

Forschungsberichte aus dem
wbk Institut für Produktionstechnik
Karlsruher Institut für Technologie (KIT)

Tobias Schlagenhauf

**Bildbasierte Quantifizierung
und Prognose des Verschleißes an
Kugelgewindetriebsspindeln**

Ein Beitrag zur Zustandsüberwachung von
Kugelgewindetrieben mittels Methoden des
maschinellen Lernens

Band 263



Forschungsberichte aus dem
wbk Institut für Produktionstechnik
Karlsruher Institut für Technologie (KIT)

Hrsg.: Prof. Dr.-Ing. Jürgen Fleischer
Prof. Dr.-Ing. Gisela Lanza
Prof. Dr.-Ing. habil. Volker Schulze

Tobias Schlagenhauf

Bildbasierte Quantifizierung und Prognose des Verschleißes an Kugelgewindetriebsspindeln

Ein Beitrag zur Zustandsüberwachung von
Kugelgewindetrieben mittels Methoden des
maschinellen Lernens

Band 263

Bildbasierte Quantifizierung und Prognose des Verschleißes an Kugelgewindetriebspindeln

Ein Beitrag zur Zustandsüberwachung von Kugelgewindetrieben
mittels Methoden des maschinellen Lernens

Zur Erlangung des akademischen Grades eines
DOKTORS DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)

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

DISSERTATION

von

M.Sc. Tobias Schlagenhaut

Tag der mündlichen Prüfung: 07.11.2022
Hauptreferent: Prof. Dr.-Ing. Jürgen Fleischer
Korreferent: Prof. Dr.-Ing. Benjamin Noack

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Zugl.: Karlsruhe, Karlsruher Institut für Technologie, Diss., 2022

Copyright Shaker Verlag 2022

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe, der Speicherung in Datenverarbeitungsanlagen und der Übersetzung, vorbehalten.

Printed in Germany.

ISBN 978-3-8440-8875-5
ISSN 0724-4967

Shaker Verlag GmbH • Am Langen Graben 15a • 52353 Düren
Telefon: 02421 / 99 0 11 - 0 • Telefax: 02421 / 99 0 11 - 9
Internet: www.shaker.de • E-Mail: info@shaker.de

Vorwort des Herausgebers

Die schnelle und effiziente Umsetzung innovativer Technologien wird vor dem Hintergrund der Globalisierung der Wirtschaft der entscheidende Wirtschaftsfaktor für produzierende Unternehmen. Universitäten können als "Wertschöpfungspartner" einen wesentlichen Beitrag zur Wettbewerbsfähigkeit der Industrie leisten, indem sie wissenschaftliche Grundlagen sowie neue Methoden und Technologien erarbeiten und aktiv den Umsetzungsprozess in die praktische Anwendung unterstützen.

Vor diesem Hintergrund wird im Rahmen dieser Schriftenreihe über aktuelle Forschungsergebnisse des Instituts für Produktionstechnik (wbk) am Karlsruher Institut für Technologie (KIT) berichtet. Unsere Forschungsarbeiten beschäftigen sich sowohl mit der Leistungssteigerung von additiven und subtraktiven Fertigungsverfahren, den Produktionsanlagen und der Prozessautomatisierung sowie mit der ganzheitlichen Betrachtung und Optimierung der Produktionssysteme und -netzwerke. Hierbei werden jeweils technologische wie auch organisatorische Aspekte betrachtet.

Prof. Dr.-Ing. Jürgen Fleischer

Prof. Dr.-Ing. Gisela Lanza

Prof. Dr.-Ing. habil. Volker Schulze

Vorwort des Verfassers

Die vorliegende Dissertation entstand während meiner Tätigkeit als akademischer Mitarbeiter am Institut für Produktionstechnik (wbk) am Karlsruher Institut für Technologie (KIT). Mein Dank gilt meinem Doktorvater Herrn Prof. Dr.-Ing. Jürgen Fleischer für die Übernahme des Hauptreferates. Ganz besonders möchte ich Prof. Fleischer für die wertvollen fachlichen Diskussionen, den überfachlichen Austausch sowie seine ständige Förderung bedanken. Mein weiterer Dank gilt Herrn Prof. Dr.-Ing. Benjamin Noack für die Übernahme des Koreferates und das mir entgegengebrachte Vertrauen. Danken möchte ich ihm speziell für den stets angenehmen fachlichen Austausch sowie seinem Engagement zur Verzahnung der (produktions-)technischen, informationstechnischen und mathematischen Domänen. Weiterhin danken möchte ich Prof. Dr.-Ing. Hans-Jörg Bauer für die Übernahme des Prüfungsvorsitzes. Mein besonderer Dank gilt allen Studierenden sowie Kolleginnen und Kollegen des wbk, welche mit mir stets in gutem fachlichem Austausch standen und ohne diese die erfolgreiche Durchführung dieser Dissertation ungleich schwerer gewesen wäre. Hier gilt mein besonderer Dank den Kollegen Jonas Hillenbrand, Malte Hansjosten, David Barton und Benedikt Klee. Für die technische Unterstützung in IT und Werkstatt gilt mein Dank allen Mitarbeitenden der Service-Center IT und Technik. Hierbei seien vor allem Aaron Schmidt und Friedbert Mathes betont, welchen ich für die Unterstützung beim Aufbau und Betrieb meines Versuchsstandes danken möchte. Abschließend möchte ich mich ganz besonders bei meiner Familie und vor allem bei meinen Eltern Eckhard und Andrea Schlagenhaut sowie meinen Schwiegereltern Stefan und Dietlinde Schick für deren fortwährende Unterstützung während des Verfassens der Dissertation bedanken. Meiner Frau Rebecca Leonie danke ich von ganzem Herzen, dass sie mir stets Stütze und Gesprächspartnerin war. Ganz besonders danke ich ihr für unsere beiden Kinder Leni Frieda und Leonard Theo, die für mich stets Leuchttürme während des Verfassens dieser Dissertation waren.

Karlsruhe, 15.07.2022

Tobias Schlagenhaut

Abstract

For the realization of autonomous production machines, it is necessary to enable them to independently assess the condition of their components and, in a medium-term step, inform the maintenance department about it. This requires methods that allow the reliable estimation of the condition of machine components. For the realization of a predictive maintenance strategy, it is also necessary to analyze and understand the wear progression of machine tool components in order to develop prognosis models on that basis.

As a commonly used and highly stressed machine tool component, the ball screw drive (BSD) plays a central role for the reliable operation of autonomous production machines. Previous work on condition monitoring of BSDs has been based on the indirect interpretation of the wear characteristics. In addition, the development of surface defects on the BSD has not been documented and investigated in image data yet, which has prevented the reliable implementation of a prognosis model based on direct wear signals.

The approach presented in this work documents the entire wear development on the surface of the BSD up to the mechanical failure of the component for the first time in image data. On this basis, the classification of the damage is investigated using machine learning methods for the first time, thus creating the basis for a reliable condition monitoring system. Based on this, a wear forecasting system is investigated. The functionality of the forecasting system could be demonstrated for the first time on image data of the BSD. The investigations are extended by experiments on data efficient classification by investigating a novel approach for data efficient classification of image data. The so called SBF-Net approach is validated on images of ball screw drives as well as on image data of other technical and non-technical domains. The superiority of the approach compared to state of the art models could be demonstrated.

Inhaltsverzeichnis

Inhaltsverzeichnis	I
Formelzeichen & Abkürzungen	IV
Begriffsklärungen & Vorbemerkungen	V
1 Einleitung	1
1.1 Motivation	1
1.2 Zielsetzung	3
1.3 Aufbau der Arbeit	4
2 Grundlagen	7
2.1 Tribologie	7
2.1.1 Reibung und Verschleiß	7
2.1.2 Verschleißmechanismen	9
2.2 Tribologisches System Kugelgewindtrieb	13
2.2.1 Verschleißmechanismen am Kugel-Laufbahn-Kontakt	15
2.2.2 Ausfallmechanismen am Kugelgewindtrieb	20
2.2.3 Lebensdauer des Kugelgewindtriebs	22
2.3 Condition Monitoring und Predictive Maintenance	29
2.3.1 Bedeutung von Condition Monitoring und Predictive Maintenance für den Maschinenbau	30
2.3.2 Ziele und Unterschiede der Ansätze	31
2.3.3 Implementierung von Condition Monitoring und Predictive Maintenance	32
2.4 Bildverarbeitung zur Klassifikation	33
2.4.1 Sensierung und Merkmalsextraktion	34
2.4.2 Bildmerkmale	35
2.4.3 Mustererkennung und Merkmalsgenerierung	38
2.4.4 Neuronale Netzwerke zur Klassifikation	40
2.4.5 Convolutional Neural Networks zur Klassifikation	46
3 Stand der Forschung und Technik	56
3.1 Bildbasierte Defekterkennung auf metallischen Oberflächen	56

3.2	Verschleißüberwachung an Maschinenelementen	61
3.2.1	Körperschall	62
3.2.2	Luftschall	63
3.2.3	Motorstrom	64
3.2.4	Temperatur	65
3.3	Relevante Arbeiten zur Verschleißüberwachung am Kugelgewindetrieb	66
3.4	Bewertung des Standes der Forschung und Technik	74
3.4.1	Fazit zum Stand der Forschung und Technik	74
3.4.2	Aktuelle Defizite im Stand der Forschung und Technik	75
4	Zielsetzung und Lösungsansatz	77
4.1	Konkretisierung der Zielsetzung	77
4.2	Übersicht zum Lösungsansatz	78
5	Experimentelle Datengenerierung	81
5.1	Lebensdauerversuchsstand	81
5.2	Versuchsdurchführung	82
5.3	Sensorsystem zur Bilddatengenerierung	84
5.3.1	Einflussgrößen und Anforderungen an das Sensorsystem	85
5.3.2	Konzeption des Sensorsystems	87
5.3.3	Aufnahme von Bilddaten	92
5.4	Bilddatensatz	94
6	Datenanalyse und Modellbildung	97
6.1	Analyse der Entwicklung von Pittings	97
6.1.1	Verschleißerscheinungsformen auf der Laufbahn des Kugelgewindetriebes	98
6.1.2	Verschleißpartikel im Schmierstoff	99
6.1.3	Die Entstehung von Pittings	102
6.1.4	Empirische Analyse der Entwicklung von Pittings	108
6.1.5	Normbasierte Präzisierung der Verschleißquantifizierung	120
6.1.6	Diskussion der Erkenntnisse im Kontext bestehender Arbeiten	122
6.1.7	Visuelle Charakteristika der Verschleißerscheinungsform Pitting	128

6.2	Klassifikation von Oberflächenzerrüttung in den Bilddaten des Kugelgewindetriebes	132
6.2.1	Grundüberlegungen zur Auswahl von Klassifikationsansätzen	133
6.2.2	Klassifikation mittels extrahierter Merkmale	136
6.2.3	Klassifikation mittels Convolutional-Neural-Network-Architekturen	151
6.2.4	Dateneffiziente Klassifikation in technischen Domänen	165
6.2.5	Fazit zu der Klassifikation von Oberflächenzerrüttung	187
7	Verschleißprognose	191
7.1	Grundüberlegungen	191
7.2	Positionsbestimmung und Quantifizierung	193
7.3	Integration von Domänenwissen	197
7.4	Prognose	200
7.4.1	Auswahl der Prognosefunktion	200
7.4.2	Validierung des Prognosesystems	204
7.5	Fazit zu der Verschleißprognose im Kontext bestehender Arbeiten	210
7.6	Erweiterung der Zustandsüberwachung am KGT	213
8	Bewertung	215
8.1	Zusammenfassung	215
8.2	Ausblick	218
	Publikationsliste des Autors	I
	Literaturverzeichnis	III
	Abbildungsverzeichnis	XXXVI
	Tabellenverzeichnis	XLV
	Anhang	XLVI
	A1 Backpropagation im Neuronalen Netzwerk	XLVI
	A2 Backpropagation in der Convolutional Base	XLVII
	A3 Ergänzende Materialien	XLIX

Formelzeichen & Abkürzungen

Alle Formelzeichen und Abkürzungen sind an den entsprechenden Stellen im Dokument erläutert. Es wird daher zugunsten einer besseren Übersichtlichkeit auf ein dediziertes Formelverzeichnis verzichtet.

Begriffsklärungen & Vorbemerkungen

Alle im Haupttext erstmalig *kursiv* geschriebenen Notationen werden hier in ihrer Bedeutung erläutert. Dies ist notwendig, da vor allem im Bereich des Maschinellen Lernens die Terminologie in englischer Sprache gehalten ist und die englischsprachige Fachterminologie in Übereinstimmung mit dem Stand der Forschung vorrangig, wenn auch nicht ausschließlich verwendet wird.

Begriff	Bedeutung
Accuracy	Klassifikationsgenauigkeit eines Modelles. ‚Validierungsgenauigkeit‘ und ‚Validierungs-Accuracy‘ werden synonym verwendet.
Convolutional Neural Networks (CNN)	Spezielle Architektur des Maschinellen Lernens, die automatisch Merkmale extrahiert sowie klassifiziert und sich unter anderem für die Klassifikation von Bilddaten als zielführend erwiesen hat.
Diskriminative Merkmale	Merkmale, die sich gut zur Unterscheidung von Objekten eignen.
(Feature-)Embeddings	Vektoren oder Matrizes, die extrahierte Merkmale eines Datenpunktes enthalten. Der Datenpunkt ist durch diesen Vektor/diese Matrix ausgedrückt bzw. darin eingebettet (engl. embedded).
Finetuning	Teilweises Anpassen der Gewichte eines Modelles mittels eines erweiternden Datensatzes.
Fully Connected Layer	Schicht-Architektur innerhalb eines Neuronalen Netzwerkes, bei dem alle Knoten der Schicht mit allen Knoten der vorherigen sowie nachfolgenden Schicht verbunden sind.
Groundtruth Data	Zugrunde liegende, wahre Daten, deren Verlauf von einer Funktion erlernt werden soll.
Hidden Layer	Zwischenschichten eines Modelles. Werden als Funktionen realisiert.

Hyperparameter	Parameter, die den Aufbau eines Machine-Learning-Modelles betreffen, bspw. die Anzahl an Schichten eines Neuronalen Netzes.
Inference Phase	Phase des Modellbetriebs (bspw. Klassifikation, Regression), nachdem das Training abgeschlossen ist.
Input Layer	Eingangsschicht eines Modells. Schnittstelle, an der Daten/Werte an ein Modell übergeben werden.
Inputs	Eingangsdaten/Eingangswerte eines Machine-Learning-Modells oder von Teilen des Modelles (bspw. von bestimmten Schichten innerhalb eines Neuronalen Netzes ¹).
Klassische Machine-Learning-Verfahren	Werden hier als diejenigen Verfahren definiert, bei denen Merkmale auf Basis von Domänenwissen extrahiert werden. Die Klassifikation kann dann entweder mittels Expertenregeln oder auf Basis von Machine-Learning-Ansätzen erfolgen.
Label	Zuordnung eines Datenpunktes zu einer Klasse $c \in C$, worüber die Identität eines Datenpunktes bestimmt wird.
Learning Rate	Lernrate oder Schrittweite bei der Optimierung eines Modelles.
Loss	Wert, der von einer Verlustfunktion (Fehlerfunktion) zurückgegeben wird und die Übereinstimmung zwischen den wahren Werten und den vorhergesagten Werten beschreibt.
Machine Learning	dt.: Maschinelles Lernen. Verfahren, das es Computern erlaubt, automatisiert zu

¹ Mit dem Begriff ‚Neuronales Netz‘ wird das künstliche neuronale Netz aus dem Bereich des Maschinellen Lernens bezeichnet, ohne hierbei die tatsächliche Wirkungsweise der biologischen Vorbilder zu berücksichtigen. Der Begriff wird ebenso wie der Begriff des Maschinellen Lernens als Eigenname verwendet und großgeschrieben.

	lernen, ohne hierfür explizit programmiert zu werden.
Mean Squared Error (MSE)	Quadrierte mittlere Abweichung der Vorhersagen einer Regressionsfunktion von den Groundtruth-Daten.
Modellparameter	Parameter eines Modelles, die während des Trainings angepasst werden. Beispiel hierfür sind die Gewichte eines Neuronalen Netzwerks.
(künstliches) Neuronales Netzwerk	Algorithmus des Maschinellen Lernens für Klassifikations- und Regressionsaufgaben. Es wird der Eigenname ‚Neuronale Netzwerke‘ bzw. ‚Neural Networks‘ verwendet.
Output Layer	Ausgangsschicht/Ausgabeschicht eines Modelles.
Outputs	Ausgabedaten/Ausgabewerte eines Machine-Learning-Modelles oder von Teilen des Modelles (bspw. von bestimmten Schichten innerhalb eines Neuronalen Netzes).
Overfitting	Überanpassung des Modelles auf die Trainingsdaten, was zu einer schlechten Generalisierung des Modelles auf Validierungsdaten führt.
Pitting	dt.: Grübchen. Verschleißerscheinungsform, hervorgerufen durch die Oberflächenzerrüttung.
Root Mean Squared Error (RMSE) / RMS-Distanz / RMS-Error	Wurzel aus der quadrierten, mittleren Abweichung der Vorhersagen einer Regressionsfunktion von den Groundtruth-Daten.
Supervised Learning	Beim Training stehen sowohl die Daten als auch die zugehörigen Klassen zur Verfügung.

Train-Test/Validation Split

Prozess zum Aufteilen der Daten in $x\%$ Trainingsdaten und $(100-x)\%$ Validierungsdaten/Testdaten. Beide Bezeichnungen sind gängig.

Unsupervised Learning

Beim Training sind die Klassen nicht bekannt. Es liegen lediglich die Daten vor.

Wortgruppen wie *Fully Connected Layer* oder *Input Layer*, die in der englischsprachigen Fachterminologie Eigennamen darstellen, werden getrennt geschrieben und nicht übersetzt. Wortgruppen, die aus einer Kombination aus englischen oder deutschen Begriffen gebildet werden, bspw. *Input-Zelle* oder *Feature-Maps*, werden mit Bindestrich geschrieben.

Es werden grundsätzlich, soweit möglich, genderneutrale Formulierungen gewählt. Der Begriff des *Expertensystems* wird in Einklang mit der Literatur im generischen Maskulinum verwendet. Ebenfalls wird für die Gruppe der *Anwender* meist das generische Maskulinum herangezogen. Genderneutrale Formulierungen werden nicht gewählt, wenn sie aufgrund grammatikalischer oder syntaktischer Gründe den Lesefluss stören würden.

Die Kapitel und Unterkapitel der Dissertation, die den eigenen Ansatz betreffen, sind grundsätzlich so aufgebaut, dass jedes Kapitel mit einer Einleitung begonnen wird. Dieser folgt der inhaltliche Teil des Kapitels. Die Kapitel schließen mit einer Zusammenfassung der Erkenntnisse und einer anschließenden Überleitung zum folgenden Kapitel.

1 Einleitung

1.1 Motivation

Die automatisierte Überwachung des Zustandes von Maschinen, Anlagen und deren Komponenten stellt einen kritischen Beitrag zur Realisierung von intelligenten und autonomen Produktionsmaschinen dar (Ran & Zhou et al. 2019).

Der unvorhergesehene Ausfall von Maschinen durch einen nicht rechtzeitig erkannten Verschleiß der Komponenten äußert sich in der Gesamtanlageneffektivität (GAE, engl. Overall Equipment Effectiveness [OEE]) als Stillstandszeit. Diese unvorhergesehenen Stillstände gilt es durch ein automatisiertes Condition Monitoring der Komponenten zu verhindern, um damit einen Beitrag zur Maximierung der GAE zu leisten (Fink & Wang et al. 2020).

Unter den Maschinenelementen stellt der Kugelgewindtrieb (KGT) eines der bedeutendsten und hauptbelasteten Maschinenelemente in heutigen Produktionsmaschinen dar (Czichos & Habig 2015), (Maier 2015) und ist dabei maßgeblich für den ungeplanten Stillstand von Werkzeugmaschinen verantwortlich (Spohrer 2019). Darüber hinaus bedeutet der Tausch eines KGT einen vergleichsweise hohen Zeitaufwand (Mean Time to Repair [MTT]) (Nesges 2005), (Schopp 2009), woraus sich das besondere Potenzial einer automatisierten Überwachung des KGT in heutigen Produktionsmaschinen ableitet.

Für den aufgrund von Verschleiß hervorgerufenen Ausfall von KGT stellen neben dem durch abrasive Verschleißeffekte hervorgerufenen Vorspannungsverlust die Oberflächenzerrüttung und die daraus resultierenden *Pittings* relevante Verschleißmechanismen dar (Haberker 1998), (Schopp 2009), (Münzing 2017). Um eine spezifische Aussage über vorherrschende Oberflächenzerrüttung auf der KGT-Spindel machen zu können, stellen die empirische Beobachtung und Analyse der Oberflächenzerrüttung am KGT sowie die darauf basierende Konzeption eines Systems zur Überwachung dieses Verschleißmechanismus einen relevanten Ansatz dar. Um den Zustand des KGT und speziell die Oberflächenzerrüttung möglichst genau abzubilden, ist es von besonderer Relevanz, die aus der Oberflächenzerrüttung resultierenden *Pittings* möglichst direkt in deren Entwicklung zu beobachten.

Ein Vorteil von bildgebenden Verfahren zur Zustandsüberwachung ist die direkte Sichtbarkeit von Objekten oder Defekten auf der Oberfläche der beobachteten Komponente.

Ein Bild der Oberfläche des KGT ermöglicht damit einen direkten Einblick und gibt den Anwender*innen eine unmittelbare Information über den Zustand der Komponente und das Vorliegen von Verschleiß. Um neben einer Momentaufnahme der Komponente die Entwicklung einer Schädigung beschreiben zu können, ist es weiterhin notwendig, die Oberfläche des KGT kontinuierlich zu überwachen. Dieses Defizit wird auch von (Forstmann 2010) explizit benannt. Er betont, dass es zum Zeitpunkt des Ausfalles von Maschinenkomponenten zwar Bilddaten gibt, jedoch keine, welche die Verschleißgeschichte dieser Komponenten abbilden. Dies ist aber für die initiale Detektion eines Fehlers sowie von Folgefehlern notwendig (Forstmann 2010). Ein System zur Aufnahme von Bilddaten der Kugelgewindetriebspindel zur Abbildung von Pittings ist zum Zeitpunkt des Verfassens dieser Dissertation noch nicht vorhanden (Schlagenhauf & Hillenbrand et al. 2019).

Neben dem Verständnis über die Entwicklung der Oberflächenzerrüttung sowie der manuellen Auswertung von Defekten in Bilddaten sind zur Realisierung eines Condition-Monitoring-Systems am KGT auch eine automatische Identifikation, Quantifizierung und Prognose des Verschleißzustandes mit Blick auf die Oberflächenzerrüttung anzustreben.

Es zeigt sich, dass sich Verfahren des Maschinellen Lernens zwar zur Verarbeitung von technischen Datensätzen etabliert haben (Samy & Foong et al. 2016), (Geogi & Umarali et al. 2020), (Djenouri & Srivastava et al. 2021) eine dedizierte Betrachtung im Kontext des KGT bisher aber noch nicht erfolgt ist (Schlagenhauf, Ruppelt & Fleischer 2020).

Zwar bestehen zur Überwachung des KGT generelle Ansätze, diese fordern jedoch häufig die manuelle Analyse der Signale durch Domänenexpert*innen. Dies schränkt deren Potenzial zur automatischen Überwachung des Verschleißzustandes des KGT ein. Gleichzeitig besteht speziell im Kontext der Verschleißüberwachung am KGT eine breite Wissensbasis über Tribologie und Verschleiß (Sommer, Heinz & Schöfer 2018) sowie das grundsätzliche Betriebs- und Ausfallverhalten von KGT (Haberkern 1998), (Spohrer 2019). Zur Identifikation und Prognose von Oberflächenzerrüttung am KGT bietet sich aus diesem Grund eine Kombination von Domänenwissen und intelligenten Algorithmen zur Auswertung von aktuell vorliegenden Daten (der Oberflächenzerrüttung) an.

Die Motivation lässt sich damit wie folgt zusammenfassen:

Um den relevanten Verschleißmechanismus der Oberflächenzerrüttung zur Zustandsüberwachung des Kugelgewindetriebes nutzbar zu machen, ist es notwendig, die Entwicklung der Verschleißerscheinungsform der Oberflächenzerrüttung empirisch zu analysieren. Zur Realisierung eines Condition-Monitoring-Systems zur Überwachung der Oberflächenzerrüttung auf der KGT-Spindel ist es weiterhin erforderlich, ein System zur automatisierten Identifikation und Prognose der Oberflächenzerrüttung zu konzipieren.

Zum Zeitpunkt des Verfassens der Dissertation liegt keine empirische Analyse zu der Entwicklung von Oberflächenzerrüttung auf der Kugelgewindetriebsspindel in Bilddaten vor. Gleichzeitig bestehen keine Ansätze, um Oberflächenzerrüttung in den Bilddaten automatisiert zu identifizieren, zu quantifizieren und zu prognostizieren.

1.2 Zielsetzung

Das Ziel der Dissertation ist die empirische Analyse der Entwicklung von Oberflächenzerrüttung auf der Kugelgewindetriebsspindel. Darauf aufbauend ist das zweite Ziel der Arbeit die Untersuchung von aus Daten lernenden Modellen zur automatischen Identifikation und Prognose der Oberflächenzerrüttung auf der Kugelgewindetriebsspindel.

Diese beiden Hauptziele gliedern sich in vier Unterziele: Das erste Unterziel ist die Generierung eines Bilddatensatzes, der die Entstehung und Entwicklung von Oberflächenzerrüttung am KGT über die gesamte Lebensdauer der Komponente abbildet. Die Grenze wird durch das Schadensmerkmal der Oberflächenzerrüttung auf der Spindel beschrieben. Schädigungen in anderen Teilsystemen des KGT sowie weitere Verschleißmechanismen stehen nicht im Fokus dieser Arbeit.

Als zweites Unterziel soll die Entwicklung der Oberflächenzerrüttung in den Bilddaten empirisch analysiert werden, um damit die Erkenntnisse bestehender Arbeiten zu konkretisieren und der Wissenschaftsgemeinschaft einen Bilddatensatz über die Entwicklung der Oberflächenzerrüttung am KGT zur Verfügung zu stellen.

Zur Realisierung von autonomen Produktionsmaschinen wird als drittes Ziel die automatisierte Identifikation von Oberflächenzerrüttung auf Bilddaten des KGT angestrebt. Der Fokus liegt hierbei auf Ansätzen, welche aus Daten lernen (maschinelle Lernverfahren), wobei die Datengrundlage sowohl durch eine manuelle, auf Domänenwissen basierende Merkmalsextraktion als auch durch eine automatisierte Merkmalsextraktion geschaffen wird.

Um die Verfügbarkeit einer Produktionsmaschine möglichst hoch zu halten, ist neben der Identifikation auch eine Prognose des Verschleißfortschrittes notwendig. Die Untersuchung einer geeigneten Prognosefunktion sowie die Identifikation einer auf die Oberflächenzerrüttung zugeschnittenen Definition der Lebensdauer des KGT stellen das vierte Unterziel dar.

Neben dem Erkenntnisgewinn in der Domäne der Zustandsüberwachung von KGT liefern die im Rahmen dieser Dissertation entstandenen Datensätze und Modelle einen Beitrag zur Domäne des Maschinellen Lernens. Die generierten Bilddatensätze sollen der Wissenschaftsgemeinschaft zur weiteren Entwicklung von Klassifikations- und Detektionsmodellen im Kontext der intelligenten Zustandsüberwachung dienen.

1.3 Aufbau der Arbeit

Basierend auf den definierten Zielen ist die Arbeit in acht Kapitel gegliedert. Der Aufbau der vorliegenden Arbeit wird in Abbildung 1-1 dargestellt.

In Kapitel 2 wird zunächst auf die Grundlagen der Tribologie sowie der Zustandsüberwachung am KGT eingegangen. Weitere Bestandteile des Grundlagenkapitels sind die Beschreibung von Bildmerkmalen und die Darstellung der Theorie zur Klassifikation als Grundlage für die Untersuchung von Klassifikationsmodellen.

In Kapitel 3 folgt der Stand der Forschung und Technik auf dem Gebiet der Zustandsüberwachung von KGT. Hier wird besonders auf die bestehenden zustands- und belastungsorientierten Ansätze eingegangen. Ein weiterer Fokus des Kapitels liegt auf der Darstellung des Standes der Forschung und Technik im Bereich der intelligenten Identifikation von Oberflächendefekten auf metallischen Oberflächen. Darauf aufbauend wird das Forschungsdefizit definiert.

In Kapitel 4 wird die Zielstellung konkretisiert und die Vorgehensweise bei der Untersuchung des Forschungsdefizites beschrieben. Weiter werden die geplanten Beiträge zum Stand der Forschung dargestellt.

In Kapitel 5 wird das Vorgehen zur Generierung des Bilddatensatzes präzisiert, welcher die Grundlage für die Untersuchungen zur Schadensentwicklung sowie für die Modellbildung zur Klassifikation von Schädigungen darstellt. Hierbei wird zunächst auf den Versuchsaufbau der Lebensdauerexperimente eingegangen und anschließend die Konzipierung des Sensorsystems zur Datenaufnahme beschrieben. Das Kapitel schließt mit einer detaillierten Betrachtung des Aufbaus des generierten Datensatzes.

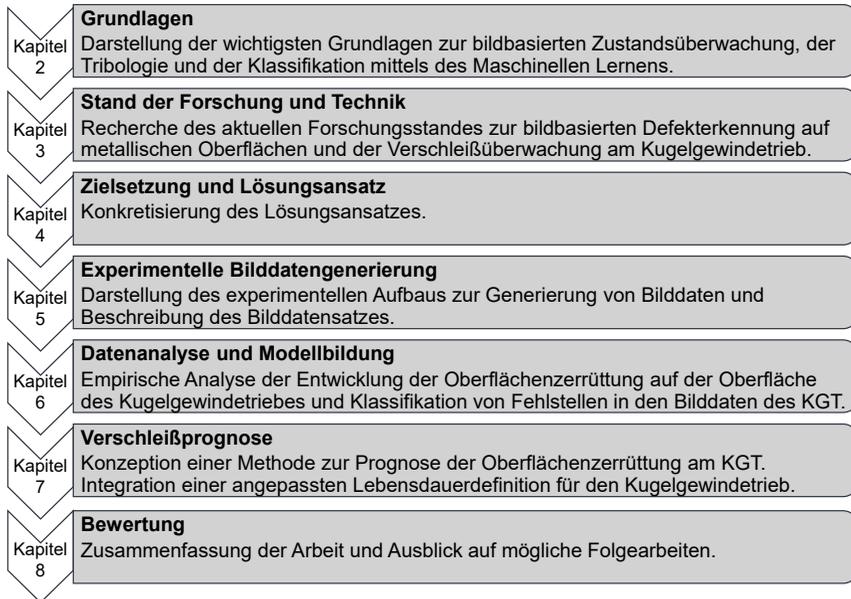


Abbildung 1-1 Aufbau der Dissertation.

In Kapitel 6 findet die Auseinandersetzung mit der empirischen Analyse und der anschließenden auf den generierten Bilddaten basierenden Modellbildung statt. Dabei werden zunächst mittels der Bilddaten die auf der Spindeloberfläche vorherrschenden Verschleißerscheinungsformen beschrieben. Anschließend werden die visuellen Charakteristika von Pittings sowie die Entwicklung der Oberflächenzerrüttung auf der Kugelgewindetriebspinde analysiert. Die Analyse der Bilddaten schließt mit einem Vorschlag zur Definition der 100 % Lebensdauer am KGT² sowie mit einer Diskussion im Kontext bestehender Arbeiten. Im Unterkapitel zur Klassifikation von Fehlstellen auf Bilddaten des KGT werden zunächst die für eine Klassifikation relevanten Merkmale von Pittings spezifiziert und darauf basierend die Klassifikation sowie die Merkmalsrelevanz untersucht. Darauf folgend werden Deep-Learning-basierte Ansätze untersucht. Hierbei wird über ein mehrstufiges Verfahren ein geeignetes Deep-Learning-Modell vorgestellt und dessen Eigenschaften werden untersucht. Der Abschnitt schließt mit der

² Für die Bedeutung sowie die Definition der 100 % Lebensdauer am KGT sei an dieser Stelle auf die Abschnitte 2.2.3 und 6.1.5 verwiesen.

Untersuchung eines dateneffizienten Verfahrens zur Klassifikation von Fehlstellen in technischen Domänen sowie mit einem Handlungsleitfaden zum Einsatz von Maschinellem Lernen zur Zustandsüberwachung.

In Kapitel 7 wird eine Methode zur Prognose der Oberflächenzerrüttung in den Bilddaten konzipiert. Das Kapitel schließt mit einer Diskussion der Erkenntnisse im Kontext bestehender Arbeiten zur Verschleißprognose am KGT.

Kapitel 8 schließt die Dissertation mit einer Bewertung ab. Hierfür wird zunächst eine Zusammenfassung der Arbeit gegeben. Darauf aufbauend werden die Limitationen der Arbeit beschrieben und Handlungs- sowie zukünftige Forschungsbedarfe abgeleitet.

2 Grundlagen

In diesem Kapitel werden die Grundlagen zur Tribologie im Allgemeinen sowie zur Tribologie am KGT beschrieben. Weiterhin werden die Begrifflichkeiten des Condition Monitoring (dt. Zustandsüberwachung) sowie der Predictive Maintenance (dt. vorausschauende Instandhaltung) im Kontext des Maschinenbaus eingeführt. Im Grundlagenkapitel werden überdies die notwendigen Elemente der Bildverarbeitung zur Klassifikation sowie relevante Grundlagen zu Klassifikationsalgorithmen behandelt.

2.1 Tribologie

„Tribologie ist die Wissenschaft und Technik von aufeinander einwirkenden Oberflächen in Relativbewegung. Sie umfasst das Gesamtgebiet von Reibung und Verschleiß, einschließlich Schmierung.“ (Gesellschaft für Tribologie 2002, S. 13)³ Ein Tribosystem, wie es in Abbildung 2-1 dargestellt ist, dient der Umsetzung von energetischen, stofflichen oder informationsbasierten Funktionen und basiert hierbei auf der Relativbewegung eines festen Körpers mit einem festen, flüssigen oder gasförmigen Gegenkörper. Bei der sogenannten Funktionstransformation der Eingangsgrößen hin zu der technischen Funktion des Systems entstehen stets Verlustgrößen. Lager und Führungen, zu denen auch der KGT gezählt werden kann, dienen „der Aufnahme und Weiterleitung von Kräften und ermöglichen dabei eine Rotations- bzw. Translationsbewegung, d. h. sie sind energiedeterminiert“ (Sauer & Steinhilper 2012, S. 5).

2.1.1 Reibung und Verschleiß

Da der Betrieb zahlreicher technischer Systeme, bspw. des KGT, der Ritzel-Zahnstangen-Antriebe oder Kupplungen, auf der Relativbewegung von aufeinander wirkenden Oberflächen beruht, tritt hier stets Tribologie auf (Czichos & Habig 2020). Der (Gesellschaft für Tribologie 2002)³ zufolge wird die Relativbewegung eines festen Körpers mit einem festen, flüssigen oder gasförmigen weiteren Medium oder Körper als tribologische Beanspruchung bezeichnet, welche die Lebensdauer eines technischen Systems begrenzt. Diese Begrenzung resultiert daraus, dass bei tribologischer Beanspruchung stets Reibung und Verschleiß auftreten. Verschleiß wird definiert als „der fortschreitende Materialverlust aus der Oberfläche eines festen Körpers, hervorgerufen durch

³ Gesellschaft für Tribologie (2002), *Verschleiß, Reibung. Definitionen, Begriffe, Prüfung*. <https://www.gft-ev.de/de/arbeitsblaetter/> [22.07.2022].

mechanische Ursachen, d. h. Kontakt und Relativbewegung eines festen, flüssigen oder gasförmigen Gegenkörpers“ (Gesellschaft für Tribologie 2002, S. 5)³. Nach (Deters 2004) sowie (Czichos & Habig 2020) wird der an einem technischen System auftretende Verschleiß durch die Materialien und Beschaffenheit der Oberflächen der wirkenden Körper sowie des verwendeten Schmierstoffes und der Umgebungsbedingungen beeinflusst. Das sogenannte tribologische System, in dem Verschleiß auftritt, ist in Abbildung 2-1 dargestellt. „Das tribologische System beinhaltet alle an einer tribologischen Beanspruchung beteiligten stofflichen Komponenten und ihre Eigenschaften mit den sich bei ihnen ergebenden Veränderungen sowie die für die Beanspruchung charakteristischen Vorgänge und Größen.“ (Gesellschaft für Tribologie 2002, S. 4)³ Auf das tribologische System wirkt das sogenannte Belastungskollektiv, was zu der oben beschriebenen tribologischen Beanspruchung führt. Dieses beschreibt die von außen wirkenden Größen, bspw. den Verlauf und die Dauer der Belastung, die Geschwindigkeit, die herrschenden Temperaturen, die Drücke sowie die Bewegungsart und -form.

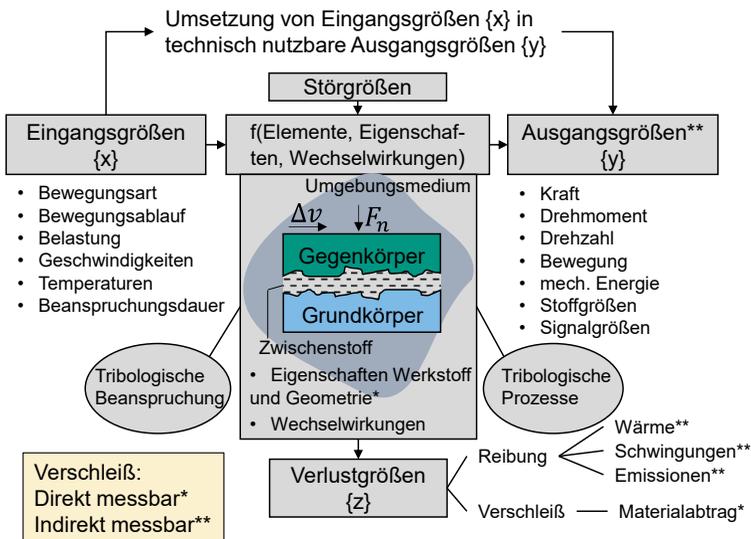


Abbildung 2-1 Technische Funktion des Tribosystems nach (Sauer & Steinhilper 2012).

Der Verschleiß lässt sich somit nicht als Konstante angeben, sondern ist eine Funktion aller im tribologischen System wirkenden Elemente und deren Eigenschaften: $\tau =$

$f(\sigma)$ mit $\tau :=$ Messbare Verschleißgröße, $\sigma :=$ Elemente des tribologischen Systems.

Hierbei ist zu betonen, dass Verschleiß keine Materialeigenschaft, sondern eine Systemeigenschaft ist, die sich abhängig vom aktuellen Kontaktzustand, Schmierungszustand, Reibungszustand sowie Verschleißzustand einstellt (Czichos & Habig 2020).

Die durch die tribologische Beanspruchung hervorgerufenen Oberflächenveränderungen werden als Verschleißerscheinungsformen bezeichnet. „Hierunter sind die sich durch Verschleiß ergebenden Veränderungen der Oberflächenschicht eines Körpers sowie Art und Form der anfallenden Verschleißpartikel zu verstehen.“ (Sauer & Steinhilper 2012, S. 24)

Bei der Definition der Verschleißerscheinungsformen wird explizit auf die visuellen Charakteristika Form und Gestalt Bezug genommen. Das Deutsche Institut für Normung (Deutsches Institut für Normung 1979) führt an, dass direkte Verschleiß-Messgrößen die direkte Form oder Masseänderung eines Körpers durch Verschleiß beschreiben. Als direkte Verschleiß-Maßzahlen können hierbei eindimensionale (Verschleiß Längen), zweidimensionale (Verschleiß Flächen) sowie dreidimensionale (Verschleiß Volumina) Veränderungen der Oberfläche herangezogen werden. Indirekt auf Verschleiß geschlossen werden kann, indem Größen gemessen werden, die sich indirekt durch Verschleiß im System ändern. Beispiele hierfür sind die Temperaturen oder (Körper-) Schallpegel.

2.1.2 Verschleißmechanismen

Die während des Verschleißes ablaufenden physikalischen und chemischen Prozesse werden als Verschleißmechanismen bezeichnet. Die in der Literatur definierten elementaren Verschleißmechanismen an technischen Systemen sind Oberflächenzerrütung, Abrasion, Adhäsion sowie tribochemische Reaktion (Sommer, Heinz & Schöfer 2018). Im Folgenden werden die einzelnen Verschleißmechanismen konkretisiert.

2.1.2.1 Adhäsion

Bei dem Verschleißmechanismus der Adhäsion, die auch als Kaltverschweißung bezeichnet wird, finden stoffliche Wechselwirkungen auf atomarer und molekularer Ebene statt. Ausgangspunkt für die Adhäsion sind Rauheitshügel, die sich auf nie ganz glatten, technischen Oberflächen zeigen. Aufgrund der Rauheitshügel kommt es zu hohen lokalen Pressungen und elastisch-plastischen Verformungen des Materials. Hierdurch

können lokal Oberflächendeckschichten zerstört werden, was zur Bildung von adhäsiven Grenzflächenverbindungen zwischen den metallisch blanken Kontaktbereichen führen kann (Sommer, Heinz & Schöfer 2018). (Achard 1980) modelliert das adhäsive Verschleißvolumen mit $W_V = \kappa * \frac{F_N * s}{H}$, wobei κ der Verschleißkoeffizient in $[\frac{mm^3}{Nm}]$, s der Gleitweg in $[mm]$ und H die Härte des Grundwerkstoffes ist. F_N beschreibt die wirkende Normalkraft. Bei der Bildung und Zerstörung der Grenzflächenverbindungen (Kaltverschweißen) kann es zu Materialübertrag zwischen Kontakt- und Grundkörper sowie zur Bildung von Verschleißpartikeln (Materialfragmenten) kommen (Czichos & Habig 2020). Die so abgetragenen Verschleißpartikel können eine höhere Festigkeit als die Ausgangsmaterialien aufweisen und wirken in Folge als abrasive Teilchen im System. Abbildung 2-2 zeigt typische Erscheinungsformen der Adhäsion.

