. Jsbrand Jan Aalbersberg, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3:160018, 2016. DOI: 10.1038/sdata.2016.18.
- [421] William J. Gordon und Christian Catalini. Blockchain Technology for Healthcare: Facilitating the Transition to Patient-Driven Interoperability. *Computational and Structural Biotechnology Journal*, 16:224–230, 2018. URL: <http://www.sciencedirect.com/science/article/pii/S200103701830028X>, DOI: 10.1016/j.csbj.2018.06.003.
- [422] Betsy White Williams, Wendy Mack, und Victor W. Henderson. Boston naming test in Alzheimer’s disease. *Neuropsychologia*, 27(8):1073–1079, 1989. DOI: 10.1016/0028-3932(89)90186-3.
- [423] Jens Wiltfang, Sarah Trost, und Harald-Jürgen Hampel. Demenz. In Hans-Jürgen Möller, Gerd Laux, und Hans-Peter Kapfhammer, editors, *Psychiatrie, Psychosomatik, Psychotherapie: Band 1: Allgemeine Psychiatrie, Band 2: Spezielle Psychiatrie*, Springer Reference Medizin, pages 1–89. Springer, Berlin and Heidelberg, 2016. ISBN: 978-3-642-45028-0. DOI: 10.1007/978-3-642-45028-0\backslash\$textunderscore.
- [424] B. Winblad, K. Palmer, M. Kivipelto, et al. Mild cognitive impairment—beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of internal medicine*, 256(3):240–246, 2004. DOI: 10.1111/j.1365-2796.2004.01380.x.
- [425] Alfred Winter, Sebastian Stäubert, Danny Ammon, et al. Smart Medical Information Technology for Healthcare (SMITH). *Methods of Information in Medicine*, 57(S 01):e92–e105, 2018. DOI: 10.3414/ME18-02-0004.
- [426] Ian H. Witten, Eibe Frank, und Mark A. Hall. *Data mining: Practical machine learning tools and techniques*. /Morgan Kaufmann series in data management systems]. Elsevier/Morgan Kaufmann, Amsterdam, 3. ed. edition, 2011. ISBN: 0080890369. URL: <http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10525052> [zuletzt abgerufen am 10.12.2021].
- [427] Thomas Wolf und Jacqueline-Helena Strohschen. Digitalisierung: Definition und Reife. *Informatik-Spektrum*, 41(1):56–64, 2018. URL: <https://link.springer.com/content/pdf/10.1007/s00287-017-1084-8.pdf> [zuletzt abgerufen am 10.12.2021], DOI: 10.1007/s00287-017-1084-8.

- [428] KongFatt Wong-Lin, Paula L. McClean, Niamh McCombe, et al. Shaping a data-driven era in dementia care pathway through computational neurology approaches. *BMC Medicine*, 18(1):398, 2020. DOI: [10.1186/s12916-020-01841-1](https://doi.org/10.1186/s12916-020-01841-1).
- [429] World Bank Country und Lending Groups. World Bank Data Help Desk, 2021. URL: <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519> [zuletzt abgerufen am 10.12.2021].
- [430] World Health Organization. ICD-11 - International Classification of Diseases for Mortality and Morbidity Statistics: The global standard for diagnostic health information, 2018. URL: <https://icd.who.int/en> [zuletzt abgerufen am 10.12.2021].
- [431] World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders: Clinical descriptions and diagnostic guidelines, 2019. URL: <https://www.who.int/classifications/icd/icdonlineversions/en/> [zuletzt abgerufen am 10.12.2021].
- [432] Hans Wouters, Aeilko H. Zwinderman, Willem A. van Gool, et al. Adaptive cognitive testing in dementia. *International journal of methods in psychiatric research*, 18(2):118–127, 2009. DOI: [10.1002/mpr.283](https://doi.org/10.1002/mpr.283).
- [433] X. Zheng, R. R. Mukkamala, R. Vatrappu, und J. Ordieres-Mere. Blockchain-based Personal Health Data Sharing System Using Cloud Storage. In *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pages 1–6, 2018. DOI: [10.1109/HealthCom.2018.8531125](https://doi.org/10.1109/HealthCom.2018.8531125).
- [434] Jianhua Xu. An extended one-versus-rest support vector machine for multi-label classification. *Neurocomputing*, 74(17):3114–3124, 2011. DOI: [10.1016/j.neucom.2011.04.024](https://doi.org/10.1016/j.neucom.2011.04.024).
- [435] Xiwei Xu, Cesare Pautasso, Liming Zhu, et al. A Pattern Collection for Blockchain-based Applications. In Unknown, editor, *EuroPLoP 2018: Proceedings of the 23rd European Conference on Pattern Languages of Programs : 4th-8th July 2018, Irsee, Germany*, ICPS: ACM international conference proceeding series, pages 1–20. ISBN: 9781450363877, New York, New York, 2018. The Association for Computing Machinery. DOI: [10.1145/3282308.3282312](https://doi.org/10.1145/3282308.3282312).
- [436] Y. Goletsis, T. P. Exarchos, N. Giannakeas, und D. I. Fotiadis. Intelligent patient profiling for diagnosis, staging and treatment selection in colon cancer. In *2008 8th IEEE International Conference on BioInformatics and BioEngineering*, pages 1–6, 2008. DOI: [10.1109/BIBE.2008.4696785](https://doi.org/10.1109/BIBE.2008.4696785).
- [437] Y. Wang, P. Li, Y. Tian, et al. A Shared Decision-Making System for Diabetes Medication Choice Utilizing Electronic Health Record Data. *IEEE Journal of Biomedical and Health Informatics*, 21(5):1280–1287, 2017. DOI: [10.1109/JBHI.2016.2614991](https://doi.org/10.1109/JBHI.2016.2614991).
- [438] Yan Hu, Cong Peng, und Guohua Bai. Sharing health data through hybrid cloud for self-management. In *2015 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, pages 1–6, 2015. DOI: [10.1109/ICMEW.2015.7169752](https://doi.org/10.1109/ICMEW.2015.7169752).
- [439] Konstantina G. Yiannopoulou und Sokratis G. Papageorgiou. Current and Future Treatments in Alzheimer Disease: An Update. *Journal of central nervous system disease*, 12:1179573520907397, 2020. DOI: [10.1177/1179573520907397](https://doi.org/10.1177/1179573520907397).