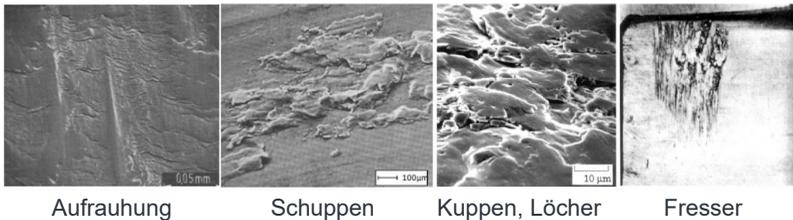


Abbildung 2-2 Verschleißerscheinungsformen der Adhäsion nach (Czichos & Habig 2020) und (Macherauch & Zoch 2011).

(Achard 1980) beschreibt, dass die Adhäsion durch Schmierung verringert werden kann. Die Zusammenhänge zwischen technischen Oberflächen, der Adhäsion sowie der Schmierung sind in Abbildung 2-3 qualitativ dargestellt.

2.1.2.2 Oberflächenzerrüttung

Der zweite in einem tribologischen System auftretende Verschleißmechanismus ist die Oberflächenzerrüttung. „Oberflächenzerrüttung ist allgemein die Folge zyklischer Beanspruchung von Festkörperoberflächen [...]“ (Sommer, Heinz & Schöfer 2018, S. 23) Sie entsteht durch eine häufig periodisch auftretende, wechselnde tribologische Belastung einer Oberfläche, wie sie bspw. beim Abrollen der Wälzkörper in einem Wälzlager vorkommt (Macherauch & Zoch 2011). Typische Verschleißerscheinungsformen der Oberflächenzerrüttung sind Pittings (dt. Grübchen) (siehe Abbildung 2-4).

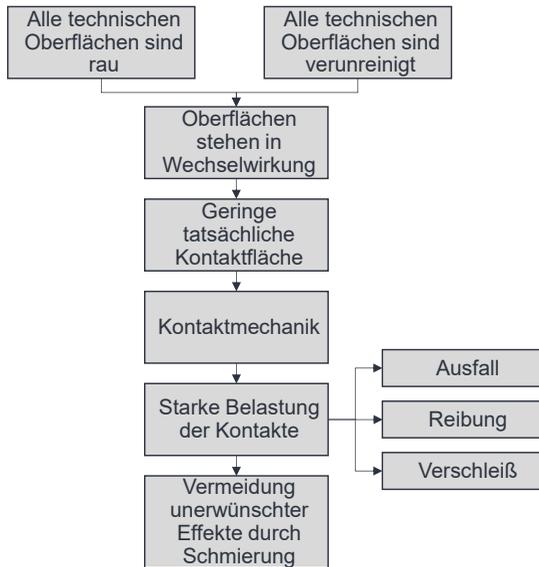


Abbildung 2-3 Qualitative Wechselwirkungen zwischen technischen Oberflächen und dem Schmieren nach (Achar 1980).

Nach (Sommer, Heinz & Schöfer 2018) entstehen Pittings ausgehend von unterhalb der Oberfläche bzw. von der Oberfläche – abhängig davon, ob ein trennender Schmierfilm vorliegt oder nicht. Auf das Entstehen von Pittings wird in Abschnitt 6.1.3 weiter eingegangen. Entsteht ein Pitting von unterhalb der Oberfläche aus, kommt es zum Wandern und Aufstau von Versetzungen im Gefüge sowie zum folgenden Risswachstum. Entsteht ein Pitting ausgehend von der Oberfläche, entsteht es ausgehend von einer Vorschädigung, die dann als Inkubator dient. Liegt ein trennender Schmierfilm vor, so resultiert die Wechselbelastung in einem Wandern von Versetzungen und in einem Versetzungsaufstau, was dann zu einem Risswachstum ausgehend von unterhalb der Oberfläche führt.

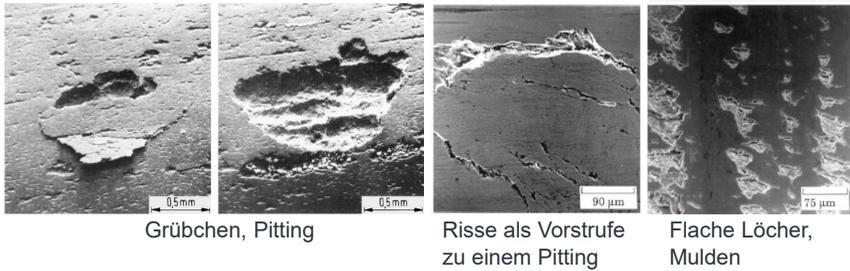


Abbildung 2-4 Verschleißerscheinungsformen des Verschleißmechanismus der Oberflächenzerrüttung nach (Czichos & Habig 2020) und (Sommer, Heinz & Schöfer 2018).

Liegt Mischreibung oder Grenzreibung vor, so kommt es an den Mikrokontakten der nie ganz glatten technischen Oberflächen (Czichos & Habig 2020) zur Werkstoffermüdung ausgehend von der Oberfläche, die dann zu einem Ausbruch von keilförmigen Verschleißpartikeln in Abrollrichtung führt. Bei nicht geschmierten metallischen Gleitvorgängen kommt es weiterhin zu Delaminationsprozessen.

2.1.2.3 Abrasion

Abrasion stellt den dritten Verschleißmechanismus dar, welcher durch die Verschleißerscheinungsformen Riefen, Kratzer, Späne, Ausbrüche und Wellen (Abbildung 2-5) gekennzeichnet ist.

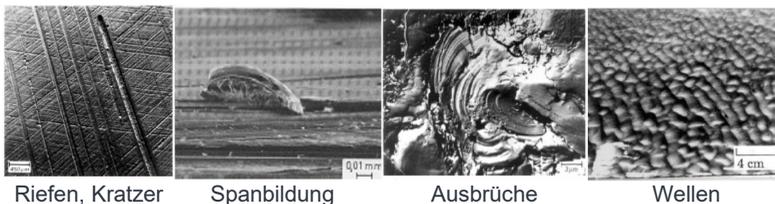


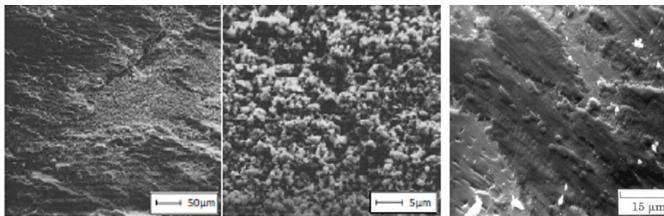
Abbildung 2-5 Verschleißerscheinungsformen des abrasiven Verschleiß nach (Macherauch & Zoch 2011) und (Czichos & Habig 2020).

Abrasion tritt auf, wenn Verschleißpartikel, Abrasivstoffe oder Rauheitsspitzen von Gegenkörpern unter Last über einen weicheren Reibpartner gleiten und dabei in das weichere Material eindringen (Sommer, Heinz & Schöfer 2018). Ist der weiche Grundwerkstoff spröde, wird in diesem Zusammenhang von Mikrobrechen gesprochen. Ähnlich

wie das Verschleißvolumen der Adhäsion, wie es von (Achard 1980) berechnet wird, berechnet (Rabinowicz 1995) das Verschleißvolumen bei der Abrasion mit: $W_V = \kappa_{ab} * \frac{F_N * s}{H}$. Hierbei steht der Verschleißkoeffizient κ_{ab} stellvertretend für den Neigungswinkel des Abrasivpartikels und die Wahrscheinlichkeit der Bildung von Partikeln. Die weiteren Indizes werden analog dem Modell für das Verschleißvolumen der Adhäsion verwendet.

2.1.2.4 Tribochemische Reaktion

Den vierten Verschleißmechanismus stellt die tribochemische Reaktion dar. „Die tribochemische Reaktion beschreibt, die infolge von tribologischer Beanspruchung aktivierte oder begünstigte, chemische Reaktion zwischen Grund- und Gegenkörper mit dem umgebenden Schmierstoff bzw. Umgebungsmedium“ (Gesellschaft für Tribologie 2002, S. 13)³ Dabei werden im Zuge der Reaktion ständig neue Reaktionsprodukte abgerieben oder generiert. Diese treten als Abrasivstoffe in das System ein. Weiterhin können durch tribochemische Effekte spröde Oxidinseln entstehen, die weitere tribologische Verschleißmechanismen nur bedingt ertragen können und dann als Abrasivpartikel in das System übergehen (Abbildung 2-6). Nach (Czichos & Habig 2020) hat die Luftfeuchtigkeit einen wesentlichen Einfluss auf den tribochemischen Verschleiß.



Änderung der Topografie durch tribochemische Reaktion

Reaktionsschicht

Abbildung 2-6 Verschleißerscheinungsform der tribochemischen Reaktion nach (Macherauch & Zoch 2011) und (Czichos & Habig 2020).

2.2 Tribologisches System Kugelgewindetrieb

Kugelgewindetriebe dienen der Wandlung von einer rotatorischen in eine translatorische Bewegung und sind aufgrund der erreichbaren hohen Genauigkeit sowie axialen Steifigkeit ein zentrales und häufig verwendetes Maschinenelement (Bender & Göhlich 2020), (Haberhauer & Kaczmarek 2014), (Murrenhoff 2010) in Werkzeugmaschinen,

Kunststoffspritzgießmaschinen, der Umformtechnik, der Medizintechnik und anderen Maschinentypen und Anwendungen, bspw. der Luftfahrttechnik (Steinmeyer 2021)⁴, (Bosch Rexroth, 2021)⁵.

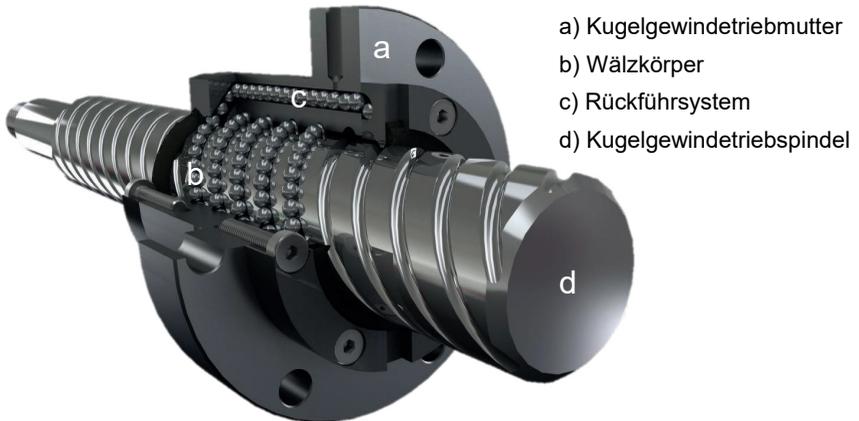


Abbildung 2-7 Komponenten des tribologischen Systems KGT nach (Neff Gewindetrieb GmbH 2018)⁶.

Durch die in (Deutsches Institut für Normung 2011b) definierten und aufeinander wirkenden Hauptkomponenten Kugelgewindtriebmutter, Kugelgewindtriebspindel, Wälzkörper und Kugelrückführsystem (siehe Abbildung 2-7) sowie das wirkende Belastungskollektiv zur Wandlung von einer rotatorischen in eine translatorische Bewegung bildet der KGT ein tribologisches System. Nach (Haberkern 1998) und (Imiela 2006) tritt hierbei Verschleiß auch am korrekt geschmierten und montierten System auf. Die Kinematik des KGT hat hierbei nach (Forstmann 2010) einen entscheidenden Einfluss auf das Verschleißverhalten des KGT, da es aufgrund der reversierenden Betriebsweise regelmäßig zu einem Abbruch des Schmierfilmes und damit, wie in Abschnitt 2.1 beschrieben, zu einem erhöhten Verschleiß an den Rauheitsspitzen kommt. Das im tribologischen System wirkende Belastungskollektiv ist hierbei je nach Einsatzgebiet

⁴ Steinmeyer (2021), *Steinmeyer Kugelgewindtriebe*. https://www.steinmeyer.com/kataloge/steinmeyer_gesamtkatalog/files/assets/common/downloads/publication.pdf [25.07.2022].

⁵ Bosch Rexroth (2021), *Gewindetriebe*. <https://www.boschrexroth.com> [09.09.2021].

⁶ Neff Gewindetrieb GmbH (2018), *Gewindetriebe Spindelhubgetriebe*. https://www.neff-gewindetriebe.de/fileadmin/Servicecenter/Kataloge/20180222_Kat_D_Downloadversion.pdf [24.05.2022].

des KGT unterschiedlich, was in der Folge zu unterschiedlichen dominierenden Verschleißmechanismen führt. Zur Veranschaulichung soll der Einsatz eines KGT in einer Spritzgießmaschine (Forstmann 2010) mit dem Einsatz zur Verstellung von Helikopter-Rotorblättern (Münzing 2017) verglichen werden. Bei Einsatz in einer Spritzgießmaschine hat der KGT immer gleichmäßige, reversierende Bewegungsabläufe zu vollziehen. Es wirken große Kräfte und die Anforderungen an die Genauigkeit sind teilweise gering. Spielfreiheit ist bspw. beim Einsatz in der Einspritzachse einer Spritzgießmaschine oft nicht erforderlich. Der Verschleiß tritt laut (Forstmann 2010) hierbei prozessbedingt immer an denselben Achspositionen auf. Prozessbedingte Kontaminierungen oder Kontaminierungen von außen wie Späne oder Verunreinigungen (Spohrer 2019) können beim industriellen Einsatz mit großer Wahrscheinlichkeit niemals ganz ausgeschlossen werden. Beim Einsatz des KGT in einem Rotorstellwerk eines Helikopters beschreibt (Münzing 2017), dass die Kugeln beim oszillierenden Betrieb bei kleinen Schwenkwinkeln (Kurzhubbetrieb) praktisch in ihrer Position verharren und sich somit bereichsabhängig unterschiedliche Lebensdauern ergeben. Aufgrund der geringen Hübe ist eine kontinuierliche Schmierung des Systems nicht immer gegeben, was zu Verschleißerscheinungsformen wie Fressen oder der in Abbildung 2-9 dargestellten Reibkorrosion führen kann. Ein ähnliches Verhalten ist auch beim Einsatz von KGT in Schleifmaschinen zu beobachten (Spohrer 2019).

Anwendungsabhängig erfahren KGT somit unterschiedliche Lastkollektive und es wirken unterschiedliche externe Einflussgrößen im Betrieb. Je nach Lastkollektiv, Einflussgrößen und Vorschädigungen des KGT können somit unterschiedliche Verschleißmechanismen im tribologischen System KGT vorherrschend sein und es ist nicht möglich, eine pauschale Aussage über einen einzigen dominierenden Verschleißmechanismus in einem KGT über alle Anwendungen hinweg zu treffen.

2.2.1 Verschleißmechanismen am Kugel-Laufbahn-Kontakt

Die in der Tribologie beschriebenen Verschleißmechanismen Abrasion, Adhäsion, Oberflächenzerrüttung und tribochemische Reaktion treten nach (Haber Kern 1998) und (Imiela 2006) auch am korrekt montierten und geschmierten KGT auf. Verschleiß am KGT führt zu den bei (Haber Kern 1998), (Imiela 2006), (Münzing 2017) und (Spohrer 2019) beschriebenen Ausfallmechanismen plötzlicher Frühausfall, langsamer Vorspannungsverlust und langsamer Spätausfall des KGT. Aufgrund der durch die Spiralform des KGT erzwungenen Kinematik der Kugeln tritt nach (Forstmann 2010) beim Betrieb

des KGT immer eine Mischung aus Wälz-, Bohr- und Gleitreibung⁷ auf, die unterschiedliche Verschleißmechanismen fördert. Hierbei treten die Verschleißmechanismen nicht einzeln, sondern in Kombination auf. Weiterhin tritt nach (Forstmann 2010) aufgrund der reversierenden Betriebsweise des KGT Mischreibung auf, die den Verschleiß durch den direkten Kontakt von Kugeln und Rauheitsspitzen der Oberflächen fördert. Nach (Haberkern 1998) geht einem Ermüdungsschaden häufig eine Schädigung der Oberfläche infolge der Verschleißmechanismen Abrasion und Adhäsion voraus. Die Vorschädigung fördert damit ein Risswachstum von der Oberfläche aus. Abrasive und adhäsive Verschleißerscheinungsformen, welche das Risswachstum ausgehend von der Oberfläche fördern, sind bspw. Fresser oder Fremdkörpereindrücke in der Oberfläche. (Münzing 2017) beschreibt in seiner Dissertation, dass bei durch Stillstandmarkierungen⁸ (engl. False Brinelling) vorgeschädigten Lagern eine gegenüber der nominellen Lebensdauer erheblich verkürzte Ermüdungslebensdauer beobachtet werden konnte. Es lässt sich somit annehmen, dass sich die Verschleißerscheinungsformen gegenseitig begünstigen. In Abbildung 2-8 werden Verschleißmechanismen und deren generelle Verschleißerscheinungsformen sowie die Erscheinungsformen an der Kugelgewindetriebspindel gegenübergestellt. (Forstmann 2010) stellt hierbei die Annahme von (Haberkern 1998), in Frage, dass tribochemischer Verschleiß aufgrund der überwiegend geringen Temperaturen von $< 200\text{ °C}$ im Maschinenbau meist vernachlässigt werden könne, und beschreibt, dass im weiteren Verlauf der tribologischen Beanspruchung durch tribochemische Reaktion gebildete Oxidschichten abgetragen werden können. Diese Oxidteilchen wirken oft als harte Abrasivteilchen im System. Auch ein Nicht-Abtragen der spröden Oxidschichten kann einen Angriffspunkt für Oberflächenzerrüttung darstellen. (Forstmann 2010) führt weiterhin aus, dass Fremdkörper durch abrasive Effekte bei zähen Grundkörperwerkstoffen Ritzspuren und bei spröden Grundkörperwerkstoffen muschelförmige Ausbrüche verursachen, welche visuell sichtbar sind und als Abrasivpartikel im System verbleiben. Das Überrollen von Rauheitsspitzen, mechanische Verklammerungseffekte sowie das Einklemmen und Einbetten von harten Partikeln fördern adhäsive Effekte. Dadurch können weitere Partikel von der Oberfläche abgelöst werden und als Abrasivstoffe im System wirken.

⁷ Zur Definition verschiedener Reibungsarten sei auf die Grundlagenliteratur (Czichos & Habig (2020) verwiesen.

⁸ Stillstandsmarkierungen werden durch Mikrooszillationen hervorgerufen (Schadow (2016).

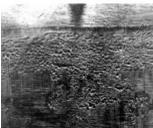
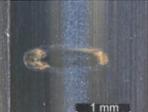
Kugel - Laufbahn Kontakt					
a	Wechselbeanspruchung der Spindellaufgänge	Läppbewegungen	Gleitanteile durch Bohr- und Schlupfbewegungen		
b	Ermüdung ①	② Abrasion und Adhäsion ③	Tribochemische Reaktion ④		
c	Materialabtragungen in Form von Grübchen	Furchen, Ritzen	Fressen, Abtragen		
d	Risse, Partikelbildung, Löcher, Mulden, Delamination, Ablätterung, Abplatzer, Narben, Ausbrüche, Eindrücke, Pitting, Graufleckigkeit, Mikropitting, Riffel	Riefen, Muschelförmige Ausbrüche, Schultern, Querwellen, Rillen, Mikropittings, Aufrauung, Eindrücke	Fresser, Riefen, Materialübertrag, Schuppen, Kuppen, Löcher, Mulden, Ansmierungen		
e	① 	① ③ 	② ③ 	② 	
f	① 	① 	④ 	① ③ 	③ ④ 
a) Kinematik b) Verschleißmechanismus c) Merkmale d) Verschleißerscheinungsformen					
e) Bildliche Verschleißerscheinungsformen auf metallischen Oberflächen					
f) Bildliche Verschleißerscheinungsformen auf der KGT Spindel					

Abbildung 2-8 Verschleißmechanismen und Verschleißerscheinungsformen am Kugel-Laufbahn-Kontakt des KGT nach (Czichos & Habig 2020), (Münzing 2017), (NSK 2020)⁹ und (Sommer, Heinz & Schöfer 2018).

⁹ NSK (2020), *Schadensdiagnose. Schadenstypen*. <https://www.nsk-europe.de/de/services/troubleshooting/damage-by-type.html> [21.04.2020].

Somit wird der abrasive Verschleiß gefördert. Neben dem Entstehen der Oberflächenzerrüttung unterhalb der Oberfläche aufgrund von periodischen Spannungsspitzen entsteht die Oberflächenzerrüttung häufig auch an im System befindlichen und überrollten Partikeln (Forstmann 2010). Ein Beispiel, bei dem ein Pitting sich aufgrund von Oberflächenzerrüttung bildet und dann mit adhäsiv bedingten Ablätterungen vermischt, ist in Abbildung 2-9 links oben dargestellt. Unten in Abbildung 2-9 sind der Übergang und die spätere Überlagerung von Passungsrost und Ausbrüchen, rechts oben ist die Überlagerung von Graufleckigkeit und Grübchen aufgeführt. Es ist erkennbar, dass die Verschleißerscheinungsformen visuell sichtbar sind.



Abbildung 2-9 Sich gegenseitig beeinflussende Verschleißmechanismen. Oben links: Innenring eines Kugellagers (Schaeffler 2000)¹⁰, oben rechts: Pendelrollenlager (Sommer, Heinz & Schöfer 2018), unten: Kugelgewindetriebspindel (Münzing 2017).

Damit kann angenommen werden, dass sich alle Verschleißmechanismen gegenseitig beeinflussen und es zu einer Überlagerung der Verschleißmechanismen und Verschleißerscheinungsformen kommt. Dies wird auch von (Sommer, Heinz & Schöfer

¹⁰ Schaeffler (2000), *Wälzlagerschäden. Schadenserkennung und Begutachtung gelauener Wälzlager*. https://www.schaeffler.de/remotemedien/media/_shared_media/08_media_library/01_publications/schaeffler_2/publication/downloads_18/wl_82102_2_de_de.pdf [25.07.2022].

2018) bestätigt, welche beschreiben, dass „bei vielen Verschleißvorgängen eine zeitliche Abfolge verschiedener Verschleißmechanismen auftritt“ (Sommer, Heinz & Schöfer 2018, S. 135), dass „die Grundmechanismen Adhäsion, Abrasion, Oberflächenzerrüttung, tribochemische Reaktion und Ablation [...] nur in seltenen Fällen einzeln auftreten“ (Sommer, Heinz & Schöfer 2018, S. 19) und dass feste Partikel durch Überrollen einen Ausgangspunkt für weiteren Verschleiß darstellen (Sommer, Heinz & Schöfer 2018). Diese Beobachtungen stützt auch (Helwig 2018) bei der Untersuchung der Verschleißmerkmale am KGT. Er äußert die Vermutung von sich gegenseitig beeinflussenden Verschleißmechanismen.

Die Wechselwirkungen lassen sich damit, wie in Abbildung 2-10 dargestellt, auf den KGT abstrahieren. Der Verschleißfortschritt des KGT wird hierbei durch das wirkende Belastungskollektiv, die Geometrie der Verschleißpartner sowie die Werkstoffeigenschaften der Verschleißpartner beeinflusst. Das Belastungskollektiv, die Geometrie sowie die Werkstoffeigenschaften führen zu den bekannten Verschleißmechanismen der Oberflächenzerrüttung, Abrasion, Adhäsion und tribochemischen Reaktion. Die Verschleißmechanismen äußern sich in den für sie charakteristischen Verschleißerscheinungsformen. Neben diesen Veränderungen der Oberfläche und der Energiedissipation führen sie auch zum Materialverlust, woraus weitere Verschleißpartikel resultieren können. Durch das Überrollen der Oberfläche und die damit einhergehende Ermüdung entstehen Änderungen im Gefüge. Hervorgerufen durch die Verschleißmechanismen Abrasion und Adhäsion kommt es in der Folge zu Einbettungen, die wiederum Ausgangspunkt für die Oberflächenzerrüttung und für darauf folgende abrasive und adhäsive Effekte darstellen. Diese Verschleißpartikel und Verschleißerscheinungsformen stellen eine Vorschädigung des Systems dar, die mit einer hohen Wahrscheinlichkeit bei weiterem Betrieb die übrigen Verschleißmechanismen beeinflusst. Diese Wechselwirkung wird durch den teilweise durchgezogenen Pfeil in Abbildung 2-10 angedeutet. Somit wirkt die Vorschädigung als weitere Eingangsgröße auf die Verschleißentwicklung und damit auch auf die Lebensdauer des Systems ein.

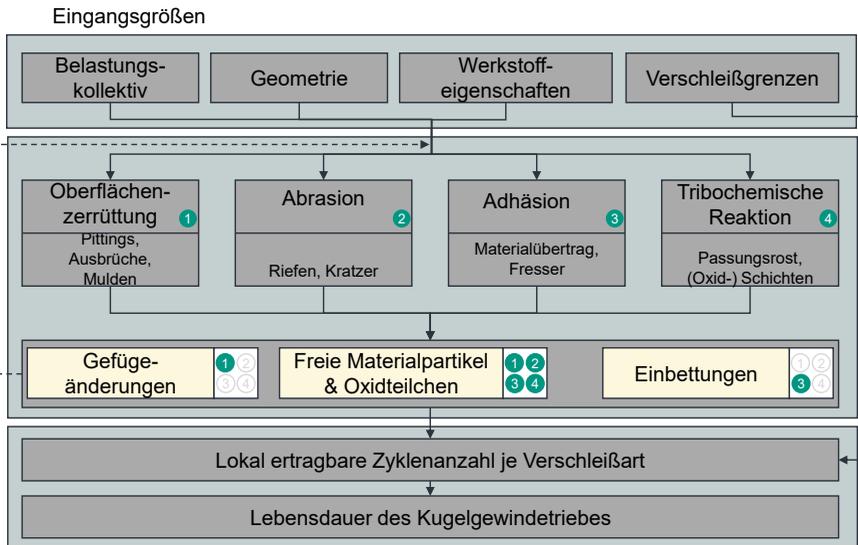


Abbildung 2-10 Qualitativer Prozess der sich gegenseitig beeinflussenden Verschleißmechanismen angelehnt an (Forstmann 2010).

2.2.2 Ausfallmechanismen am Kugelgewindetrieb

(Haberker 1998) definiert in seiner Dissertation die drei dem Ausfall des KGT zugrunde liegenden Ausfallmechanismen plötzlicher Frühausfall, langsamer Vorspannungsverlust und langsamer Spätausfall. Diese Ausfallmechanismen wurden in nahezu allen nachfolgenden Arbeiten zur Beschreibung des Ausfallverhaltens von KGTs aufgegriffen und verwendet.

Die drei Ausfallmechanismen lassen sich anschaulich im Kontext der Badewannenkurve in Abbildung 2-11 diskutieren. Der plötzliche Frühausfall manifestiert sich relativ früh zu Beginn der Betriebszeit. Im Kontext der Badewannenkurve lässt sich die hohe Ausfallrate mit bspw. Einbaufehlern, Fehlern im Material oder Fertigungsfehlern, die zu einem Verklemmen des Systems führen, erklären. (Haberker 1998) beschreibt als Ursache für den plötzlichen Frühausfall Umformmechanismen im Rückführsystem, die im Weiteren zu einem Verklemmen und anschließenden Bruch des Systems führen. Den Frühausfällen folgt eine längere Zeit, in der sich unter normalem Betrieb nur wenige

Ausfälle ereignen. Diese Phase wird als Konsolidierungsphase bezeichnet und beschreibt ein funktionsfähiges und eingefahrenes System. Dies bedeutet, dass keine initialen Fehler vorliegen und das System durch den natürlichen Verschleiß noch nicht so weit geschädigt ist, dass es ausfällt. Die dritte Phase der Badewannenkurve beschreibt die durch den natürlichen Verschleiß eines Systems auftretenden Spätausfälle. Dieser Phase lassen sich die Ausfallmechanismen langsamer Spätausfall sowie langsamer Vorspannungsverlust zuordnen. Der langsame Spätausfall wird nach (Haberkern 1998) ausschließlich durch die Oberflächenzerrüttung und die daraus resultierende Verschleißerscheinungsform der Pittings verursacht. Nach einer Phase des unauffälligen Betriebs (Konsolidierungsphase der Badewannenkurve) entsteht eine durch Oberflächenzerrüttung auf den Kugeln und Laufbahnen verursachte Laufunruhe, die schlussendlich zum Verklemmen und Ausfall des Systems führt.

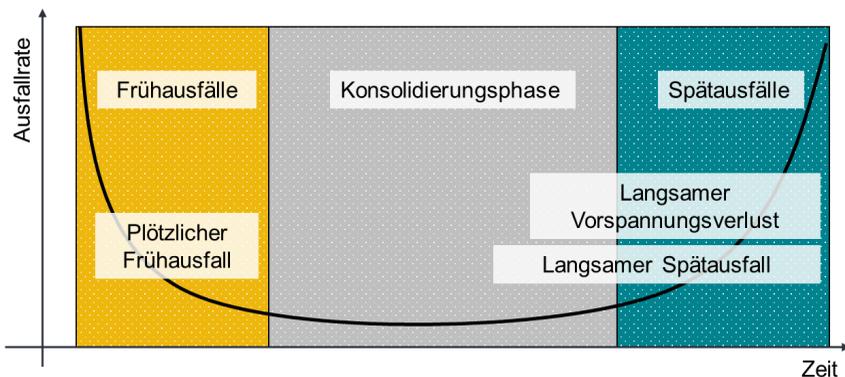


Abbildung 2-11 Badewannenkurve nach (Heise 2002).

(Haberkern 1998) beschreibt hierbei zwar, dass der langsame Spätausfall erst nach Überschreiten der nominellen Lebensdauer des KGT auftritt, relativiert dies allerdings damit, dass Betriebsbedingungen vorausgesetzt werden, die andere Schadensmechanismen sowie Verschmutzungen ausschließen. Der dritte von (Haberkern 1998) beschriebene Ausfallmechanismus ist der langsame Vorspannungsverlust, welcher zu einer Vorspannungsabnahme im System und schließlich zum Ausfall der Achse führt. Bei diesem Ausfallmechanismus werden die Kugeln durch Regelkreisschwingungen oder ständigen Kurzhubbetrieb, wie er bspw. in Schleifmaschinen auftritt, geläpft. Der Ausfallmechanismus des langsamen Vorspannungsverlustes wird nach (Haberkern 1998)

durch den Verschleißmechanismus der Abrasion verursacht. Der Verschleißmechanismus der tribochemischen Reaktion wird in den Beschreibungen der Ausfallmechanismen nicht berücksichtigt, was von (Forstmann 2010) kritisiert wird. (Münzing 2017) beobachtet tribochemische Reaktionen am KGT in Form von Reibkorrosion.

2.2.3 Lebensdauer des Kugelgewindetriebs

KGT zählen zu den am stärksten belasteten und am häufigsten ausfallenden Maschinenelementen in Werkzeugmaschinen und tragen durch ihren Ausfall zu einer reduzierten OEE bei (Dehli 2019), (Fleischer & Schopp et al. 2007), (Schopp 2009), (Spohrer 2019). In der Industrie werden vermehrt zustandsorientierte Instandhaltungsstrategien eingesetzt, um auf Basis des aktuellen Zustandes einer Maschine deren Lebensdauer abschätzen und Instandhaltungsmaßnahmen rechtzeitig einleiten zu können (Deutsches Institut für Normung 2018a).

In den folgenden Unterkapiteln werden die Grundlagen zur Lebensdauer von KGT dargestellt. Zunächst wird der Begriff des Abnutzungsvorrates diskutiert, bevor auf die normativen Grundlagen zur Lebensdauerberechnung eingegangen wird. Das Unterkapitel schließt mit der Einführung der Begrifflichkeiten ‚Brauchbarkeitsdauer‘ und ‚Betriebszeit‘. Im Rahmen der Arbeit von (A_Hofmann 2021) wurden die (normativen) Grundlagen im Kontext der Lebensdauerberechnung des KGTs analysiert. Die Erkenntnisse sind im folgenden Abschnitt zusammengefasst dargestellt.

2.2.3.1 Abnutzungsvorrat

Nach (Gronau 2008) wird der sogenannte Abnutzungsvorrat einer Maschine verwendet, um deren Zustand abzuschätzen. Der Abnutzungsvorrat beschreibt den Vorrat an noch zu ertragender Abnutzung, bevor eine Maschine eine unter festgelegten Bedingungen definierte Funktionserfüllung nicht mehr gewährleisten kann (Deutsches Institut für Normung 2019). Er ist zusammen mit der sogenannten Abnutzungsgrenze in Abbildung 2-12 dargestellt. Die Abnutzungsgrenze beschreibt einen anwendungsspezifisch definierten Wert des Abnutzungsvorrates, bei dessen Unterschreiten die Maschine in Stand gesetzt werden muss. In Abbildung 2-12 sind drei unterschiedliche Abnutzungsgrenzen dargestellt. Das Ziel der Instandhaltung muss es sein, eine maximale Zeit zwischen zwei Ausfällen (engl. Mean Time between Failure, MTBF) sowie eine minimale Zeit zur Instandhaltung (engl. Mean Time to Restoration, MTTR) zu erreichen (Fleischer

& Broos et al. 2009). Es ist ersichtlich, dass die Wahl der Abnutzungsgrenze maßgeblich die MTBF- und MTTR-Zeiten beeinflusst. Wird eine Maschine (zu) früh instandgesetzt, entspricht dies einer (zu) hohen Abnutzungsgrenze (Fall 1 - gelb). In diesem Fall ist noch ein hoher Abnutzungsvorrat vorhanden, was den weiteren Betrieb der Maschine erlaubt. Durch die rechtzeitig planbaren Instandhaltungsmaßnahmen und die selten auftretenden ungeplanten Ausfälle kann die MTTR-Zeit minimiert werden. Wird eine Komponente instandgesetzt, wenn der Abnutzungsvorrat fast aufgebraucht ist, entspricht dies einer Instandhaltungsstrategie, die kurz vor einer reaktiven, ausfallbedingten Instandhaltung angesiedelt ist (Reichel, Müller & Haeffs 2018) (Fall 3 - rot). In diesem Fall wird die MTBF maximiert, die MTTR ist allerdings aufgrund der häufigen, ungeplanten Ausfälle hoch. Dies resultiert aus der Tatsache, dass ein kurz bevorstehender Ausfall der Komponente nicht zweifelsfrei bestimmt werden kann. Diese Strategie kann im Kontext der Produktion als zu spät und damit ebenfalls suboptimal angesehen werden.

Die optimale Abnutzungsgrenze ist mit dem Fall 2 (grün) abgebildet. Hierbei wird die MTBF möglichst hochgehalten und gleichzeitig versucht, die Komponenten rechtzeitig vor einem Ausfall geplant zu warten. Die OEE der Maschine wird durch die hohe MTBF und die niedrige MTTR maximiert (Dehli 2019). Hierbei profitiert ein Unternehmen, wenn ein bevorstehender Maschinenausfall möglichst zweifelsfrei festgestellt werden kann. Damit ist es möglich, Abnutzungsgrenze 2 nahe der Abnutzungsgrenze 3 (siehe Abbildung 2-12) zu halten und gleichzeitig einen ungeplanten Ausfall zu vermeiden.

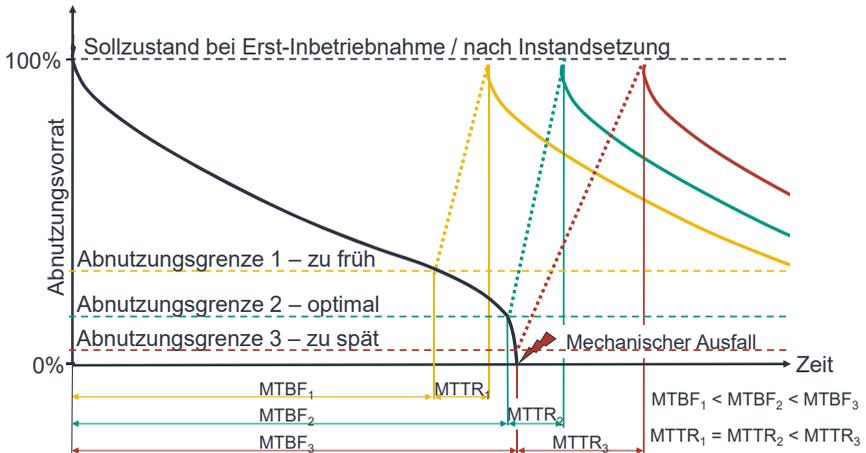


Abbildung 2-12 Abnutzungsvorrat und Abnutzungsgrenze (A_Hofmann 2021).

2.2.3.2 Normative Basis zur Lebensdauerberechnung des Kugelgewindetriebes

Die Lebensdauer des KGTs wird durch (Deutsches Institut für Normung 2011a, S. 13) definiert als „diejenige Anzahl von Umdrehungen, die eine Kugelgewindemutter (oder eine Kugelgewindespindel) in Bezug auf die Kugelgewindespindel (die Kugelgewindemutter) ausführt, bevor sich erste Anzeichen von Materialermüdung an einem der beiden Teile oder am Wälzkörper zeigen“. Da diese Definition keine Abschätzung der noch verfügbaren Restlebensdauer zulässt, definiert dieselbe Norm die sogenannte L_{10} oder nominelle Lebensdauer als „für einen einzelnen KGT oder für eine Gruppe offensichtlich gleicher KGT, die unter gleichen Bedingungen laufen, diejenige Lebensdauer, die mit 90 % Erlebenswahrscheinlichkeit erreicht wird“ (Deutsches Institut für Normung 2011a, S. 13). Um mittels dieser Definition eine Abschätzung über die tatsächliche Lebensdauer des Systems zu erhalten, wird die sogenannte dynamische axiale Tragzahl C_a definiert, als die „konstante zentrische axiale Last, die ein KGT theoretisch mit einer nominellen Lebensdauer von 10^6 Umdrehungen aufnehmen kann“ (Deutsches Institut für Normung 2011a, S. 13). Die dynamische axiale Tragzahl beschreibt damit indirekt, dass 90 % aller KGT, die mit einer Kraft äquivalent zur definierten dynamischen axialen Tragzahl belastet werden, 10^6 Umdrehungen aufnehmen können, bevor die Lebensdauer erreicht ist. Mittels der definierten dynamischen axialen Tragzahl lässt sich die nominelle Lebensdauer in Umdrehungen nun einfach nach folgender Formel basierend

auf den Experimenten von (Lundberg & Palmgren 1952) berechnen: $L_{10} = \left(\frac{C_a}{f_w * F_a}\right)^3 * 10^6$ mit L_{10} [N] als der nominellen Lebensdauer, C_a der dynamischen axialen Tragzahl [N], F_a [N] der Axialbelastung und f_w dem Belastungsfaktor. Durch die Definition der nominellen Lebensdauer ist impliziert, dass nicht jeder KGT bei Erreichen der nominellen Lebensdauer mechanisch defekt ist, sondern es ist dann lediglich keine Berechnung mittels dieser Berechnungsvorschrift mehr möglich. Die Norm beschreibt also ein Mindesthaltbarkeitsdatum, das 90 % der KGT erreichen. Die Norm stellt dabei aber keine Methode zur Abschätzung des tatsächlichen Verschleißes bereit. So ist es durchaus möglich, dass ein individueller KGT aufgrund eines Frühausfalles vor Erreichen der nominellen Lebensdauer ausfällt bzw. die nominelle Lebensdauer überschreitet. Auch das Erreichen der definierten Lebensdauer muss nicht den technischen Ausfall des Systems bedeuten (Klein 2011). Zu betonen ist, dass die Normenwerke für KGT lediglich die Oberflächenzerrüttung als Ursache für das Lebensdauerende definieren und die Berechnungsvorschriften in Teilen an die Systematik zur Wälzlagerauslegung angelehnt sind (Hilmer 1978). Weitere Schädigungsmechanismen bleiben unberücksichtigt. Hierin unterscheidet sich die Norm für KGT nicht von den lebensdauerbestimmenden Normen weiterer wälzgelagerter Systeme, bspw. Wälzlager (Deutsches Institut für Normung 2010c) oder Linear-Wälzlager (Deutsches Institut für Normung 2018b). Weiterhin werden zur Lebensdauerberechnung weder die Art des Schmierstoffes noch eine Belastung durch Abrasivpartikel berücksichtigt (Forstmann 2010). Diese Vereinfachungen können nach (Spohrer 2019) zu einer signifikanten Abweichung der Lebensdauer des KGTs von der L_{10} Lebensdauer führen.

2.2.3.3 Praktische Ergänzungen zur Lebensdauer des Kugelgewindetriebes

Die in Abschnitt 2.2.3.2 eingeleiteten Grundlagen zur normativen Bestimmung der Lebensdauer eines KGT sollen in diesem Unterkapitel um weitere praxisrelevante, in der Norm (Deutsches Institut für Normung 2011a) nicht spezifizierte Grundlagen ergänzt werden.

In der Theorie beschreibt die 100-%-Lebensdauer eines Systems die erreichte Abnutzungsgrenze bei dem mechanischen Ausfall desselben. Wie in Abbildung 2-12 dargestellt, geht daraus hervor, dass die Definition einer solchen Abnutzungsgrenze für den

Einsatz eines KGT für die meisten industriellen Anwendungen ungeeignet ist. (Drescher 2015)¹¹, (Imiela 2006) und (Münzing 2017) beschreiben zudem, dass KGT durch Verunreinigungen im System, fehlerhaften Betrieb oder überhöhten Verschleiß oft vor Erreichen der theoretischen Lebensdauer ausgetauscht werden müssen, da das System die geforderte Funktionalität nicht mehr zuverlässig erbringen kann. Die praktische Definition der 100 % Lebensdauer weicht von der theoretischen systembedingten Lebensdauer in den jeweiligen Normen für wälzgelagerte Systeme ab. Die Normen, welche die Lebensdauer von KGT (Deutsches Institut für Normung 2011b), von Wälzlagern (Deutsches Institut für Normung 2010c), Profilschienenführungen (Deutsches Institut für Normung 2010a) sowie von Linear-Wälzlagern (Deutsches Institut für Normung 2018b) definieren, beziehen sich bei der Definition der Lebensdauer auf eine Anzahl an zu ertragenden Umdrehungen (Wälzlager, KGT) oder eine zurücklegbare Strecke (Profilschienenführungen, Linear-Wälzlager). Die Lebensdauer einer Komponente wird als erreicht definiert, wenn sich Anzeichen von Materialermüdung auf den Teilkomponenten des Systems zeigen. Damit auf dieser Basis eine Dimensionierung für wälzgelagerte Systeme vorgenommen werden kann, wird bspw. für den KGT die oben beschriebene sogenannte L_{10} Lebensdauer eingeführt. Als Ausfallkriterium wird hier, wie auch in den Normenwerken zu weiteren wälzgelagerten Systemen, lediglich das Erscheinungsbild der Oberflächenzerrüttung berücksichtigt. (Huf 2012) kritisiert an der nominellen Lebensdauer des KGTs, dass deren Erreichen nach Norm nicht zwangsläufig die praktische Funktionsunfähigkeit des Systems bedeutet, und auch (Schaeffler 2000)¹⁰ beschreibt, dass das System nach Auftreten eines Defektes meist noch eine gewisse Zeit weiterbetrieben werden kann. (Spohrer 2019) hebt außerdem hervor, dass in der Berechnungsvorschrift für die nominelle Lebensdauer des KGTs adhäsive und abrasive Effekte nicht direkt berücksichtigt werden. Wie in Abschnitt 2.2.1 beschrieben, ist bekannt, dass sich die Verschleißmechanismen ab einem bestimmten Zeitpunkt überlagern und nicht mehr komplett getrennt betrachtet werden können. Dies erklärt auch, dass die Lebensdauern der KGT von (Fleischer & Lanza et al. 2013), (Haber Kern 1998) und (Schopp 2009) als teils über und teils unter der nominellen Lebensdauer des KGTs eingestuft wurden. Die Autoren (Haber Kern 1998), (Münzing 2017) und (Spohrer

¹¹ Drescher, U. (2015), *Leitfaden für Konstrukteure – Schmiermittelauswahl für Kugelgewindetriebe*. <http://www.konstruktionspraxis.vogel.de/leitfaden-fuer-konstrukteure-schmiermittelauswahl-fuer-kugelgewindetriebe-a-506405/> [22.07.2022].

2019) stufen die in der Norm zur Berechnung der L_{10} Lebensdauer des KGTs vorgegebenen Berechnungen als für die reale Anwendung unzureichend ein. In den Arbeiten von (Forstmann 2010), (Jarosch 2008) und (Münzing 2017) wurde versucht, die Definition der Lebensdauer des KGTs zu erweitern, ohne dass diese Erweiterungen in der Norm übernommen wurden. Diese Arbeiten beschäftigen sich als Erweiterung der Norm zwar mit einer Erweiterung der Berechnungsvorschrift, sie beantworten allerdings nicht die Frage nach einer anwendungsübergreifenden Definition der 100 % Lebensdauer. Es fehlt ein „allgemeingültiges, allgemein akzeptiertes und quantifizierbares Abbruchkriterium“ (Schopp 2009, S. 85).

2.2.3.4 Brauchbarkeitsdauer und Betriebszeit

Um die Lebensdauer mit der Festlegung einer Abnutzungsgrenze in Einklang zu bringen, werden im Folgenden die Begriffe der Brauchbarkeitsdauer und der Betriebszeit eingeführt. Die Betriebszeit einer Komponente ist definiert durch die Zeit, innerhalb welcher die Komponente „die geforderte Funktion erbringt“ (Deutsches Institut für Normung 2018a, S. 33). Die Betriebszeit ist damit unabhängig von etwaigen Schädigungen im System lediglich über die zuverlässige Erfüllung einer spezifischen Aufgabe definiert. Sie ist damit für jede Anwendung individuell festzulegen. Zur Überprüfung der Betriebszeit wird das Erreichen der definierten Systemanforderungen als Kriterium herangezogen. Die Betriebszeit kann erreicht sein, obwohl das System die Grundfunktionalität nach wie vor erfüllt, und damit von der theoretischen Lebensdauer einer Komponente abweichen. Unter diesem Blickwinkel kann die Betriebszeit theoretisch als ideale Größe zur Bestimmung der Lebensdauer einer Komponente betrachtet werden, sie muss allerdings für jede Anwendung geeignet definiert werden. In der Praxis ist der Zusammenhang zwischen dem tatsächlichen Zustand der Komponenten und dem Prozessergebnis allerdings oft nicht deterministisch. So schreibt (Schopp 2009, S. 133): „[A]ufgrund der stabilen Lageregelung und des vorhandenen direkten Messsystems konnten die vorhandenen Schädigungen der KGT-Mutter hinreichend kompensiert werden. [Es] [...] kann allein aus dem Betriebsverhalten und der Werkstückqualität auch bei Erreichen von ca. 95 % der maximalen Lauflistung nicht auf den Zustand des KGTs geschlossen werden.“ Auch (Huf 2012) stellt fest, dass ein KGT nach dem Auftreten der ersten Schädigungen meist weiter betrieben werden kann. Da der tatsächliche Schädigungszustand schwer zu ermitteln ist, besteht somit die Gefahr eines schwer geschädigten Systems, ohne Kenntnis darüber zu haben. Dies führt im schlimmsten Fall zu

einem ungeplanten Maschinenstillstand durch Überschreiten der Lebensdauer der Komponenten.

Werden nun konkrete Merkmale überwacht, über die direkt auf den Verschleiß einer Komponente geschlossen werden kann, kann darüber die Brauchbarkeitsdauer einer solchen definiert werden. Die Brauchbarkeitsdauer ist hierbei definiert als „Zeitspanne beginnend ab der ersten Verwendung und endend mit dem Zeitpunkt, an dem der Grenzzustand erreicht ist“ (Deutsches Institut für Normung 2018a, S. 22). Der Grenzzustand kann hierbei über ein zuvor definiertes Verschleißmaß oder bspw. die Betriebszeit definiert werden. Die Brauchbarkeitsdauer lässt damit das Prozessergebnis außer Acht und betrachtet lediglich den Zustand der Komponente. Abbildung 2-13 zeigt qualitativ drei beispielhafte Brauchbarkeitsdauern, wie sie bspw. für eine Messmaschine (L_{Betrieb1}), eine Werkzeugmaschine (L_{Betrieb2}) sowie eine Spritzgussmaschine (L_{Betrieb3}) definiert werden können. Es ist ersichtlich, dass die Prozessanforderungen je nach Anwendung schwanken und damit auch die definierten Brauchbarkeitsdauern unterschiedlich sind. Dies wird auch von (Forstmann 2010) bestätigt: „Im Gegensatz zu Werkzeugmaschinen ist es beim Einsatz von Kugelgewindtrieben an Einspritz- oder Schließachse einer Spritzgießmaschine nicht erforderlich Spielfreiheit sicher zu stellen.“ (Forstmann 2010, S.95) Damit liegt die sinnvolle Brauchbarkeitsdauer des KGTs einer Messmaschine häufig unter der einer Werkzeugmaschine und diese in der Regel unter der einer Spritzgussmaschine. Eine allgemeingültige Aussage kann hier allerdings nicht getroffen werden, da die Brauchbarkeitsdauer anhand einer spezifischen Anforderung gemessen wird.

Für die hier zu untersuchende Fragestellung nach der Zustandsüberwachung der Komponente KGT scheint die Definition einer Brauchbarkeitsdauer geeigneter als die Definition einer zulässigen Betriebszeit. Die Herausforderung besteht in einer Definition der Brauchbarkeitsdauer, die einer optimalen Abnutzungsgrenze (siehe Abbildung 2-12) entspricht. In Abbildung 2-13 ist die Brauchbarkeitsdauer an der Stelle der L_{Betrieb2} aufgetragen. Dies entspricht qualitativ einer optimalen Abnutzungsgrenze. Durch kontinuierliche Überwachung und Analyse des Verschleißzustandes der Komponente gilt es die Komponente nach dieser optimalen Brauchbarkeitsdauer auszutauschen.

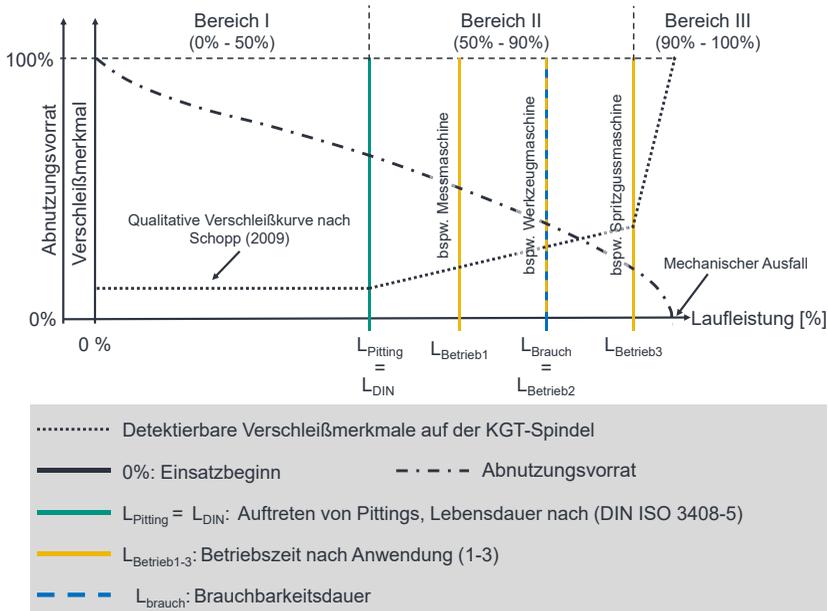


Abbildung 2-13 Qualitative Definition des Abnutzungsvorrates für verschiedene Anwendungen nach (A_Hofmann 2021).

Die Betriebszeit und die Brauchbarkeitsdauer müssen also nicht zwangsläufig übereinstimmen. Die von beiden Größen nicht zu überschreitende Größe ist das mechanische Versagen der Komponente. Da die Betriebszeit für jede Anwendung individuell bestimmt und anhand des Prozessergebnisses meist nicht genau zu überprüfen ist, ist es unmöglich, eine für alle Anwendungen passende Betriebszeit zu definieren, welche die Betriebszeit gleich der Brauchbarkeitsdauer festsetzt. Aus diesem Grund wird im Rahmen dieser Arbeit die Definition einer 100%-Lebensdauermarke (Abnutzungsgrenze) angestrebt, die an die Brauchbarkeitsdauer angelehnt ist, auch wenn eine allgemeingültige Definition außer dem mechanischen Versagen der Komponente über alle Anwendungen hinweg ausgeschlossen ist.

2.3 Condition Monitoring und Predictive Maintenance

Condition Monitoring (dt. Zustandsüberwachung) bezeichnet den Prozess der rechnerbasierten Feststellung des technischen Zustands von Maschinen und Anlagen auf Basis von Zustandsdaten und hat nach (Deutsches Institut für Normung 2010b) das Ziel,

auftretende Zustandsänderungen zu detektieren, um damit Instandhaltungsmaßnahmen (präventiv) einleiten zu können (Hübner 2018)¹². Dieser Prozess kann in die vier Aktivitäten Überwachung, Zustandsbeurteilung, Diagnostik und Prognostik unterteilt werden. Die Überwachung beschreibt die Erfassung von Prozess- und Maschinendaten. Im Rahmen der Zustandsbeurteilung werden diese Daten mit Bezugswerten verglichen, um den Schädigungszustand des Beurteilungsobjektes festzustellen. Darauf aufbauend wird eine Diagnose über Schadensart, -ort und -ausmaß vorgenommen. Abschließend werden der Schadensfortschritt, die Ausfallwahrscheinlichkeit und das Risiko weiterer Schäden prognostiziert (Deutsches Institut für Normung 2018c). Unter Predictive Maintenance (dt. voraussagende, vorausschauende oder vorhersagende Instandhaltung) wird eine zustandsorientierte Instandhaltungsstrategie verstanden, bei der die Instandhaltungsmaßnahmen nicht nur auf Basis des Ist-Zustands, sondern mithilfe einer Zustandsvorhersage der zu wartenden Objekte eingeleitet werden. Die Vorhersage wird im Rahmen des Condition Monitoring anhand von „wiederholter Analyse oder bekannten Eigenschaften und [...] wichtigen Parametern, welche den Abbau des Objektes kennzeichnen“ (Deutsches Institut für Normung 2018a, S. 35), getroffen.

2.3.1 Bedeutung von Condition Monitoring und Predictive Maintenance für den Maschinenbau

Die Wirtschaftlichkeit der Produktion ist in modernen wettbewerbsorientierten Industrien ein kritischer Faktor aller produzierenden Unternehmen. Aus diesem Grund ist eine kosteneffiziente Instandhaltung der Anlagen von großer Bedeutung (Davies 2012). Insbesondere ein unvorhergesehener Ausfall einer Komponente resultiert in ungeplanten Stillstandzeiten, die häufig mit erhöhten Instandsetzungszeiten und höheren Kosten einhergehen (Dehli 2019). Die moderne Instandhaltung hat sich in der Folge von vorbeugenden hin zu vorhersagenden Ansätzen gewandelt. Hierzu werden die Technologien des Condition Monitoring um fortschrittliche statistische Methoden zur Schätzung der Komponentenlebensdauer ergänzt, sodass zuverlässige Lebensdauer- und Ausfallvorhersagen möglich werden. Eine voraussagende Instandhaltung mit optimierter Ersatzteillagerhaltung und störungsarmen Instandhaltungsaktivitäten, die an ohnehin geplante Anlagenstillstände angepasst werden können, sind die Folge. Die verlängerte Zeitspanne zwischen geplanten Stillständen ermöglicht es der Instandhaltung nicht nur

¹² Hübner, T. (2018), *Die Zustandsüberwachung bei Kugelgewindtrieben im Blick*. <https://www.maschinenmarkt.vogel.de/die-zustandsueberwachung-bei-kugelgewindtrieben-im-blick-a-695846/> [29.09.2021].

an der Korrektur, sondern auch an der Verbesserung der Anlagen und der Verlängerung ihrer Lebensdauer zu arbeiten. Die Folge sind Instandhaltungsstrategien, die eine Maximierung der OEE forcieren (Stamboliska, Rusiński & Moczko 2015).

2.3.2 Ziele und Unterschiede der Ansätze

2.3.2.1 Condition Monitoring

Spontanausfälle, bspw. Gewaltbrüche, können auf Basis von Condition-Monitoring-Maßnahmen nicht prognostiziert werden. Verschleiß- bzw. Ermüdungsausfälle sind hingegen Phänomene, bei denen die Ausfallwahrscheinlichkeit berechenbar ist und die als Funktion der Betriebsdauer in den meisten Fällen der sogenannten Badewannenkurve folgen (siehe Abbildung 2-11) (Heise 2002). Demnach besteht direkt nach dem Einbau einer Komponente eine erhöhte Ausfallwahrscheinlichkeit, bedingt durch Material- oder Montagefehler (Kolerus & Wassermann 2008), die durch Condition Monitoring erkannt werden können. Während der darauffolgenden Konsolidierungsphase sind Ausfälle sehr selten und die Komponente unterliegt ihrem natürlichen, belastungsabhängigen Verschleiß (Mobley 2002). In dieser Phase kann das Condition Monitoring besonders zur Sicherung der Anlagenverfügbarkeit beitragen, indem ein sich verschlechternder – jedoch noch nicht ausfallkritischer – Komponentenzustand erfasst wird. Der fortschreitende Verschleiß kann im Rahmen eines Condition Monitorings erfasst und zur Vorhersage der dritten Phase (Spätausfälle) herangezogen werden. Da der Ausfall einer Komponente häufig Folgeschäden verursacht, kann eine rasche Fehlerfrüherkennung im Rahmen eines Condition Monitorings zum Schutz vor fatalen Schäden an der Maschine, Umwelt oder dem Menschen beitragen (Kolerus & Wassermann 2008). Um über den Zeitverlauf der Ausfallwahrscheinlichkeit zu einer Lebensdauerabschätzung zu gelangen, wird der oben eingeführte Abnutzungsvorrat betrachtet (siehe Abbildung 2-12 und Abbildung 2-13). Die Bestimmung des Abnutzungsvorrates ist Aufgabe des Condition Monitorings.

2.3.2.2 Predictive Maintenance

Eine vorausschauende Instandhaltungsstrategie (engl. Predictive Maintenance) verfolgt das Ziel, die Abnutzungsgrenze anhand des (durch Condition Monitoring) festgestellten Abnutzungsvorrates im Voraus auslastungs-, personal- und materialoptimal zu wählen, sodass sich ein möglichst wirtschaftlicher Instandsetzungszeitpunkt ergibt (Carnero 2006). Die vorausschauende Instandhaltung grenzt sich somit dadurch von

der klassischen zustandsorientierten Instandhaltung ab, dass nicht nur der Ist-Zustand, sondern auch eine (z. B. aus Condition-Monitoring-Aktivitäten resultierende) Prognose über Schadensfortschritt, Ausfallwahrscheinlichkeit und das Risiko weiterer Schäden sowie die Personal-, Ersatzteil- und Produktionsplanung zur Planung der Instandhaltungsmaßnahmen herangezogen werden (Deutsches Institut für Normung 2018a). So werden ungeplante Ausfälle (Abnutzungsgrenze zu spät), die häufig in einer erhöhten Mean Time to Repair (vgl. Abbildung 2-12 MTTR₃) resultieren (Dehli 2019), reduziert und gleichzeitig unnötig häufige Instandsetzungen (Abnutzungsgrenze zu früh), die zu einer reduzierten Mean Time between Failures (vgl. Abbildung 2-12 MTBF₁) führen (Wang & Wang 2018), vermieden. Darüber hinaus wird mit einer ganzheitlichen Predictive-Maintenance-Philosophie eine Erhöhung der OEE angestrebt (Mobley 2002).

2.3.3 Implementierung von Condition Monitoring und Predictive Maintenance

Ein allgemeines Verfahren zur Implementierung eines ganzheitlichen Condition-Monitoring-Programmes in Anlehnung an (Deutsches Institut für Normung 2018c) ist in Abbildung 2-14 dargestellt. Nach einer erfolgreichen Implementierung beschränken sich die Condition-Monitoring-Aktivitäten vorrangig auf die kontinuierliche Durchführung des Moduls Datenaufnahme und -analyse, das die vier Teilschritte Überwachung/Zustandserfassung, Zustandsbeurteilung, Diagnose und Prognose beinhaltet. Im Rahmen dieses Moduls können dann auch maschinelle Lernverfahren eingesetzt werden. Die Schwierigkeiten bei der Umsetzung von Condition-Monitoring-Maßnahmen sind im Wesentlichen die Suche nach aussagekräftigen Parametern (Zustandsgrößen) für die Schädigung der zu überwachenden Komponenten, die Auswahl geeigneter Messstellen und Sensoren, die gezielte Anwendung von Signalanalyse- und Mustererkennungsmethoden (Klein 2008) sowie der Umgang mit sowohl sehr großen als auch (zu) geringen Datenmengen. Nach (Roda, Macchi & Fumagalli 2018) ergeben sich darüber hinaus Herausforderungen betriebswirtschaftlicher und organisatorischer Natur.

(Lughofer & Sayed-Mouchaweh 2019) unterteilen die Umsetzung einer vorausschauenden Instandhaltungsstrategie in fünf Module:

1. Datenerfassung und Vorverarbeitung
2. Entwicklung von Zustandsindikatoren
3. Erkennung und Lokalisierung von Anomalien

4. Diagnose des System-/Maschinenzustands sowie Prognose des Abnutzungsvorates und potenziell auftretender Fehler
5. Auswahl und Planung von Abhilfe- oder Instandhaltungsmaßnahmen, wenn unerwünschte Systemzustände erkannt oder vorhergesagt werden.

Zur Überwindung der obigen Herausforderungen bedarf es einer Gesamtstrategie für eine digitale Fertigung, die eine robuste digitale Unternehmenskultur voraussetzt (Bousdekis & Apostolou et al. 2020). Hierfür ist ein umfangreiches Veränderungsmanagement auf allen Unternehmensebenen notwendig (Bousdekis & Apostolou et al. 2020), das beispielsweise die Bereitschaft für anfängliche Investitionen in Systeme zur Datenerfassung und Datenauswertung fördert (Dehli 2019).

2.4 Bildverarbeitung zur Klassifikation

Nachdem in den vorherigen Abschnitten die Grundlagen des tribologischen Systems KGT sowie der Ansätze des Condition Monitorings und der Predictive Maintenance dar-

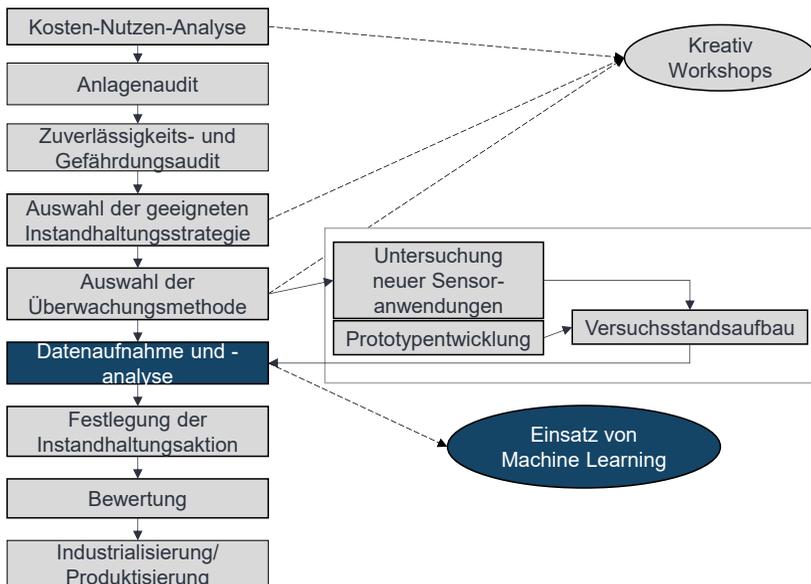


Abbildung 2-14 Prozess zur Durchführung industrieller Zustandsüberwachung nach (Deutsches Institut für Normung 2018c).

gestellt wurden, sollen im Folgenden die Grundlagen der Realisierung eines bildbasierten Systems zur Zustandsüberwachung am KGT aufgezeigt werden. Hierzu wird zunächst auf den Prozess der Verarbeitung von Bilddaten und die in der Bildverarbeitung grundlegenden Bildmerkmale eingegangen.

Abbildung 2-15 stellt den Ablauf der Bildverarbeitung zur Klassifikation qualitativ dar. Im Prozess werden zunächst mittels eines Sensorsystems Bilder des zu beobachtenden Objektes aufgenommen und damit die Objekte der realen Welt in sogenannte Muster überführt. Aus diesen Mustern werden in einem nächsten Schritt mittels ausgewählter Merkmalsextraktionsalgorithmen Merkmale extrahiert, die dann in einem vierten Schritt zur Klassifikation verwendet werden können. Die Klassifikation ordnet den Merkmalen wiederum ein Objekt der realen Welt (bzw. dessen Klasse, bspw. Fahrrad) zu.

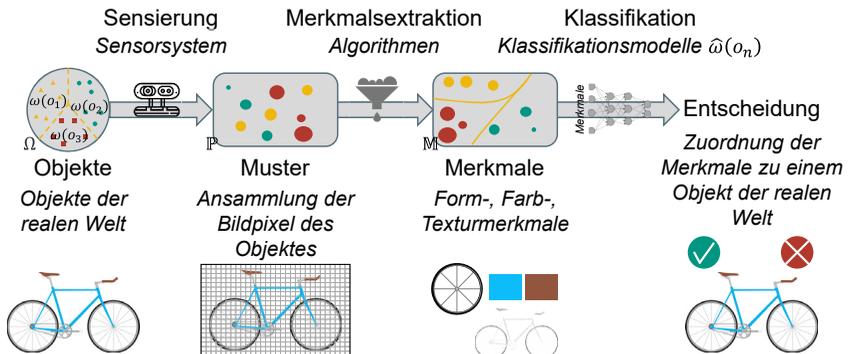


Abbildung 2-15 Qualitativer Ablauf der Bildverarbeitung anhand von extrahierten Bildmerkmalen nach (Richter 2018).

Zusammenfassend kann der Prozess der Sensierung und Merkmalsextraktion als Übersetzung eines realen Objektes in (mathematische) Merkmale verstanden werden. Mittels der Klassifikation werden diese Merkmale dann wiederum einem realen Objekt zugeordnet. Der Prozess stellt damit eine Verbindung zwischen einem Objekt der realen Welt und einem Klassifikationsalgorithmus dar. Auf die einzelnen Aspekte wird im Folgenden eingegangen.

2.4.1 Sensierung und Merkmalsextraktion

Bei der Sensierung mittels einer Kamera wird ein Objekt $o \in \Omega$ der realen Welt in ein Muster $p \in \mathbb{P}$ umgewandelt und somit (optisch) erfasst. Es ergibt sich eine Ansammlung

von Bildpixeln, die Informationen zur Verteilung der Lichtintensität, zum Spektrum, zur Polarisation, Kohärenz, Phase und Zeitabhängigkeit enthalten können. Da für die Sensierung messtechnische Unsicherheiten gelten, wird ein Objekt nicht immer dem gleichen Muster zugeordnet, sodass mehrere Muster für das gleiche Objekt entstehen können.

Im darauffolgenden Schritt der Merkmalsextraktion ist es das Ziel, aus der Ansammlung der Pixel (dem Bild) Merkmale zu extrahieren, die dieses Bild beschreiben. Der Schritt der Merkmalsextraktion umfasst die konkrete Extrahierung und Zusammenfassung bestimmter Merkmale aus dem Muster p eines Objekts. Hier können z. B. Farbmerkmale mithilfe der Analyse des Farbraums oder Konturmerkmale mit der Bestimmung eines minimal umgebenden Rechtecks gewonnen werden. Diese werden zu einem Merkmalsvektor $x \in \mathbb{M}$ im Merkmalsraum zusammengefasst. Die extrahierten Merkmale sollten für das betrachtete Objekt möglichst charakterisierend sein, um die Klassifizierung zu erleichtern (Richter 2018). Im konkreten Beispiel des Fahrrades entspricht dies Merkmalen wie der Struktur und Farbe der Speichen oder dessen länglicher Form sowie dessen Farben. Diese Merkmale müssen aus den Pixelwerten des Bildes extrahiert werden. Theoretisch können die Pixel des Bildes auch direkt als Merkmale verwendet werden. Hierbei entspricht jedes Pixel einem Merkmal.

Die Merkmalsextraktion lässt sich entweder manuell (klassisch) mittels definierter Merkmalsextraktionsalgorithmen oder automatisiert über den Einsatz von Machine-Learning- und speziell von Deep-Learning-Verfahren erreichen (Chollet 2018).

Zusammenfassend wird ein Objekt mittels eines Sensorsystems sensiert, das dann ein digitales Abbild des Objektes, bspw. ein Bild bei Verwendung einer Kamera, erstellt. Dieses digitale Abbild kann entweder direkt verwendet oder durch Merkmalsextraktion können weitere relevante Merkmale aus dem Bild extrahiert werden.

Im Folgenden wird der Begriff der Bildmerkmale weiter konkretisiert.

2.4.2 Bildmerkmale

Bevor auf die einzelnen Merkmale (Formmerkmale, Farbmerkmale, Texturmerkmale) eingegangen wird, die zur Klassifikation verwendet werden können, sollen zunächst Merkmale im Allgemeinen beschreiben werden.

Merkmale beschreiben nach (Richter 2018) die charakterisierenden Eigenschaften eines Objekts und sind eine grundlegende Voraussetzung für die Klassifikation. Hierunter

sind zum einen vom Menschen interpretierbare Merkmale wie geometrische Eigenschaften oder Farben (sogenannte semantische Merkmale) und zum anderen abstrakte Merkmale wie Energien in einem Spektralbereich oder berechnete Texturcodes zu verstehen.

In der Mustererkennung werden alle Merkmale eines Objekts, die erkannt werden können, in einem Merkmalsvektor x zusammengefasst. Die Merkmalsvektoren befinden sich in einem Merkmalsraum \mathbb{M} , dessen Dimension der Anzahl an Merkmalen entspricht. Werden bspw. die Struktur der Speichen sowie die Form eines Fahrrades verwendet, so hat der Merkmalsraum die Dimension zwei. Bei Verwendung der Pixel entspricht die Dimension des Merkmalsraums der Anzahl der Pixel im Bild. Metrische Merkmale, deren Werte reelle Zahlen sind, können direkt in den Merkmalsvektoren genutzt werden. Die Darstellung nichtmetrischer Merkmale (z. B. kategorialer Merkmale) erfolgt mittels Dummy-Variablen. Über diese werden binäre Kodierungen für jede Ausprägung eines nichtmetrischen Merkmals eingeführt und in den Merkmalsraum integriert. Beispielweise können vier kategoriale Eigenschaften als Dummy-Vektoren $[0,0,0]$, $[0,0,1]$, $[0,1,0]$ und $[1,0,0]$ dargestellt werden. Ein Beispiel hierfür sind verschiedene Farben, welche in einem Bild vorkommen. Mittels dieser numerischen Darstellung kann weitergerechnet werden.

Für die Mustererkennung in Bildern ist die Unterteilung der Merkmale nach Form-, Farb-, Kontur- und Texturmerkmalen besonders relevant, wobei die Konturmerkmale häufig den Formmerkmalen zugeordnet werden.

Im Folgenden wird ein kurzer Überblick über die relevantesten Merkmale und die in diesem Kontext verwendeten Algorithmen zur Merkmalsextraktion gegeben. Für eine ausführliche mathematische Einführung zu den Algorithmen sei auf die Grundlagenarchitektur (Beyerer & Puente León et al. 2016), (Bredies & Lorenz 2018) verwiesen.

2.4.2.1 Formmerkmale

Formmerkmale umfassen die geometrischen Eigenschaften eines Objekts und lassen sich weiter in Flächen- und Konturmerkmale unterscheiden. Flächenmerkmale beziehen sich auf globale Eigenschaften, die sich nicht aus einzelnen Teilen eines Objekts ableiten. Hierzu zählen unter anderem der Durchmesser, die Länge, die Hauptachsen, die Kompaktheit, der Füllgrad und die Konvexität eines Objektes. Die Bestimmung des Füllgrads ergibt sich z. B. aus dem minimal umgebenden Rechteck des Objekts (Beyerer, Richter & Nagel 2018). Der Füllgrad ist dann als $m = \text{Fläche des Objekts} /$

Fläche des minimal umgebenden Rechtecks definiert. Da globale Flächenmerkmale oftmals nur eine grobe Unterscheidung ermöglichen, werden sie meistens zur groben Eliminierung von Objekten genutzt oder müssen mit feineren Merkmalen kombiniert werden (Zhang & Lu 2004).

2.4.2.2 Konturmerkmale

Konturmerkmale werden aus dem Objektprofil abgeleitet und basieren oft auf der Unterteilung des Objekts in einzelne Elemente. Ein Beispiel hierfür ist die Beschreibung eines Fahrrades als Objekt, das aus zwei Rädern, einem Sattel, einem Lenker und einem Rahmen, also aus insgesamt fünf Elementen besteht. Diese werden dann z. B. weiter auf Tangentenwinkel, Krümmung, Kettencodes oder Fourier-Konturdeskriptoren untersucht, um eine feinere Beschreibung zu erzielen. Ein Vorteil dieser Methoden ist, dass Objekte stückweise verglichen werden können. Demgegenüber steht allerdings das Problem der Aufteilung in einzelne Elemente, da es für Konturen keine allgemeingültige Aufteilung gibt und mit der Aufteilung eine Steigerung der benötigten Rechenleistung einhergeht (Zhang & Lu 2004). (Czimmermann & Ciuti et al. 2020) stellen hierzu im Kontext von Fehlstellen eine anschauliche Übersicht an visuell feststellbaren Fehlertypen sowie deren optische Beschreibung dar (siehe Abbildung 0-2 im Anhang).

2.4.2.3 Farbmerkmale

Farbmerkmale beschreiben die Verteilung der Farben eines Objekts, wodurch Rückschlüsse auf mechanische, thermische, chemische und biologische Einwirkungen möglich sind (Richter 2018). Für die Mustererkennung mit Farbmerkmalen können Farbmomente wie die mittlere Objektfarbe und die Farbstreuung genutzt werden, die sich für ein Bild mit $n * m$ Pixeln wie folgt bestimmen: *Mittlere Objektfarbe* = $\sum_{i=1}^n \sum_{j=1}^m x_{ij} / mn$ und *Farbstreuung* = $\frac{1}{mn} * \sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \text{Mittlere Objektfarbe})^2$ mit x_{ij} als Farbwert für das Pixel in der Zeile i und Spalte j . Neben den Farbmomenten können Histogramme zur Analyse der Farbverteilung genutzt werden, welche die Verteilung von Pixeln im Bild beschreiben (Kodituwakku & Selvarajah 2010). Die Annahme ist hierbei, dass bspw. das Farbhistogramm eines Fahrrades dem Farbhistogramm eines zweiten Fahrrades ähnlicher ist als dem eines anderen Objektes, weshalb das Farbhistogramm zur Klassifikation genutzt werden kann.

2.4.2.4 Texturmerkmale

Die letzte Kategorie der Bildmerkmale sind Texturmerkmale. Sie umfassen Informationen zu den Oberflächendetails und sind oft die diskriminativsten Merkmale (Richter 2018). Eine Möglichkeit zur Untersuchung von Texturmerkmalen ist die Berechnung von Grauwertmatrizen (Haralick, Shanmugam & Dinstein 1973). Hierbei ist die zentrale Annahme, dass sich die Textur eines Objekts aus der durchschnittlichen räumlichen Beziehung einzelner benachbarter Grautöne ergibt. Dazu wird für jedes Pixel der Grauton bestimmt und anschließend werden verschiedene Relationsmatrizen berechnet. Je größer die Anzahl an unterschiedlichen Grautönen in einem Bildbereich, desto dominanter ist die Textur für diesen Bereich. Eine weitere Möglichkeit ist die Nutzung von Local Binary Patterns (Ojala, Pietikainen & Maenpaa 2002). Für jedes Pixel werden die Differenzen aus den Grauwerten der benachbarten Pixel und dem Grauwert des betrachteten Pixels bestimmt. Anschließend werden die Differenzen x mit der Transformation $s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$ binär codiert. Das so entstehende Local Binary Pattern ist invariant gegenüber der Grauwerteskala und gegenüber der Rotation des Objekts, was die Erkennung universeller Muster ermöglicht.

Zusammenfassend lassen sich Objekte mittels Verwendung von Merkmalsextraktionsalgorithmen durch Merkmale beschreiben. Diese Merkmale sollten zur Klassifikation eines Objektes möglichst charakteristisch für dieses sein. Für die Auswahl der Merkmale ist häufig Domänenwissen notwendig (Beyerer, Richter & Nagel 2018).

Im Folgenden wird auf die Erkennung von Mustern (in Merkmalen) sowie die Generierung von neuen Merkmalen eingegangen.

2.4.3 Mustererkennung und Merkmalsgenerierung

Nachdem in den vorherigen Abschnitten die Eigenschaften von Form-, Farb- und Texturmerkmalen beschrieben wurden, soll nachfolgend das generelle Vorgehen bei der Erkennung von Mustern sowie bei der Generierung von neuen Merkmalen eingeführt werden. Die Klassifikation eines Objektes folgt dann als Schritt der Mustererkennung. Wird ein bestimmtes Muster erkannt, so wird diesem im Prozess der Klassifikation eine bestimmte Klasse zugewiesen.

2.4.3.1 Mustererkennung

Bei der Mustererkennung werden die im vorherigen Abschnitt beschriebenen Merkmale eines Objekts erkannt und anschließend für die Klassifizierung genutzt. Die mathematische Repräsentation von Objekten in der Welt ist für diesen Prozess grundlegend. Ein Objekt o ist ein Element der Welt Ω und zwischen zwei Elementen $o, p \in \Omega$ existiert eine Äquivalenzrelation $o \sim p$. Eine Klasse ω ist eine von einem Prototyp $p \in \Omega$ und einer über Ω definierten Äquivalenzrelation $p \sim o$ induzierte Untermenge der Welt.

Ziel der Mustererkennung ist es, Objekte $o \in \Omega$ verschiedenen Klassen ω zuzuweisen.

Dabei werden die Schritte Sensierung, Merkmalsextraktion, Feature Engineering und Klassifizierung durchgeführt (Richter 2018), wie sie in Abbildung 2-15 dargestellt sind. Die Klassifikation ist also der finale Schritt der Mustererkennung. Das Feature Engineering stellt eine gesonderte Vorgehensweise zur Merkmalsextraktion dar und wird daher in einem eigenen Abschnitt betrachtet.

2.4.3.2 Feature Engineering

Nach (Richter 2018) bezeichnet Feature Engineering (dt. Merkmalsgenerierung) das Entwickeln von für ein gegebenes Problem angepassten Verfahren zur Merkmalsextraktion. Das Ziel ist es Merkmale zu finden, die eine diskriminative und robuste Klassifikation ermöglichen. Gleichzeitig ergeben sich aus der Reduktion des Merkmalsraums auf ausschließlich relevante Merkmale niederdimensionale Merkmalsvektoren, was wiederum den Rechenaufwand der Klassifikation reduziert. Ein weiterer Vorteil ist, dass ausreichende Ergebnisse mit weniger Daten für die Klassifikation erzielt werden können (Chollet 2018).

Für das Feature Engineering wird domänenspezifisches Wissen benötigt und es gibt kein standardisiertes generelles Vorgehen für die Entwicklung von Merkmalen. Mittels Domänenwissen werden für die Klassifikationsaufgabe charakteristische Merkmale zur Lösung der Klassifikationsaufgabe definiert und extrahiert. Nachteilig ist hierbei, dass eine Übertragung von Merkmalen, die für eine bestimmte Aufgabe extrahiert wurden, auf eine andere Aufgabe nicht immer ohne Weiteres möglich ist.

Im Kontext des Maschinellen Lernens und speziell bei der Verwendung von *Convolutional Neural Networks (CNN)* findet der Prozess des Feature Engineering sowie der Merkmalsextraktion automatisiert statt, mit dem Nachteil, dass hier nur eingeschränkt Domänenwissen für die Merkmalsdefinition und Merkmalsextraktion verwendet wird.

Nachdem in den vorherigen Abschnitten der Prozess der Datenaufnahme und Generierung von Merkmalen beschrieben wurde, werden im Folgenden zwei Ansätze zur Klassifikation von Bilddaten vorgestellt.

2.4.4 Neuronale Netzwerke zur Klassifikation

Neuronale Netzwerke haben sich als performantes Mittel zur Klassifikation von Daten und speziell von Bilddaten erwiesen, weshalb der Fokus in dieser Arbeit auf Algorithmen aus der Familie der Neuronalen Netze gelegt wird (Chollet 2018), (Aggarwal 2019).¹³

Unter einer Klassifikation wird die Abbildung $f: \mathbb{R}^n \rightarrow \{0, 1, \dots, k\}$ verstanden. Eine relevante Vorbemerkung ist hierbei, dass ein Neuronales Netzwerk lediglich eine mathematische Funktion zur Abbildung des Definitionsbereiches auf den Wertebereich darstellt. Neben den fixen Eingangswerten und Modellparametern des Neuronalen Netzes sind es lediglich die Parameter der Funktion (Gewichte), die das Funktionsergebnis bestimmen. Es gilt diese Gewichte so einzustellen, dass das vorhergesagte mit dem gewünschten Ergebnis übereinstimmt. Dieser Prozess wird im Kontext des Maschinellen Lernens als Lernen bzw. Training bezeichnet.

Als einführendes Beispiel soll die Ausgabe eines Getränkeautomaten dienen. Je nachdem, ob Wasser oder Milch in den Automaten gegeben werden, soll automatisch entweder ein Glas oder eine Tasse befüllt werden. Der Getränkeautomat besteht aus einer Vielzahl an Röhren und Ventilen und muss zur korrekten Befüllung die Ventile automatisch so steuern, dass das Getränk in das richtige Gefäß geleitet wird. Das automatische Steuern der Ventile als Folge der Flüssigkeit, die eingegeben wird, wird als Lernen bezeichnet. Der Getränkeautomat kann automatisch Getränke in passende Gefäße füllen, ohne eine weitere Eingabe des Menschen zu erhalten.

Im Kontext der Klassifikation von Bilddaten werden die Flüssigkeiten durch Bilder respektive durch die Pixel der Bilder ersetzt. Diese werden nun stellvertretend für das Bild in das Neuronale Netz gegeben. Als konkretes Beispiel lässt sich die Klassifikation zweier Fahrzeuge nennen. Die Aufgabe des Neuronalen Netzwerkes ist es, einem Bild die Klasse Fahrrad oder PKW zuzuordnen, d. h. zu erkennen bzw. zu klassifizieren, ob es sich bei dem Objekt um ein Fahrrad oder um einen PKW handelt.

¹³ Das Thema der Regression ist in dieser Arbeit nicht von Relevanz, daher werden die Grundlagen für Neuronale Netze anhand der Klassifikation dargestellt.

Die folgenden Betrachtungen sollen anhand des angeführten Beispiels der binären Klassifikation erfolgen, wie sie auch in dieser Arbeit verwendet wird. Die binäre Klassifikation ist definiert mit: $f: \mathbb{R}^n \rightarrow \{0,1\}$. Aus dieser Abbildung ergeben sich für das Neuronale Netz drei grundlegende Fragestellungen, die im Folgenden beschrieben werden sollen.