- 
- [440] Ziming Yin, Yinhong Zhao, Xudong Lu, und Huilong Duan. A hybrid intelligent diagnosis approach for quick screening of Alzheimer's disease based on multiple neuropsychological rating scales. *Computational and mathematical methods in medicine*, 2015:258761, 2015. DOI:10.1155/2015/258761.
- [441] Vahab Yousofzadeh, Bernadette McGuinness, Liam P. Maguire, und KongFatt Wong-Lin. Multi-Kernel Learning with Dartel Improves Combined MRI-PET Classification of Alzheimer's Disease in AIBL Data: Group and Individual Analyses. *Frontiers in human neuroscience*, 11:380, 2017. DOI:10.3389/fnhum.2017.00380.
- [442] Panayiotis Zaphiris, Sri Kurniawan, und Mariya Ghiawadwala. A systematic approach to the development of research-based web design guidelines for older people. *Universal Access in the Information Society*, 6(1):59–75, 2007. DOI:10.1007/s10209-006-0054-8.
- [443] Fubao Zhu, Xiaonan Li, Haipeng Tang, et al. Machine Learning for the Preliminary Diagnosis of Dementia. *Scientific Programming*, 2020:1–10, 2020. DOI:10.1155/2020/5629090.
- [444] Gokhan Zorluoglu, Mustafa E. Kamasak, Leyla Tavacioglu, und Pinar O. Ozanar. A mobile application for cognitive screening of dementia. *Computer methods and programs in biomedicine*, 118(2):252–262, 2015. DOI:10.1016/j.cmpb.2014.11.004.



# Abkürzungsverzeichnis

<b>AD</b>	Alzheimer Demenz	<b>CDS</b>	Clinical Decision Support
<b>Ada-WHIPS</b>	Adaptive-Weighted High Importance Path Snippets	<b>CERAD</b>	Consortium to Establish a Registry for Alzheimer’s Disease
<b>ADL</b>	Activities of Daily Living	<b>CHIRPS</b>	Collection of High Importance Random Path Snippets
<b>ANN</b>	Artificial Neural Networks	<b>CNN</b>	Convolutional Neural Network
<b>ANOVA</b>	Analysis of Variance	<b>CRISP-DM</b>	Cross-industry Standard Process for Data Mining
<b>API</b>	Application Programming Interface	<b>CT</b>	Computertomographie
<b>App</b>	Applikation	<b>CV</b>	Computer Vision
<b>AUC</b>	Area Under the Curve	<b>D</b>	dement
<b>BAGP</b>	Bundesarbeitsgemeinschaft der PatientInnenstellen	<b>dApps</b>	Decentralized Applications
<b>BDSG</b>	Bundesdatenschutzgesetz	<b>DAS</b>	Deutsche Arbeitsgemeinschaft Selbsthilfegruppen e.V.
<b>BfArM</b>	Bundesinstitut für Arzneimittel und Medizinprodukte	<b>DaTraV</b>	Datentransparenz-Verordnung
<b>BGB</b>	Bürgerlichen Gesetzbuch	<b>DBR</b>	Deutscher Behindertenrat
<b>BGH</b>	Bundesgerichtshof	<b>DHCQ</b>	Digital-Health-Compliance-Questionnaire
<b>BMBF</b>	Bundesministerium für Bildung und Forschung	<b>DNN</b>	Deep Neural Network
<b>BMG</b>	Bundesministerium für Gesundheit	<b>DGPPN</b>	Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde
<b>BNT</b>	Boston Naming Test	<b>DiGA</b>	Digitale Gesundheitsanwendung
<b>BPMN</b>	Business Process Model and Notation	<b>DiGAV</b>	Digitale-Gesundheitsanwendungen-Verordnung
<b>BSN</b>	Body Sensor Network	<b>DiGAVZ</b>	Verzeichnis für Digitale Gesundheitsanwendungen
<b>CARET</b>	Classification And REgression Training	<b>DICOM</b>	Digital Imaging and Communications in Medicine
<b>CART</b>	Classification And Regression Tree		
<b>CDR</b>	Clinical Dementia Rating		