1. Wie ist die Funktion $f(\cdot)$ (das Neuronale Netzwerk) aufgebaut?
2. Wie lässt sich die Klassifikationsgüte bestimmen?
3. Wie kann die Funktion automatisiert angepasst werden, damit das gewünschte Klassifikationsergebnis erzielt wird?

2.4.4.1 Architektur des Neuronalen Netzwerkes

Die Funktion $f(\cdot)$ wird hier als ein Neuronales Netzwerk gewählt, wie es in Abbildung 2-16 schematisch dargestellt ist. Es besitzt sogenannte Eingänge (engl. *Inputs*), welche die Eingangsdaten (\mathbb{R}^n) darstellen, sowie (im binären Fall) einen sogenannten Ausgang (engl. *Output*), der die Klasse $\{0,1\}$ annimmt. Die Werte 0 und 1 werden hier stellvertretend für die Klassen Fahrrad und PKW bzw. bei der Klassifikation von Oberflächendefekten für Pitting und kein Pitting verwendet.

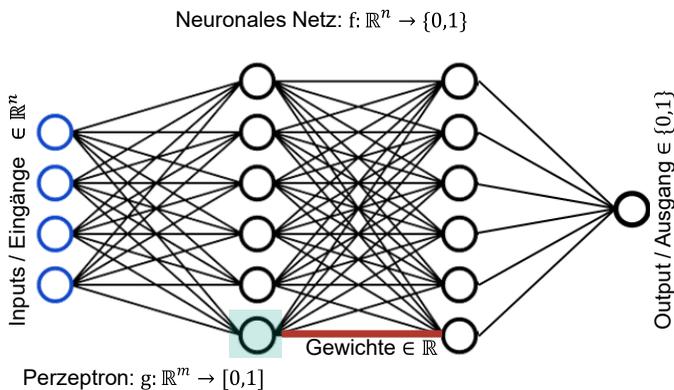


Abbildung 2-16 Qualitative Abbildung eines künstlichen Neuronalen Netzwerkes nach (Zhou 2019)¹⁴.

¹⁴ Zhou, V. (2019), *Neural Networks From Scratch*. <https://victorzhou.com/series/neural-networks-from-scratch/> [26.01.2022].

Die Eingangswerte können bei der Klassifikation von Bilddaten entweder direkt die Pixelwerte eines Bildes oder vorverarbeitete Merkmale (engl. *Features*) sein (siehe Abbildung 2-17). Im Kontext der Klassifikation von Fahrzeugen sind vorverarbeitete Merkmale, welche die Objekte kennzeichnen, bspw. die Größe des Objektes, das Vorhandensein von Speichen oder das Vorhandensein eines Lenkers. Diese semantischen Merkmale müssen mittels geeigneter Algorithmen aus dem Bild extrahiert und so dem Neuronalen Netz zugänglich gemacht werden.

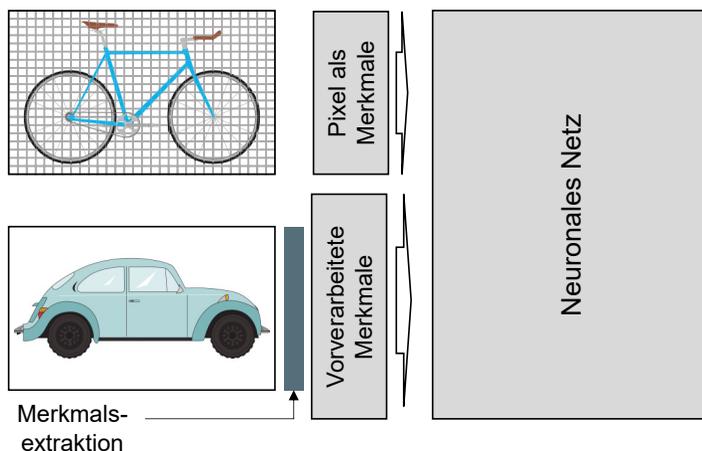


Abbildung 2-17 Klassifikation auf Basis von Pixelwerten oder vorverarbeiteten Merkmalen.

Das Neuronale Netz besteht aus einem Zusammenschluss einfacher Funktionseinheiten (sogenannter Perzeptronen), die ursprünglich von (Rosenblatt 1958) eingeführt wurden (siehe Abbildung 2-16). Diese Perzeptronen sind ihrerseits Funktionen, die Eingangswerte auf einen Ausgangswert zwischen 0 und 1 abbilden: $g: \mathbb{R}^m \rightarrow [0,1]$. Abbildung 2-18 zeigt den Aufbau eines einzelnen Perzeptrons. Die Perzeptronen des Neuronalen Netzes werden als Neuronen bezeichnet.

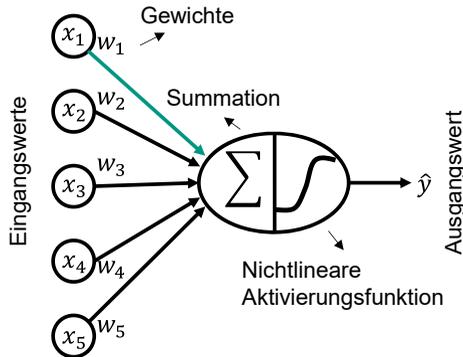


Abbildung 2-18 Kernelemente eines einzelnen Perzeptrons.

Die Fähigkeit des Neuronalen Netzes zur Klassifikation entstammt hauptsächlich diesem Zusammenschluss an einfachen Einheiten (Aggarwal 2019) und wird im Folgenden beschrieben.

2.4.4.2 Berechnung der Klassifikationsgüte

Das Neuronale Netzwerk empfängt numerische Werte als Eingänge entweder direkt in Form von Pixelwerten oder als vorverarbeitete Merkmale. Die Klassifikation erfolgt durch die Verarbeitung dieser Werte durch eine Aneinanderreihung von Perzeptronen.

Ein Perzeptron empfängt Eingangswerte x_n . Die Eingangswerte werden mittels Gewichten w_i gewichtet und dann innerhalb des Perzeptrons addiert, bevor die Summe durch eine nichtlineare Funktion verarbeitet wird. Der Ausgang eines Perzeptrons ergibt sich somit als: $g: \mathbb{R}^m \rightarrow [0,1], x \mapsto \sigma(\sum_{i=1}^k x_i w_i)$ bzw. $\sigma(\mathbf{x}\mathbf{w}^T)$. Die Funktion $\sigma(v) = \frac{1}{1+e^{-v}} = o$ (Sigmoid-Funktion) wird hierbei als sogenannte Aktivierungsfunktion gewählt. Die Gründe hierfür sind, dass die Funktion zwar nicht linear, aber stetig und damit differenzierbar ist. Die Nichtlinearität der Funktion führt in Konsequenz dazu, dass das Neuronale Netz auch komplizierte nichtlineare Funktionen erlernen kann. Weiterhin ist die Ableitung der Funktion einfach zu berechnen ($\frac{\partial o}{\partial v} = o(1-o)$) und die Funktion nimmt Werte zwischen 0 und 1 an. Zusammenfassend ermöglicht es die Verwendung einer nichtlinearen Funktion, die Eingangswerte (Merkmale) performanter zu verarbeiten.

In einem nächsten Schritt werden mehrere dieser Perzeptronen in Schichten (engl. *Layer*) hintereinander geschaltet. Die Ausgänge von Perzeptronen in früheren Schichten bilden die Eingänge für Perzeptronen in tieferen Schichten. Dadurch, dass jedes Perzeptron mit einer nichtlinearen Funktion abgeschlossen wird, lassen sich komplizierte Funktionen bilden und somit Bilddaten klassifizieren. Aufgrund dieser hochdimensionalen Verkettung von Perzeptronen wird bei Neuronalen Netzwerken umgangssprachlich von sogenannten Black-Box-Modellen gesprochen, da es nicht mehr ohne Weiteres möglich und semantisch nachvollziehbar ist, wie das Neuronale Netz zu einer Entscheidung gelangt.

Das Neuronale Netz besteht also aus einer Verkettung mehrerer Funktionen

$f(\mathbf{x}) = g_o \left(g_{n-1} \left(g_k \left(\dots \left(g_1(\mathbf{x}) \right) \right) \right) \right) = \hat{y}$, worüber die Eingangswerte auf die Ausgangswerte abgebildet werden. Dieser Vorgang wird Forward-Pass genannt.

Initial ist das Neuronale Netzwerk mit zufällig gewählten Gewichten versehen und hat daher noch nichts aus den Daten gelernt.

Das Ziel ist es, einem Bild eines Fahrzeuges die korrekte Klasse Fahrrad oder PKW zuzuweisen. Um einem Modell näher zu kommen, das zuverlässig Eingangswerte auf korrekte Ausgangswerte abbildet, muss zunächst die zweite oben gestellte Frage nach der Bestimmung der Klassifikationsgüte beantwortet werden. Hierzu bedarf es einer Möglichkeit, die Klassifikationsgüte des Modelles zu bewerten, d. h. zu entscheiden, wie gut die vom Modell zurückgegebenen Prognosen mit den bekannten Klassenwerten übereinstimmen. Die Klassifikationsgüte wird nach dem ersten Forward-Pass schlecht sein, da das Neuronale Netz die Eingangsbilder zunächst nur ein einziges Mal gesehen hat und damit noch nicht genügend Zeit zum Erlernen der relevanten Kombinationen der Merkmale hatte. Das Erlernen der Kombinationen erfolgt durch das wiederholte Präsentieren der Bilddaten in sogenannten Iterationen. Eine hohe Anzahl an Iterationen, d. h. das wiederholte Präsentieren der Bilddaten und der zugehörigen Label, veranlasst das Neuronale Netzwerk zu lernen und damit zunehmend bessere Klassifikationsergebnisse zu erzielen.

Zur Berechnung der Klassifikationsgüte wird klassischerweise die sogenannte Binary-Cross-Entropy-Funktion (Log Loss) mit $\mathcal{L}_{ln} = -y * \ln \hat{y} - (1 - y) * \ln(1 - \hat{y})$ verwendet. Hierzu sollen zwei Beispiele gemacht werden. Sei der wahre Wert y , der zu einem bestimmten Input gehört, 1. Wenn nun angenommen wird, dass das Modell einen Wert \hat{y}

von 0.2 zurückgibt, dann ergibt sich ein Fehler von: $\mathcal{L}_{ln} = -1 * \ln(0.2) - (1 - 1) * \ln(1 - 0.2) = -1 * \ln(0.2) = 1.61$. Es ist ersichtlich, dass der zweite Term der Verlustfunktion verschwindet. Im zweiten Beispiel wird der wahre Wert y mit 0 und die Vorhersage mit 0.1 gewählt. In diesem Fall ergibt sich: $\mathcal{L}_{ln} = -0 * \ln(0.1) - (1 - 0) * \ln(1 - 0.1) = -\ln(0.9) = 0.11$, wobei der erste Term der Verlustfunktion verschwindet. Es ist ersichtlich, dass der Wert der Verlustfunktion sinkt, wenn der prognostizierte Wert nahe an dem wahren Wert liegt. Die Funktion eignet sich also zur Messung der Güte der Klassifikation.

2.4.4.3 Training des Neuronalen Netzwerkes

Das Ziel des Neuronalen Netzwerkes ist es, alle Eingangswerte so zu verarbeiten, dass der prognostizierte Wert \hat{y} mit dem wahren Wert y übereinstimmt und die Verlustfunktion zu $\mathcal{L}_{ln} = -1 * \ln(1) = 0$ wird. Da die einzigen nicht fixen Parameter des Neuronalen Netzes dessen Gewichte w_i sind, sind diese für die Änderung der Fehlerfunktion verantwortlich. Es ist bekannt, dass eine Funktion dann minimal oder maximal ist, wenn die erste Ableitung der Funktion 0 ist. Da es sich bei dem Neuronalen Netz um die Verkettung einer Vielzahl an Funktionen handelt, ist es aus rechnerischer Sicht schwer möglich, den Gradienten der Verlustfunktion in vertretbarer Zeit für jedes Gewicht mittels der Kettenregel zu berechnen (Aggarwal 2019). Es wird daher der sogenannte Backpropagation-Algorithmus (siehe bspw. (Rojas 1996)) zur Anpassung der Gewichte angewandt. Mittels des Backpropagation-Algorithmus werden die partiellen Ableitungen $\frac{\partial L}{\partial w_i}$ der Verlustfunktion zu den Gewichten des Netzwerkes berechnet und die Einzelgewichte mittels $w_i^{new} = w_i^{old} - \alpha * \frac{\partial L}{\partial w_i^{old}}$ angepasst. Hierdurch ist sichergestellt, dass der Fehler sinkt bzw. die Klassifikationsgüte steigt. α ist dabei ein Parameter zur Gewichtung der Anpassung. Das Vorgehen zur Berechnung der Verlustfunktion $\mathcal{L}_{ln} = -y * \ln \hat{y} - (1 - y) * \ln(1 - \hat{y})$ ist in den folgenden Schritten beschrieben.

(1) Der in der Verlustfunktion verwendete Wert \hat{y} berechnet sich mit (siehe Abbildung

$$2-18) \hat{y} = \sigma(\sum_{i=1}^k x_i w_i) \text{ und damit ist}$$

$$(2) \mathcal{L}_{ln} = -y * \ln(\sigma(\sum_{i=1}^k x_i w_i)) - (1 - y) * \ln(1 - \sigma(\sum_{i=1}^k x_i w_i)).$$

(3) Die Ableitung $\frac{\partial L}{\partial w_i}$ lässt sich nach der Kettenregel berechnen mit $\frac{\partial L}{\partial w_i} = \frac{\partial L}{\partial o} * \frac{\partial o}{\partial v} * \frac{\partial v}{\partial w_i}$,

$$\text{mit } o = \sigma(v) \text{ und } v = \sum_{i=1}^k x_i w_i.$$

(4) Die Ableitungen von $\frac{\partial o}{\partial v} = o(1 - o)$ und $\frac{\partial v}{\partial w_i} = x_i$ sind bereits bekannt.

(5) Die Ableitung des natürlichen Logarithmus ist $\ln(x)' = \frac{1}{x}$.

(6) Damit ergibt sich die Ableitung von $\frac{\partial L}{\partial o}$ zu: $-\left(\frac{y}{o} - \frac{(1-y)}{(1-o)}\right)$.

(7) Durch Erweiterung mit $o(1-o)$ ergibt sich $\frac{o-y}{o(1-o)}$ und damit: $\frac{\partial L}{\partial w_i} = \frac{o-y}{o(1-o)} * o(1-o) * x_i = x_i(o-y)$.

Mittels dieser Schritte wurde die Ableitung zu dem grün hervorgehobenen Input in Abbildung 2-18 berechnet. Die Anpassung des Gewichtes erfolgt nun mittels der oben beschriebenen Regel $w_i^{neu} = w_i^{alt} - \alpha * \frac{\partial L}{\partial w_i^{alt}}$. Um nun $\frac{\partial L}{\partial w_i}$ für alle w_i im Netz zu berechnen, wird die soeben berechnete Ableitung verwendet und der Fehler rückwärts durch das Netz geschleust. Davon leitet sich der Name ‚Backpropagation‘ ab. Der Prozess der Backpropagation des Fehlers zu einem Perzeptron in der vorfinalen Schicht ist im Abschnitt *Backpropagation im Neuronalen Netzwerk* im Anhang dargestellt.

Im folgenden Abschnitt werden diese Überlegungen auf Convolutional Neural Networks übertragen.

2.4.5 Convolutional Neural Networks zur Klassifikation

Convolutional Neural Networks (CNN) stellen eine Erweiterung der oben gezeigten Neuronalen Netzwerke dar. Ein CNN kann als Kombination aus einem Vorverarbeitungsschritt, der aus den Bilddaten Merkmale extrahiert, und einem Neuronalen Netz zur Klassifikation verstanden werden (siehe Abbildung 2-17, Merkmalsextraktion). Somit werden für die Klassifikation mittels des Neuronalen Netzwerkes nicht direkt die Pixel eines Bildes verarbeitet, sondern zunächst aus diesen Pixeln semantisch relevante Merkmale extrahiert, ähnlich wie dies auch bei der klassischen Merkmalsextraktion der Fall ist. Der Unterschied besteht darin, dass CNNs Merkmale automatisiert aus den Bilddaten extrahieren. Diese automatische Extraktion von Merkmalen ist eine Hauptaufgabe des CNN.

Bildlich gesprochen besteht, wie im vorherigen Abschnitt angedeutet, ein Bild eines Fahrrades aus semantisch charakteristischen Elementen wie den Rädern oder dem Lenker. Diese Merkmale werden auch vom Menschen als charakteristische Merkmale beschrieben, mittels derer eine Klassifikation durchgeführt wird. Sie sind in allen Objekten einer bestimmten Klasse vorhanden. CNN haben zunächst die Aufgabe, diese semantischen Merkmale aus den Pixelwerten zu extrahieren.

Die Motivation für die Verwendung von CNN entstammt ursprünglich einer Arbeit von (Hubel & Wiesel 1959), die festgestellt haben, dass Bilddaten im visuellen Kortex von Katzen hierarchisch verarbeitet werden und unterschiedliche Regionen für die Erkennung von unterschiedlichen Merkmalen (bspw. Linien) verantwortlich sind. Die auf diese Arbeit folgenden Arbeiten wurden schließlich von (Lecun & Boser et al. 1989) und (Lecun & Bottou et al. 1998) zu ersten CNNs weiterentwickelt und die Überlegenheit der Klassifikation von handgeschriebenen Zahlen im Vergleich zu klassischen Neuronalen Netzwerken wurde aufgezeigt. Die Architekturen wurden seither stetig verbessert und die Überlegenheit gegenüber dem Menschen bei der Klassifikation von Bilddaten wurde von (He & Zhang et al. 2016) nachgewiesen.

2.4.5.1 CNN – Erweiterung gegenüber klassischen Neuronalen Netzwerken

Ebenso wie das klassische Neuronale Netzwerk ist ein CNN zur binären Klassifikation von Bildern eine $f: \mathbb{N}_{\geq 0, \leq 255}^{n \times n} \rightarrow \{0,1\}$ mit dem Unterschied, dass das CNN als Eingangswerte ein Bild (eine Matrix) der Dimension $n \times n$ und nicht einen Vektor der Dimension $n * n$ erhält. Diese Transformation muss zunächst durch das CNN vorgenommen werden. Jedes Pixel des Bildes kann eine natürliche Zahl im Bereich von 0 bis 255 annehmen. Die Abbildung $f: \mathbb{N}_{\geq 0, \leq 255}^{n \times n} \rightarrow \{0,1\}$ erfolgt hierbei durch zwei Teile: die sogenannte Convolutional Base und einem klassischen Neuronales Netzwerk (siehe Abbildung 2-19).

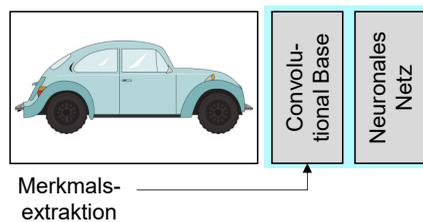


Abbildung 2-19 Bestandteile eines CNN.

Die Convolutional Base hat hierbei die Aufgabe, Merkmale aus den Bilddaten zu extrahieren und dem Neuronales Netzwerk als Vektor x zur Verfügung zu stellen. Der Vektor x ist dann identisch dem Eingangsvektor, der im vorherigen Abschnitt diskutiert wurde, allerdings codiert er die semantischen Informationen der Eingangsbilddaten (Sattel, Lenker, Speichen, Größe) und nicht die originalen Werte der Pixel, wie sie im klassischen Neuronales Netzwerk verwendet werden. An dieser Stelle wird das Eingangsbild bzw.

die Pixel-Matrix des Eingangsbildes in einen Vektor überführt. Die Klassifikation wird also nach wie vor von einem klassischen Neuronales Netzwerk durchgeführt, die Convolutional Base dient lediglich der Vorverarbeitung der Bilddaten. Der große Vorteil des CNN ist, dass die Extraktion dieser Merkmale automatisch während des Trainings des CNN erfolgt. Hierin ist auch der größte Unterschied zu klassischen Ansätzen zu sehen, bei denen die Merkmale auf Basis von Domänenwissen und Feature-Engineering-Techniken definiert und konstruiert werden. Dieser Prozess der Merkmalsdefinition, Merkmalsextraktion sowie Merkmalskombination wird nun durch die Convolutional Base automatisiert.

Die Funktion $f: \mathbb{N}_{\geq 0, \leq 255}^{n \times n} \rightarrow \{0,1\}$ lässt sich also aufteilen in die Convolutional Base $f_1: \mathbb{N}_{\geq 0, \leq 255}^{n \times n} \rightarrow \mathbb{R}^m$ und das folgende Neuronale Netz $f_2: \mathbb{R}^m \rightarrow \{0,1\}$. Damit ist $f(X) = f_2(f_1(X))$.

f_2 ist dabei mit dem oben diskutierten Neuronales Netz identisch, weshalb sich die Diskussionen hier auf die Convolutional Base f_1 konzentrieren sollen.

Analog zum Neuronales Netz stellen sich auch beim Training des CNN drei Fragen:

1. Wie ist die Funktion $f_1(\cdot)$ (die Convolutional Base) gestaltet?
2. Wie werden Bilddaten verarbeitet?
3. Wie kann die Funktion automatisiert angepasst werden, damit das gewünschte Klassifikationsergebnis erreicht wird?

Abbildung 2-20 zeigt die Gesamtarchitektur eines CNN mit seinen Einzelkomponenten.

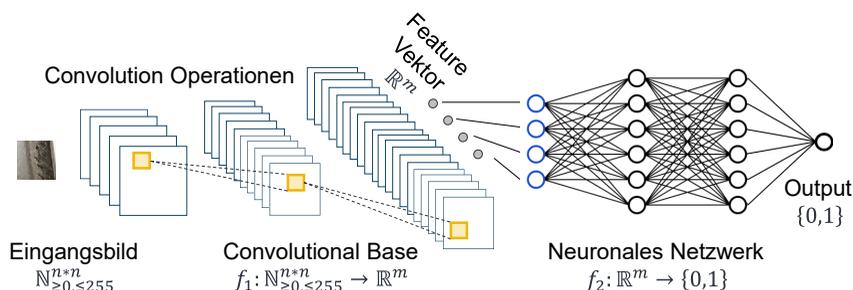


Abbildung 2-20 Grundelemente eines Convolutional Neural Network.

2.4.5.2 Architektur des CNN

Die grundlegende Operation, die innerhalb der Convolutional Base durchgeführt wird, ist die sogenannte Convolution-Operation. Neben der Convolution-Operation bestehen im Kontext von CNN noch die Operationen Pooling, Padding und Strides. Diese sind allerdings aus Effektivitätsgründen eingeführt worden und für die grundsätzliche Funktionsfähigkeit eines CNN nicht zwingend. Für eine Übersicht zu diesen Operationen und weiteren Modell-Hyperparametern, die im Zuge der Modellbildung und des Modelltrainings gesetzt werden können, sei auf die Grundlagenliteratur (Aggarwal 2019) und (Goodfellow, Bengio & Courville 2016) verwiesen. Hieran sind auch die folgenden Ausführungen angelehnt.

2.4.5.3 Convolution-Operation

Die Convolution-Operation ist verantwortlich für die Extraktion von Merkmalen aus den Eingangsdaten und damit die Bereitstellung der Merkmale für die Klassifikation mittels des Neuronalen Netzwerkes. Neben dem Input sind an der Convolution-Operation der sogenannte Filter (engl. *Kernel*) sowie eine Aktivierungsfunktion beteiligt, wie sie auch aus den obigen Diskussionen zum klassischen Neuronalen Netzwerk bekannt ist.

Der Input stellt zu Beginn das Bild dar, das klassifiziert werden soll. Die Filter haben die Aufgabe, in dem Input nach speziellen Merkmalen zu suchen, d. h. den Input nach diesen Merkmalen zu filtern. Für die Filter gilt hierbei $\gamma \in \mathbb{R}^{n \times n \times d}$. Sie stellen also dreidimensionale Matrizen dar. Die nachfolgenden Diskussionen sollen für den Fall $d = 1$ (der Farbkanal hat eine Dimension) ausgeführt werden. Die grundlegende Aufgabe der Filter ist es, relevante Merkmale aus den Eingangsdaten zu extrahieren. Ein konkretes Beispiel ist hierbei das Hervorheben von markanten horizontalen oder vertikalen Linien wie sie bspw. bei den Speichen eines Fahrrades vorkommen. Abbildung 2-21 stellt die grundsätzliche Funktionsweise eines Filters angewandt auf einen Input dar. Der Filter wird über den Input (bspw. das Eingangsbild) gelegt und spaltenweise (Pixelspalte) nach rechts geschoben, bis das Ende des Inputs erreicht ist. Danach wird der Kernel um eine Zeile (Pixelzeile) nach unten verschoben und erneut über den Input bewegt.

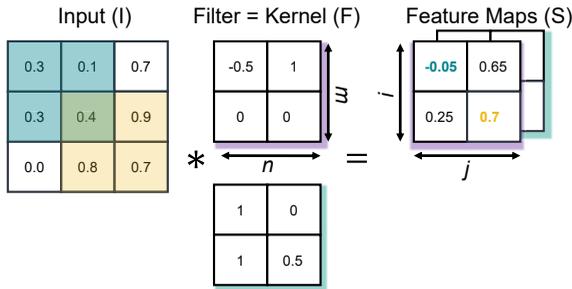


Abbildung 2-21 Funktionsweise der Kernel-Operation. Die Eingangswerte wurden hier auf Werte zwischen 0 und 1 normalisiert. Die farbig hervorgehobenen Zahlen in den Feature-Maps werden aus den gleichfarbigen Bereichen des Inputs durch Multiplikation der Elemente und anschließender Addition berechnet. Zwei Feature-Maps resultieren, da zwei Filter (lila und grün) über den Input gelegt werden.

Es ist ersichtlich, wie sich der Kernel zeilen- und spaltenweise über den Input bewegt und damit Feature-Maps generiert. Das Filtern, d. h. die Berechnungsvorschrift, mit welcher der Filter den Input verarbeitet, lautet: $S_{i,j} = \sum_m \sum_n I_{i+m,j+n} F_{m,n}$. I stellt den Input, F den Filter und S die Feature-Map dar. Die Elemente in F werden ebenfalls als Gewichte $w_{m,n}$ bezeichnet. Analog zum klassischen Neuronalen Netzwerk werden die Elemente des Inputs mit diesen Gewichten multipliziert und es sind diese Gewichte, die zur Extraktion von hilfreichen Merkmalen angepasst werden müssen.

Abhängig von der Schrittweite, mit der die Filter über den Input bewegt werden, ergibt sich die Größe der resultierenden Feature-Maps. Mit einer Schrittweite von 1 (größere Schrittweiten sind möglich) ergibt sich die Dimension der Feature-Map mit Höhe: $Höhe(I) - Höhe(F) + 1$ und Breite: $Breite(I) - Breite(F) + 1$.

Die Anzahl der resultierenden Feature-Maps ist abhängig von der Anzahl der verwendeten Filter (siehe Abbildung 2-21). Jeder Filter durchläuft unabhängig den Input und generiert eine neue Feature-Map. Je mehr Filter verwendet werden, desto mehr Merkmale werden aus dem Input extrahiert. Analog zu den Schichten im klassischen Neuronalen Netz werden die Feature-Maps als Inputs für spätere Filter verwendet. Hiermit bildet sich auch beim CNN ein Aufbau in Schichten (Layer) aus.

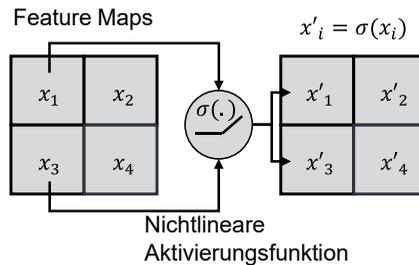


Abbildung 2-22 Anwendung einer nichtlinearen Funktion auf die Elemente der Feature-Maps.

Bevor die Feature-Maps an spätere Filter weitergegeben werden, wird auch hier eine Aktivierungsfunktion auf die Elemente der Feature-Maps angewandt (siehe Abbildung 2-22). Dieser Vorgang verläuft analog zum klassischen Neuronalen Netz.

Im Falle von CNN hat sich hierbei die sogenannte Rectified-Linear-Unit(ReLU)-Funktion mit $\sigma(S) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}, x \in S$ durchgesetzt. Diese ist ebenfalls eine nichtlineare Funktion, die sich als mathematisch vorteilhaft für die Verwendung in CNN erwiesen hat.¹⁵ Die Ableitung der ReLu-Funktion ist 1 für $x > 0$ und 0 für $x \leq 0$ und damit einfach zu berechnen. Diese Tatsache ist hilfreich bei der Berechnung der Gradienten mittels des Backpropagation-Algorithmus.

Die Vorgehensweise, bei welcher die Filter über alle Bereiche des Inputs bzw. der Feature-Maps geschoben werden, garantiert, dass ein Merkmal, das von einem Filter extrahiert wird, unabhängig von der Position im Bild gefunden wird. Bildlich gesprochen sucht jeder Filter nach einem spezifischen Merkmal im Input. Am Beispiel der Klassifikation eines Fahrrades ist es im Voraus nicht klar, an welcher Stelle im Bild sich die Speichen befinden. Das Verschieben der Filter führt dazu, dass unabhängig von der Position im Bild nach Speichen gesucht wird, da diese unabhängig von deren Position ein charakteristisches Merkmal für ein Fahrrad darstellen.

Da die Filter für die Extraktion von Merkmalen verantwortlich sind, bestimmt die Anzahl an verwendeten Filtern pro Schicht zwar automatisch die Kapazität des Netzwerkes, erhöht allerdings auch die Anzahl an Parametern, die trainiert werden müssen. Die

¹⁵ Für detaillierte Ausführungen sei auf (Aggarwal (2019)) verwiesen.

Convolution-Operation wird wiederholt auf die aus den Schichten hervorgehenden Feature-Maps angewandt. Hierdurch entsteht eine hierarchische Extraktion von Merkmalen, die mit zunehmender Tiefe an Komplexität gewinnen. In den frühen Convolution-Schichten werden einfache Merkmale wie Linien und Helligkeitsunterschiede gefiltert. Diese werden mittels der Filter und der darin enthaltenen Gewichte in tieferen Schichten zu zunehmend komplexeren Strukturen kombiniert. Auch dies erfolgt analog zum Vorgehen im klassischen Neuronalen Netzwerk. Am Beispiel der Klassifikation eines Fahrrades werden also in frühen Schichten zunächst einfache Linien extrahiert und dann in späteren Schichten zu einer charakteristischen Speichenstruktur kombiniert, die ein charakteristisches Merkmal von Fahrrädern ist. Wird in einem Bild diese charakteristische Speichenstruktur gefunden, besteht eine hohe Wahrscheinlichkeit, dass es sich bei dem Objekt im Bild um ein Fahrrad handelt. Werden zusätzlich ein Lenker und ein Sattel gefunden, so erhöht sich die Wahrscheinlichkeit entsprechend. Mit zunehmender Tiefe nehmen die extrahierten Merkmale anwendungsspezifischere Formen an. Übertragen auf die Klassifikation eines Pittings ist dieser Prozess qualitativ in Abbildung 2-23 dargestellt.

In der letzten Schicht der Convolutional Base werden alle Merkmale in einen Vektor x überführt (siehe Abbildung 2-20). An dieser Stelle entspricht der Vektor exakt der Form der Eingänge, wie sie auch in einem klassischen Neuronales Netz vorhanden sind. Damit ist es möglich, diesen Vektor als Eingangsvektor an ein klassisches Neuronales Netz zur Klassifikation (siehe Abbildung 2-20) zu übergeben.

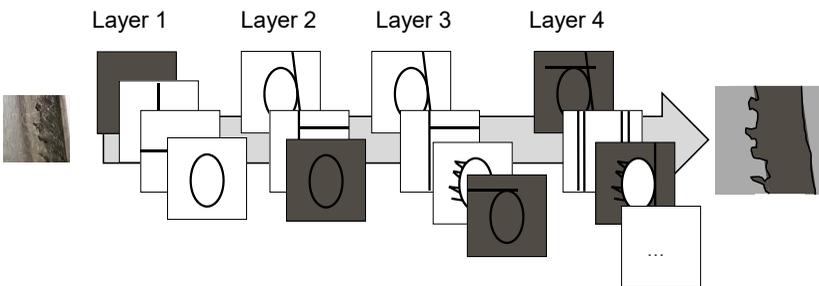


Abbildung 2-23 Qualitative Darstellung der Kombination von extrahierten Merkmalen zu zunehmend komplexeren Merkmalen.

Da das Neuronale Netz mittels dieser Merkmale eine Klassifikation durchführen muss, entscheidet die Qualität der extrahierten Merkmale über die Klassifikationsgüte des

Neuronalen Netzes. Sind die Merkmale diskriminativ, so ist der Klassifikationsfehler gering. Wurde als Merkmal bspw. die Speicherstruktur korrekt extrahiert, kann eine Klassifikation korrekt durchgeführt werden. Die Fähigkeit von CNN zur effektiven Klassifikation leitet sich also vom Umstand ab, dass die Gewichte der Filter in der Convolutional Base aufgrund des vom Neuronalen Netzwerk berechneten und zurückgegebenen Klassifikationsfehlers automatisiert angepasst werden können.

Im Folgenden soll die dritte Frage, wie die Filter der Convolutional Base automatisch angepasst werden können, beantwortet werden.

2.4.5.4 Training der Convolutional Base

Ein Eingangsbild wird durch die Convolutional Base verarbeitet, indem mehrere Schichten an Filtern über das Bild laufen und zunehmend komplexere Merkmale extrahieren. Diese Merkmale werden dann einem Neuronalen Netz zur Klassifikation übergeben. Damit das Neuronale Netz eine Klassifikation mittels dieser Merkmale durchführen kann, müssen die Filter in der Convolutional Base so angepasst werden, dass charakteristische Merkmale extrahiert werden.

Es ist von Neuronalen Netzwerken bekannt, dass zur Anpassung der Filtergewichte die partiellen Ableitungen des Fehlers für diese Gewichte berechnet werden müssen: $\frac{\partial L}{\partial w_{m,n}}$.

Die Herangehensweise zur Ableitung des Fehlers nach den Eingangsgewichten des Neuronalen Netzwerkes mittels des Backpropagation-Algorithmus wurde im vorangehenden Abschnitt diskutiert. Die Vorgehensweise zur Anpassung der Filtergewichte der Convolutional Base kann in leicht abgewandelter Form wiederverwendet werden. Zur Berechnung der partiellen Ableitungen innerhalb der Convolutional Base zeigt Abbildung 2-24 die Verknüpfung der Input-Zellen mit den Feature-Map-Zellen über die Gewichte. Die Farben der Gewichte finden sich hierbei sowohl in den Zellen der Feature-Map als auch in den Input-Zellen.

Input (I)	Filter (F)	Feature Map (S)																	
<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="border: 1px solid black; padding: 5px; text-align: left;">1 0.3</td> <td style="border: 1px solid black; padding: 5px; text-align: left;">2 1 0.1</td> <td style="border: 1px solid black; padding: 5px; text-align: left;">2 0.7</td> </tr> <tr> <td style="border: 1px solid black; padding: 5px; text-align: left;">3 1 0.3</td> <td style="border: 1px solid black; padding: 5px; text-align: left;">4 3 2 1 0.4</td> <td style="border: 1px solid black; padding: 5px; text-align: left;">2 4 0.9</td> </tr> <tr> <td style="border: 1px solid black; padding: 5px; text-align: left;">3 0.0</td> <td style="border: 1px solid black; padding: 5px; text-align: left;">4 3 0.8</td> <td style="border: 1px solid black; padding: 5px; text-align: left;">4 0.7</td> </tr> </table>	1 0.3	2 1 0.1	2 0.7	3 1 0.3	4 3 2 1 0.4	2 4 0.9	3 0.0	4 3 0.8	4 0.7	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="border: 1px solid black; padding: 5px; text-align: left;">-0.5</td> <td style="border: 1px solid black; padding: 5px; text-align: left;">1</td> </tr> <tr> <td style="border: 1px solid black; padding: 5px; text-align: left;">0</td> <td style="border: 1px solid black; padding: 5px; text-align: left;">0</td> </tr> </table>	-0.5	1	0	0	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="border: 1px solid black; padding: 5px; text-align: left;">1 1 1 1 -0.05</td> <td style="border: 1px solid black; padding: 5px; text-align: left;">2 2 2 2 0.65</td> </tr> <tr> <td style="border: 1px solid black; padding: 5px; text-align: left;">3 3 3 3 0.25</td> <td style="border: 1px solid black; padding: 5px; text-align: left;">4 4 4 4 0.7</td> </tr> </table>	1 1 1 1 -0.05	2 2 2 2 0.65	3 3 3 3 0.25	4 4 4 4 0.7
1 0.3	2 1 0.1	2 0.7																	
3 1 0.3	4 3 2 1 0.4	2 4 0.9																	
3 0.0	4 3 0.8	4 0.7																	
-0.5	1																		
0	0																		
1 1 1 1 -0.05	2 2 2 2 0.65																		
3 3 3 3 0.25	4 4 4 4 0.7																		

Abbildung 2-24 Filterelemente, welche die Input-Zellen mit den Feature-Map-Zellen verbinden. Die Farbe der Filterzelle verbindet die Elemente aus Input- und Feature-Map. Zahlen ordnen die Input-Zellen den Feature-Map-Zellen zu.

Es ist ersichtlich, dass jede Zelle der Feature-Map eine Summe von vier Inputs darstellt, die mittels der Filtergewichte gewichtet wurden. Die Feature-Map-Zelle oben links berechnet sich bspw. aus der Summe der Multiplikation der Filterelemente mit den Input-Elementen, welche die Ziffer 1 in den Farbmarkierungen tragen. Daraus resultiert, dass jedes Gewicht des Filters in jeder Zelle der Feature-Map verwendet wird (siehe hierzu Farben in den Zellen der Feature-Map).

Ein relevanter Aspekt ist hierbei, dass ein Gewicht im Filter somit nicht einem einzelnen Element in der Feature-Map, sondern immer allen Elementen zugleich zugewiesen ist.

Der Aspekt, dass ein Gewicht nicht einer einzelnen Zelle der Feature-Map zuzuschreiben ist, sondern immer mehreren, muss bei der Anpassung der Gewichte beachtet werden. Der Grund hierfür ist, dass eine Änderung der Gewichte eine Auswirkung auf alle Elemente in der Feature-Map hat. Klar wird dies, wenn die Ableitung des Fehlers nach einem Filtergewicht mit $\frac{\partial L}{\partial w_{m,n}}$ berechnet werden soll. Dazu muss $\frac{\partial L}{\partial S_{i,j}} * \frac{\partial S_{i,j}}{\partial w_{m,n}}$ berechnet

werden. Es ist nun ersichtlich, dass es nicht ein einzelnes $\frac{\partial S_{i,j}}{\partial w_{m,n}}$ gibt, sondern vier: $\frac{\partial S_{0,0}}{\partial w_{m,n}}$,

$\frac{\partial S_{0,1}}{\partial w_{m,n}}$, $\frac{\partial S_{1,0}}{\partial w_{m,n}}$, $\frac{\partial S_{1,1}}{\partial w_{m,n}}$, die alle das Gewicht $w_{m,n}$ beinhalten. Dies ist auch anhand der Farben in der Feature-Map in Abbildung 2-24 ersichtlich, da jede Zelle alle vier Farben der

Gewichte beinhaltet. Damit ist $\frac{\partial L}{\partial w_{m,n}} = \sum_{S_{i,j} \in S} \frac{\partial L}{\partial S_{i,j}} * \frac{\partial S_{i,j}}{\partial w_{m,n}}$ bzw. unter Einbeziehung der

Aktivierungsfunktion σ , $\frac{\partial L}{\partial w_{m,n}} = \sum_{S_{i,j} \in S} \frac{\partial L}{\partial o_{i,j}} * \frac{\partial o_{i,j}}{\partial S_{i,j}} * \frac{\partial S_{i,j}}{\partial w_{m,n}}$. Analog zum Vorgehen beim

klassischen Neuronalen Netz werden so die Fehler mittels dieser Berechnungsvorschrift durch die Convolutional Base geschleust und die Gewichte angepasst.

Die Anpassung der Gewichte erfolgt anschließend analog zum klassischen Neuronalen Netzwerk mit: $w_{0,0}^{neu} = w_{0,0}^{alt} - \alpha * \frac{\partial L}{\partial w_{0,0}^{alt}}$. Der Abschnitt *Backpropagation in der Convolutional Base* im Anhang beinhaltet analog zum Neuronalen Netz ein konkretes Berechnungsbeispiel.

Es sei abschließend nochmals betont, dass die Leistungsfähigkeit des Ansatzes in der Fähigkeit liegt, semantisch relevante Merkmale automatisiert aus den Bilddaten zu extrahieren, sodass diese möglichst optimal für die Klassifikation der Bilddaten geeignet sind.

2.4.5.5 Berechnung der Ähnlichkeit von Bilddaten

Der angesprochene Vektor, welcher der letzten Schicht in der Convolutional Base folgt und als Input für das Neuronale Netz verwendet wird, kann ebenfalls zur Berechnung der Ähnlichkeit von Bilddaten verwendet werden.

Es ist aus dem vorherigen Abschnitt bekannt, dass dieser Vektor die semantischen Informationen eines Bildes codiert. Sind zwei Bilder ähnlich, so sind auch die Einträge der Vektoren ähnlich, d. h. die Distanz der Vektoren ist gering. Am Beispiel der Klassifikation von Fahrzeugen sind die Vektoren zweier Fahrräder ähnlich, wohingegen sich die Vektoren von Fahrrädern und PKWs unterscheiden. Die Distanz der Vektoren kann mittels der Wurzel der summierten quadratischen Abweichung der Vektorelemente (euklidische Distanz) mit $\sum_{i \in \text{Vektoren}} \sqrt{(x_i^1 - x_i^2)^2}$ berechnet werden, wobei x_i^1 und x_i^2 die Einträge der beiden Vektoren darstellen.