<b>DIMDI</b>	Deutsches Institut für Medizinische Dokumentation und Information	<b>EN</b>	Europäische Norm
<b>DIN</b>	Deutsches Institut für Normung	<b>ePA</b>	Elektronische Patientenakte
<b>DiPA</b>	Digitale Pflegeanwendung	<b>ePF</b>	Elektronisches Patientenfach
<b>DiPAV</b>	Digitale-Pflegeanwendungen-Verordnung	<b>EU</b>	Europäische Union
<b>DKG</b>	Deutsche Krankenhausgesellschaft	<b>EUDAMED</b>	European Database on Medical Devices
<b>DL</b>	Deep Learning	<b>EUSTM</b>	European Society for Translational Medicine
<b>DLT</b>	Distributed Ledger Technology	<b>FA</b>	funktionalen Anforderungen
<b>DM</b>	Data Mining	<b>FAQ</b>	Frequently Asked Questions
<b>DPVCG</b>	W3C Data Privacy Vocabularies and Controls Community Group	<b>FAIR</b>	Findability, Accessibility, Interoperability, and Reusability
<b>DPV</b>	Data Privacy Vocabulary	<b>FCSRT</b>	Free and Cued Selective Reminding Test
<b>DSGVO</b>	Datenschutz-Grundverordnung	<b>FDA</b>	Food and Drug Administration
<b>DSM</b>	Diagnostic and Statistical Manual of Mental Disorders	<b>FFS</b>	Filter Feature Selection
<b>DSRM</b>	Design Science Research Methodology	<b>FHIR</b>	Fast Healthcare Interoperability Resources
<b>DVG</b>	Digitale-Versorgung-Gesetz	<b>FN</b>	Falsch-Negativ
<b>DVPMG</b>	Digitale-Versorgung-und-Pflege-Modernisierungs-Gesetz	<b>FNR</b>	Falsch-Negativ-Rate
<b>eAA</b>	elektronische Arztakte	<b>FP</b>	Falsch-Positiv
<b>EEG</b>	Elektroenzephalografie	<b>FPM</b>	Frequent Pattern Mining
<b>eEPA</b>	einrichtungsübergreifende elektronische Patientenakte	<b>FPR</b>	Falsch-Positiv-Rate
<b>eFA</b>	elektronische Fallakte	<b>G</b>	gesund
<b>EFS</b>	Embedded Feature Selection	<b>G-BA</b>	Gemeinsame Bundesausschuss
<b>eGA</b>	elektronische Gesundheitsakte	<b>gematik</b>	Gesellschaft für Telematikanwendungen der Gesundheitskarte mbH
<b>eGK</b>	elektronischen Gesundheitskarte	<b>GDS</b>	Geriatrische Depressions-Skala
<b>eHA</b>	elektronischen Heilberufsausweis	<b>GKV</b>	Gesetzliche Krankenversicherung
<b>eHealth</b>	Electronic Health	<b>GmbH</b>	Gesellschaft mit beschränkter Haftung
<b>EHR</b>	Electronic Health Record	<b>SGKV</b>	Spitzenverband Bund der Krankenkassen
<b>EMR</b>	Electronic Medical Record	<b>GMLT</b>	Groton Maze Learning Test
		<b>GUI</b>	grafische Benutzerschnittstelle

<b>HGDF</b>	Helmholtz-Gemeinschaft Deutscher Forschungszentren e.V.	<b>KI</b>	Künstliche Intelligenz
<b>HIMSS</b>	Healthcare Information and Management Systems Society	<b>KIS</b>	Klinisches Informationssystem
<b>HIRPS</b>	High Importance Random Path Snippets	<b>KIT</b>	Karlsruher Institut für Technologie
<b>HL7</b>	Health Level Seven	<b>k-NN</b>	k-nearest-neighbours- Algorithmus
<b>HTTP</b>	Hypertext Transfer Protocol	<b>KV</b>	Kassenärztliche Vereinigung
<b>HPMLT</b>	Hidden Pathway Maze Learning Task	<b>KWT</b>	Kruskal-Wallis-Test
<b>ICD</b>	International Statistical Classification of Diseases and Related Health Problems	<b>KZBV</b>	Kassenzahnärztliche Bundesvereinigung
<b>ID</b>	Identität	<b>LE</b>	Leistungserbringer
<b>IEEE</b>	Institute of Electrical and Electronics Engineers	<b>LIME</b>	Local Interpretable Model-agnostic Explanations
<b>IEC</b>	International Electrotechnical Commission	<b>LOINC</b>	Logical Observation Identifiers Names and Codes
<b>IGeL</b>	Individuelle Gesundheitsleistung	<b>LOOCV</b>	Leave-One-Out- Kreuzvalidierung
<b>IHE</b>	Integrating the Healthcare Enterprise	<b>MCI</b>	Mild Cognitive Impairments
<b>IKT</b>	Informations- und Kommunikationstechnologie	<b>MDR</b>	Medical Device Regulation
<b>IMDRF</b>	International Medical Device Regulators Forum	<b>MDSW</b>	Medical Device Software
<b>IML</b>	Interpretable Machine Learning	<b>mHealth</b>	Mobile Health
<b>IoT</b>	Internet of Things	<b>MHMD</b>	MyHealthMyData
<b>ISO</b>	International Organization for Standardization	<b>MII</b>	Medizininformatik-Initiative
<b>IT</b>	Informationstechnologie	<b>MIO</b>	Medizinische Informationsobjekte
<b>IVDR</b>	In-Vitro Diagnostic Medical Devices Regulation	<b>ML</b>	Maschinelles Lernen
<b>JSON</b>	JavaScript Object Notation	<b>MMSE</b>	Mini-Mental State Examination
<b>KBV</b>	Kassenärztliche Bundesvereinigung	<b>MoCA</b>	Montreal Cognitive Assessment
<b>KDD</b>	Knowledge Discovery in Databases	<b>MoRF</b>	Most Relevant First
		<b>MPG</b>	Medizinproduktgesetz
		<b>MRT</b>	Magnetresonanztomographie
		<b>MSIBW</b>	Ministerium für Soziales und Integration Baden-Württemberg
		<b>N3</b>	Notation 3
		<b>NFA</b>	nicht-funktionalen Anforderungen
		<b>NG</b>	nicht gesund

<b>NLP</b>	Natural Language Processing	<b>ROC</b>	Receiver Operating Characteristic
<b>NPW</b>	Negativer Prädiktiver Wert		
<b>ODRL</b>	Open Digital Rights Language	<b>RUDAS</b>	Rowland Universal Dementia Assessment Scale
<b>OLG</b>	Oberlandesgericht	<b>RPR</b>	Richtig-Positiv-Rate
<b>OWL</b>	Web Ontology Language	<b>RNR</b>	Richtig-Negativ-Rate
<b>P3P</b>	Platform for Privacy Preferences Project	<b>RNN</b>	Recurrent Neural Networks
<b>PC</b>	Personal Computer	<b>SAGE</b>	Self-administered Gerocognitive Examination
<b>PDSG</b>	Patientendaten-Schutzgesetz	<b>SaMD</b>	Software as a Medical Device
<b>PEPA</b>	Persönliche einrichtungsübergreifende Elektronische Patientenakte	<b>SC</b>	Support Count
<b>PE</b>	persönliche Einstellung gegenüber Technik	<b>SEMMA</b>	Sample, Explore, Modify, Model, and Assess
<b>PET</b>	Positronen-Emissions-Tomographie	<b>SET</b>	Story-based Empathy Task
<b>PHR</b>	Personal Health Record	<b>SGB</b>	Sozialgesetzbuch
<b>PP.rt</b>	Klinik für Psychiatrie, Psychotherapie und Psychosomatik in Reutlingen	<b>SHAP</b>	Shapley Values Additive Explanations
<b>PRO</b>	Patient Reported Outcome	<b>SMART</b>	Substitutable Medical Applications, Reusable Technologies
<b>PVS</b>	Praxisverwaltungssystem	<b>SNOMED CT</b>	Systematized Nomenclature of Medicine Clinical Terms
<b>RBF</b>	Radial Basis Funktion	<b>SPECIAL</b>	Scalable Policy-aware Linked Data Architecture For Privacy, Transparency and Compliance
<b>RDF</b>	Ressource Description Framework	<b>SPL</b>	SPECIAL Usage Policy Language
<b>RDFS</b>	Ressource Description Framework Schema	<b>SPLog</b>	SPECIAL Policy Log Vacabulary
<b>REST</b>	Representational State Transfer	<b>SPARQL</b>	SPARQL Protocol and RDF Query Language
<b>RF</b>	Random Forest	<b>SVM</b>	Support Vector Machine
<b>RFE</b>	Recursive Feature Elimination	<b>SWT</b>	Shapiro-Wilk-Test
<b>RKI</b>	Robert Koch-Institut	<b>TAM</b>	Thinking Aloud Methode
<b>RN</b>	Richtig-Negativ	<b>TE</b>	Erfahrung im Umgang mit Technik
<b>RP</b>	Richtig-Positiv	<b>TSVG</b>	Terminservice- und Versorgungsgesetz
<b>PCA</b>	Principal Component Analysis		
<b>PKV</b>	Private Krankenversicherung		
<b>PPW</b>	Positiver Prädiktiver Wert		

---

<b>TI</b>	Telematikinfrastruktur
<b>TM</b>	Translationale Medizin
<b>TMT</b>	Trail Making Test
<b>TOM</b>	Theory of Mind
<b>TÜV</b>	Technischer Überwachungsverein
<b>URI</b>	Uniform Resource Identifier
<b>URL</b>	Uniform Resource Locator
<b>VB</b>	Verbraucherzentrale Bundesverband e.V.
<b>IADL</b>	Instrumental Activities of Daily Living
<b>QOL-AD</b>	Quality of Life-Alzheimer's Disease
<b>QS</b>	Quantified Self
<b>UI</b>	User Interface
<b>UML</b>	Unified Modeling Language
<b>USA</b>	Vereinigten Staaten von Amerika
<b>UX</b>	User Experience
<b>VPKV</b>	Verband der Privaten Krankenversicherung
<b>WFS</b>	Wrapper Feature Selection
<b>WHO</b>	World Health Organization
<b>WRT</b>	Wilcoxon-Rangsummen-Test
<b>WVRT</b>	Wilcoxon-Vorzeichen-Rang-Test
<b>WWW</b>	World Wide Web
<b>XAI</b>	Explainable Artificial Intelligence
<b>XDS</b>	Cross-Enterprise Document Sharing
<b>xDT</b>	XML Data Transfer
<b>Xk-NN</b>	Explainable k-NN
<b>XML</b>	Extensible Markup Language
<b>ZLG</b>	Zentralstelle der Länder für Gesundheitsschutz bei Arzneimitteln und Medizinprodukten
<b>6SQuID</b>	Six Steps in Quality Intervention Development