Über diesen Prozess kann eine ähnlichkeitsbasierte Klassifikation vorgenommen werden, wie sie in Abschnitt 6.2.4 zur dateneffizienten Klassifikation verwendet wird. Hierbei wird der Vektor eines unbekanntes Objektes mit den Vektoren von bekannten Objekten verglichen und es wird überprüft, zu welchen Objekten die geringste Distanz besteht. Am Beispiel der Klassifikation von Fahrzeugen wird ein neues Objekt bspw. mit je drei Fahrrädern und drei PKWs verglichen. Stellen die Bilder mit den drei geringsten Distanzen zum unbekanntes Objekt allesamt Fahrräder dar, so kann das Objekt als Fahrrad klassifiziert werden. Gleiches gilt, wenn sich unter den drei geringsten Distanzen auch ein Bild eines PKWs befindet, wenngleich die Klassifikationswahrscheinlichkeit dann etwas geringer ausfällt.

3 Stand der Forschung und Technik

In diesem Kapitel wird zunächst der aktuelle Stand zur bildbasierten Defekterkennung auf metallischen Oberflächen dargestellt. Daran anschließend folgt die Darstellung von Arbeiten im Bereich der Verschleißüberwachung an Maschinenelementen, bevor dediziert auf die Arbeiten zur Verschleißüberwachung am KGT und an den sich daraus ergebenden Defiziten eingegangen wird.

3.1 Bildbasierte Defekterkennung auf metallischen Oberflächen

Der Stand der Forschung zur bildbasierten Defekterkennung zeigt, dass seit der Renaissance von Deep-Learning-Methoden im Jahre 2006 (LeCun, Bengio & Hinton 2015) auch im Bereich der Bilderkennung die Mehrzahl an performanten Ansätzen auf Methoden des Deep Learning basiert. Diese Ansätze werden meist bei Vorhandensein ausreichend großer Datenmengen eingesetzt. Zuvor stützten sich Ansätze auf *klassische Machine-Vision-Verfahren*, die seit den 1970er Jahren auf manuell definierten Merkmalen und manueller Merkmalsextraktion aufbauen (Beyerer & Puente León et al. 2016).

In neueren Forschungsarbeiten finden sich vielzählige Ansätze zur Detektion und Klassifikation von Fehlstellen auf technischen Oberflächen unterschiedlicher Domänen, bspw. auf Eisenbahnschienen (Faghieh-Roohi & Hajizadeh et al. 2016), (Soukup & Huber-Mörk 2014), auf Betonbauteilen (Koch & Georgieva et al. 2015), (Cha, Choi & Büyükköztürk 2017), (Cha & Choi et al. 2018), (Ren & Huang et al. 2020), auf Holzwerkstoffen (He & Liu et al. 2019), (Zhang & Liu et al. 2015) oder auf Textilien (Park & Oh et al. 2019), (Hu & Long et al. 2021). Die aktuellen Ansätze haben hierbei gemeinsam, dass sie größtenteils auf rein Deep-Learning-basierte Verfahren setzen. Die Firma Cognex bietet mit ihrem Produkt Cognex-Deep Learning (Cognex 2021)¹⁶ eine Lösung zur bildbasierten Überwachung individueller industrieller Produkte an. Die notwendigen Datenmengen sowie die spezifischen Fähigkeiten und Grenzen des Systems werden nicht weiter beschrieben.

Neben den genannten Ansätzen in technischen Domänen sind für die vorliegende Arbeit besonders solche Ansätze relevant, die sich zum Ziel setzen, Fehlstellen auf me-

¹⁶ Cognex (2021), *Cognex Deep Learning für Life Sciences*. <https://connect.cognex.com/> [31.01.2022].

tallischen Oberflächen zu identifizieren. Bisherige Forschungsarbeiten lassen sich hierbei neben der zu überwachenden Komponente (metallische Bauteile, metallische Werkstoffe sowie Maschinenelemente und Werkzeuge) vor allem hinsichtlich der verwendeten Methoden gliedern. Hierbei kann eine Untergliederung in klassische Verfahren vorgenommen werden, die hauptsächlich auf manueller Merkmalsdefinition und -extraktion basieren (auch unter Verwendung von Verfahren des Maschinellen Lernens), und solchen, bei denen eine automatische Merkmalsextraktion und Klassifikation vorgenommen werden (Deep-Learning-basierte Verfahren).

Speziell im Bereich der Identifikation von Fehlstellen auf Oberflächen von Rohstahlprodukten (Bleche, Bandstahl, kaltgewalzter Stahl, warmgewalzter Stahl, Stahlbrammen) findet sich eine Vielzahl an Arbeiten, die versuchen, relevante Defekte wie Risse, Kratzer oder Einschlüsse zu identifizieren. Dies lässt sich auf die Tatsache zurückführen, dass die Stahlindustrie die Grundmaterialien für viele nachfolgende Prozesse liefert und damit hohen Qualitätsanforderungen genügen muss.

Relevante Datensätze, die Defekte auf metallischen Oberflächen zeigen und in einer Vielzahl von weiteren Arbeiten für die Entwicklung und Validierung von Modellen verwendet werden, sind der Severstal-Datensatz¹⁷, der NEU-Datensatz (Song & Yunhui 2019) sowie der NEU-DET-Datensatz (He & Song et al. 2020) und der GC10-DET-Datensatz (Lv & Duan et al. 2020). Diese Datensätze wurden vom Autor um den KGT-Datensatz (Schlagenhauf 2021) ergänzt.

Ansätze, die klassische Machine-Vision-Techniken wie Gaborfilter oder Wavelet-Transformationen auf die Bilddaten anwenden und die so entstehenden Merkmale entweder schwellwertbasiert (Yun & Kim et al. 2017), (Liu & Wang et al. 2017), (Zheng, Kong & Nahavandi 2002) oder auf Basis einfacher Machine-Learning-Verfahren klassifizieren, finden sich bei (Zheng, Kong & Nahavandi 2002), (Ghorai & Mukherjee et al. 2013) sowie (Kang & Liu 2005). (Luo & He 2016) identifizieren Oberflächendefekte wie Oxidstellen, Rollmarken und Kratzer auf warmgewalztem Stahl mittels einer Heuristik und eines klassischen Machine-Vision-Verfahrens. Hierzu wird ein kostenattraktives Kamerasystem verwendet. Zur Identifikation von Verschleiß auf Werkzeugen verwenden (Li & An 2016) unter anderem klassische Schwellwertverfahren zur Untersuchung der Bildentropie und einer darauf basierenden Klassifikation. Im Bereich der kamerabasierten

¹⁷ Severstal (2019), *Severstal: Steel Defect Detection*. <https://www.kaggle.com/c/severstal-steel-defect-detection/data> [08.12.2020].

Überwachung metallischer Zugproben finden sich Arbeiten von (Medved, Breton & Irving 2004), (Zamber & Hillberry 1999) sowie (Vanderwalde & Hillberry 2008), bei denen Oberflächendefekte bildbasiert erfasst werden und deren Einfluss auf die Ermüdungslebensdauer der Zugproben untersucht wird.

(Galan & Orta et al. 2018) verwenden ein schwellwertbasiertes Verfahren zur Identifikation von Verschleiß auf Gussbauteilen, bei dem Pixel, die einen Schatten auf der beleuchteten Bauteiloberfläche darstellen, gezählt werden. Ähnlich wird in der früheren Arbeit von (Alegre & Alaíz-Rodríguez et al. 2009) vorgegangen, in welcher der Verschleiß an den Schneidkanten von Schneidwerkzeugen schwellwertbasiert abgeschätzt wird. Der so entstehende Konturvektor wird in ein Diagramm überführt und der Verschleiß wird anhand dessen Gestalt interpretiert.

Im Bereich der Überwachung von Maschinenelementen und verwandten Bauteilen untersuchen (Chen, Tsai & Hsu 2016) einen kamerabasierten Ansatz zur Identifikation von Oberflächendefekten auf Kugellager-Kugeln. (Kattelus, Miettinen & Lehtovaara 2018) und (Vanoverberghe 2015) untersuchen Oberflächendefekte auf Zahnflanken von Zahnrädern, wobei letzterer einen rein bildbasierten Ansatz wählt und erstere einen Abgleich der Bilddaten mit den Ergebnissen einer Ölanalyse durchführen. Ein früherer Ansatz zur stichprobenbasierten Beobachtung von Ermüdungsschäden in einer Schrägkugellager-Innenringlaufbahn findet sich bei (Schaeffler 2000)¹⁰.

Die Herausforderung der Ansätze, bei denen mittels Vorverarbeitungsschritten Merkmale manuell extrahiert werden, die dann regelbasiert oder auf Basis von einfachen Machine-Learning-Modellen klassifiziert werden, besteht darin, dass sich diese häufig einer komplexen Pipeline an aufeinanderfolgenden, speziell für einen Anwendungsfall angepassten Schritten bedienen. In Fällen, in denen sich die Randbedingungen ändern oder die Daten eine hohe Varianz aufweisen, können Deep-Learning-basierte Ansätze Vorteile aufweisen (Chollet 2018).

(Masci & Meier et al. 2012), (Tao & Zhang et al. 2018) und (Liu, Wang & Chen 2019) nutzen Deep-Learning-basierte Ansätze zur Klassifikation von Fehlstellen auf gewaltem Stahl bzw. Stahlbändern. (Hatab, Malekmohamadi & Amira 2020) beschreiben Detektionsalgorithmen zur Detektion von Oberflächendefekten auf Stahloberflächen, wobei durchschnittliche Ergebnisse erzielt werden, was von den Autoren auf die geringe Menge der zur Verfügung stehenden Trainingsdaten zurückgeführt wird. (Kumar & Va-

shishtha et al. 2021) untersuchen einen Ansatz, in dem Körperschallsignale eines Kugellagers im Frequenzbereich als Bild einem CNN zum Training übergeben werden. Das Modell wird hierbei zunächst auf Daten eines verwandten Rollenkugellagers trainiert und dann in einem zweiten Transfer-Learning-Ansatz auf das vorliegende Kugellager angepasst. Damit kann die Wirksamkeit des Transfer-Learning gezeigt werden. (Lv & Duan et al. 2020) stellen ein Deep-Learning-basiertes Klassifikations- und Detektionsmodell zur Identifikation von Fehlstellen auf metallischen Oberflächen vor. Hierzu erstellen die Autoren den sogenannten GC10-DET-Datensatz, der in seinen Charakteristika dem schon bestehenden NEU-DET-Datensatz ähnelt.

(Di He, Xu & Zhou 2019) nehmen eine Umgestaltung von Detektionsalgorithmen vor, wie sie bspw. bei (Ren & He et al. 2016) vorgestellt werden, um bessere Ergebnisse bei der Detektion von Fehlstellen auf heißgewalztem Stahl zu erreichen. (Liu, Xu & Xu 2019) folgen einem ähnlichen Ansatz und modifizieren das VGG16-Netzwerk. Explizit werden hierbei Rollmarken gesucht, die mit einer Genauigkeit von 96 % detektiert werden können. Zusätzlich werden Erweiterungen des Netzwerkes um eine Long-Short-Term-Memory-Komponente sowie eine Attention-Funktion vorgenommen, mittels welcher die Fehlerrate weiter gesenkt werden konnte. (Li & Su et al. 2018) untersuchen den Einsatz des You-Only-Look-Once (YOLO)-Algorithmus zur Detektion und Kategorisierung von sechs Fehlervarianten auf kaltgewalztem Stahl. Es wird ein eigens aufgebauter Datensatz verwendet, der allerdings nicht veröffentlicht wurde. Das Modell ist in der Lage, eine Genauigkeit von 97,55 % zu erreichen. (Youkachen & Ruchanurucks et al. 2018) verwenden zur Identifikation von Fehlstellen auf warmgewalztem Stahl einen Convolutional Autoencoder. Die Autoren verwenden hierfür den bekannten NEU-Datensatz. (He & Song et al. 2019) adressieren mittels der Generative-Adversarial-Netzwerk(GAN)-Methode nach (Goodfellow & Pouget-Abadie et al. 2014) die Herausforderung der geringen Datenmengen, indem sie versuchen, künstliche Daten zu generieren. Dieser Ansatz wurde auch in (A_Sun 2019) untersucht, wobei festgestellt wurde, dass die künstliche Generierung von KGT-Oberflächendefekten mittels GAN-Architekturen zwar eine Verbesserung der Klassifikationsgenauigkeit für Bilddaten des KGTs zulässt, die Ergebnisse unter der Einschränkung der verwendeten Methodik aber nicht signifikanter sind als jene Ergebnisse, die mittels klassischer Data Augmentation erreicht werden können. (Niu & Li et al. 2020) verwenden ebenfalls GAN-basierte Methoden, sie erweitern diese allerdings und versprechen damit signifikantere Ergebnisse. (Kim &

Park et al. 2019) und (Deshpande & Minai et al. 2020) verwenden Deep-Learning-basierte Siamese-Neural-Network-Strukturen zur Klassifikation von Defekten auf Stahloberflächen, wobei letztere eine deutliche Reduktion der Merkmalsanzahl verfolgen und die dabei erzielte Genauigkeit dennoch deutlich über jener in der Arbeit von (Kim & Park et al. 2019) liegt.

Eine Zusammenfassung unterschiedlicher Ansätze zur Identifikation von Defekten auf metallischen Oberflächen findet sich bei (Neogi, Mohanta & Dutta 2014) und (Luo & Fang et al. 2020), wobei die neuere Arbeit von (Luo & Fang et al. 2020) eine klare Tendenz zu Verfahren erkennen lässt, welche Machine-Learning- und Deep-Learning-Verfahren implementieren.

Aus dem Stand der Forschung zur kamerabasierten Defekterkennung auf metallischen Oberflächen lassen sich grundsätzlich vier Kernaussagen ableiten:

1. Neuere Arbeiten setzen vermehrt auf die Anwendung von Verfahren des Maschinellen Lernens und insbesondere des Deep Learnings.
2. Obwohl eine große Anzahl an Deep-Learning-Verfahren sowie weitreichende Weiterentwicklungen derselben bestehen, reduzieren sich die in diesen Arbeiten verwendeten und untersuchten Datensätze weitestgehend auf eine geringe Anzahl an öffentlich zugänglichen Stand-der-Technik-Datensätze. Diese Datensätze repräsentieren größtenteils Bilddaten, die unter nicht verschmutzten Laborbedingungen aufgenommen wurden. Dies ist im Bereich der Maschinenelemente allerdings oft nicht der Fall. Der später vorgestellte KGT-Datensatz (Schlagenhauf 2021) stellt damit eine für diese Aufgabe zielführende Erweiterung dar.
3. Die Einordnung in ein Modell zur Abschätzung des Verschleißzustandes und zur Prognose eines Verschleißmerkmals auf Basis eines extrahierten Schadens erfolgt in den Arbeiten nicht. Weiterhin findet keine Analyse der Bilddaten hinsichtlich der Entstehung und des Wachstums eines Verschleißmerkmals statt.
4. Es lässt sich feststellen, dass klassische Ansätze vermehrt dann eingesetzt werden, wenn sich das Umfeld und die Umgebungsbedingungen nicht ändern sowie wenn die zu erwartenden Merkmale am Überwachungsobjekt gleich bleiben und damit eindeutige Merkmale zur Klassifikation definiert werden können. Andernfalls werden Deep-Learning-Verfahren eingesetzt.

3.2 Verschleißüberwachung an Maschinenelementen

Nach der bildbasierten Überwachung von metallischen Oberflächen wird in diesem Abschnitt der aktuelle Stand der Forschung im Bereich der generellen Verschleißüberwachung an Maschinenelementen dargestellt. Da es hierbei nicht möglich ist, sich wie bei den Bilddaten auf ein Sensorsignal zu beschränken, werden lediglich die wichtigsten Arbeiten dargestellt.

Die gängigsten Signale, die im aktuellen Stand der Forschung und Technik zur Überwachung des Verschleißes verwendet werden, sind: Signale, die direkt aus der Maschinensteuerung verwendet werden können (Positionsdaten, Motorströme und Motormomente), Luft- und Körperschall-Signale sowie Temperatur-, Moment- und Kraftsignale. Grundsätzlich kann hierbei nach (Henrich 2013) und (Munzinger, Schopp & Henrich 2010) zwischen zustandsorientierten und belastungsorientierten Ansätzen unterschieden werden. Eine genaue Beschreibung der Verfahren erfolgt im Kontext der Zustandsüberwachung am KGT und soll daher nur kurz angeschnitten werden. Bei zustandsorientierten Verfahren wird über die Aufnahme einer Sensorgröße auf den vorliegenden Schädigungszustand einer Komponente geschlossen, während bei belastungsorientierten Ansätzen über die Historie der Lastkollektive der Zustand der Komponente abgeschätzt wird. Es kann damit auch von heuristischen und analytischen Verfahren zur Abschätzung des Verschleißzustandes gesprochen werden.

Je nach zu überwachender Maschinenkomponente eignen sich verschiedene Messgrößen zur Analyse des Verschleißfortschritts. Beispielsweise hat sich der Körperschall (Metzele 2008), (Albers, Dickerhof & Burger 2009), (Wang & Han et al. 2019), (Lee & Yu et al. 2021) zur Überwachung von Wälzlagern, Getrieben und Riementrieben als probates Signal etabliert. Die Auswertung von Temperaturdaten wird vermehrt verwendet, um Hauptspindeln und weitere elektrische Komponenten wie elektrische Motoren und Generatoren zu überwachen (Guo & Infield et al. 2012), (Gupta & Peroulis 2013), (Abdusamad, Gao & Muljadi 2013). Der Motorstrom eines Antriebes wird hingegen unter anderem für die Zustandsüberwachung an Wälzlagern, Getrieben, dem KGT und dem Motor verwendet (Kar & Mohanty 2006), (Bradley & Ebrahimi et al. 2014), (Dalvand, Kalantar & Safizadeh 2016), (Triyono & Prasetyo et al. 2021). Es lassen sich also Überschneidungen zwischen Komponenten und Sensorsignalen erkennen. Ein Trend bei der Verarbeitung der Daten hin zu Methoden des Deep Learnings wird von (Serin

& Sener et al. 2020) in einer Übersichtsrecherche zur Zustandsüberwachung von Fräs-
werkzeugen beobachtet.

Im Folgenden sollen relevante Arbeiten mit besonderem Bezug zu Maschinenelemen-
ten näher betrachtet werden.

3.2.1 Körperschall

Die Überwachung von Kugellagern auf Basis klassischer Prognoseverfahren wurde von (deBotton, Ben-Ari & Sher 2000) unter Beobachtung des von der Maschine erzeugten Körperschalls untersucht. Über sechs Monate hinweg zeichneten die Autoren hierfür die Vibrationen an unterschiedlichen Stellen am Gehäuse auf. Die Frequenzen wurden gefiltert und bei Überschreiten von Grenzwerten in bestimmten Frequenzbändern wurde auf eine Fehlfunktion geschlossen. (Brecher, Klein & Paepenmüller 2006) untersuchten den Körperschall einer Maschine zur Überwachung von Linearführungssystemen. Unter Verwendung statistischer Methoden werden die Schadensarten Grübchen und Spankontamination unterschieden. (Ellinger & Benker et al. 2019) untersuchen die Komponente auf dieselbe Art, schätzen mit den Daten allerdings den Vorspannungsverlust und daraus den Zustand der Linearführung ab. (Ehrmann, Isabey & Fleischer 2016) heben die Wichtigkeit der Zustandsüberwachung von Ritzel-Zahnstangenantrieben hervor. Basierend auf dieser Arbeit untersuchte (Ehrmann 2019) in seiner Dissertation den Einsatz von Acoustic Emission zur Früherkennung von Defekten auf Ritzel-Zahnstangenantrieben. Die generelle Machbarkeit des Ansatzes konnte gezeigt werden. (Al-Arbi 2012) verwendet den Körperschall zur Analyse des Zustandes eines Getriebes. In der Dissertation wird neben den Forschungsergebnissen zur Signalaufnahme und Vorverarbeitung ein Vergleich statistischer Methoden zur Auswertung angestellt. Es werden die generelle Machbarkeit der Fehlererkennung mittels der Betrachtung von Schwingungssignalen sowie die Adaption und Optimierung stochastischer Funktionen auf die Anwendung beschrieben. Eine Gegenüberstellung der klassischen Hüllkurvenanalyse mit zwei Ansätzen des Maschinellen Lernens beschreibt (Lessmeier 2017). Er bestätigt eine prinzipielle Eignung der Methoden des Maschinellen Lernens und leitet aus den Schwächen Handlungsempfehlungen für weitere Arbeiten ab. (Geister, Hohwieler & Fraunhofer-Institut für Produktionsanlagen und Konstruktionstechnik 2019) stellen die Erkennungsmöglichkeiten des Abnutzungsgrades einer Schleifscheibe anhand verschiedener Messgrößen und Algorithmen gegenüber. Zur Detektion

des Körper- und Luftschalls sowie der Beschleunigung werden dabei Sensoren unterschiedlicher Hersteller verwendet und durch Algorithmen der Klassen K-Nearest Neighbours (KNN), Support Vector Machine (SVM), Random Forrest und Naive Bayes analysiert. Als Vergleichssieger etablieren sich die Körper- und Luftschalldaten in Kombination mit einer Random-Forrest Methode, die eine Genauigkeit von 90 % erzielen kann. Der Einsatz Neuronaler Netze zur Körperschallanalyse wird in der Arbeit von (Alqatawneh & Rabeyee et al. 2021) beschrieben, wobei eine geeignete Netzwerkarchitektur zur allgemeinen Schwingungsanalyse vorgestellt wird und deren spezifische Anpassungen diskutiert werden. Die Anwendung wird an einer Getriebeeinheit validiert. (Krishnakumar, Rameshkumar & Ramachandran 2018) testen die vier Ansätze Decision Tree, Naive Bayes, SVM sowie Neuronales Netz zur Interpretation von Schwingungsdaten einer Fräsmaschine zur Bestimmung des Werkzeugzustandes. Die Aufgabe besteht darin, die Beschaffenheit des Fräsaufsatzes in die drei Zustandskategorien gut, gebraucht und verschlissen einzuordnen, was dem künstlichen Neuronalen Netz mit 99,3 % am besten gelingt. (Lee & Yu et al. 2021) untersuchen die Eignung von Deep-Learning-Methoden für die Klassifikation von Körperschalldaten im Hinblick auf den Zustand eines Kugellagers. Die gemessenen Signale werden nach dem Prinzip der Spracherkennung verarbeitet und einer eingeschränkten Version des VGG19-Netzwerkes übergeben. Als Resultat wird der Zustand des Kugellagers den vier Kategorien Unwucht, Fehlausrichtung, Reibung und Normal zu 99,5 % richtig zugeordnet. Ein Ansatz zur Verwendung von Transfer-Learning wird in der Arbeit von (Cao, Zhang & Tang 2018) untersucht. Die Autoren untersuchen aufbauend auf dem Netzwerk von (Krizhevsky, Sutskever & Hinton 2012) verschiedene Stufen des Transfer-Learnings und erzielen bei der Untersuchung von Schwingungsdaten eines Zahnrades sehr gute Ergebnisse. (Hillenbrand & Fleischer 2020) verwenden eine Autoencoder-Architektur zur Merkmalsextraktion von Körperschalldaten von Radialkugellagern. Die extrahierten Merkmale werden mittels eines Clusterverfahrens verarbeitet und so werden drei Fehlertypen beim Normalbetrieb einer Maschinenkomponente unterschieden.

Eine weiterführende Aufstellung zu Ansätzen der körperschallbasierten Überwachung von Maschinenkomponenten ist bei (Toh & Park 2020, S. 17) zu finden.

3.2.2 Luftschall

(Krishnakumar, Rameshkumar & Ramachandran 2018) verwenden analog zu den Untersuchungen auf Basis des Körperschalls die vier Ansätze Decision Tree, Naive

Bayes, SVM sowie Neuronales Netz zur Untersuchung des Luftschalls einer Fräsmaschine. Auch bei diesen Untersuchungen setzt sich das Neuronale Netz gegenüber den anderen Verfahren ab. (Rafezi, Akbari & Behzad 2012) und (Kothuru, Nooka & Liu 2019) analysieren die Schallentwicklung an Fräswerkzeugen, wobei erstere auf klassische Verfahren und letztere auf Deep-Learning-basierte Verfahren zur Datenanalyse setzen. (Lee & Jwo et al. 2020) untersuchen ebenfalls den Luftschall während des Fräsbearbeitungsprozesses und versuchen den Zustand des Werkzeuges mittels Deep-Learning-Verfahren abzuleiten. Die Autoren bauen auf die von (Valenti & Squartini et al. 2017) konzipierte CNN-Architektur auf und erreichen eine Genauigkeit von 97,44 %. Einen generelleren Ansatz zur Interpretation des Schallpegels einer Maschine zu finden, wird im Rahmen der DCASE2020 Challenge von (Koizumi & Kawaguchi et al. 2020) angestrebt. Ziel ist es, mittels eines nicht überwachten Lernverfahrens (engl. *Unsupervised Learning*) Anomalien in den Geräuschen einer Maschine aufgrund von Verschleiß zu erkennen. Dabei erreichen die Autoren eine maximale Genauigkeit von 85 %. (van Truong & Chi Hieu et al. 2021) verfolgen einen ähnlichen Ansatz wie bei der DCASE2020 Challenge und entwickeln das von (Ronneberger, Fischer & Brox 2015) ursprünglich für medizinische Zwecke entwickelte U-Net-Verfahren weiter. Das System soll Anomalien in Maschinengeräuschen unterschiedlicher Herkunft erkennen, wofür Trainingsbeispiele von sechs verschiedenen Maschinenkomponenten verwendet werden. Im Schnitt erreicht das Netzwerk eine Genauigkeit von 83 %. (Caggiano & Caggiano 2018) beschreiben einen Ansatz, in dem Luftschall, Körperschall und Kraft kombiniert werden, um den Zustand eines Drehmeißels zu überwachen. Zur Datenverarbeitung werden Deep-Learning-Architekturen verwendet. (Serin & Sener et al. 2020) geben einen Überblick über unterschiedliche Arbeiten, die unter anderem Luftschall, Körperschall sowie den Motorstrom kombinieren und mittels Deep-Learning-Verfahren analysieren.

3.2.3 Motorstrom

(Huf 2012) verwendet Beschleunigungs- sowie Motorstromwerte und ermittelt das Belastungsintegral von Kugelumlaufmuttern über die Zeit. Die verbleibende Betriebsdauer wird anhand der zum Zeitpunkt gefahrenen Lastkurve und des resultierenden Abnutzungsvorrates prognostiziert. Der Ansatz zählt zu den belastungsorientierten Methoden. (Henrich 2013) verwendet in seiner Dissertation ebenfalls den Motorstrom und

ergänzt diesen um Stichprobenmessungen weiterer Größen wie Schwingung, Temperatur und Drehzahl eines KGT. Hiermit wird eine Kombination aus einem belastungs- und zustandsorientierten Ansatz verfolgt. Die untersuchten Spindeln können dabei mit einer Genauigkeit von ca. 93 % vor dem mechanischen Ausfall ausgetauscht werden. (Yang & Li et al. 2020) wählen für die Überwachung des Zustandes eines KGT zehn Prognosefunktionen auf Basis des Maschinellen Lernens aus und analysieren damit die Motorstromdaten während des Betriebes eines KGT. Den Autoren zufolge befähigt diese Information drei der zehn Algorithmen zu einer Genauigkeit von 100 %. Untersucht werden zwei Schadensstadien sowie der Normalzustand. Die Ergebnisse sind aufgrund der außerordentlichen Güte allerdings kritisch zu hinterfragen. Den Einsatz einer Unsupervised-Learning-Methode zur Frequenzanalyse des Motorstroms zur Identifikation von Schäden an Kugellagern beschreiben (Cipollini & Oneto et al. 2019). Der Ansatz wird ohne konkrete Angaben der Leistungsfähigkeit als machbar bestätigt.

3.2.4 Temperatur

(Touret & Changenet et al. 2018) geben einen Überblick über unterschiedliche Möglichkeiten zur Zustandsüberwachung einer Getriebeeinheit anhand der Temperatur. Allgemein wird zwischen der Anomalieerkennung durch Messung der Temperatur einer Baugruppe sowie der Infrarotüberwachung konkreter Bauteile unterschieden. Erstere Herangehensweise ist einfach in der Umsetzung und dient der Ausfallvorhersage wenige Stunden oder Tage vor Ausfall der Komponente. Die aufwändigeren Infrarot-Methoden erlauben dagegen detailliertere Aussagen über den Zustand einzelner Komponenten. Die Anwendung der temperaturgestützten Zustandsüberwachung betrachten (Fu & Chu et al. 2019) an Getrieben in Windturbinen. Zur Analyse der Temperaturdaten wird ein CNN verwendet. Die Ergebnisse zeigen die Funktionsweise des Ansatzes und nach Angaben der Autoren kann der Ansatz auch auf ähnliche Getriebeeinheiten übertragen werden. Ähnliche Resultate erzielen (Cheng, Wang & Zhang 2019) bei der Überwachung der Kugellager in Radlagern von Zügen. So kann die Anomalie spätestens zwei Tage vor Versagen der Komponente anhand der Temperaturdaten detektiert werden. (Kuntoğlu & Sağlam 2021) untersuchen einen Multisensor-Ansatz zur Werkzeuganalyse an Drehmaschinen. Sie kommen zu dem Schluss, dass die Verwendung der Temperaturmessungen in Kombination mit den Schwingungsdaten die für die Aufgabe höchste Aussagekraft ermöglicht. Insgesamt werden dabei die fünf Sensorgrößen Schall, Vibration, Temperatur, Schneidkraft und Motorspannung verwendet. (Lee &

Huang et al. 2018) konzentrieren sich bei der Analyse von Schneidwerkzeugen auf Temperaturdaten. Die durch eine Infrarot-Kamera erzeugten Bilder werden unter Zuhilfenahme Neuronaler Netze analysiert und der Zustand des Werkzeugs wird mit einer Fehlerrate von ca. 5 % richtig eingeschätzt.

Zusammenfassend lässt sich analog zu den vorherigen Sensorgrößen festhalten, dass auch bei der Auswertung von Temperaturdaten vermehrt Methoden des Maschinellen Lernens zum Einsatz kommen. Speziell die Architektur des Neuronalen Netzwerkes erzielt anwendungsübergreifend gute Ergebnisse.

Insgesamt lässt sich die Popularität von maschinellen Lernverfahren zur Sensordatenanalyse im Kontext der Zustandsüberwachung mit den nicht vollständig erforschten Möglichkeiten der Verfahren und den damit einhergehenden Potenzialen begründen. Weiterhin lassen sich maschinelle Lernverfahren oftmals aufwandsminimaler einsetzen und untersuchen, da einfache Sensorsysteme für eine Vielzahl von Anwendungen eingesetzt werden können und häufig keine aufwändigen, speziell zugeschnittenen Lösungen notwendig sind. Eine Einschränkung ist, dass häufig große Mengen an Daten für das Training der Modelle notwendig sind. Die zunehmende Anzahl an öffentlich zugänglichen Datensätzen aus unterschiedlichen Domänen verringert allerdings die Einstiegshürde zur Entwicklung von maschinellen Lernverfahren und verstärkt damit diesen Trend, wenngleich die Datensätze die in der Realität vorliegenden Gegebenheiten häufig nur eingeschränkt wiedergeben.

3.3 Relevante Arbeiten zur Verschleißüberwachung am Kugelgewindetrieb

Nachdem in den vorherigen Abschnitten relevante Arbeiten zur bildbasierten Defekterkennung auf metallischen Oberflächen sowie zur Verschleißüberwachung an Maschinenelementen dargestellt wurden, wird in diesem der Fokus auf die relevanten Arbeiten zur Verschleißüberwachung am KGT gelegt.

Im Folgenden werden relevante Arbeiten zur Verschleißüberwachung und Verschleißminimierung am KGT chronologisch dargestellt. Diese Arbeiten sind für die vorliegende Arbeit insofern von besonderer Relevanz, als sie die bereits untersuchten Ansätze zur Verschleißüberwachung am KGT zusammenfassen und daraus ersichtlich wird, dass die bildbasierte Überwachung der Oberflächenzerrüttung eine bisher nicht betrachtete Ergänzung zu den bestehenden Arbeiten darstellt.

Wie in Abschnitt 2.1.1 dargelegt wurde, kann der am KGT auftretende Verschleiß entweder direkt über ein Verschleißmerkmal oder indirekt über sich durch den Verschleiß ändernde Größen gemessen werden. Weiterhin lässt sich die Verschleißanalyse am KGT in belastungsorientierte und zustandsorientierte Ansätze unterteilen (Henrich 2013). Dabei beschränken sich belastungsorientierte Ansätze ausschließlich auf die Abschätzung des vorliegenden Lastkollektives und die Berechnung der Lebensdauer bzw. des Verschleißgrades anhand (Deutsches Institut für Normung 2011b). Hierbei wird also keine Messung von Verschleißmerkmalen vorgenommen. Da der genaue Zustand einer Komponente nicht gemessen wird, kann die Genauigkeit zur Bestimmung des tatsächlich vorliegenden Verschleißes demnach als einschränkt angenommen werden.

Die Aufteilung in direkte und indirekte Verfahren zur Verschleißüberwachung beschränkt sich damit auf die zustandsorientierten Ansätze, da hierbei anders als bei den belastungsorientierten Ansätzen der Verschleiß im System anhand einer aufgenommenen Größe abgeleitet wird (siehe hierzu Abbildung 2-1 in Abschnitt 2.1.1). Dadurch besteht auf der einen Seite ein erhöhter Aufwand zur Aufnahme der Messgrößen, auf der anderen Seite erlauben zustandsorientierte Ansätze die Feststellung des aktuell vorliegenden Systemzustandes.

Im Folgenden werden die Arbeiten chronologisch dargestellt.

(Golz 1990) widmet sich in seiner Arbeit der Optimierung des Betriebsverhaltens des KGTs. Besonderes Augenmerk legt er hierbei auf das Kugelrückführsystem sowie die Vorspannkräfte und Reibungsverhältnisse am KGT. Im Rahmen der Arbeit wird ein Vorspannungsregler zur Einstellung der Vorspannkraft am KGT entwickelt, um damit das Betriebsverhalten des KGTs gezielt anzupassen. Die Einstellung der Vorspannung hat damit Auswirkungen auf das Reibungs- und Verschleißverhalten des KGTs. (Haberkern 1998) legt in seiner Arbeit den Fokus auf den Abrasivverschleiß am KGT und entwickelt ein Modell zur Berechnung des abrasiven Verschleißes der Wälzkörper. Nach (Haberkern 1998) wird der Großteil des Kugelverschleißes durch Kugel-Kugel-Kontakt bzw. durch Gleitvorgänge im Rückführsystem hervorgerufen. Auf Basis dieser Erkenntnis wird eine Beschichtung der Kugeln sowie der Umlenkstücke vorgenommen und damit eine Reduzierung der Reibung von bis zu 60 % gegenüber einem handelsüblichen KGT erreicht. Es wird weiterhin beobachtet, dass die Reduzierung der Reibung negativ mit der Drehzahl korreliert.

(Dambacher 2005) setzt sich im Rahmen seiner Dissertation das Ziel, die Betriebseigenschaften des KGTs durch Anpassung der Geometrie der Spindellaufbahn zu verbessern. Der Fokus liegt hierbei auf der Erhöhung des Druckwinkels und der damit einhergehenden höheren Steifigkeit des Systems. Eine höhere Genauigkeit des Systems wird angestrebt. (Dambacher 2005) führt hierzu Simulationen durch. Die Validierung mittels Prototypen mit einem Druckwinkel von 62° zeigt verbesserte Steifigkeitseigenschaften gegenüber herkömmlichen KGT bei vergleichbaren Laufeigenschaften. Durchgeführte Körperschallmessungen lassen zusätzlich auf verbesserte Schwingungseigenschaften schließen.

(Imiela 2006) untersucht den Vorspannungsverlust einer KGT-Mutter und versucht darüber auf den am KGT vorliegenden Verschleiß zu schließen. Die Abschätzung des Verschleißes basiert hierbei auf einer Mischung aus einem zustandsorientierten und belastungsorientierten Ansatz. Für die Abschätzung des Verschleißes wird ein Prognosemodell der Muttervorspannung entworfen (belastungsorientierter Ansatz), das durch die zyklische Messung von Betriebsparametern re-parametrisiert wird. Die Messung der Betriebsparameter kann hierbei als indirekte Messung des vorliegenden Verschleißes aufgefasst werden.

(Schopp 2009) stützt sich auf ein zustandsorientiertes, indirektes Verfahren zur Bestimmung des Verschleißzustandes eines KGT. Weiterhin wird ein Ansatz für ein Prognosemodell zur Ableitung der Resteinsatzdauer vorgeschlagen. Als Größen zur Zustandsabschätzung werden hierfür der Luftschall, der Körperschall, die Spindeleigenfrequenz und die Antwort des KGTs auf einen Drehzahlsprung aufgezeichnet. Der Fokus der Zustandsabschätzung sowie Zustandsprognose stützt sich auf Körper- und Luftschall. Schopp verschleißt KGT über Lebensdauerversuche und leitet aus den Ergebnissen die in Abbildung 3-1 dargestellte dreiphasige Verschleißkurve am KGT ab, die bereits in der gemeinsamen Arbeit von (Munzinger, Schopp & Broos 2009) vorgestellt wurde.

(Schopp 2009) interpretiert, dass ab einer Laufleistung von ca. 60 % Verschleiß im System vorhanden sein muss, was in einem signifikanten Ansteigen der Kurve resultiert. Bei Überschreiten von ca. 90 % der Lebensdauer des Systems steigt die Kurve nochmals stark an und mündet schließlich im mechanischen Ausfall des Systems. (Schopp 2009) beschreibt, dass vor dem Erreichen von 60 % Laufleistung nicht sicher auf das

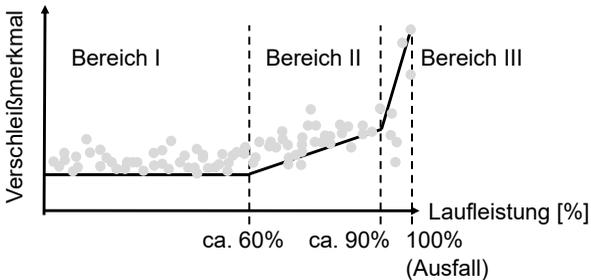


Abbildung 3-1 Dreiphasige Verschleißkurve nach (Schopp, 2009).

Vorhandensein von Schädigungen geschlossen werden kann. Weiterhin nehmen das Rauschen der Sensorsignale und damit die Unsicherheit mit steigender Laufleistung zu. (Schopp 2009) definiert eine lineare Prognosefunktion, die auf Basis der letzten drei Messungen das Signal für die Zukunft prognostiziert. Eine Einschränkung ist an dieser Stelle, dass die Prognosefunktion nur bei gleichbleibenden Betriebsbedingungen angewandt werden kann, da die Änderung im Schallniveau somit auf sich ändernde Betriebsbedingungen zurückzuführen ist. Eine Validierung der Prognosefunktionen mittels experimenteller Daten findet nicht statt.

(Verl & Heisel et al. 2009) untersuchen ein System zur sensorlosen, automatischen Überwachung des Verschleißes an KGT. Das System misst den Verschleiß hierbei indirekt, indem Positionsdaten, Geschwindigkeitsdaten sowie Motorstrom überwacht werden. Der Vorteil des Systems besteht in der sensorlosen Ausführung, indem lediglich Signale aus der vorhandenen Steuerungsinfrastruktur verwendet werden. Verschleißerscheinungen konnten mit dem vorgeschlagenen Ansatz festgestellt werden. Weiterhin zeigen (Verl & Frey 2010) die Korrelation zwischen der Vorspannung des KGTs und der Vorschubkraft. Die Formel für die äquivalente Last am KGT wird um einen Korrekturfaktor erweitert und es wird vorgeschlagen, diese Erkenntnis für die Zustandsüberwachung des KGTs zu nutzen, was allerdings nicht weiter untersucht wird.

(Schmid & Gädeke et al. 2010) konzipieren ein drahtloses System zur Überwachung des Körperschalls von KGT. Der Fokus liegt dabei stärker auf der industriellen Einsetzbarkeit des Systems als auf der Datenauswertung. Besonders der drahtlose Ansatz schafft hierbei Flexibilität beim Einsatz. (Stockinger 2011) untersucht in seiner Arbeit Ansätze zur Zustandsüberwachung von Werkzeugmaschinen mit KGT und schlägt als Ergebnis ein Verfahren vor, das auf Basis eines Modelles Reibungs- und Vorspannungsänderungen berechnet. (Walther 2011) setzt sich zum Ziel, Verschleiß am KGT ohne die Verwendung von zusätzlicher Sensorik lediglich über die steuerungsinternen Signale der Position, der Geschwindigkeit sowie des Motorstroms zu identifizieren und hierbei zwischen periodischen Schadensarten und nicht periodischen Schadensarten zu unterscheiden. Dabei weist er eine Korrelation von Verschleiß und Motorstrom nach. Die Signale des Motorstroms decken sich im Verlauf mit dem von (Schopp 2009) abgeleiteten Verlauf. (Walther 2011) stellt weiterhin hilfreiche Ursache-Wirkungs-Ketten für unterschiedliche Schäden am KGT dar.

(Broos 2012) stützt sich bei der Durchführung seiner Arbeit auf einen vollständig belastungsorientierten Ansatz, bei dem der Verschleiß an einer Maschinenkomponente auf Basis von realitätsnahen Simulationen abgeleitet wird. Hierzu werden die aus dem Simulationsmodell resultierenden Belastungen in die gängigen Formeln zur Bestimmung der nominellen Lebensdauer integriert. (Broos 2012) validiert die Simulation durch Vergleichsmessungen an einem Achsprüfstand und kommt dabei zu einer guten Übereinstimmung. Die Ansätze werden anschließend auf ein Horizontalbearbeitungszentrum übertragen.