# Abbildungsverzeichnis

2.1	DSRM nach Peffers et. al.. . . . .	10
2.2	CRISP-DM nach Shearer. . . . .	12
2.3	Beispiel einer ROC-Kurve. . . . .	17
2.4	Akteure im selbstverwaltenden Gesundheitswesen. . . . .	19
2.5	Akteure in der Medizinproduktherstellung. . . . .	20
2.6	Anatomische Auswirkungen einer AD. . . . .	22
2.7	Vergleich kognitiver Abbauprozesse. . . . .	22
2.8	Neurokognitive Domänen nach der DSM-V . . . . .	23
2.9	Anzahl von Menschen mit AD im Ländergruppenvergleich. . . . .	25
2.10	Geschätzte Prävalenz in Europa und Inzidenzrate. . . . .	25
2.11	Beispiele für Intentions-, Emotionsattribution und Kausalitätsreaktion von links aus Dodich et al.. . . . .	36
2.12	Fragebogen zur Beurteilung des Vorliegens von MCI nach Nasreddine et al.. . . . .	37
2.13	Chronologie relevanter Gesetzes- und Verordnungseinführungen im Zeitraum von 2015 bis 2021. . . . .	38
2.14	Datenspende ab Januar 2023 nach DVG, PDSG und DaTraV. . . . .	41
2.15	Zusammenhänge der Anforderungen und harmonisierten Normen der MDR erweitert nach Hastenteufel und Renaud. . . . .	43
2.16	Gebrauchstauglichkeitsorientierter Entwicklungsprozess nach IEC 62366-1 und Zusammenhang mit dem Risikomanagement nach ISO 14971 aus Hastenteufel und Renaud. . . . .	44
2.17	Anforderungen an den Entwurf einer gebrauchstauglichen Benutzerschnittstelle nach Hastenteufel und Renaud. . . . .	45
2.18	Konzeptioneller Aufbau einer Blockchain. . . . .	53
2.19	Ebenen der Interoperabilität nach der HIMSS. . . . .	55
2.20	Schematische Darstellung der eingesetzten Klassifikationsalgorithmen. . . . .	59
2.21	Beispiele für Bewertungsschemata. . . . .	63
2.22	Optimierung der Modelleistung anhand von Parameter-Tuning. . . . .	64
3.1	Schematische Darstellung von LIME nach Ribeiro et al.. . . . .	81
3.2	Schematische Darstellung von CHIRPS nach Hatwell et al.. . . . .	85
4.1	Schematische Darstellung des Datenkreislaufs nach Schinle et al., 2021. . . . .	92
4.2	Medizinproduktlebenszyklus nach Hastenteufel und Renaud. . . . .	93
4.3	Abgeleitetes Vorgehen erweitert nach Schinle et al., 2021. . . . .	94
4.4	Verfahren zur Erkennung in Abhängigkeit des Verlaufs nach Ritchie et al.. . . . .	97
4.5	Würfel-Test. . . . .	101
4.6	Modell der latenten Variablen (links) und der Tests (mittig) sowie relevante Leistungskriterien (rechts) nach Schinle et al., 2021. . . . .	102
5.1	Aktive Parametererfassung. . . . .	106
5.2	Testdurchlauf FCSRT aus Testteilnehmersicht. . . . .	107
5.3	Unterprozess - Kontrolliertes Lernen aus Testteilnehmersicht. . . . .	107

5.4	Beispielhafte Vierergruppe von Objekten für kontrolliertes Lernen im FCSRT. . . . .	108
5.5	Komponentendiagramm der grundlegenden Systemarchitektur. . . . .	109
5.6	Verknüpfung von zielgruppenspezifischen Anforderungen und Normen für eine gebrauchstaugliche Benutzerschnittstelle. . . . .	115
5.7	Evolution der GUI am Beispiel des Test E in der Tablet-Version. . . . .	117
5.8	Beispiel neu gestalteter Bedienelemente in <i>DemPredict</i> im Vergleich zwischen Iteration III (links) und Iteration IV (rechts) der Tablet-Version. . . . .	118
5.9	Beispiel hinzugefügter Bedienelemente in <i>DemPredict</i> im Vergleich zwischen Iteration III (links) und Iteration IV (rechts) in der Tablet-Version. . . . .	118
5.10	Beispiel hinzugefügter Bedienelemente in <i>DemPredict</i> im Vergleich zwischen Iteration III (links) und Iteration IV (rechts) in der Tablet-Version. . . . .	119
5.11	Erweitertes SPECIAL-Modell nach Kirrane et al.. . . . .	123
5.12	Antworten auf Frage 3. . . . .	125
5.13	Antworten auf Frage 6. . . . .	125
5.14	Antworten auf Frage 9. . . . .	126
5.15	Antworten auf Frage 15. . . . .	126
5.16	Antworten auf Frage 10. . . . .	126
5.17	Antworten auf Frage 7. . . . .	126
5.18	Antworten auf Frage 1. . . . .	127
5.19	Antworten auf Frage 13. . . . .	127
5.20	Schematische Darstellung der Datenintegration am Beispiel von <i>DemPredict</i> auf FHIR. . .	128
5.21	Technisches Konzept zur Erreichung der Interoperabilität. . . . .	129
5.22	Vergleich zwischen Ergebnissen der CERAD- und <i>DemPredict</i> -Version für Test J. . . . .	134
5.23	Korrelation zwischen Würfeltest-Punktzahl in CERAD- und <i>DemPredict</i> -Version nach Pearson. . . . .	135
5.24	Korrelationsanalyse zwischen den Würfeltest-Ergebnissen der CERAD- und <i>DemPredict</i> -Version mit der <i>TE</i> . . . . .	137
5.25	Würfeltest-Ergebnisse der CERAD- und <i>DemPredict</i> -Version pro Diagnosegruppe. . . . .	139
6.1	Visuelle Wahrnehmung und Rekonstruktion eines Würfels. . . . .	143
6.2	{ <b>AE,EB,CF,FD</b> } in äußere Linien zerlegt. . . . .	144
6.3	Extraktion der äußeren räumlich orientierten Kanten. . . . .	144
6.4	Ergebnis eines gesunden Testteilnehmers (rot markierte Abweichungen vom idealen Würfel), eines Testteilnehmers mit MCI und Demenz. . . . .	144
6.5	Vergleich der ursprünglichen Matrix (links) und den algorithmisch erfassten zusammenhängenden Komponenten (rechts) . . . . .	145
6.6	Vergleich der zusammengesetzten Kanten <b>AE</b> vor (links) und nach (rechts) der heuristischen Filterung am Beispiel des gesunden Testteilnehmers. . . . .	145
6.7	Vergleich der ursprünglichen Zeichnung (vgl. Abbildung 6.4) mit dem Ergebnis nach der Kantendetektion für Testteilnehmer mit den Labels G, MCI und D. . . . .	145
6.8	Definition von Parametern zur automatisierten Eckendetektion in Test J. . . . .	146
6.9	Ergebnis der Eckendetektion für den in Abbildung 6.4 dargestellten Würfel eines gesunden Testteilnehmers (G) nach Schritt 1 (links) und Schritt 2 (rechts). . . . .	147
6.10	Ergebnis nach der Eckendetektion für einen gesunden sowie von Testteilnehmern mit MCI und Demenz nach Schritt 2 . . . . .	147
6.11	Ergebnis der Kombinationen aus Kanten- und Eckendetektion für den in Abbildung 6.4 dargestellten Würfel eines gesunden Testteilnehmers (G) in der Reihenfolge Kante-Ecken (links) und Ecken-Kanten (mittig) im Vergleich mit einer alleinigen Eckendetektion (rechts). . . . .	148
6.12	Charakteristiken der Dreidimensionalität. . . . .	149

6.13	Charakteristiken der inneren Kanten. . . . .	149
6.14	Charakteristiken der Frontseite. . . . .	149
6.15	Charakteristiken der Parallelität. . . . .	149
6.16	Merkmalspezifisches Scoring nach Punkteschema. . . . .	155
6.17	Auf kognitiven Domänen basiertes Scoring nach Punkteschema. . . . .	156
6.18	Auf kognitiven Domänen basiertes Scoring nach Anzahl auffälliger Domänen. . . . .	156
6.19	Modell der gefilterten latenten Variablen (links) und der Tests (Mitte) sowie relevante Leistungskriterien (rechts) nach Schinle et al., 2021. . . . .	160
6.20	Gegenüberstellung des mittleren Informationsgehalts der CERAD-Merkmale für die Klassifikationswahrscheinlichkeit. . . . .	164
6.21	PCA für Trainingsdaten des Xk-NN. . . . .	165
6.22	PCA als Erklärung für Mensch mit D. . . . .	165
6.23	Berechnungszeit der XAI-Methoden. . . . .	167
6.24	Variationskoeffizient der wichtigsten Merkmale. . . . .	168
6.25	MoRF-Störungskurve. . . . .	169
7.1	Umfrageergebnisse bezüglich Durchführungsfrequenz. . . . .	172
7.2	Umfrageergebnisse der Akzeptanzkriterien zu den Fragen 29, 30, 34, 35, 37, 38, 39, 40, 41 und 42. . . . .	175
7.3	Umfrageergebnisse zu Frage 45., 49. und 51. . . . .	176
7.4	Entwicklung digitaler Präventions- und Behandlungsprogramme aus der Kontextsicht. . . . .	177
7.5	Prototyp der Anwendung für Leistungserbringer. . . . .	179
7.6	Mockup der Anwendung für Leistungsempfangende. . . . .	180
7.7	Architektur der Präventions- und Behandlungsplattform. . . . .	181
A.1	Antworten auf Umfrage aus Abschnitt A.3. . . . .	208
A.5	Vergleich zwischen aggregierten Testergebnissen der CERAD- und <i>DemPredict</i> -Version. . . . .	223
A.7	Korrelationsanalyse zwischen individuellen Testergebnissen der CERAD- und <i>DemPredict</i> -Version nach Pearson. . . . .	224
A.9	Antworten auf die Fragen zur Beantwortung der übergeordneten Fragen <b>F1-F4</b> und <b>F5</b> . . . . .	251
A.11	Antworten auf die Fragen zur Beantwortung der übergeordneten Fragen <b>F2,F7</b> und <b>F8</b> . . . . .	252
A.12	Umfrageergebnisse zu den wichtigsten Akzeptanzkriterien (Frage 44). . . . .	253



# Tabellenverzeichnis

2.1	Stärke der Übereinstimmung für $\kappa$ -Werte nach Landis et al. . . . .	15
2.2	Klassifikation nach der IMDRF für Software als Medizinprodukt. . . . .	42
2.3	Übersicht über deutsche Aktenlösungen für elektronische Gesundheitsdatenverwaltung. . .	50
2.4	Ausgewählte interpretierbare Modelle und deren Eigenschaften nach Molnar. . . . .	65
4.1	Zusammengefasste Anforderungen an den Umgang mit Gesundheitsdaten in Gesundheitsdatenökosystemen nach Schinle et al., 2021. Details in Tabelle A.1. . . . .	91
4.2	Übersicht über verwendete psychometrische Testverfahren nach Schinle et al., 2018. . . . .	99
4.3	Identifizierte Anforderungen an die Umsetzung digitaler Gesundheitsanwendungen. . . . .	104
5.1	Auswirkungen zielgruppenspezifischer Einschränkungen auf die Gestaltung einer Benutzerschnittstelle. . . . .	112
5.2	Leitfaden mit zielgruppenspezifischen Anforderungen an die Benutzerschnittstelle. . . . .	115
5.3	Ergebnisse der Rang-Korrelationsanalyse nach Spearman für vergleichbare Parameter. . . . .	136
5.4	Ergebnisse des WRTs zum paarweisen Vergleich der Diagnosegruppen und deren <i>TE</i> . . . . .	137
5.7	Ergebnisse des WRTs zum paarweisen Vergleich der Diagnosegruppen und Würfel-Punktzahl pro Version. . . . .	139
6.1	Gegenüberstellung zwischen menschlicher und algorithmischer Entscheidung für Abbildung 6.10 mit dem Ziel der Maximierung des Cohen's Kappa bei $X = 75\%$ . . . . .	151
6.2	Performanz des Automatisierungsansatzes unter Verwendung der Eckendetektion über alle Diagnosegruppen hinweg mit dem Ziel der Maximierung des Cohen's Kappa. . . . .	152
6.3	Performanz des Automatisierungsansatzes unter Verwendung der Eckendetektion über alle Diagnosegruppen hinweg mit dem Ziel der Minimierung des mittleren absoluten Fehlers bei $X = 60\%$ und $X = 50\%$ . . . . .	152
6.4	Performanz des Automatisierungsansatzes unter Verwendung der Eckendetektion über alle Diagnosegruppen hinweg mit dem Ziel der Maximierung des Cohen's Kappa und 5-fach kreuzvalidierten Algorithmenparametern. . . . .	153
6.5	Parameterselktion anhand von FFS. . . . .	155
6.15	Performanz der verschiedenen Scoring-Systeme ohne Merkmalsselektion. . . . .	157
6.16	Performanz der verschiedenen Scoring-Systeme mit FFS. . . . .	157
6.17	Performanz der verschiedenen Scoring-Systeme für CERAD-Merkmale. . . . .	158
6.18	Durchschnittliche Modell-Leistungen ohne Merkmalsselektion erweitert nach Schinle et al., 2021. . . . .	161
6.19	Durchschnittliche Modell-Leistungen mit statistischem Filter für die Merkmalsselektion (FFS) nach Abschnitt 6.5. . . . .	161
6.20	Durchschnittliche Modell-Leistungen mit WFS erweitert nach Schinle et al., 2021. . . . .	162
6.21	Durchschnittliche Modell-Leistungen mit RFE. . . . .	162
6.22	Mittelwerte der Ergebnisse des CHIRPS-Verfahrens. . . . .	164
6.23	Verwendete Pakete und selbst geschriebene Funktionen für die Implementierung der XAI-Methoden in R und Python. . . . .	165
6.24	Zusammenhang zwischen XAI-Verfahren und Wahl der Evaluationsmetriken. . . . .	167