(Möhring & Bertram 2012) entwickeln ein sensorbasiertes Doppelmutterssystem zur Messung der Vorspannung am KGT, bei dem die notwendige Sensorik in einem Gehäuse radial um den KGT angebracht ist. Die Funktionsweise des Systems wird dabei in Lebensdauerexperimenten validiert. Die Art der Anbringung des Sensorsystems dient als Grundlage für das von (Schlagenhauf & Hillenbrand et al. 2019) prototypisch vorgestellte Sensorsystem, auf dem diese Dissertation basiert.

(Henrich 2013) untersucht in seiner Dissertation einen Ansatz, bei dem eine zuverlässige zustandsorientierte Methode mit einer aufwandsärmeren belastungsorientierten Methode kombiniert wird. Das Ziel ist die Ableitung und Prognose des Verschleißzustandes des KGTs. Der zustandsorientierte Ansatz basiert hierbei auf der Messung der axialen Schwingung, während der belastungsorientierte Ansatz auf der Ableitung der Belastung mittels des Motorstroms erfolgt. (Henrich 2013) entwickelt auf Basis von

gefitteten Exponentialfunktionen ein Prognosemodell und bestätigt in seinen Untersuchungen, dass eine für einen bestimmten KGT-Typ gebildete Prognosefunktion auch auf andere KGT derselben Bauart angewendet werden kann. (Hennrich 2013) kombiniert die beiden Ansätze und nutzt den zustandsorientierten Ansatz zur Validierung der Prognose des belastungsorientierten Ansatzes, bevor ein KGT als defekt gekennzeichnet wird. (Hennrich 2013) stellt ähnlich wie (Schopp 2009) fest, dass eine signifikante Änderung des Zustandssignals erst ab einer Laufleistung von ca. 70 % beobachtet werden kann. Die Methoden werden an KGT validiert.

(Herder 2013) widmet sich ähnlich wie (Golz 1990) der Einstellung der Vorspannung des KGTs. Hierzu werden ein piezoelektrischer Self-Sensing-Aktor sowie eine aktive Vorspannungsregelung entwickelt. Ziel der Dissertation ist es, den durch Verschleiß bedingten Vorspannungsverlust am KGT zu kompensieren. (Herder 2013) gibt zwar einen Ausblick über ein vorstellbares Prognosesystem, setzt dieses allerdings nicht um. Mittels der konzipierten Vorspannungseinstellung kann eine Verringerung der Vorspannungsschwankung von ca. 80 % nachgewiesen werden.

(Yagmur 2014) bemängelt in seiner Arbeit die unzureichenden Berechnungsvorschriften für den KGT, die lediglich auf der Ermüdungslebensdauer beruhen und auch hier keine zuverlässigen Ergebnisse liefern. Dies begründet (Yagmur 2014) damit, dass bei dem Betrieb eines KGT unterschiedliche Verschleißmechanismen wirken, die zu unterschiedlichen Verschleißerscheinungsformen und Ausfallursachen führen. Ziel seiner Arbeit ist es, die unterschiedlichen Effekte und Einflussgrößen im Betrieb zu analysieren. Als Größen für die Beschreibung des Verschleiß verwendet (Yagmur 2014) die Größen Leerlaufdrehmoment und axiale Steifigkeit. Er konzipiert ein Sperrluft-Dichtungssystem, durch das Spankontamination verhindert werden soll. Die Wirksamkeit kann er in seinen Versuchen bestätigen.

(Münzing 2017) untersucht in seiner Dissertation ein Verfahren zur Auslegung von KGT unter Kurzhubbelastung, wie sie bspw. im Stellwerk von Helikopterrotoren auftritt. (Münzing 2017) stellt hierbei fest, dass mit kleiner werdenden Bewegungen die Schmierverhältnisse zunehmend negativ beeinflusst werden und damit das Erreichen der Ermüdungslebensdauer zunehmend unwahrscheinlicher wird. Basierend auf dieser Erkenntnis entwickelt (Münzing 2017) ein Verfahren zur Auslegung von KGT, wobei der Fokus vor allem auf der Auswahl des Schmierstoffes, des Wälzkörperwerkstoffes sowie des Laufbahnwerkstoffes liegt. Die Untersuchungen werden mittels Lebensdaueruntersuchen validiert.

Zur Verschleißüberwachung am KGT untersucht (Helwig 2018) einen Ansatz zur Multi-Sensor-Analyse von Motorstrom, Luft- und Körperschall. Die mittels Merkmalsextraktionsverfahren vorverarbeiteten Signale werden im Anschluss durch ein auf Maschinellem Lernen basierendes Verfahren klassifiziert. Bei Versuchsende werden großflächige Pittings an der Spindel beobachtet und „[h]ierbei ist bemerkenswert, dass trotz des erheblichen Verschleißes der KGT noch bis zum Ausfall funktionstüchtig blieb“ (Helwig 2018, S. 153). (Helwig 2018) äußert die Vermutung von sich gegenseitig beeinflussenden Verschleißmechanismen, ohne dies jedoch zu belegen.

(Spohrer 2019) widmet sich ähnlich den frühen Arbeiten nicht primär der Bestimmung des Verschleißes am KGT, sondern vielmehr der Optimierung der Betriebseigenschaften und damit der Lebensdauer des KGTs. Im Gegensatz zu bisherigen Arbeiten optimiert (Spohrer 2019) dabei nicht direkt die Komponente des KGTs, sondern die Strategie, mit der die Komponente geschmiert wird. Hierzu wird ein Modell zur adaptiven Schmierung der Komponente auf Basis einer indirekten Zustandsmessung des KGTs entwickelt. Dabei wird der Ist-Schmierzustand über die Messung der Temperatur und des Reibmomentes abgeleitet und mit dem modellbasierten Soll-Schmierzustand verglichen. Daraus werden dann eine optimale Schmiermenge und Schmierstrategie errechnet. Neben der Temperatur und dem Reibmoment fließt ebenfalls die Schmierhistorie mit in das Modell ein. (Spohrer 2019) führt Lebensdauerversuche von KGT durch und validiert damit die Funktionsfähigkeit des Ansatzes. Weiterhin wird das Ansteigen der Frequenz der Schmierzyklen als Indikator für vorherrschenden Verschleiß vorgeschlagen, wenngleich daraus in der Arbeit kein bevorstehender Komponentenausfall abgeleitet werden kann. Hierzu schlägt (Spohrer 2019) die Verwendung der ansteigenden Temperatur sowie des ansteigenden Reibmomentes vor. Die prinzipielle Eignung dieser Signale wird zwar in Form einer Abbildung dargestellt, allerdings nicht weiter verfolgt.

(Benker, Kleinwort & Zäh 2019) wählen einen wahrscheinlichkeitsbasierten Ansatz mittels einer Beschleunigungsmessung am KGT. Die Autoren bringen hierzu zunächst einen Beschleunigungssensor zentral an einer Maschine an. Im Anschluss werden die Beschleunigungssignale eines neuen sowie eines verschlissenen KGT mittels eines Gaussian-Mixture-Ansatzes klassifiziert. Die prinzipielle Eignung des Ansatzes kann zwar gezeigt werden, es wird allerdings betont, dass weitere Validierungsversuche notwendig sind.

(Xi & Kehne et al. 2020) verfolgen einen Ansatz, bei dem die erste Eigenfrequenz des KGTs überwacht wird, und stellen hierbei eine messbare Veränderung bei Verschleiß fest. Das vorgeschlagene Verfahren wird auf Basis einer FE-Simulation entwickelt, die mit realen Maschinendaten parametrisiert wurde. Eine Validierung an realen KGT erfolgt nicht.

Die Arbeiten von (Riaz & Shah et al. 2020) und (Riaz & Shah et al. 2021) untersuchen den Einsatz von CNN auf Motorstromsignalen zur Überwachung des Zustandes des KGTs. Es wird sowohl das Motorstromsignal im Zeitbereich mittels eines 1D-CNN als auch das in ein Bild transformierte Motorstromsignal mittels eines 2D-CNN verarbeitet. Die Autoren zeigen eine hohe Genauigkeit des Ansatzes, wenngleich die Validierung nur knapp beschrieben ist.

(Veith & Zimmermann et al. 2020) stellen das sogenannte Guard-Plus-System vor, bei dem der Vorspannungsverlust am KGT mittels eines in der Kugelgewindtriebmutter integrierten Dehnmessstreifens sowie Temperatursensors festgestellt wird. Es werden Verschleißversuche durchgeführt, bei denen ein zusätzliches Abrasivmedium verwendet wird. Dadurch sollen die Versuchszeiten reduziert und abrasiver Verschleiß verstärkt werden. In allen Versuchen kann hierbei eine Korrelation der Vorspannung mit der Lebensdauer der Komponente festgestellt werden. Mittels eines Regressionsansatzes wird die verbleibende Lebensdauer der Komponente prognostiziert. Die Untersuchungen sind besonders für solche Domänen interessant, in denen abrasiver Verschleiß der lebensdauerbegrenzende Verschleißmechanismus ist.

Die (A. Mannesmann Maschinenfabrik GmbH, 2022)¹⁸ bietet integrierte Temperatursensorik für KGT zur Verschleißüberwachung an.

Zusammenfassend lässt sich aus den Arbeiten zur Verschleißüberwachung am KGT die Relevanz der Überwachung und Optimierung des Betriebszustandes und der Betriebseigenschaften des KGTs ableiten. Bei der Betrachtung der Arbeiten ist erkennbar, dass sich frühe Arbeiten vornehmlich mit der Optimierung der Betriebseigenschaften des KGTs befassen, während sich spätere Arbeiten auf die Überwachung des Betriebszustandes des KGTs konzentrieren. In neueren Arbeiten ist außerdem ein verstärkter Einsatz von Methoden des Maschinellen Lernens erkennbar. Die Arbeiten zur Zustand-

¹⁸ A. Mannesmann Maschinenfabrik GmbH (2021), *Sensorik für Kugelgewindtriebe*. <https://www.amannesmann.de/produkte/kugelgewindtriebe/sensorik/> [21.01.2021].

süberwachung konzentrieren sich einerseits auf die Überwachung des Vorspannungsverlustes aufgrund abrasiver Verschleißeffekte und andererseits auf die Überwachung des generellen Verschleißzustandes des KGTs. Hierbei fokussieren die Arbeiten nicht ein spezifisches Verschleißmerkmal, sondern sie versuchen, aus der Änderung eines indirekten Sensorsignals auf den Verschleiß der Komponente zu schließen.

3.4 Bewertung des Standes der Forschung und Technik

Im Folgenden sollen die Erkenntnisse aus dem Stand der Forschung und Technik bewertet werden. Zunächst wird ein Gesamtfazit gegeben und daraus dann die aktuellen Defizite im Stand der Forschung und Technik abgeleitet.

3.4.1 Fazit zum Stand der Forschung und Technik

Es wurde aufgezeigt, dass zur bildbasierten Defekterkennung auf metallischen Oberflächen zahlreiche Arbeiten bestehen, welche die grundsätzliche Machbarkeit der Identifikation von Oberflächendefekten auf metallischen Oberflächen bestätigen. Aus den Beiträgen wird ersichtlich, dass sich der Großteil der aktuellen Arbeiten auf Methoden des Maschinellen Lernens und speziell auf Deep-Learning-basierte Methoden zur Klassifikation der Merkmale stützt. Dies verdeutlicht die grundsätzliche Eignung von Deep-Learning-Verfahren, wenngleich der Großteil der Arbeiten industrielle Bedingungen wie Verschmutzungen unbeachtet lässt.

Die Arbeiten an Maschinenelementen sowie an KGT zeigen einen Einsatz von unterschiedlichen Sensorsignalen und Systemen, wobei sich die Ansätze in belastungsorientierte und zustandsorientierte Ansätze gliedern lassen. Der Vorteil der belastungsorientierten Ansätze besteht in der sensorlosen Zustandsüberwachung, wobei die Herausforderung in einer möglichst realistischen Modellierung der Komponenten und Einflussgrößen liegt. Da der Großteil der Modelle die aktuell vorliegenden Umweltbedingungen sowie den tatsächlichen Verschleißzustand unbeachtet lässt, ist eine eingeschränkte Genauigkeit der Modelle zur Verschleißbewertung zu vermuten. Systeme, die sich selbst durch zyklische Messungen mit Realdaten updaten, scheinen hierbei erfolgsversprechend zu sein. Bei zustandsorientierten Ansätzen wird versucht, direkt und indirekt durch sensorbasierte Messungen auf den Verschleiß zu schließen. Hierbei finden sich im Stand der Forschung größtenteils indirekte Verfahren, bei denen über eine sekundäre Größe auf den Verschleiß geschlossen wird. Direkte Verfahren, bei

denen eine Verschleißgröße direkt durch Messen der Form oder Masse von Verschleißpartikeln bewertet wird, sind kaum vorhanden. Aufgrund des indirekten Schließens auf Verschleiß zeichnen sich die bisherigen Arbeiten dadurch aus, dass die Bandbreite an möglichen Defekten, die sich im Signal abzeichnen, zwar groß ist, dies allerdings mit einer erhöhten Wahrscheinlichkeit zu Lasten der Genauigkeit der Systeme einhergeht.

Autorenübergreifend wird kritisiert, dass die aktuellen Berechnungsvorschriften für den KGT unzureichend sind, um den tatsächlichen Verschleiß und daraus die verbleibende Lebensdauer des Systems abzuschätzen. Aus den Arbeiten wird weiterhin ersichtlich, dass nicht ein einzelner Verschleißmechanismus wirkt, sondern immer eine Kombination an Verschleißmechanismen, wobei die Autoren je nach Arbeit unterschiedliche Verschleißmechanismen als lebensdauerbegrenzend ansehen. Somit liegt die Schlussfolgerung nahe, dass je nach Betriebsszenario und primär wirkenden Verschleißmechanismen anwendungsindividuelle Verschleißmechanismen primär lebensdauerbegrenzend sind und sich damit ein spezielles Sensorsignal und Sensorsystem zu deren Überwachung besonders eignet. Hierbei finden sich einerseits Ansätze, die speziell den Vorspannungsverlust (Abrasivverschleiß) beobachten und damit auf den Verschleiß der Komponente schließen. Andererseits gibt es Ansätze, die breiter angelegt sind und über die Messung eines Sensorsignals indirekt auf (unterschiedliche) Verschleißmechanismen schließen. Beispiele hierfür sind Motorstrom, Schall und Temperatur.

Die Arbeiten von (Möhring & Bertram 2012), (Spohrer 2019) und (Veith & Zimmermann et al. 2020) am KGT zeigen, dass sich Sensorsysteme an der KGT-Mutter implementieren lassen und darüber Verschleiß im System identifiziert werden kann.

3.4.2 Aktuelle Defizite im Stand der Forschung und Technik

Aus dem Stand der Forschung und Technik lassen sich vier konkrete Defizite ableiten, die den Rahmen für diese Arbeit vorgeben:

1. Es besteht bisher für den KGT kein direktes bildbasiertes Sensorsystem zur Analyse von Oberflächendefekten im Allgemeinen und der Oberflächenzerrüttung im Speziellen.
2. Es besteht aktuell kein Datensatz, der die Entwicklung der Oberflächenzerrüttung am KGT zeigt und für weitere Condition-Monitoring-Systeme verwendet werden kann. Darüber hinaus liefern die bisherigen Arbeiten am KGT für den Menschen schwer interpretierbare Signale, wodurch sich vorhandene Schädigungen auf der Oberfläche erst spät zweifelsfrei belegen lassen.

3. Eine empirische Analyse der Verschleißentwicklung der Oberflächenzerrüttung am KGT wurde in der bisherigen Forschung nicht durchgeführt.
4. Ansätze zur automatischen Identifikation und Prognose der Oberflächenzerrüttung am KGT mittels aus Bilddaten lernenden Verfahren bestehen nicht.

Insgesamt lässt sich das Defizit im Stand der Forschung und Technik wie folgt zusammenfassen:

Ein empirische Analyse der Entwicklung der Oberflächenzerrüttung auf der Kugelgewindetribspindel sowie eine automatisierte Identifikation und Prognose der Oberflächenzerrüttung auf Basis von Bilddaten sind bisher nicht vorhanden.

Dies ist aber für das Verständnis über die Verschleißentwicklung der Oberflächenzerrüttung sowie zur automatisierten Identifikation und Prognose von Oberflächenzerrüttung im Kontext der Zustandsüberwachung von Relevanz und soll daher im Rahmen dieser Arbeit untersucht werden.

4 Zielsetzung und Lösungsansatz

Im nachfolgenden Kapitel soll die Zielsetzung der vorliegenden Dissertation basierend auf dem Defizit im Stand der Forschung und Technik konkretisiert werden. Aufbauend auf der Konkretisierung der Zielstellung soll anschließend eine Übersicht zum Lösungsansatz gegeben werden, in der das Vorgehen zur Zielerfüllung ganzheitlich dargestellt wird.

4.1 Konkretisierung der Zielsetzung

Einleitend soll festgehalten werden, dass der zu untersuchende Ansatz keinen Anspruch darauf erhebt, die bestehenden Ansätze und Sensorsysteme gänzlich zu substituieren. Vielmehr sollen mit dieser Arbeit die Erkenntnisse der bisherigen Arbeiten (empirisch) konkretisiert und die bestehenden Ansätze zur Verschleißüberwachung am KGT um einen direkten, speziell auf die Oberflächenzerrüttung ausgelegten Ansatz ergänzt werden. Dieser Ansatz soll in seiner Genauigkeit bestehende Ansätze zur Quantifizierung der Oberflächenzerrüttung übertreffen.

Die Vorteilhaftigkeit der Anwendung des vorgestellten Ansatzes gegenüber bereits bestehenden Ansätzen leitet sich ebenso wie bei den schon bestehenden Ansätzen aus den Anforderungen eines Anwendungsfalles ab.

Ziel der vorliegenden Dissertation sind die Konzeption eines Sensorsystems zur Bildaufnahme der Kugelgewindetriebsspindel sowie die empirische Analyse der Entwicklung der Oberflächenzerrüttung in den Bilddaten. Daran anschließend soll eine Methodik zur automatisierten Identifikation und Prognose der Oberflächenzerrüttung mithilfe von lernenden Verfahren konzipiert und untersucht werden.

Die Dissertation soll damit einen Beitrag zur Untersuchung und Überwachung der Entstehung und Entwicklung der Oberflächenzerrüttung auf der Kugelgewindetriebsspindel leisten.

Die Generierung eines Bilddatensatzes, der die Entwicklung der Oberflächenzerrüttung zeigt, ist in der Forschung bisher nicht vorhanden und soll der Wissenschaftsgemeinschaft zur weiteren Untersuchung von Condition-Monitoring-Verfahren dienen. Auch die Untersuchung von lernenden Verfahren zur Identifikation und Prognose der Oberflächenzerrüttung in Bilddaten ist bisher nicht erfolgt und soll als Basis für weitere intelligente Verfahren zur Verschleißüberwachung in Bilddaten dienen.

Die Arbeit soll damit einen Beitrag zur automatisierten Zustandsüberwachung von Maschinenelementen leisten und die bestehenden Systeme um ein weiteres Sensorsystem ergänzen. Durch die Integration, Untersuchung und Weiterentwicklung von Verfahren zur automatisierten Zustandsüberwachung soll eine Brücke zur breiten Anwendung von Verfahren des Maschinellen Lernens im Maschinenbau geschlagen werden.

Durch die kostengünstige Integration von Sensoren und Software in ein einfach zu betreibendes Gesamtsystem soll den praktischen Anwender*innen ein Werkzeug zur Automatisierung der Zustandsüberwachung der Kugelgewindetriebspindel an die Hand gegeben werden. Der damit generierte Bilddatensatz soll sowohl in der Domäne des Maschinenbaus zur Untersuchung der bildlichen Entwicklung von Schädigungen auf Maschinenelementen als auch in der Domäne des Maschinellen Lernens zum Training von Fehler-Klassifikations-, Fehler-Detektions- und Prognosesystemen verwendet werden können.

4.2 Übersicht zum Lösungsansatz

Die in Abschnitt 4.1 konkretisierten Ziele sollen im Rahmen eines dreistufigen Lösungsansatzes untersucht und erarbeitet werden. Der Lösungsansatz ist in die Stufen der experimentellen Datengenerierung, der Datenanalyse und Modellbildung sowie der Verschleißprognose gegliedert und in Abbildung 4-1 dargestellt.

Im Rahmen der experimentellen Datengenerierung wird zunächst ein Sensorsystem zur Datenaufnahme basierend auf einer theoretischen Analyse des Standes der Technik sowie auf Versuchen zur Konzeptabsicherung konzipiert. Mittels dieses Sensorsystems werden dann in Lebensdauerversuchen Bilddaten der Oberfläche des KGTs aufgenommen (Kapitel 5).

Im darauffolgenden Schritt der Datenanalyse und Modellbildung werden Oberflächenfehler zunächst theoretisch aus Sicht der Tribologie betrachtet. Die Entstehung und Entwicklung speziell der Oberflächenzerrüttung auf der Kugelgewindetriebspindel wird empirisch analysiert und bewertet (Kapitel 6).

Zur automatisierten Identifikation und Prognose der Oberflächenzerrüttung in den Bilddaten werden durch ebendiese zunächst relevante Merkmale zur Verschleißquantifizierung generiert (Kapitel 6).

Darauf basierend werden Lösungsprinzipien zur Modellbildung aus dem Stand des Maschinellen Sehens sowie des Maschinellen Lernens abgeleitet und praktisch auf die

vorliegende Aufgabe übertragen. Zunächst wird die Möglichkeit der Klassifikation der Bilddaten mit Merkmalen untersucht, die auf Basis von Domänenwissen definiert wurden. In einem weiteren Schritt werden Deep-Learning-Ansätze untersucht, die Merkmale automatisiert aus den Bilddaten extrahieren, um damit einen möglichst leistungsfähigen Ansatz zur Identifikation von Oberflächenzerrüttung in den Bilddaten zu schaffen (Kapitel 6).

Der Aspekt der dateneffizienten Klassifikation wird in einem Unterkapitel betrachtet, wobei eine dateneffiziente Modellarchitektur konzipiert und anhand von Daten aus dem Stand der Forschung validiert wird (Kapitel 6).

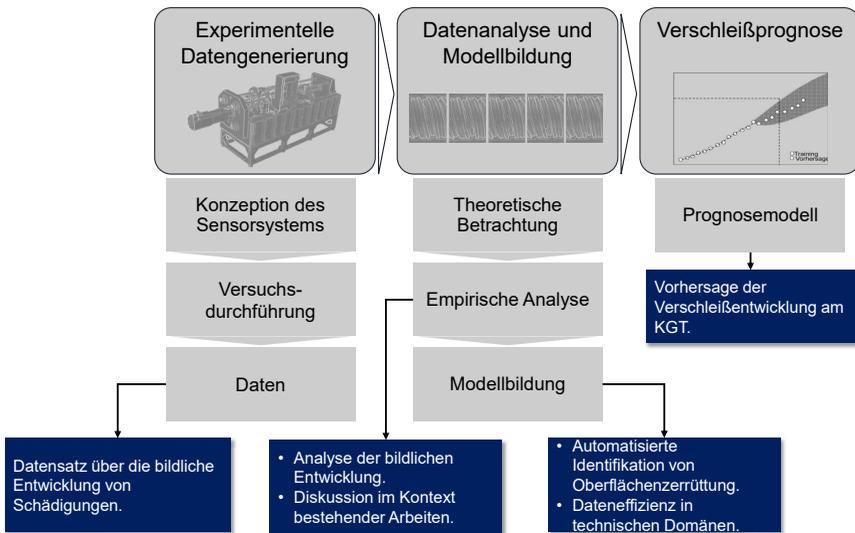


Abbildung 4-1 Vorgehen zur Zielerfüllung. Blau: Markante Inhalte und Ergebnisse der Kapitel.

In Kapitel 7 wird eine Methode zur Prognose der Verschleißentwicklung der Oberflächenzerrüttung untersucht. Hierbei wird die Verschleißgrenze angelehnt an bestehende Normen definiert. Die modellbasierte Vorhersage wird mit einem Expertensystem verknüpft, um zu einer finalen Prognose zu gelangen. Das Expertensystem baut hierbei auf Domänenwissen über die Entwicklung der Oberflächenzerrüttung aus Kapitel 6 auf.

Zusammenfassend wurden die Einzelschritte zur Zielerfüllung sowie markante Inhalte und Ergebnisse vorgestellt. Im Folgenden Abschnitt wird auf die experimentelle Datengenerierung und damit auf den ersten Schritt zur Zielerfüllung eingegangen.

5 Experimentelle Datengenerierung

In diesem Kapitel werden der experimentelle Versuchsaufbau, der Prozess zur Bildaufnahme, die Qualifizierung des Sensorsystems zur Bilddatengenerierung sowie der aufgenommene Bilddatensatz beschrieben.

Hierbei bildet das Sensorsystem die Grundlage für die Bilddatengenerierung und die folgende Datenanalyse. Zunächst wird der Versuchsstand beschrieben, auf dem Lebensdauerversuche durchgeführt werden und das Sensorsystem integriert wird. Darauf aufbauend werden die Randbedingungen zur Versuchsdurchführung dargestellt. Im nächsten Unterkapitel wird auf das Sensorsystem und dessen Komponenten sowie den Prozess der Bilddatengenerierung eingegangen. Das Kapitel schließt mit einer Beschreibung des Bilddatensatzes. Dieser stellt die Grundlage für die Datenanalyse sowie die folgende Modellbildung dar.

5.1 Lebensdauerversuchsstand

Zur Durchführung der Lebensdauerversuche wird ein am wbk Institut für Produktionstechnik vorhandener 5-Achs-Versuchsstand erweitert. Auf diesem Versuchsstand können KGT mit einer Gesamtlänge von bis zu 2500 mm und einem Durchmesser von bis zu 52 mm integriert werden. Hiermit kann eine große Anzahl an gängigen, in Produktionsmaschinen verbauten KGT getestet werden. Jeder KGT ist in zwei unabhängige Teilsysteme¹⁹ unterteilt, die unabhängig voneinander von je einer Mutter überrollt werden. Der Versuchsstand ist hierbei so ausgelegt, dass eine Überschneidung der Bereiche, die von den beiden Müttern pro Spindel überrollt werden, ausgeschlossen ist. Die Teilsysteme können als unabhängige KGT angesehen werden. Der Versuchsstand ist in Abbildung 5-1 dargestellt. Den Kern des 5-Achs-Versuchsstandes bilden die fünf Spindelaufnahmen, wobei die mittlere Spindel mittels eines Leitantes angetrieben wird. Die vier symmetrisch um die Leitspindel angeordneten Aufnahmen werden über ein Verteilergetriebe angetrieben und folgen der Leitspindel. Ein weiteres Kernmerkmal des Versuchsstandes sind die sogenannten Zugstäbe zur Aufbringung von Last. Jeder auf dem Versuchsstand integrierte KGT ist mit zwei gegenläufigen Müttern ausgestattet, die über je zwei Zugstäbe gegeneinander verspannt sind. Die Zugstäbe sind mit Kraftmessensoren ausgestattet, um eine definierte axiale Belastung einzubringen. Die

¹⁹ Nachfolgend KGT genannt.

Kraft wird im Versuch kontinuierlich überwacht, um einen ungewollten Drift der Axiallast zu verhindern.

Die KGT können so in Lebensdaueruntersuchungen unter kontrollierten Bedingungen verschlissen und der Verschleißfortschritt kann durch die Integration von Sensorsystemen vom Neuzustand bis zum mechanischen Versagen der Komponenten beobachtet werden.

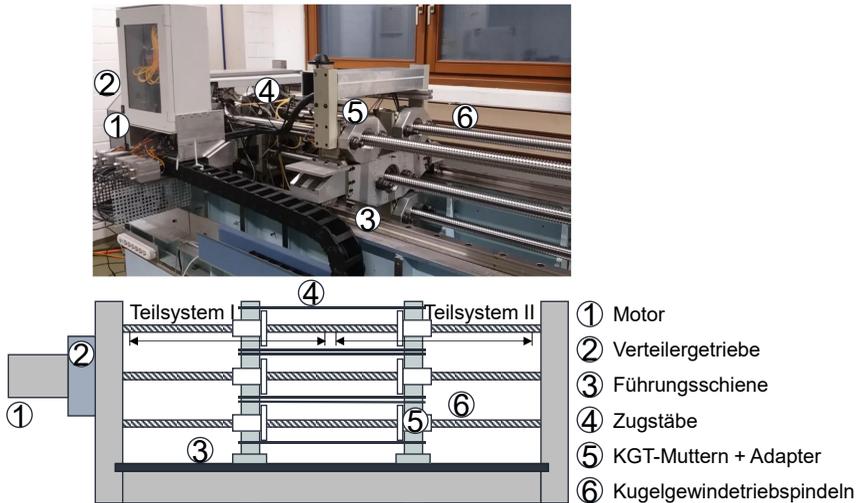


Abbildung 5-1 Hauptkomponenten des 5-Achs-Versuchsstandes angelehnt an (Hilfenbrand, Spohrer & Fleischer 2018).

Die Versuchsdurchführung wird im folgenden Abschnitt beschrieben, bevor auf die Konzeption und Integration des Kamerasystems im Versuchsstand eingegangen wird.

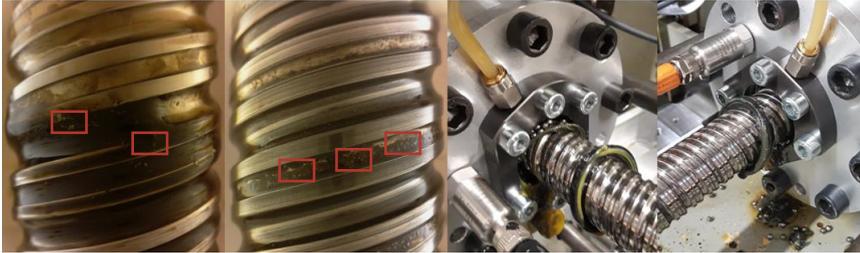
5.2 Versuchsdurchführung

Auf dem beschriebenen Versuchsstand werden Lebensdaueruntersuchungen zur Generierung des Bilddatensatzes durchgeführt.

Bei der Versuchsdurchführung werden in mehreren Testläufen insgesamt zwölf Kugelgewindetrieb-Teilsysteme verschlissen. Für die Versuche werden KGT mit einer Länge von 2500 mm und einem Durchmesser von 32 mm der Firma Bosch Rexroth verbaut. Die genauen Spezifikationen sind in Abbildung 0-9 im Anhang dargestellt. Die KGT

haben eine dynamische axiale Tragzahl (C_{dyn}) von 23.6 kN und werden bei Raumtemperatur (20 °C) mit einer axialen Last von $0.4 * C_{dyn}$ beaufschlagt. Angelehnt an die Arbeiten von (Spohrer 2019) und (Veith & Zimmermann et al. 2020) wurden die Belastungen so gewählt, dass Verschleißversuche in vertretbarer Zeit unter realistischen Bedingungen durchführbar sind. Die Reibung in den Systemen führt zusammen mit deren Fertigungstoleranzen zu einer Erwärmung auf 50 bis 55 °C bei Betriebstemperatur, was die anliegenden Kräfte auf Werte zwischen $0.6 * C_{dyn}$ und $0.7 * C_{dyn}$ ansteigen lässt. Unter diesen Bedingungen werden die Spindeln kontinuierlich mit 400 U/min verschlissen. In Intervallen von vier Stunden erfolgt eine Aufnahme der Spindeloberflächen (Kamerafahrt) durch das Sensorsystem. Dieser Zyklus wird bis zum mechanischen Versagen der Komponenten beibehalten. Während einer Kamerafahrt geht die anliegende Kraft wieder auf die Kraft bei Raumtemperatur zurück. Die Kraft schwankt also zwischen zwei Kamerafahrten und es entsteht ein periodischer Kraftverlauf. Bei der Versuchsdurchführung werden die KGT gemäß Herstellerangaben geschmiert. Temperatursensoren sind an den Mutternflanschen angebracht, um die Betriebstemperatur zu überwachen. Die Versuchsläufe werden unterbrochen, wenn die Temperatur eines Sensors 60 °C übersteigt oder der in der SPS gemessene Motorstrom einen kritischen Schwellwert von 12 Ampere übersteigt. Diese Grenzen beschreiben das mechanische Versagen der Komponente bzw. das Verklemmen des Systems und wurden empirisch auf Basis von Expertenwissen festgesetzt.

Bei der Versuchsdurchführung werden die Versuche in allen Fällen entweder aufgrund des Überschreitens des Motorstroms oder des Überschreitens der Temperatur und anschließender Experteninspektion abgebrochen. Abbildung 5-2 zeigt exemplarische Aufnahmen von Spindeloberflächen sowie des Abstreifsystems bei Überschreiten einer Abbruchgrenze. Durch die klar sichtbaren Metallpartikel auf den Laufbahnschultern ist erkennbar, dass eine massive Schädigung des Systems vorliegt.



Abbruch durch Überschreiten der Temperatur Abbruch durch Überschreiten der Kraft

Abbildung 5-2 Links: Metallpartikel auf der Kugelgewindetriebspindel (rot), rechts: defekte Abstreifsysteme bei Versuchsende.

Als Ergebnis können mittels des beschriebenen Versuchsaufbaus Lebensdauerversuche unter konstanten und realistischen Bedingungen durchgeführt werden.

Im Folgenden werden der Aufbau und die Konzeption des Sensorsystems beschrieben, das zur Versuchsdurchführung auf dem 5-Achs-Versuchsstand integriert wurde.

5.3 Sensorsystem zur Bilddatengenerierung

In diesem Abschnitt wird das Sensorsystem zur Datengenerierung vorgestellt. Das Vorgehen lehnt sich dabei an die Entwicklungsmethodik für cyberphysische mechatronische Systeme (Verein Deutscher Ingenieure & Verband der Elektrotechnik Elektronik Informationstechnik 2020) an. Die Komponenten des Sensorsystems wurden in den Arbeiten von (A_Hansjosten 2020) und (A_Schlund 2019) befähigt. Die Ergebnisse werden nachfolgend zusammengefasst.

In einem ersten Schritt werden mittels einer Anforderungsanalyse die Anforderungen an das Sensorsystem abgeleitet. Da die vorliegende Arbeit die bildbasierte Analyse der Kugelgewindetrieboberfläche in den Fokus nimmt und ein im Vergleich zu den Kosten eines KGT hochpreisiges Sensorsystem aus betriebswirtschaftlichen Aspekten nachteilig ist, werden bei der Anforderungsanalyse des Sensorsystems lediglich kostengünstige Standardkomponenten berücksichtigt. Im darauffolgenden Abschnitt wird die Konzeption des Sensorsystems beschrieben und dessen Komponenten werden vorgestellt. Darauf aufbauend werden der Prozess der Bildaufnahme mittels des konzipierten Sensorsystems sowie die Bild- und Datenvorverarbeitung beschrieben. Der Abschnitt schließt mit einer Vorstellung und ersten Analyse der aufgenommenen Bilddaten.

5.3.1 Einflussgrößen und Anforderungen an das Sensorsystem

Das Ziel der Datenaufnahme ist es, Bilddaten von Schädigungen auf der Kugelgewindetriebspindel mit einer möglichst hohen Auflösung zu generieren und hierbei ein betriebswirtschaftlich attraktives Sensorsystem zu verwenden. Unter hoher Auflösung wird verstanden, dass das System Bilddaten liefert, die möglichst alle Pittings auf der Oberfläche des KGTs in einer Qualität abbilden, die sowohl frühzeitige Pittings als auch den Schadensfortschritt erkennen lassen und die zur Untersuchung von Klassifikationsmodellen geeignet sind. Die Bildaufnahme soll dabei möglichst unabhängig von Umgebungsbedingungen und gleichbleibend über die Zeit erfolgen.

Zur Ableitung von Anforderungen an das Sensorsystem werden zunächst mittels eines Ishikawa-Diagrammes angelehnt an (Kern 2009) die Haupteinflussgrößen auf die Bildaufnahme abgeleitet. Diese Haupteinflussgrößen dienen als Grundlage für die Anforderungsdefinition an das Sensorsystem. Abbildung 5-3 stellt die fünf definierten Haupteinflussgrößen auf die Bilddatengenerierung dar.

Bei der Wahl des Kamerasystems beeinflusst die Auflösung des Bildsensors maßgeblich die Genauigkeit, mit der Schädigungen auf der Spindel erfasst werden können. Da der Bauraum in Werkzeugmaschinen oftmals begrenzt ist, muss außerdem eine Bildaufnahme unter beschränktem Platzangebot möglich sein. Die zweite Haupteinflussgröße sind die Umgebungsbedingungen.

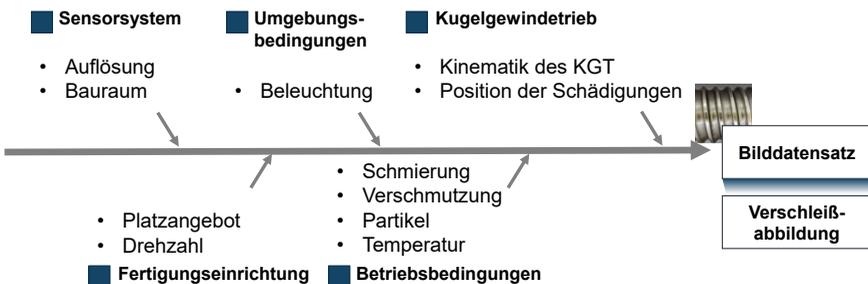


Abbildung 5-3 Einflussgrößen auf den Bilddatensatz.

Hierbei ist als Einflussgröße in erster Linie die Umgebungsbeleuchtung zu nennen, die orts- und prozessbedingt stark schwanken kann. Der KGT bildet die zweite Haupteinflussgröße. Die Abmessungen und Kinematik des KGTs beeinflussen die Relativbewegung seiner Komponenten und damit die Bildaufnahme. Die dritte Einflussgröße des

KGTs auf die Bilddatengenerierung ist die Position der Schädigungen auf der Kugelgewindetribspindel. Die vierte Haupteinflussgröße ist die Fertigungseinrichtung selbst. Hier haben vor allem das Platzangebot und die Drehzahl, mit der die KGT angetrieben werden, einen Einfluss auf die Bilddatengenerierung. Beispielsweise sind in einer Werkzeugmaschine die Platzverhältnisse oft eingeschränkt, was durchaus Auswirkungen hinsichtlich der Abmessungen des Sensorsystems sowie des Fokusabstandes haben kann. Die fünfte maßgebliche Haupteinflussgröße auf die Bilddatengenerierung stellen die Betriebsbedingungen dar. Hierbei sind besonders der Schmierstoff, prozessbedingte Verunreinigungen des KGTs sowie des Sensorsystems, Verschleißpartikel sowie Partikel aus dem Prozess (bspw. Späne) und die vorherrschende Temperatur zu nennen. Besonders Schmierstoff und Verunreinigungen können sich auf dem Kamerasystem absetzen und somit die Sicht auf die Kugelgewindetribspindel beeinflussen.

Basierend auf den Einflussgrößen werden die folgenden Anforderungen an das Kamerasystem abgeleitet. Das Ziel der Bilddatengenerierung ist die bildbasierte Beobachtung des Verschleißfortschrittes der Oberflächenzerrüttung auf dem KGT. Um den Verschleißfortschritt möglichst vollständig beobachten zu können, muss das Kamerasystem in der Lage sein, über die gesamte Betriebszeit des KGTs automatisiert Bilder von dessen Oberfläche aufzunehmen. Die Aufnahme der gesamten Spindeloberfläche sowie die Möglichkeit, zu definierten Zeitpunkten Bilddaten automatisiert zu generieren, stellen die beiden ersten Anforderungen an das Sensorsystem dar.

Da das Sensorsystem in die Anwendung integriert sein soll, stellt eine platzsparende Integration die dritte Anforderung dar.

Als vierte (mit der dritten Anforderung verwandte) Anforderung kann die Möglichkeit einer Inline-Integration des Kamerasystems möglichst nahe am KGT definiert werden.

Wie bereits beschrieben unterscheiden sich die Betriebsbedingungen in Fertigungsanlagen hinsichtlich des verwendeten Schmierstoffes sowie der verwendeten Schmierstoffmenge und weiterer Betriebsbedingungen wie Verschmutzung, Temperatur und Partikel. Eine weitere Einflussgröße stellt die Beleuchtungssituation in der Fertigungseinrichtung bzw. die Umgebungsbeleuchtung dar. Es wird von dem Kamerasystem gefordert, dass es von diesen Größen unabhängige Bilddaten in hoher Qualität liefert. Dies stellt die fünfte Anforderung an das Sensorsystem dar.

Zur Beobachtung der Oberflächenzerrüttung auf der Kugelgewindetribspindel ist es notwendig, auch frühzeitige und damit kleine Schädigungen in den Bilddaten erkennen

zu können. Eine Auflösung, welche die Darstellung von frühzeitigen Pittings erlaubt, ist damit die sechste Anforderung an das Sensorsystem.

Als Ergebnis werden die Anforderungen an ein Sensorsystem abgeleitet, das eine zuverlässige Überwachung der Oberflächenzerrüttung ermöglicht. Basierend auf diesen Anforderungen wird im folgenden Abschnitt die Konzeption des Sensorsystems beschrieben.

5.3.2 Konzeption des Sensorsystems

Im vorherigen Abschnitt wurden insgesamt sechs Anforderungen an das Sensorsystem abgeleitet. Basierend hierauf werden nun in Anlehnung an die Entwicklungsmethodik (Verein Deutscher Ingenieure & Verband der Elektrotechnik Elektronik Informationstechnik 2020) diesen Anforderungen Funktionsträger zugeordnet, die für die Erfüllung dieser Anforderungen verantwortlich sind. Diese Funktionsträger werden im Weiteren detailliert und zu einem elektromechanischen Kamerasystem kombiniert. Dies entspricht dem Schritt der Systemintegration in der (Verein Deutscher Ingenieure & Verband der Elektrotechnik Elektronik Informationstechnik 2020).