A.1	Detaillierte Anforderungen an den Umgang mit Gesundheitsdaten in Gesundheitsdatenökosystemen nach Schinle et al., 2021. . . . .	199
A.2	Studienteilnehmerdaten der Umfrage unter Leistungserbringern. . . . .	207
A.3	Studienteilnehmerdaten der Beobachtungsstudie nach Iteration II. . . . .	214
A.4	Übersicht über die möglichen digitalen Parameter. . . . .	219
A.5	Ergebnisse des SWTs für mit <i>DemPredict</i> erhobene Parameter. . . . .	220
A.6	Ergebnisse des WRTs für die mit <i>DemPredict</i> domänenspezifischen Parameter in T1 und T2. . . . .	222
A.7	Korrelation zwischen Technologie-Erfahrung und den domänenspezifischen Parametern. . . . .	226
A.8	Diskriminationsfähigkeit zwischen Diagnosegruppen anhand von KWT und WRT. . . . .	228
A.9	Übersicht über die individuellen Schwellwerte der digitalen Parameter. . . . .	230
A.10	Berechnungszeiten in Sekunden. . . . .	231
A.11	MoRF-Störungswerte des RF LIME. . . . .	232
A.12	MoRF-Störungswerte des RF SHAP. . . . .	232
A.13	MoRF-Störungswerte des k-NN LIME. . . . .	233
A.14	MoRF-Störungswerte des k-NN SHAP. . . . .	233
A.15	Gewichtung der Merkmale für RF LIME. . . . .	234
A.16	Gewichtung der Merkmale für RF SHAP. . . . .	234
A.17	Gewichtung der Merkmale für k-NN LIME. . . . .	234
A.18	Gewichtung der Merkmale für k-NN SHAP. . . . .	234
A.19	Demografische Daten der Studie zur Adhärenz. . . . .	235
A.20	Studienteilnehmerdaten der Umfrage unter Leistungserbringern. . . . .	250
A.21	Freitext-Antworten auf die Frage 44. . . . .	259
A.22	Freitext-Antworten auf die Frage 48. . . . .	260
A.23	Antworten auf den Fragebogen aus Anhang A.11. . . . .	268
A.24	Demografische Daten und Antworten auf den Fragebogen aus Anhang A.7 sowie berechnete PE und TE. . . . .	268

# Eigene Veröffentlichungen

## Journalartikel

- [445] Markus Schinle, Christina Erler, Mayumi Kaliciak, Christopher Milde, Simon Stock, Marius Gerdes, Wilhelm Stork. Digital Health Apps in the Context of Dementia: Questionnaire Study to Assess the Likelihood of Use among Physicians. *JMIR Formative Research*, 2022. doi: 10.2196/35961.

## Konferenzbeiträge

- [446] Marianne Behrends, Mathias Witte, Ralf Eckert, Markus Schinle, Matthias Gietzelt. Die Wohnung als Standort der Gesundheitsversorgung: Informatik in der guten Stube. In *Health — Exploring Complexity*. URL: <https://www.egms.de/static/en/meetings/gmds2016/16gmds126.shtml>, doi:10.3205/16GMDs126.
- [447] Timon Blöcher, Johannes Schneider, Markus Schinle, Wilhelm Stork. An online PPGI Approach for Camera Based Heart Rate Monitoring using Beat-To-Beat Detection. In *2017 IEEE Sensors Applications Symposium*, pages 1–6, Piscataway, NJ, USA, 2017. IEEE. doi:10.1109/SAS.2017.7894052.
- [448] Christina Erler, Markus Schinle, Michael Dietrich, Wilhelm Stork. Decision model to design a blockchain-based system for storing sensitive health data. In *An Online AIS Conference (Hg.) 2022 – In ECIS 2022 Proceedings*. URL: [https://aisel.aisnet.org/ecis2022\\_rp/151](https://aisel.aisnet.org/ecis2022_rp/151).
- [449] Markus Lücking, Raphael Manke, Markus Schinle, Lukas Kohout, Stefan Nickel, Wilhelm Stork. Decentralized patient-centric data management for sharing IoT data streams. In *2020 International Conference on Omni-layer Intelligent Systems (COINS)*, pages 1–6. IEEE, 8/31/2020 - 9/2/2020. doi:10.1109/COINS49042.2020.9191653.
- [450] Markus Schinle, Christina Erler, Philip Nicolai Andris, Wilhelm Stork. Integration, Execution and Monitoring of Business Processes with Chaincode. In *2020 2nd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS)*, pages 63–70. IEEE, 9/28/2020 - 9/30/2020. doi:10.1109/BRAINS49436.2020.9223283.
- [451] Markus Schinle, Christina Erler, Maximilian Hess, Wilhelm Stork. Explainable Artificial Intelligence in Ambulatory Digital Dementia Screenings. *Studies in health technology and informatics*, 294:123–124, 2022. doi:10.3233/SHTI220411.

- [452] Markus Schinle, Christina Erler, Simon Leenstra, Simon Stock, Marius Gerdes, Wilhelm Stork. A Decision Process Model for De-Identification Methods on the Example of Psychometric Data. In *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pages 1–6. IEEE, 10/7/2021 - 10/8/2021. doi:10.1109/ICECCME52200.2021.9591139.
- [453] Markus Schinle, Christina Erler, Timon Schneider, Joana Plewnia, Wilhelm Stork. Data-driven Development of Digital Health Applications on the Example of Dementia Screening. In *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6. IEEE, 6/23/2021 - 6/25/2021. doi:10.1109/MeMeA52024.2021.9478676.
- [454] Markus Schinle, Christina Erler, Wilhelm Stork. Distributed Ledger Technology for the systematic Investigation and Reduction of Information Asymmetry in Collaborative Networks. In Tung Bui, editor, *Proceedings of the 53rd Hawaii International Conference on System Sciences*, Proceedings of the Annual Hawaii International Conference on System Sciences. Hawaii International Conference on System Sciences, 2020. doi:10.24251/HICSS.2020.651.
- [455] Markus Schinle, Christina Erler, Wilhelm Stork. Data Sovereignty in Data Donation Cycles: Requirements and Enabling Technologies for the Data-driven Development of Health Applications. In Tung Bui, editor, *Proceedings of the 54rd Hawaii International Conference on System Sciences*, Proceedings of the Annual Hawaii International Conference on System Sciences. Hawaii International Conference on System Sciences, 2021. doi:10.24251/HICSS.2021.482.
- [456] Markus Schinle, Christina Erler, Adrian Rudolf Vetter, Wilhelm Stork. How to disclose selective Information from permissioned DLT-based Traceability Systems? In *2020 IEEE International Conference on Decentralized Applications and Infrastructures (DAPPS)*, pages 153–158. IEEE, 8/3/2020 - 8/6/2020. doi:10.1109/DAPPS49028.2020.00020.
- [457] Markus Schinle, Thomas Födisch, Wilhelm Stork. Assistierte Kommunikation in ambulanten Betreuungsformen: Teilhabe mit AAL. In *AAL Kongress*. URL: [https://hci.w-hs.de/wp-content/uploads/2018/11/pub\\_AAL\\_2018\\_Tagungsband\\_web.pdf](https://hci.w-hs.de/wp-content/uploads/2018/11/pub_AAL_2018_Tagungsband_web.pdf).
- [458] Markus Schinle, Ioannis Papantonis, Wilhelm Stork. Personalization of Monitoring System Parameters to support Ambulatory Care for Dementia Patients. In *2018 IEEE Sensors Applications Symposium*, pages 1–6, Piscataway, NJ, 2018. IEEE. doi:10.1109/SAS.2018.8336724.
- [459] Markus Schinle, Johannes Schneider, Timon Blöcher, Jochen Zimmermann, Sebastian Chiriac, Wilhelm Stork. A Modular Approach for Smart Home System Architectures Based on Android Applications. In *MobileCloud 2017*, pages 153–156, Piscataway, NJ, 2017. IEEE. doi:10.1109/MobileCloud.2017.20.
- [460] Markus Schinle, Daniel Wyszka, Frank Schwärzler, Kirsten Volz, Marie-Anne Ruby, Emir Sejdinovic, Wilhelm Stork. An Approach to digitalize Psychological Tests to support Diagnosis of Alzheimer's Disease in Ambulatory Care. In *2018 IEEE International Symposium on Medical Measurements and Applications*, pages 1–6, Piscataway, NJ, 2018. IEEE. doi:10.1109/MeMeA.2018.8438761.
- [461] Frank Schwärzler, Markus Schinle, Anne Röhm. Innovation durch Digitalisierung? Ein Erfahrungsbericht aus dem Projekt "Pflegeteaching für die optimalen Unterstützung von Menschen mit

---

Demenz -PflegeCoDe". In *Deutsche Gesellschaft für Gerontopsychiatrie und -psychotherapie e.V. (DGGPP) 2021*. URL: [https://dggpp-kongress.de/dggpp2021/Abstracts/abs\\_34.htm](https://dggpp-kongress.de/dggpp2021/Abstracts/abs_34.htm).

- [462] Simon Claus Stock, Marius Gerdes, Lucas Maximilian Martini, Julia Veloso de Oliveira, Lars Hauptmann, Markus Schinle, Wilhelm Stork. Evaluation of Innovative SSVEP Stimulation Patterns for Neuro-Ophthalmology. doi:10.13140/RG.2.2.27298.76483.
- [463] Simon Stock, Florian Mazura, Fernando Gomez de La Torre, Marius Gerdes, Markus Schinle, Wilhelm Stork. MMAI - Mobile Moods AI; Electroencephalography Artifact Detection; Towards Objective Assessment of Mental States. In *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pages 01–06. IEEE, 10/7/2021 - 10/8/2021. doi:10.1109/ICECCME52200.2021.9590972.