Zur Erfüllung der sechs Anforderungen werden fünf Funktionsträger definiert. Die Zuordnung der Funktionsträger zu den Anforderungen ist in Tabelle 5-1 dargestellt.

Tabelle 5-1 Zuordnung von Anforderungen zu Funktionsträgern.

Anforderung	Funktionsträger
Aufnahme der gesamten KGT-Oberfläche	Gehäuse
Automatisierte Aufnahme zu definierten Zeitpunkten	Kommunikationsarchitektur
Platzsparende Integration	Gehäuse
Inline-Integration	Montageadapter
Bilddaten unabhängig von Umgebungsbedingungen	Beleuchtungssystem
Zuverlässiges Abbilden der Spindeloberfläche	Kamera

Die erste Anforderung, die gesamte Oberfläche des KGTs aufnehmen zu können, soll durch geschicktes Ausnutzen der Kinematik des KGTs erfüllt werden. Der Funktionsträger ist hierbei als das Gehäuse definiert, in das die Kamera integriert und von oben auf die Kugelgewindetriebspindel gerichtet ist. Durch die Positionierung der Kamera

und das Anbringen des Gehäuses an der Kugelgewindetriebmutter läuft die Spindel unter der Kamera hindurch. Damit durchläuft jeder Abschnitt der Spindel die Kamera. Die zweite Anforderung des Kamerasystems bezieht sich auf die automatisierte Aufnahme der Oberfläche zu definierten Zeitpunkten. Diese Anforderung wird dem Funktionsträger der Kommunikationsarchitektur zugeschrieben. Die dritte Anforderung der platzsparenden Integration in Fertigungseinrichtungen wird abermals dem Funktionsträger Gehäuse zugewiesen. Um die vierte Anforderung (Inline-Integration) zu ermöglichen, wird ein Montageadapter als weiterer Funktionsträger konzipiert. Der Montageadapter stellt sicher, dass das Sensorsystem an allen gängigen Kugelgewindetriebmutter-Geometrien befestigt werden kann, ohne hierfür das Gehäuse anpassen zu müssen. Die fünfte Anforderung besteht darin, Bilddaten zu liefern, die von den Umgebungsbedingungen unabhängig sind. Dieser Anforderung wird der Funktionsträger des Beleuchtungssystems zugewiesen. Durch geschickte Auswahl, Anordnung und Integration der Lichtquelle zusammen mit einem für die Anwendung konzipierten Diffusor werden gleichbleibende Beleuchtungsverhältnisse geschaffen. Die sechste Anforderung, zuverlässig Pittings auf der Spindeloberfläche zu erkennen, wird von dem Funktionsträger Kamera erfüllt. Durch Auswahl einer kompakten und dennoch hochauflösenden Standardkamera können frühe Stadien der Schädigung abgebildet und die Kamera kann platzsparend integriert werden.

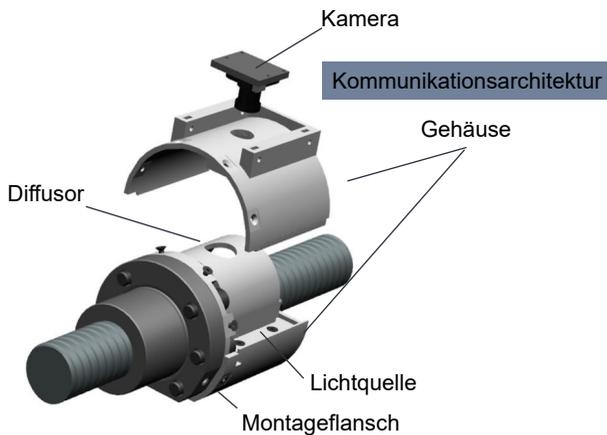


Abbildung 5-4 Komponenten des Sensorsystems nach (Schlagenhauf, Ruppelt & Fleischer 2020).

Abbildung 5-4 stellt die Komponenten des Sensorsystems dar. Im Folgenden wird auf die einzelnen Funktionsträger im Detail eingegangen.

Das Sensorsystem wurde in einer ersten Version von (Schlagenhauf & Hillenbrand et al. 2019) vorgestellt und besteht aus den Komponenten Gehäuse, Lichtquelle, Diffusor, Montageadapter, der Kamera sowie der Kommunikationsarchitektur. Das Gehäuse übernimmt mehrere Aufgaben: Zum einen dient es der Befestigung des Kamerasystems an dem Montageadapter und zum anderen als Integrationseinheit für die Kamera und die Lichtquelle. Es fixiert weiterhin den Diffusor. Das Gehäuse ist so gestaltet, dass das Kamerasystem möglichst platzsparend integriert werden kann. Weiterhin wurde bei der Konzipierung des Gehäuses die Beleuchtungssituation berücksichtigt. Aufgrund der meist reflektierenden metallischen Oberfläche von KGT muss eine möglichst diffuse Beleuchtung angestrebt werden. Die Außenwand des Gehäuses dient damit als Oberfläche, an der das Licht gestreut wird. Durch Kombination mit dem Diffusor wird damit sichergestellt, dass das Licht möglichst gleichmäßig auf die Oberfläche des KGTs geleitet wird. Eine weitere Anforderung, die durch das Gehäuse erfüllt wird, ist die Abschirmung des Systems gegenüber den Betriebs- und Umgebungsbedingungen. Durch das weitestgehend geschlossene Gehäuse wird ein Eindringen von Verschmutzung und Spänen von außen verhindert. Lediglich vom KGT selbst, der durch das Gehäuse läuft, kann Verschmutzung in das System eingebracht werden. Auf diese Thematik wird mit dem unten beschriebenen Diffusor eingegangen. Neben der Abschirmung vor ungewollter Verschmutzung trägt das Gehäuse zu einer gleichmäßigen Beleuchtungssituation bei, da es die Komponente KGT vor störendem und schwankendem Umgebungslicht abschirmt.

Die zweite Komponente des Sensorsystems ist der Montageadapter, der eine einfache Integration in unterschiedliche KGT ermöglicht.

Die dritte Komponente des Kamerasystems stellt die Lichtquelle dar. Hierbei wird eine LED-Beleuchtung aufgrund ihrer Robustheit und Langlebigkeit gewählt. Die Lichtquelle schafft eine konstante Beleuchtungssituation.

Die vierte Komponente des Kamerasystems ist der Diffusor. Dieser nimmt eine zentrale Rolle ein und erfüllt zwei Hauptaufgaben: Erstens sorgt er für diffuses Licht, indem das von der Gehäusewand reflektierte Licht weiter gestreut wird. Zweitens umschließt der Diffusor den KGT und dient damit als Schutz der Kamera und des Objektivs vor Schmierstoff und Verschmutzung, die aufgrund der Rotation der Spindel und der dabei

entstehenden Zentrifugalkräfte nach außen getragen werden können.²⁰ Damit dennoch eine Aufnahme der Kugelgewindetriebspindel möglich ist, ist eine kreisförmige Aussparung vorgesehen, durch welche die Kamera auf den KGT blickt. Der Diffusor ist aus halbtransparentem Kunststoff gefertigt. Um eine möglichst gute Beleuchtungssituation zu schaffen, wurden zur Auswahl des Diffusors in (A_Schlund 2019) Versuche mit unterschiedlichen Diffusor-Geometrien durchgeführt. Alle Konzepte verfolgten hierbei das Ziel, eine möglichst diffuse Beleuchtung zu generieren, und sind an die Arbeiten von (Beyerer & Punte León et al. 2016) angelehnt. Die Diffusor-Geometrien sind zusammen mit den Ergebnissen der Versuche in Abbildung 5-5 dargestellt. Die Aufnahmen werden hinsichtlich des Schattenwurfes und der sogenannten Bright Spots (dt.: überbelichtete Stellen) bewertet. Das Konzept, in dem der Diffusor als eine Art Tunnel mit kreisförmiger Aussparung für die Kamera gestaltet ist, liefert hinsichtlich der Bewertungskriterien die besten Ergebnisse und wurde für die Bildaufnahmen ausgewählt.

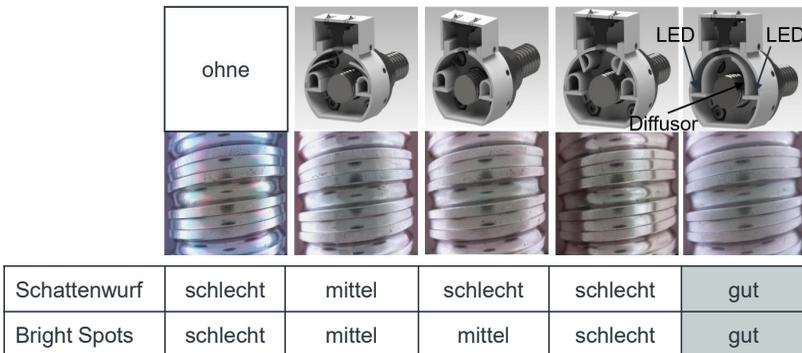


Abbildung 5-5 Versuchsergebnisse unterschiedlicher Diffusor-Geometrien nach (A_Schlund 2019).

Die fünfte Komponente des Kamerasystems ist die Kamera zur Bildaufnahme. Als Kameralösung wird eine handelsübliche, hochauflösende Mikrocontrollerkamera mit einem 8-Megapixel-Sony-IMX219-Chip verwendet. Bei der Kameraauswahl wurden die Aspekte Kosten, benötigter Platz zum Einbau, Auflösung sowie die Datenbereitstellung berücksichtigt. Die Kamera liefert Bildaufnahmen mit einer maximalen Auflösung von 3280×2464 Pixel.

²⁰ Vorgegriffen sei hier angemerkt, dass während der gesamten Verschleißversuche die Kameras nicht durch Fremdpartikel beeinflusst wurden und auch die Verschmutzungen an den Diffusoren zu vernachlässigen waren.

Die letzte Komponente des Kamerasystems ist die Kommunikationsarchitektur, welche die automatisierte Aufnahme von Bilddaten ermöglicht. Die Kommunikationsarchitektur ist grundsätzlich von der Hardware des Kamerasystems unabhängig. Um Bilddaten zur Abbildung des Verschleißfortschrittes zu generieren, ist eine automatisierte Bildaufnahme zu definierten Zeitpunkten notwendig. Hierfür muss das Kamerasystem mit der speicherprogrammierbaren Steuerung (SPS) der Anlage kommunizieren. Die Kommunikation wird dabei durch eine OPC-UA-Client-Server-Schnittstelle hergestellt. Der OPC-UA-Client wird durch ein Matlab-Skript erzeugt, das die Bewegung der Achse sowie die Bildaufnahme zu definierten Zeitpunkten automatisch startet. Abbildung 5-6 zeigt das verbaute Kamerasystem, wie es auf der Hannover Messe 2021 ausgestellt wurde.

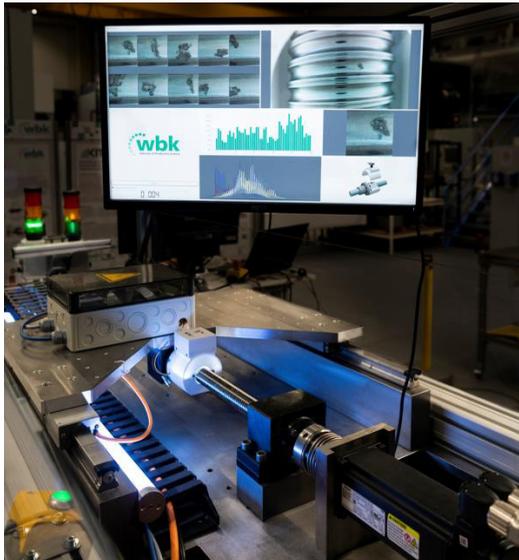


Abbildung 5-6 Sensorsystem im eingebauten Zustand.

Als Ergebnis wurde ein Kamerasystem konzipiert, mit dem automatisch Bilddaten der Spindeloberfläche in Verschleißversuchen aufgezeichnet werden können.

Der Prozess der Bildaufnahme sowie eine Absicherung der Eigenschaften des Kamerasystems durch Vorversuche werden im folgenden Abschnitt beschrieben.

5.3.3 Aufnahme von Bilddaten

In diesem Abschnitt werden der Prozess der Bildaufnahme sowie die mittels des konzipierten Sensorsystems erzielten Ergebnisse der Bildaufnahmen vorgestellt. Weiterhin wird in diesem Abschnitt das Sensorsystem hinsichtlich der im vorherigen Abschnitt definierten Anforderungen validiert.

Mit dem Kamerasystem wird ein Bilddatensatz generiert, der die in Kapitel 6 vorgestellte Verschleißentwicklung von KGT von der Inbetriebnahme bis zum mechanischen Ausfall der Komponente abbildet. Dieser Bilddatensatz dient als Grundlage für die Analyse der Verschleißentwicklung. Zur Bilddatengenerierung wird das Sensorsystem am 5-Achs-Versuchsstand integriert. Die Systeme werden an den Kugelgewindtriebmuttern befestigt und per Ethernet-Schnittstelle mit der Auswerteeinheit verbunden (siehe Abbildung 5-7).

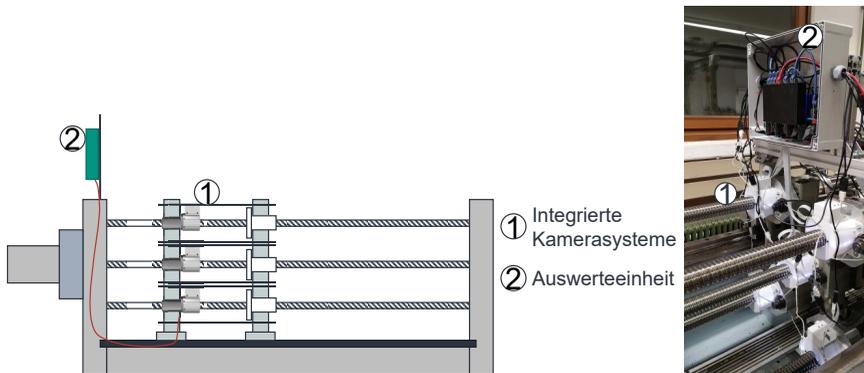


Abbildung 5-7 Integrierte Sensorsysteme am 5-Achs-Versuchsstand.

Die Auswerteeinheit besteht hierbei aus Mikrocontrollern, in welche die oben beschriebene Kommunikationsarchitektur implementiert wurde. Diese übernehmen die Kommunikation mit der SPS und starten die Kameraaufnahmen automatisiert alle vier Stunden. Das Kamerasystem liefert Aufnahmen von der Oberfläche der Kugelgewindtriebspindel, wie sie in Abbildung 5-8 beispielhaft dargestellt sind. Stellen mit Oberflächendefekten in Form von Pittings sind grün umrahmt dargestellt.



Abbildung 5-8 Beispielhafte Aufnahmen des Kamerasystems. Links: Oberfläche mit Schädigungen; rechts: nicht geschädigte Oberfläche mit betriebsbedingten Verunreinigungen.

Das Kamerasystem wurde in der obigen Anordnung in einem Vorversuch zur Eigenchaftsabsicherung in den Versuchsstand integriert und ein Dauerversuch wurde gestartet. Abbildung 5-9 zeigt denselben Spindelabschnitt, der automatisiert zu diskreten Zeitpunkten über einen Zeitraum von 40 Tagen aufgenommen wurde. Aus den Aufnahmen ist erkennbar, dass das Kamerasystem über den betrachteten Zeitraum hinweg Bilddaten in gleichbleibender Qualität liefert. Der Schmierstoff, der für die Schmierung des KGTs verwendet wird, beeinflusst die Bildaufnahmen nicht. Ebenfalls ist die Verschmutzung in der Mitte der Spindel sehr gut zu erkennen, was sich auch bei näherem Betrachten des Bildausschnittes, der die Verschmutzung zeigt, bestätigt. Trotz der geringen Größe der Verschmutzung lässt sich diese auf den Bilddaten sehr detailliert erkennen. Die Eignung des Kamerasystems für eine automatisierte Generierung von Bilddaten zur Abbildung von Schädigungen lässt sich damit bestätigen. Die Bildaufnahmen werden alle vier Stunden automatisiert gestartet. Nach jedem Bild wird die Spindel um 22.5° weitergedreht. Damit wird sichergestellt, dass alle Bereiche der Spindel abgebildet sind. Die einzelnen Bilder werden mit einer Auflösung von 2592×1944 Pixel aufgenommen. Die Auflösung wurde als Trade-off zwischen der maximal möglichen Auflösung, der Zeit zur Aufnahme und des benötigten Speicherplatzes gewählt. Es entstehen so pro Kamerafahrt und Spindel alle vier Stunden 521 Bilder, welche die gesamte Spindeloberfläche abbilden.

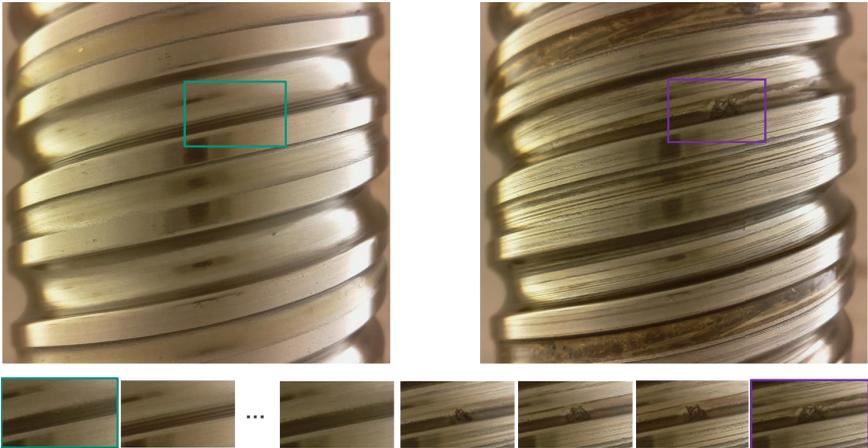


Abbildung 5-10 Verschleißfortschritt eines Spindelabschnittes. Oben: Spindelabschnitt zu Versuchsbeginn und Versuchsende; unten: Entwicklung einer Schädigung.

Die Bilddaten der Spindeln werden in Kapitel 6 für die Analyse der Verschleißentwicklung und die Untersuchung von Klassifikationsmodellen verwendet. Zur Verwendung in einem Modell zur Klassifikation von Pittings wird die Datenbasis weiter vorverarbeitet. Dazu werden aus den Originalbilddaten 150 x 150 Pixel große Ausschnitte ausgeschnitten. Die Größe der Ausschnitte ist so gewählt, dass sowohl kleine als auch große Schädigungen darüber abgebildet werden können. Insgesamt wurden aus den Originalbilddaten mittels dieses Verfahren 21 853 Bilder ausgeschnitten und manuell mit einem Label $\mathcal{L} \in \{P, N\}$ versehen, wobei P für *Positiv* := *Pitting* und N für *Negativ* := *Kein Pitting* steht. Der Bilddatensatz enthält 11 075 Bilder der Klasse N sowie 10 778 Bilder der Klasse P und ist damit nahezu gleichmäßig aufgeteilt. Die Bilddaten wurden in (Schlagenhauf 2021) veröffentlicht, um damit einen Beitrag zum Stand der Forschung zu leisten, der die Forschungsgemeinschaft in die Lage versetzen soll, weitere Modelle zur Klassifikation von Fehlstellen auf metallischen Oberflächen unter Realbedingungen zu untersuchen.

In dem Bilddatensatz sind Bilder aller möglichen KGT-Zustände enthalten. So finden sich Bilder, die eine Spindel ohne Verschmutzung und ohne Schädigung zeigen, und solche, die eine verschmutzte Spindel ohne Schädigung zeigen. Weiterhin sind Bilder

enthalten, die eine geringe Schädigung ohne Verschmutzung sowie eine geringe Schädigung mit Verschmutzung zeigen. Gleiche Bilder sind für große Schädigungen vorhanden (Abbildung 5-11).

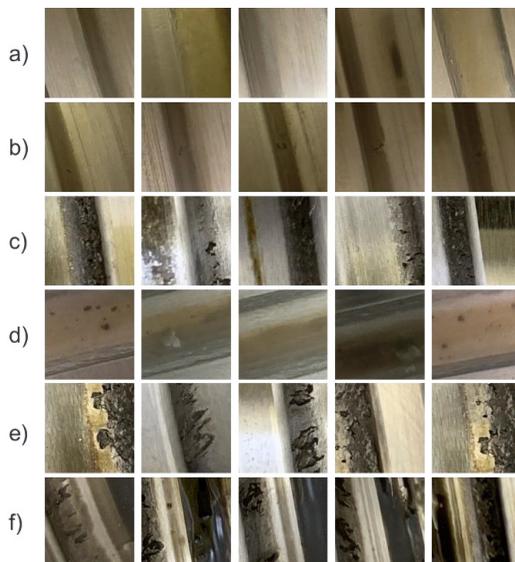


Abbildung 5-11 Beispiele der aufgenommenen Bilddaten. a) Keine Schädigung, keine Verschmutzung, b) Geringe Schädigung, keine Verschmutzung, c) Geringe Schädigung, Verschmutzung, d) Keine Schädigung, Verschmutzung, e) Große Schädigung, keine Verschmutzung, f) Große Schädigung, Verschmutzung. (Schlagenhauf & Landwehr 2021).

Abbildung 0-3 und Abbildung 0-4 im Anhang erweitern die in Abbildung 5-11 gezeigte Menge an Bilddaten um Bilddaten mit und ohne Schädigung.

Als Ergebnis konnte in den Versuchen ein Bilddatensatz generiert werden, der zum einen die gesamte Schädigungshistorie aller Pittings auf der Spindel beinhaltet und zum anderen zur Untersuchung von Klassifikationsmodellen zur automatischen Identifikation von Oberflächenzerrüttung in den Bilddaten verwendet werden kann. Die in diesem Bilddatensatz enthaltenen Entwicklungen der Oberflächenzerrüttung werden im folgenden Kapitel empirisch analysiert.

6 Datenanalyse und Modellbildung

In den vorherigen Abschnitten wurden der Versuchsaufbau sowie das Sensorsystem zur experimentellen Datengenerierung beschrieben und das Sensorsystem wurde in seinen Eigenschaften abgesichert. Weiter wurden die Merkmale des in den Versuchen entstandenen Bilddatensatzes beschrieben. Diese Bilddaten werden in diesem Kapitel empirisch analysiert.

Zunächst werden die in den Bilddaten vorherrschenden Verschleißerscheinungsformen auf der KGT-Laufbahn sowie im Schmierstoff beschrieben. Anschließend wird die Entstehung von Pittings aufgrund von Oberflächenzerrüttung zunächst theoretisch eingeführt und dann empirisch untersucht. Der Untersuchung der Entstehung von Pittings folgt die Analyse der Entwicklung von Pittings über die Lebensdauer der Komponenten hinweg. Das Kapitel schließt mit der Ableitung von charakteristischen Merkmalen eines Pittings.

Im Teilkapitel der Klassifikation von Fehlstellen in Bilddaten des KGTs werden die definierten Charakteristika zur Klassifikation von Pittings verwendet und schließlich unterschiedliche Deep-Learning-basierte Ansätze zur Klassifikation der Bilddaten untersucht. Das Kapitel schließt mit einem Exkurs zur Dateneffizienz in technischen Domänen.

6.1 Analyse der Entwicklung von Pittings

In diesem Abschnitt wird die Entwicklung des Schadensbildes der Oberflächenzerrüttung am KGT über die Lebensdauer des KGTs hinweg analysiert.

Zunächst wird auf die in den Bilddaten grundsätzlich zu identifizierenden Verschleißmerkmale auf der Kugelgewindetriebspindel sowie im Schmierstoff eingegangen. Anschließend wird der Fokus auf die Analyse des betrachteten Verschleißmechanismus der Oberflächenzerrüttung sowie der daraus hervorgehenden Pittings und deren visueller Charakteristika gelegt. Der Abschnitt schließt mit einem Vorschlag zur Erweiterung der Norm zur Bestimmung der Lebensdauer des KGTs sowie mit einer kritischen Diskussion der Erkenntnisse im Kontext bestehender Arbeiten zur Zustandsüberwachung am KGT. Die Analysen wurden im Rahmen der Arbeiten von (A_Hofmann 2021), (A_Krasnikow 2021) und (Schlagenhauf & Scheurenbrand et al. 2022) durchgeführt. Die Ergebnisse werden im Folgenden zusammenfassend dargestellt.

6.1.1 Verschleißerscheinungsformen auf der Laufbahn des Kugelgewindetriebes

Pittings, wie sie in unterschiedlichen Stadien auf Kugelgewindetriebspindeln entstehen, sind an vielen Stellen (siehe bspw. Abschnitt 2.2.1, Abschnitt 5.3.3 sowie Abbildung 0-3 und Abbildung 0-4 im Anhang) dieser Dissertation abgebildet. Die visuellen Charakteristika von Pittings werden in Abschnitt 6.1.7 detailliert beschrieben.

Im Grundlagenabschnitt 2.2.1 wurde detailliert auf die vorherrschenden Verschleißmechanismen am Kugel-Laufbahn-Kontakt des tribologischen Systems KGT eingegangen, weshalb dieser Abschnitt den Grundlagenteil lediglich durch markante Beispiele erweitern soll. In Abschnitt 2.2.1 wurde aufgrund der in der Literatur beschriebenen Vorgänge auf sich gegenseitig beeinflussende Verschleißmechanismen am KGT geschlossen. (Forstmann 2010) schreibt, dass „ein initialer Verschleißmechanismus bei fortgesetzter Beanspruchung zur Ausprägung weiterer Verschleißmechanismen“ (Forstmann 2010, S. 22) führt.

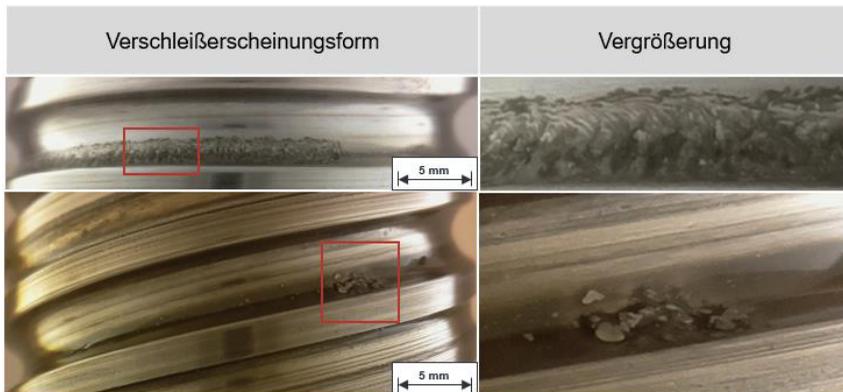


Abbildung 6-1 Sich überlagernde Verschleißmechanismen auf der Kugelgewindetriebspindel. Oben: Adhäsion; unten: Abrasion.

Es ist davon auszugehen, dass bei großflächigen Schädigungen auf der Oberfläche nicht ausschließlich der Verschleißmechanismus der Oberflächenzerrüttung zum Tragen kommt, sondern ebenfalls der Verschleißmechanismus der Adhäsion, der dann zu den in Abbildung 6-1 gezeigten großflächigen Ausbrüchen und Abschälungen aus der Oberfläche führt. Weiterhin ist ersichtlich, dass der Schmierstoff, der eine Oberflächenzerrüttung auf der Spindel bedeckt, mit klar erkennbaren Verschleißpartikeln durchsetzt

ist. Diese abrasiven Verschleißpartikel können dann wiederum den Ausgangspunkt für weitere Oberflächenzerrüttungen darstellen (siehe Abbildung 6-8 in Abschnitt 6.1.3.2).

Als Ergebnis kann hiermit die Annahme, dass sich überlagernde und gegenseitig beeinflussende Verschleißmechanismen am KGT vorliegen, anhand von Bilddaten des KGTs empirisch geprüft werden. Weitere empirische Belege finden sich bei der Betrachtung der Verschleißpartikel im Schmierstoff sowie bei der Analyse der Entwicklung der Schädigungen, auf die im Folgenden eingegangen wird.

6.1.2 Verschleißpartikel im Schmierstoff

Über die Farbe des Schmierstoffes sowie die vorliegenden Verschleißpartikel soll eine Korrelation mit dem Verschleißzustand des KGTs hergestellt werden. Da die verschleißbehafteten Kugeln, die KGT-Mutter sowie die Rückführsysteme nicht direkt mit dem präsentierten Ansatz eingesehen werden, soll über die Farbe des Schmierstoffes sowie die Verschleißpartikel im Schmierstoff indirekt auf den Zustand dieser Komponenten geschlossen werden. Hierbei sei erwähnt, dass Partikel, wie in Abschnitt 2.2.1 beschrieben, durch Abrasion, Adhäsion, tribochemische Reaktion oder Oberflächenzerrüttung entstehen und neben den genannten Komponenten ebenfalls auf der Kugelgewindetriebspindel auftreten können.

Mit dem vorgestellten System kann damit lediglich eine Abschätzung vorgenommen werden, eine klare Zuordnung der Partikel zum Entstehungsort ist dagegen nicht möglich. Abbildung 6-2 zeigt unterschiedliche Partikel im Schmierstoff des KGTs. Eine praktische Möglichkeit zur Identifikation von Partikeln auf der Kugelgewindetriebspindel bietet die sogenannte positive Fingerprobe (Spohrer 2019) bzw. der Fingertest (Deutsches Institut für Normung 2010a). Dieser Test wird oftmals zur einfachen Feststellung von Verschleiß im System eingesetzt, ist in seiner Durchführung aber nicht weiter definiert. Beim Fingertest wird mit dem Finger undefiniert über die Spindel gestrichen und der Abstrich wird auf metallische Partikel untersucht. Sind metallische Partikel sichtbar, so liegt Verschleiß im System vor. Ein mehrfaches Durchführen des Tests an unterschiedlichen Stellen der Spindel sichert den Test ab.

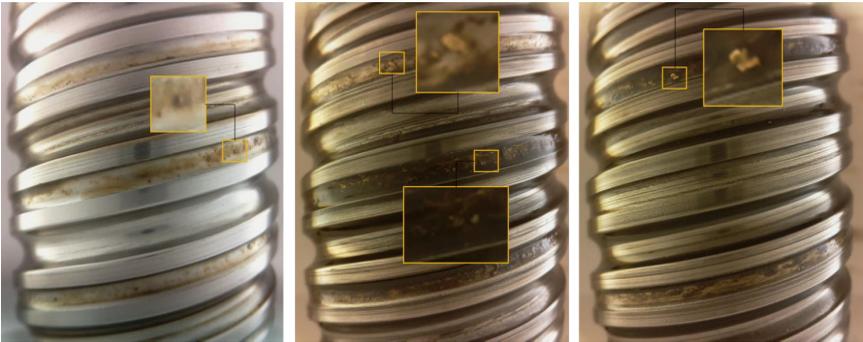


Abbildung 6-2 Späne im Schmierstoff des KGTs als Indikator für vorliegenden Verschleiß im System.

Über die Größe und Anzahl der Partikel im System kann eine grobe Abschätzung über den Verschleißzustand getroffen werden. Der Vorteil des Fingertests ist, dass dieser sehr einfach durchzuführen ist. Zwar ist keine Zuordnung zu spezifischen Komponenten möglich, allerdings lässt sich bei Vorhandensein von Partikeln zweifelsfrei auf Verschleiß im System schließen (Abbildung 6-3).



Abbildung 6-3 Fingertest mit kleinen Metallpartikeln.

Neben dem Auftreten von Partikeln im Schmierstoff lässt sich eine gewisse Korrelation der Farbe des Schmierstoffes mit der Betriebszeit des KGTs feststellen. Es ist zu beobachten, dass sich der visuelle Gesamtzustand der Spindel über die Einsatzzeit des KGTs hinweg dahingehend verändert, dass der verwendete Schmierstoff eine zuneh-

mend dunkle Färbung aufweist. Diese Entwicklung lässt sich in den Bilddaten gut beobachten und kann als schwacher Indikator für den Verschleißzustand der Spindel gewertet werden.

	Entwicklung des Schmierstoffes über die Lebensdauer (Ä) des KGT					
						Verschleiß auf der Spindellaufbahn
Schulter						
Lebensdauer (Ä)	6%	55%	86%	91%	97%	
Farbe	Hellbraun	Hellbraun	Braun & Schwarz	Schwarz	Schwarz	
Partikel und Späne	-	-	+	++	+++	

Abbildung 6-4 Korrelation des Schmierstoffzustandes mit dem Verschleißzustand des KGTs. Markierungen stellen Verschleißpartikel dar.

Die in dieser Arbeit ausgewerteten Bilddaten zeigen bei allen vorliegenden ausgefallenen Spindeln eine dunkelbraune bis schwarze Verfärbung des Schmierstoffes, wobei der Farbumschlag von Hellbraun zu Dunkelbraun bzw. Schwarz teilweise erst bei > 85 % der erreichten Lebensdauer eintritt (Abbildung 6-4).

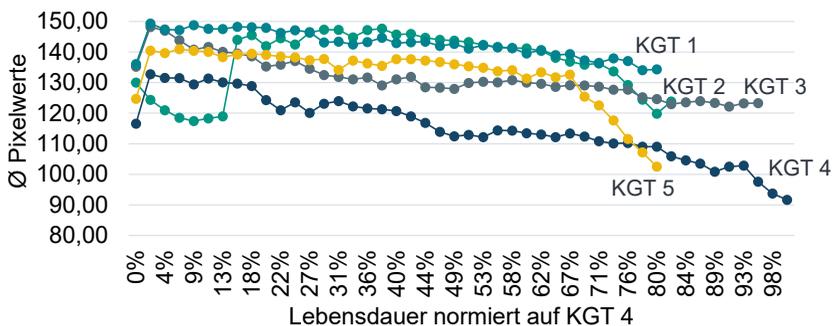


Abbildung 6-5 Entwicklung der Schmierstofffärbung über die Lebensdauer der KGT hinweg.

Bei allen Versuchen konnten bei Ausfall der Komponente Metallpartikel im Schmierstoff nachgewiesen werden. Abbildung 6-5 stellt diesen Zusammenhang anhand von fünf KGT dar. Hierbei wurde der Durchschnitt der Pixelwerte (Helligkeit) zu unterschiedlichen Zeitpunkten berechnet und über der Lebensdauer der Komponenten aufgetragen. Es ist eine klare Korrelation zwischen dunklen Verfärbungen und zunehmender Lebensdauer zu beobachten.

Ein wesentlicher Aspekt, der die Farbe des Schmierstoffes als generellen Indikator einschränkt, ist, dass der Schmierstoff nach (Murrenhoff 2010) Alterungsprozessen unterliegt, die unabhängig vom Betrieb des KGTs eintreten. Diese Alterungsprozesse führen damit zu einer Verfärbung des Schmierstoffes, die durch korrodierende Metallpartikel im System zu einem späteren Zeitpunkt im Betrieb noch verstärkt wird. Da bei der Dauer der Einzelversuche (wenige Monate) der starke Einfluss von Alterungseffekten jedoch weitestgehend vernachlässigt werden kann, lässt sich aufgrund der Verfärbung im Schmierstoff folgender Zusammenhang ableiten: Die Farbe des Schmierstoffes ist ein Indiz für den Verschleißzustand des KGTs, wobei vorliegende Partikel im System ein Merkmal für Verschleiß sind.

Damit kann festgehalten werden, dass sich über die Verfärbung des Schmierstoffes eine schwache Korrelation mit dem Verschleißzustand einer Komponente herstellen lässt. Da sich Schmierstoff auch aufgrund von Alterungseffekten verfärbt, ist die Farbe des Schmierstoffes allerdings nicht ausreichend, um zweifelsfrei auf Verschleiß im System zu schließen. Hierzu ist ebenfalls das Vorhandensein von Verschleißpartikeln im Schmierstoff notwendig.

Im Anschluss an die Betrachtung des Schmierstoffes folgt die Untersuchung der Entstehung von Pittings auf der Spindeloberfläche des KGTs.

6.1.3 Die Entstehung von Pittings

Im Folgenden wird die Entstehung von Pittings aufgrund von Oberflächenzerrüttung zunächst theoretisch diskutiert und anschließend anhand der aufgenommenen Bilddaten empirisch analysiert.

6.1.3.1 Theoretische Betrachtung

In der vorliegenden Arbeit wird die Identifikation und Quantifizierung des Verschleißmerkmals Pitting auf der Oberfläche der Kugelgewindetriebspindel fokussiert. Nach

(Sommer, Heinz & Schöfer 2018) sowie (Czichos & Habig 2015) entsteht das Verschleißmerkmal Pitting entweder ausgehend von der Oberfläche oder von unterhalb der Oberfläche des KGTs.

Bei vollständiger Trennung der Wälzpartner durch Schmierstoff und eine damit vorliegende elastohydrodynamische Schmierung entstehen die Pittings ausgehend von unterhalb der Oberfläche durch Aufstau von Versetzungen an Korngrenzen und das daraus resultierende Risswachstum. Bei vorliegenden Einschlüssen im Werkstoff oder Vorschädigungen, bspw. durch Eindrücke, wachsen Pittings ausgehend von der Oberfläche. Ebenfalls tritt ein Wachstum ausgehend von der Oberfläche bei zunehmender Misch- oder Grenzreibung auf, das in einem Herausbrechen von keilförmigen Partikeln resultiert. Ein Beispiel für die Entstehung eines Pittings von der Oberfläche (keilförmiger Ausbruch) ist in Abbildung 6-6 oben dargestellt.

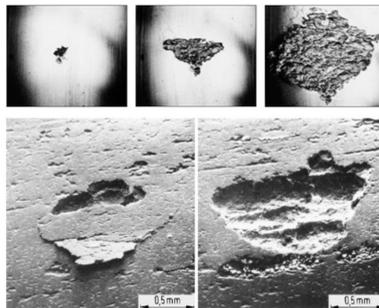


Abbildung 6-6 Oben: Pitting ausgehend von adhäsiver Vorschädigung der Oberfläche durch Materialeinschluss (Sommer, Heinz & Schöfer 2018); unten: Ausbrechen von Materialpartikeln (Verein Deutscher Ingenieure 2017).

Die Entstehung eines Pittings an der Oberfläche durch Herausbrechen von Materialpartikeln ist in Abbildung 6-6 unten dargestellt. Bei der Entstehung eines Pittings ausgehend von der Oberfläche ohne den Einfluss von Fremdpartikeln wird davon ausgegangen, dass an der Oberfläche durch Überschreiten der zulässigen Beanspruchung sogenannte Schwingrisse entstehen, die dann weiter wachsen. Durch weitere Belastung entstehen Pittings als keilförmige Ausbrüche. Ist die Oberfläche durch Formabweichungen, Kerben oder Mulden bzw. Fremdkörpereinschlüsse vorgeschädigt, so wird dieser Effekt verstärkt. Bei der Entstehung von Pittings an der Oberfläche und vorliegendem negativem Schlupf werden die Risse in der Oberfläche vor der Kugel geöffnet und es dringt Schmierstoff ein. Dadurch bilden sich sogenannte Ölkavernen. Überrollt

eine Kugel diese Ölkavernen, wird der Riss verschlossen und der darin befindliche Schmierstoff komprimiert, was eine weitere Vergrößerung des Risses zur Folge hat und schließlich in Pittings auf der Oberfläche resultiert (Sommer, Heinz & Schöfer 2018).

Entsteht ein Pitting unterhalb der Oberfläche, lässt sich dieser Prozess nach (Czichos & Habig 2015) in die folgenden vier Subphasen unterteilen: Inkubationsphase, Mikrorissphase, Rissausbreitungsphase und Phase des endgültigen Bruches. Die Annahme ist hierbei, dass der Grundwerkstoff durch Überrollen der Kugel einer periodischen Last unterliegt und es zu einer Akkumulation der Gitterversetzungen kommt. Diese vereinigen sich zu Mikrorissen, die sich später in Form von Pittings an der Oberfläche ausdrü-

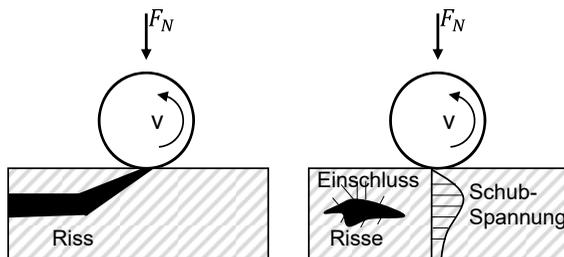


Abbildung 6-7 Mechanismus der Entstehung eines Pittings. Links: Ausgehend von der Oberfläche; rechts: Ausgehend von unterhalb der Oberfläche nach (Czichos & Habig 2015).

cken. Bei der Entstehung von Versetzungen wird davon ausgegangen, dass die Schubspannung am Punkt der Versetzung größer ist als die kritische Schubspannung des Materials. Dies ist qualitativ in Abbildung 6-7 rechts dargestellt.

Zusammenfassend entstehen Pittings also entweder ausgehend von der Oberfläche durch Vorschädigungen oder von unterhalb der Oberfläche ausgehend durch das Wandern und den Aufstau von Versetzungen.

Im folgenden Abschnitt wird auf Frühindikatoren in den Bilddaten eingegangen, die auf eine Entstehung von Pittings von unterhalb der Spindeloberfläche oder von der Spindeloberfläche ausgehend schließen lassen, und die theoretische Betrachtung wird damit empirisch untermauert.

6.1.3.2 Frühindikatoren für die Entstehung von Pittings in den Bilddaten

Wie im vorherigen Abschnitt angedeutet, lassen sich Pittings ihrer Entstehung nach gliedern in jene, die ausgehend von der Oberfläche als Resultat einer Vorschädigung entstehen, und solche, die von unterhalb der Oberfläche ausgehend entstehen und sich damit erst nach Durchbrechen der Oberfläche in den Bilddaten zeigen. Im Rahmen der Verschleißversuche sind insgesamt 54 Pittings entstanden, welche die Grundlage für die weiteren Untersuchungen darstellen.

6.1.3.3 Entstehung ausgehend von einer Oberflächenschädigung der Spindel

Vorschädigungen auf der Spindel deuten eine Entstehung von Pittings ausgehend von der Oberfläche an. Dies wird auch von (Sommer, Heinz & Schöfer 2018) bestätigt, welche beschreiben, dass vorgeschädigte Bereiche aufgrund der höheren Belastung potenzielle Ausgangsorte für Pittings darstellen. Abbildung 6-8 zeigt das Wachstum eines Pittings ausgehend von einem Abrasivpartikel, der durch Überrollen in die Oberfläche der KGT-Spindel eingedrückt wurde. Hierbei ist in einem frühen Stadium der normierten

Pitting	Laufleistung	Bilddaufnahme
Mit Vorschädigung als Frühindikator	25%	
	26%	
	80%	
	100%	

Abbildung 6-8 Beispiel für die Entstehung eines Pitting ausgehend von einem eingedrückt Abrasivpartikel in der Oberfläche nach (A_Hofmann 2021).

Lebensdauer ein Span auf der Spindellaufbahn zu beobachten. Dieser Span wird dann unmittelbar nach dieser Beobachtung überrollt und in die Spindeloberfläche einge-

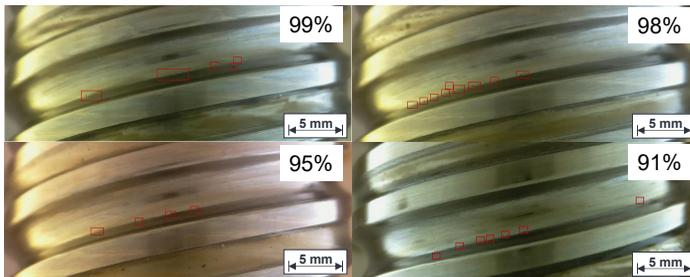


Abbildung 6-9 Beispiele von Vorschädigungen auf der Spindeloberfläche, aus denen auch bei fortgeschrittener Lebensdauer keine Pittings entstanden sind.

drückt (abrasiver Verschleiß). An diesem eingedrückten Span entsteht ein Pitting aufgrund von Oberflächenzerrüttung. Dieses Pitting wächst und es lassen sich adhäsive Effekte nicht mehr ausschließen bzw. diese sind aufgrund der freiliegenden Oberfläche sehr wahrscheinlich.

Im Rahmen der Untersuchungen konnte festgestellt werden, dass Pittings mit einer erhöhten Wahrscheinlichkeit an Stellen mit Vorschädigungen entstehen. Es konnte allerdings nicht bestätigt werden, dass während der Versuche jede Vorschädigung in einem Pitting mündet. So zeigt Abbildung 6-9 Vorschädigungen (rot umrandet), aus denen auch bei fortgeschrittener Lebensdauer keine Pittings entstanden sind. Dies schränkt die Verwendung von eingebetteten Partikeln als sichere Vorstufe zum Pitting als Merkmal ein.

Sind keine Vorschädigungen auf der Spindel zu beobachten, so stellen Verschleißpartikel im System dennoch ein Indiz für sich ankündigende Oberflächenzerrüttung dar. Dies ist damit zu begründen, dass eine Zunahme an Verschleißpartikeln die Wahrscheinlichkeit erhöht, dass Verschleißpartikel überrollt werden und diese wiederum als Ausgangspunkt für Pittings dienen. Weiterhin deuten auch nicht überrollte Verschleißpartikel im System auf eine Vorschädigung hin. Liegt eine Vorschädigung bspw. an den Wälzkörpern vor, so führen die scharfen Kanten zu verschleißfördernden Festkörperkontakten. Sind bereits Pittings auf der Spindel sichtbar, dann ist auch dies ein Indiz dafür, dass sich weitere Schädigungen auf der Spindel manifestieren.

Zusammenfassend konnte die Entstehung von Pittings ausgehend von der Oberfläche in den Bilddaten gezeigt werden. Im nächsten Abschnitt wird die Entstehung von Pittings ausgehend von unterhalb der Oberfläche betrachtet.

6.1.3.4 Entstehung ausgehend von unterhalb der Spindeloberfläche

In den Untersuchungen konnte beobachtet werden, dass der größere Teil der Pittings (34) von unterhalb der Oberfläche ausgehend entstanden ist, wohingegen bei 20 Pittings das Wachstum ausgehend von einer Vorschädigung beobachtet wurde. Abbildung 6-10 zeigt die plötzliche Entstehung eines Pittings ausgehend von unterhalb der Oberfläche.

Pitting	Laufleistung	Bildaufnahme
Ohne Vorschädigung / ohne Frühindikator	70%	
	71%	
	100%	

Abbildung 6-10 Beispiel für die Entstehung eines Pittings ausgehend von unterhalb der Oberfläche nach (A_Hofmann 2021).

Zu beobachten ist, dass das Pitting plötzlich an der Oberfläche der Spindel sichtbar wird und im Voraus keinerlei Verschleißpartikel auf der Oberfläche zu beobachten waren. Nach Entstehen des Pittings ist kein Unterschied mehr festzustellen und es wächst analog zu den Pittings, die ausgehend von der Oberfläche entstanden sind. Hierbei ist kritisch anzumerken, dass mittels des Sensorsystems nicht vollkommen ausgeschlossen werden kann, dass eine Menge der Pittings, von denen angenommen wird, dass sie ausgehend von unterhalb der Oberfläche entstanden sind, durch plötzliche Effekte ebenfalls ausgehend von der Oberfläche entstanden ist. Da allerdings bei keinem der Pittings im Voraus Verschleißpartikel an den entsprechenden Stellen identifiziert werden konnten, ist mit überwiegender Wahrscheinlichkeit von einer Entstehung ohne oberflächliche Vorschädigung auszugehen.

Als Ergebnis konnte in den Bilddaten die Entstehung von Pittings ausgehend von unterhalb der Oberfläche gezeigt werden. Als Nebeneffekt zur Beobachtung von Frühin-

dikatoren für die Entstehung von Pittings können hiermit die Aussagen aus den vorherigen Abschnitten bestärkt werden, dass sich Verschleißmechanismen gegenseitig beeinflussen.

Anschließend an die Entstehung wird im nächsten Abschnitt das Wachstum der Pittings über die Lebensdauer der Komponenten hinweg betrachtet.

6.1.4 Empirische Analyse der Entwicklung von Pittings

In den vorangehenden Abschnitten wurde gezeigt, dass Verschleiß auf der Oberfläche des KGTs mittels des konzipierten Sensorsystems zuverlässig aufgenommen und in Bilddaten abgebildet werden kann. Ebenfalls kann die Entstehung von Pittings in den Bilddaten sowohl ausgehend von einer Vorschädigung an der Oberfläche als auch ausgehend von unterhalb der Oberfläche beobachtet werden. In diesem Abschnitt wird die zeitliche Entwicklung von Pittings über die Lebensdauer des KGTs hinweg analysiert. Im Zuge der Analyse wird detailliert auf die Entwicklung der Anzahl an Pittings auf den Spindeln, die Entwicklung der Schadensfläche sowie die Entwicklung der Ausdehnungen der Pittings eingegangen. Das Ziel dieses Abschnitts ist es, die Entwicklung des Verschleißmerkmals Pitting erstmals in Bilddaten zu analysieren und somit Erkenntnisse für den Stand der Forschung zu schaffen. Weiterhin sollen die im Stand der Forschung bestehenden Erkenntnisse zur Verschleißentwicklung am KGT erstmals anhand von Bilddaten empirisch beleuchtet werden.



Abbildung 6-11 Beispiele für die Entwicklung unterschiedlicher Pittings (P1–P7) über die Lebensdauer von Komponenten hinweg nach (Schlagenhauf & Scheurenbrand et al. 2022). Erste Spalte: Bilder zu Versuchsbeginn; letzte Spalte: Bilder bei Lebensdauerende. Dazwischen wurden Beispiele gleichmäßig aus den Bilddaten gezogen. P2, P3 und P5 zeigen Partikel auf der Spindel, bevor ein Pitting entsteht.

Auf insgesamt sechs der zwölf untersuchten Teilsystemen konnte Verschleiß identifiziert werden. Auf den Spindeln konnten die Entstehung und das Wachstum von insgesamt 54 Pittings beobachtet werden. Tabelle 6-1 zeigt die Schwankungen in der tatsächlich erreichten Lebensdauer der Komponenten bei nahezu gleichen Betriebsbedingungen und unterstreicht damit die Notwendigkeit einer zuverlässigen Verschleißüberwachung. Von besonderer Relevanz sind die Fälle, die noch vor Erreichen der nominalen Lebensdauer ausgefallen sind, da selbst bei einer konservativen Instandhaltungsstrategie auf Basis der L_{10} Lebensdauer ein ungeplanter Maschinenstillstand die Folge gewesen wäre.

Tabelle 6-1 Erreichte Lebensdauer der KGT, normiert auf die L_{10} Lebensdauer.

Teilsystem	Erreichte nominelle Lebensdauer
KGT1	410,93 %
KGT2	228,48 %
KGT3	1348,45 %
KGT4	78,99 %
KGT5	475,57 %
KGT6	51,64 %

Abbildung 6-11 stellt die Entwicklung von sieben Pittings über die gesamte Lebensdauer dar. Das erste Bild stellt jeweils die Oberfläche zu Versuchsbeginn und das letzte die Oberfläche bei Versuchsende dar. Zwischen dem jeweils ersten und zweiten Bild liegt ein größerer Zeitraum, in dem keine Veränderung der Oberfläche sichtbar ist. Da hauptsächlich das Wachstum der Schädigungen von Interesse ist, wurde dieser Zeitraum aufgrund der Übersichtlichkeit nicht dargestellt. Abbildung 0-6 im Anhang zeigt den gesamten Verlauf der Schädigungen aus Abbildung 6-11. Die Entstehung von Schädigungen ist auf den Bildern gut sichtbar und die Entwicklung gut nachvollziehbar. Schwere Verschmutzungen der Spindeln haben erst bei Erreichen von nahezu 100 % Lebensdauer einen Einfluss auf die Sichtbarkeit der Schädigungen in den Bilddaten. In der Entwicklung von P2 und P3 ist gut sichtbar, dass diese durch eine Vorschädigung der Spindel in Form eines überrollten Abrasivpartikels entstanden sind. Weiterhin ist für P1 und P6 gut sichtbar, dass diese plötzlich entstanden sind und keine sichtbare Vorschädigung vorhanden ist.

Für alle Schädigungen ist mit zunehmender Lebensdauer eine Ausbreitung entlang der Spindellaufbahn zu beobachten, wobei sich das Pitting axial nicht weiter ausbreitet. Dies ist mit der Kinematik des KGTs sowie den Kugeldurchmessern zu begründen, da eine Schädigung, die durch Überrollen entstanden ist, den Kugeldurchmesser nicht stark überschreiten kann. In tangentialer Richtung ist das Wachstum hingegen nicht begrenzt.

Zusammenfassend können die große Schwankung in den Versuchszeiten bei identischen Betriebsbedingungen und damit die Notwendigkeit einer Zustandsüberwachung unterstrichen werden. Weiterhin ist in den Bilddaten das Wachstum der Pittings von Versuchsbeginn bis zum Ende der Lebensdauern der Komponenten gut sichtbar.

Im Folgenden wird auf die Entwicklung der Anzahl, der Fläche sowie der Ausdehnung der Pittings auf den Komponenten sowie auf die lokalen Charakteristika der Pittings detailliert eingegangen. Die Vorarbeiten dazu wurden in den Arbeiten von (A_Krasnikow 2021) und (A_Hofmann 2021) geleistet und werden im Folgenden zusammenfassend dargestellt.

6.1.4.1 Anzahl von Pittings

Als erste charakteristische Größe wird die Entwicklung der Anzahl an Pittings untersucht.

Abbildung 6-12 zeigt die Entwicklung der Anzahl an Pittings über die Lebensdauer der Komponenten hinweg. Hierbei ist zu beobachten, dass vereinzelt bereits bei Erreichen einer geringen Lebensdauer (ca. 30 %) erste Pittings entstehen und diese Entstehung zuverlässig mithilfe der Bilddaten identifiziert werden kann.

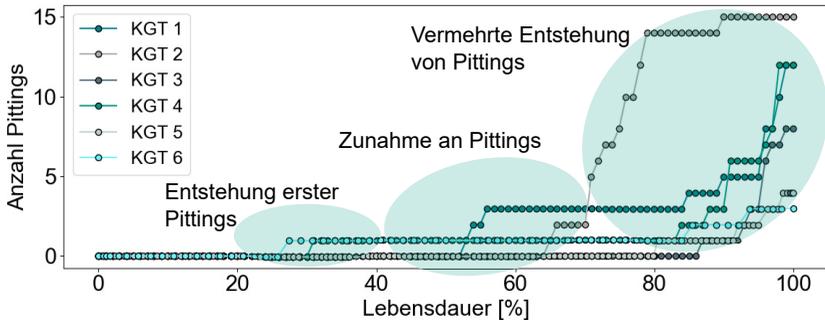


Abbildung 6-12 Anzahl an Pittings auf den Oberflächen der Spindeln über die Lebensdauer hinweg.

Dieser Phase folgt eine zweite Phase, beginnend bei ca. 60 bis 70 % der Lebensdauer, in der eine Zunahme in der Anzahl der Pittings beobachtet werden kann. Dies deckt sich mit den Erkenntnissen von (Schopp 2009), der ebenfalls bei Erreichen von ca. 60 % der Lebensdauer erste Schädigungen vermutet. Dieser zweiten Phase folgt eine dritte Phase bei Erreichen von > 80 % der Lebensdauer, in der eine vermehrte Zunahme an Pittings beobachtet werden kann. Diese starke Zunahme deutet das bevorstehende Lebensdauerende der Komponenten an und lässt sich mit den sich gegenseitig beeinflussenden Verschleißmechanismen erklären. Als übergreifendes Verhalten ist zu beobachten, dass der initialen Entstehung von Pittings eine Phase folgt, in der kein erneutes Entstehen beobachtet werden kann. Die Länge dieser Phase schwankt, ebenso wie der Entstehungszeitpunkt der ersten Pittings, zwischen den Komponenten stark. Insgesamt durchlaufen also alle Komponenten die beschriebenen drei Phasen, wobei sich die Übergänge zwischen den Phasen komponentenabhängig unterscheiden.

Als Fazit zur Entwicklung der Anzahl an Pittings lässt sich festhalten, dass die Entwicklung einen dreiphasigen Verlauf beschreibt, wobei Pittings vereinzelt schon in einem frühen Stadium entstehen.

Die reine Anzahl an Schädigungen lässt allerdings nur eine bedingte Aussage über die tatsächliche Schädigung der Komponenten zu, da keine Aussage über die Ausdehnung der Pittings gemacht wird. Aus diesem Grund wird das Wachstum der Schädigungen nachfolgend näher betrachtet.

6.1.4.2 Fläche

Abbildung 6-13 stellt die Gesamtfläche der Pittings auf den Spindeln dar. Hier ist ebenfalls ein dreiphasiger Verlauf zu beobachten. Die Schädigungsfläche steigt bis ca. 60 % der Lebensdauer kaum an und nimmt dann kontinuierlich leicht zu, bis ca. 80 % der Lebensdauer erreicht sind. Danach folgt eine Phase der starken Zunahme der Schädigungsfläche. Die Entwicklung entspricht damit qualitativ der Entwicklung, wie sie von (Schopp 2009) beschrieben wurde. Dieses Verhalten ist durch Betrachten des Wachstums der einzelnen Schädigungen (Abbildung 6-14) zu erklären.

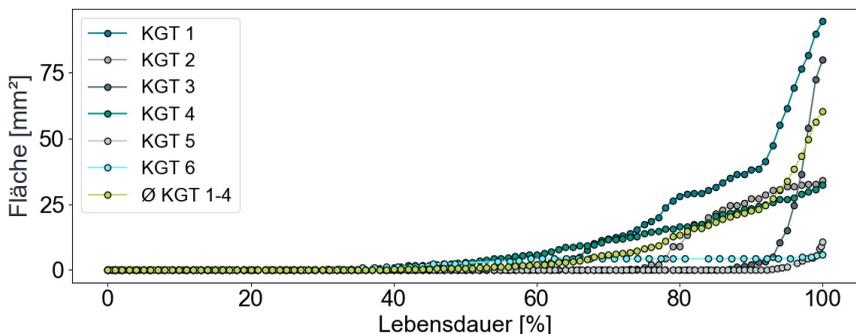


Abbildung 6-13 Entwicklung der Gesamtfläche der Pittings auf den Spindeloberflächen über die Lebensdauer der Komponenten.

Es ist ersichtlich, dass vereinzelte Pittings bereits ab ca. 30 % der Lebensdauer entstehen, wie dies auch in der Entwicklung der Anzahl an Pittings ersichtlich war. Die Entwicklungen folgen ebenfalls dem beschriebenen dreiphasigen Verlauf. Bei Überschreiten von ca. 80 % der Lebensdauer ist eine starke Zunahme an Pittings zu erkennen. Zwar ist zu diesem Zeitpunkt jedes der entstandenen Pittings in seiner Größe noch begrenzt, die starke Zunahme in der Anzahl führt jedoch zu einer starken Zunahme der Gesamtfläche der Pittings. Die Gesamtfläche der Schädigungen ergibt sich als Summe der Einzelschädigungen. Sind dabei eine große Anzahl an Pittings mit geringer Fläche und nur wenige Pittings mit größerer Fläche vertreten, so erfährt die Summe der kleinen

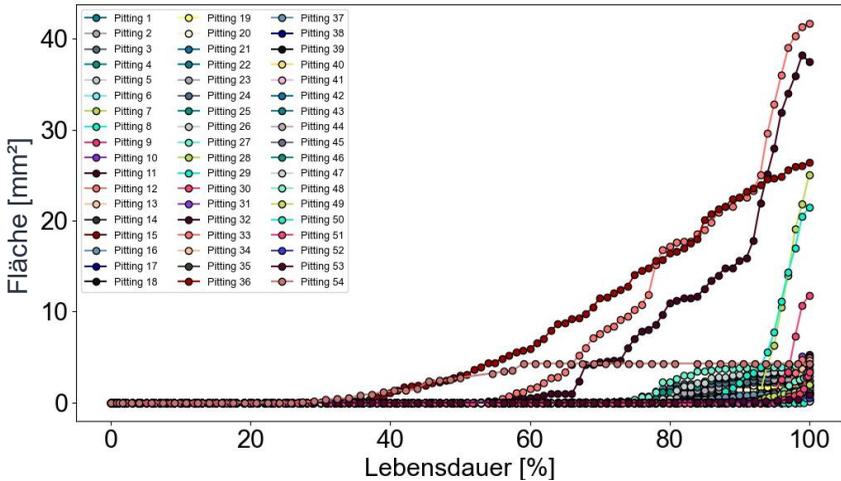


Abbildung 6-14 Entwicklung der Einzelflächen von Pittings auf den Spindeloberflächen.

Pittings eine Zunahme an Gewicht und führt zu einer starken Zunahme der Schädigungsfläche. Es zeigt sich, dass der Entstehungszeitpunkt der Pittings durch die Lebensdauer der Komponenten begrenzt ist. Je näher am Lebensdauerende der Komponente ein Pitting entsteht, desto weniger Zeit bleibt für das Wachstum, weshalb nicht bei allen Pittings der beschriebene Verlauf beobachtet werden kann. Es ist allerdings zu beobachten, dass das Wachstum reziprok zum Entstehungszeitpunkt ist. Besonders spät entstandene Pittings wachsen besonders schnell, wobei die angedeutete zweite Phase im Verhältnis zu früher entstandenen Pittings nur sehr kurz ist und diese direkt in ein sehr starkes Wachstum übergehen. Dieser Effekt lässt sich analog zur überproportionalen Entstehung von Pittings gegen Ende der Lebensdauer mit sich gegenseitig verstärkenden Verschleißmechanismen erklären.

Zusammenfassend lässt sich die Entwicklung der Gesamtfläche der Pittings mit einem exponentiellen Wachstum beschreiben. Dieses exponentielle Wachstum lässt sich durch die Entwicklung der Einzelflächen und das vermehrte Entstehen von Pittings gegen Ende der Lebensdauer der Komponenten erklären. Dies lässt sich wiederum auf sich gegenseitig beeinflussende und verstärkende Verschleißmechanismen zurückführen.

Die Entwicklung der Flächen beschreibt zusammen mit der Entwicklung der Anzahl an Pittings relevante Aspekte der Schädigungsentwicklung am KGT. Zu einer vollständigen Analyse soll weiterhin die Entwicklung der Ausdehnungen der Pittings untersucht werden.

6.1.4.3 Ausdehnung

In diesem Abschnitt werden zunächst die tangentielle und axiale Gesamtausdehnung betrachtet. Daran schließt die Betrachtung der axialen und tangentialen Ausdehnungen der einzelnen Pittings an.

Die Gesamtausdehnung der Pittings in tangentialer Richtung (siehe Abbildung 6-15) weist einen ähnlichen Verlauf wie das Flächenwachstum auf. Es ist ersichtlich, dass Pittings erstmals bei ca. 30 % der Lebensdauer entstehen und dann dem bekannten Wachstumsverlauf folgen. Auch hier ist ersichtlich, dass die Steigung der tangentialen Ausdehnung mit zunehmender Lebensdauer steigt. Dies ist auch der Fall, wenn wie bei KGT 5 erst spät ein Wachstum zu verzeichnen ist.

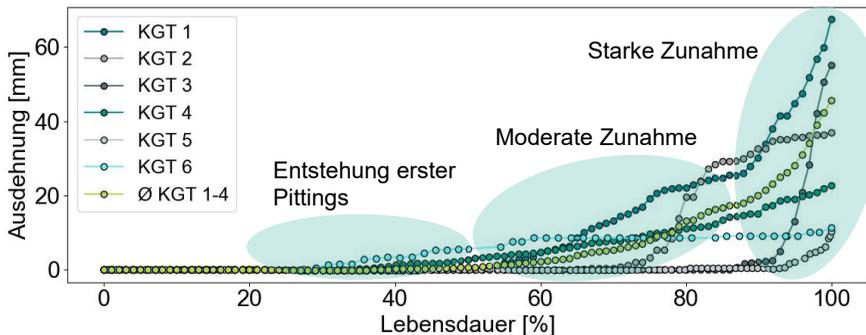


Abbildung 6-15 Tangentielle Gesamtausdehnung der Schädigungen auf den Komponenten.

Wird zusätzlich die axiale Gesamtausdehnung betrachtet (siehe Abbildung 6-16), so zeigt sich ein analoges Bild mit der Ausnahme, dass die Gesamtausdehnung auf KGT 2 in eine Sättigung zu laufen scheint. Ein weiterer Unterschied ist die maximale summierte Ausdehnung der Pittings. Sie beträgt bei der tangentialen Gesamtausdehnung in Summe ca. 60 mm, wobei sie bei der axialen Ausdehnung um den Faktor 3 geringer ist.

Dieses begrenzte Wachstum lässt sich durch den limitierenden Faktor des Kugeldurchmessers auf das Wachstum in axialer Richtung erklären (siehe Abbildung 6-16). Dieses Bild wird klar, wenn die tangentialen und axialen Ausdehnungen der einzelnen Pittings betrachtet werden. Abbildung 6-17 zeigt die tangentialen Ausdehnungen der Pittings auf den Spindeln.

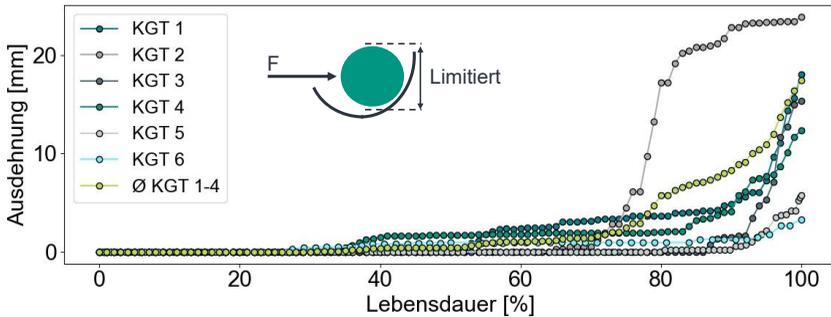


Abbildung 6-16 Axiale Gesamtausdehnung der Schädigungen auf den Komponenten.

Bei einer erreichten Lebensdauer zwischen 60 und 80 % entsteht eine Vielzahl an Oberflächendefekten. Bei > 80 % beginnen diese stärker zu wachsen. Dies führt im Resultat analog zur Flächenentwicklung dazu, dass am Ende der Lebensdauer der Komponente ein exponentielles Wachstum in der tangentialen Gesamtausdehnung zu beobachten ist.

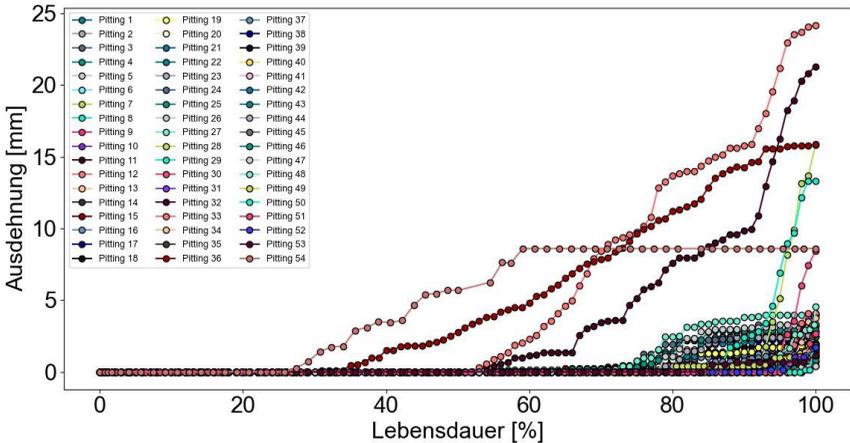


Abbildung 6-17 Entwicklung der tangentialen Ausdehnungen von Pittings auf den Spindeloberflächen.

Das Wachstum der Pittings in axialer Richtung ist in Abbildung 6-18 dargestellt. Es beginnt mit einer Phase des starken Wachstums und wird dann gefolgt von einer Phase des abflachenden linearen Wachstums.

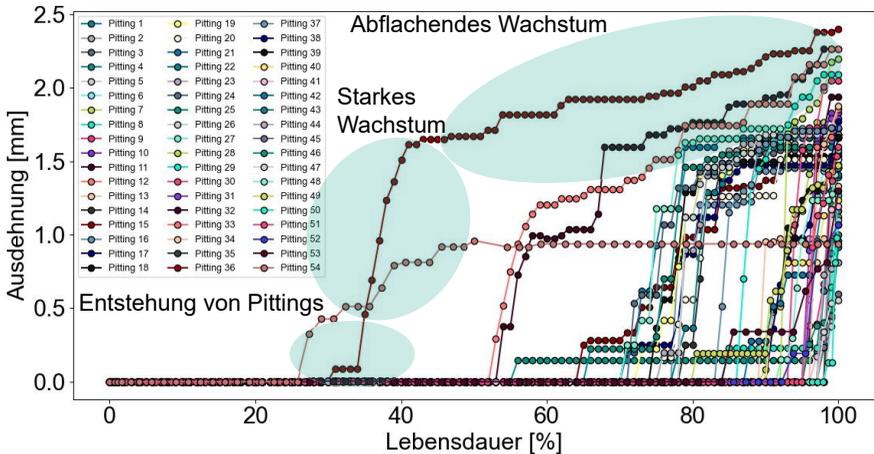


Abbildung 6-18 Entwicklung der axialen Ausdehnungen von Pittings auf den Spindeloberflächen.

Dieses Verhalten lässt sich mit der Kinematik und den am KGT wirkenden Lasten erklären. Bei der Entstehung von Pittings ist die Schadensfläche gering. Die über die Wälzkörper auf die Spindel wirkenden Kräfte bleiben nahezu konstant. Aufgrund der Kugelform der Wälzkörper ergibt sich somit initial eine hohe Kraft pro Flächeneinheit (= Druck), woraus eine hohe Materialbelastung resultiert. Je größer die Schadensfläche, desto geringer ist der Druck, der auf die nicht geschädigten Spindelbereiche wirkt – dadurch kommt es zu einer langsameren axialen Ausdehnung, die dann durch den Kugeldurchmesser begrenzt wird. Dies ist qualitativ in Abbildung 6-19 dargestellt. Aufgrund der vorangegangenen Beobachtungen wird vermutet, dass dieser Effekt auch durch Fremdstoffe im Schmiermedium sowie durch vorgeschädigte Wälzkörper beeinflusst wird und daher weiter untersucht werden muss. Das Phänomen des abflachenden Wachstums in axialer Richtung bei konstanten Lastbedingungen kann allerdings hierdurch erklärt werden.

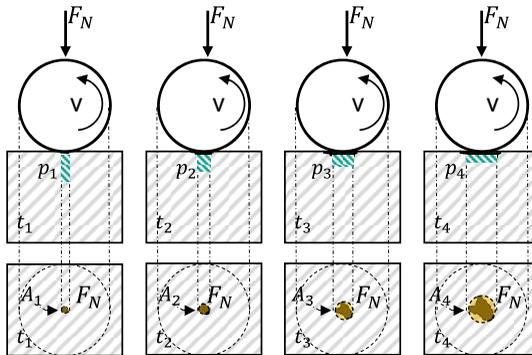


Abbildung 6-19 Qualitative Darstellung des wirkenden Drucks auf die nicht geschädigten Spindelbereiche.

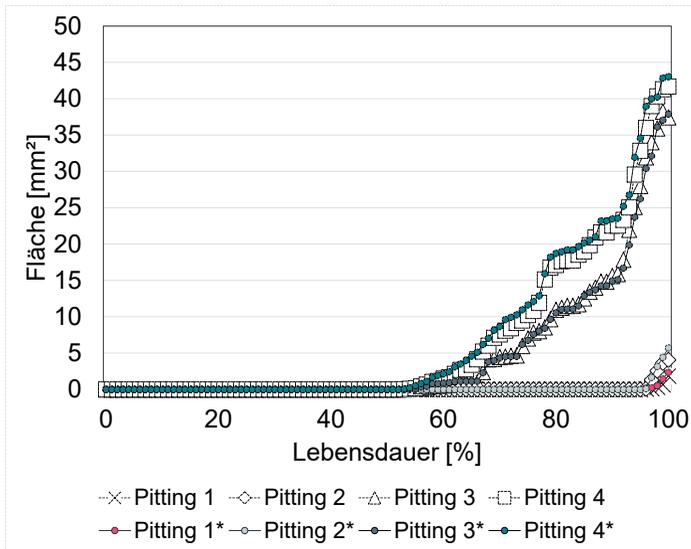


Abbildung 6-20 Gegenüberstellung der Fläche von vier zufällig ausgewählten Pittings mit deren Annäherung als Ellipse. Pitting n^* ist hierbei durch die Formel $\frac{Tangential_{pitting}}{2} * \frac{Axial_{pitting}}{2} * \pi$ angenähert.

Dieser Analyse folgend kann die Form des Pittings aufgrund der axialen Begrenzung als elliptisch angenommen werden. Damit ist $A_{Pitting} = a * b * \pi$ mit a, b , den Halbachsen der Ellipse. Zu Beginn des Wachstums kann die Länge der Haupt- und Nebenachse als gleich angenommen werden. Dies ist auch in den Verläufen in Abbildung 6-11 ersichtlich. Nach Annahme, dass a, b gleich sind, ergibt sich: $A_{Pitting} = \pi r^2$. Nachdem die axiale Ausdehnung durch die Kugeldurchmesser begrenzt ist, ergibt sich mit a als Konstante: $A_{Pitting} = c * b * \pi$. Danach wird die Fläche des Pittings hauptsächlich von der tangentialen Ausdehnung des Pittings bestimmt. Dies ist auch in Abbildung 6-20 verdeutlicht, in der für vier zufällig ausgewählte Pittings die Fläche der Pittings $A_{Pitting}$ deren Annäherung über eine Ellipse $\frac{Tangential_{pitting}}{2} * \frac{Axial_{pitting}}{2} * \pi$ gegenübergestellt ist. Die Ergebnisse zeigen eine nahezu perfekte Übereinstimmung der beobachteten und angenäherten Flächenverläufe. Somit ist davon auszugehen, dass das Pitting als Ellipse modelliert werden kann.

Im Ergebnis können die tangentialen und axialen Gesamtausdehnungen der Pittings auf der Spindel ebenfalls mit einem dreiphasigen Verlauf beschrieben werden. Dabei unterscheiden sich die Verläufe dadurch, dass das Wachstum der Pittings in axialer Richtung durch den Wälzkörperdurchmesser begrenzt ist. Durch diese Begrenzung lässt sich die Fläche eines fortgeschrittenen Pittings als Ellipse annähern, was durch einen Abgleich mit der Fläche von Pittings gezeigt wurde.

Im Folgenden werden die Erkenntnisse der empirischen Analyse der Entwicklung der Oberflächenzerrüttung zusammengefasst und es wird ein Fazit gezogen.

6.1.4.4 Fazit zu der empirischen Analyse der Entwicklung von Pittings

Die folgenden Kernaspekte fassen die empirische Analyse der Entwicklung von Pittings zusammen.

Die empirische Analyse von Pittings zeigt, dass in der Ausdehnung der Pittings der von (Schopp 2009) vermutete dreiphasige Verlauf zu erkennen ist. Hierbei sei angemerkt, dass – anders als von (Schopp 2009) prognostiziert – erste Schädigungen bereits ab ca. 30 % der Lebensdauer auftreten.

Der Flächenverlauf kann aufgrund der Anschaulichkeit zwar durchaus wie auch von (Schopp 2009), (Hennrich 2013) und (Spohrer 2019) als dreiphasiger Verlauf interpretiert werden, die Untersuchungen zeigen aber, dass der tatsächliche Verlauf anstelle eines dreiphasigen Verlaufes ebenso durch einen exponentiellen Verlauf angenähert werden kann. Dies deckt sich auch mit den Messdaten von (Walther 2011). Die Vermutung, dass bestehende Defekte die Entstehung von neuen Defekten fördern, wird vor allem durch die stark ansteigende Zahl an Schädigungen gegen Ende der Lebensdauer der Komponenten bestärkt.

Die Flächenausdehnung der Pittings lässt sich über die Formel für eine Ellipse annähern, wobei die Fläche bei Entstehen der Pittings zunächst kreisförmig ist und dann in axialer Richtung schnell wächst. Da das Wachstum in axialer Richtung durch die Kugeldurchmesser begrenzt ist, stellt die tangentiale Ausdehnung die verschleißbestimmende Größe dar.

Aus den Versuchen wird ersichtlich, dass Pittings in den Bilddaten unabhängig vom Ort und Entstehungszeitpunkt zuverlässig abgebildet sind und identifiziert werden können. Der Entstehungszeitpunkt variiert in den Versuchen zwischen 30 und 90 %, wobei das

Wachstum der Schädigungen als zunehmend stärker beobachtet wurde, je später diese entstanden sind.

Aus den Analysen geht hervor, dass der Schädigungszustand der Spindeloberfläche mit Blick auf die Oberflächenzerrüttung in den Bilddaten genau abgebildet werden kann. Dies ermöglicht bei geeigneter Definition der 100 % Lebensdauer eine maximale Ausnutzung des Abnutzungsvorrates. Nach Berechnungen der (A. Mannesmann Maschinenfabrik GmbH, 2022)²¹ beträgt die Lebensdauer eines durchschnittlichen KGT im Einsatz mehr als 20 000 Stunden bzw. 833 Tage. Damit stehen selbst bei Erreichen von 90 % der Lebensdauer noch 83 Tage für den Tausch der Komponenten zur Verfügung. Bei einer großzügig ausgelegten Lieferzeit von sechs Wochen stehen noch 41 Tage zur Umsetzung der Wartungsmaßnahmen zur Verfügung.

Wird die Oberflächenzerrüttung als lebensdauerbegrenzendes Verschleißmerkmal identifiziert, so ist es für eine maximale Ausnutzung des Abnutzungsvorrates relevant, dass die Definition der 100 % Lebensdauer die aktuell vorliegende Schwere der Oberflächenzerrüttung am KGT berücksichtigt. Dieser Aspekt soll im folgenden Abschnitt aufgegriffen werden.

6.1.5 Normbasierte Präzisierung der Verschleißquantifizierung

Im vorherigen Abschnitt konnte gezeigt werden, dass die Entwicklung von Pittings in den Bilddaten zuverlässig abgebildet werden kann. Im Folgenden soll die Herausforderung der Definition einer Abnutzungsgrenze auf Basis der Bilddaten diskutiert werden. Hierzu wird auf die Vorarbeit von (Münzing 2017) und die (Deutsches Institut für Normung 2010a) zurückgegriffen und deren Erkenntnisse werden adaptiert.

Basierend auf den in Abschnitt 2.2.3 dargestellten Grundlagen zur Bestimmung der Abnutzungsgrenze für den Betrieb eines KGT und den Diskussionen zur praktischen Ergänzung der Lebensdauer des KGTs aus Abschnitt 2.2.3.3 wird zunächst ein Vorgehen zur Definition einer Abnutzungsgrenze auf Basis der Oberflächenzerrüttung vorgeschlagen, die den Zeitpunkt zur Instandhaltung möglichst optimal und vor allem anwendungsspezifisch ermöglichen soll. Ziel ist es, Anwenden die Definition einer Abnutzungsgrenze zu ermöglichen, die 1. die Oberflächenzerrüttung als lebensdauerbestimmendes Verschleißmerkmal integriert, 2. quantifizierbar ist und eine möglichst hohe

²¹ A. Mannesmann Maschinenfabrik GmbH (2021), *Kugelgewindetriebe für Einsätze im Maschinenbau, bei denen Genauigkeit zählt*. <https://www.amannesmann.de/produkte/kugelgewindetriebe/> [20.10.2021].

Verfügbarkeit der Komponente sicherstellt und 3. ausreichend Zeit zur Instandhaltung der Komponente gewährleistet (siehe Abnutzungsgrenze 2 in Abbildung 2-12, Abschnitt 2.2.3).

Wie dargestellt wurde, führt die Verwendung der Lebensdauer nach (Deutsches Institut für Normung 2011b) dazu, dass ein Großteil der KGT ausgetauscht wird, obwohl deren mechanische Funktionsfähigkeit noch gegeben ist. Die Ursache hierfür ist, dass neben dem mechanischen Versagen des Systems keine für alle Anwendungen universell gültige Definition der Lebensdauer bestehen kann, die darüber hinaus die Bestimmung einer optimalen Abnutzungsgrenze erlaubt. Da die Lebensdauer des KGTs aber stark von dessen Betrieb abhängig ist und individuelle Anforderungen an ihn gestellt werden, sollten Unternehmen dazu in der Lage sein, diese Größe individuell zu definieren.

Im Folgenden wird eine Definition der Lebensdauer des KGTs vorgeschlagen, welche die obigen drei Anforderungen erfüllt.

In den Versuchen konnte gezeigt werden, dass es möglich ist, die Ausdehnungen von Pittings auf dem KGT zuverlässig in den Bilddaten zu quantifizieren. Hierüber erhalten die Anwender eine messbare Größe zur Definition eines anwendungsspezifischen Abnutzungsvorrates. Damit sind die Anforderungen 1 und 2 weitestgehend erfüllt.

Weiterhin definiert (Deutsches Institut für Normung 2010a) die Lebensdauer von Profilschienenführungen. Diese Norm wurde bereits von (Münzing 2017) zur Definition der Lebensdauer des KGTs verwendet. Hierbei wird das Ausfallkriterium nach Abbildung 6-21 mit $d_s \geq D_w * 0.3$ berechnet. Die Lebensdauer der Komponente ist erreicht, wenn ein Pitting das 0.3-fache des Kugeldurchmessers überschreitet. Diese Größe ist damit nicht absolut, sondern von den Komponenteneigenschaften abhängig, was sie flexibler macht. Die Versuchsergebnisse und der Stand der Forschung zeigen, dass es möglich ist, einen KGT über das initiale Entstehen von Pittings hinaus zu betreiben, und dass diese Wahl damit zielführend ist. Damit ist auch die Forderung 2 vollständig erfüllt.

Um sicherzustellen, dass die Komponente rechtzeitig getauscht werden kann, wird vorgeschlagen, die Definition um den Skalierungsfaktor α zu erweitern. Dieser Faktor erlaubt es, die zulässige Ausdehnung anwendungsspezifisch anzupassen und so eine optimale Abnutzungsgrenze zu definieren. Damit ist auch die Forderung 3 erfüllt.

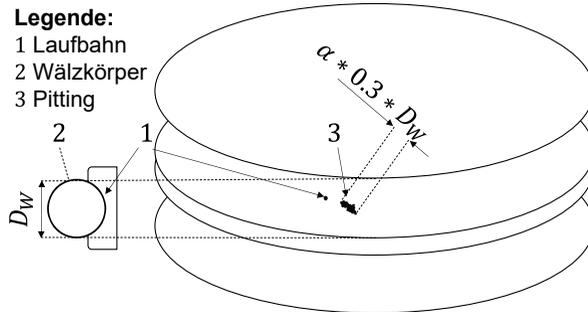


Abbildung 6-21 Vorgeschlagenes Ausfallkriterium für einen KGT angelehnt an das Ausfallkriterium einer Linearschienenführung nach (Deutsches Institut für Normung 2010a, S. 13).

Zusammenfassend wurde in diesem Abschnitt angelehnt an eine bestehende Norm die Lebensdauer des KGTs definiert, sodass die Oberflächenzerrüttung als Merkmal zur Lebensdauerberechnung verwendet werden kann. Durch die Fähigkeit, die Oberflächenzerrüttung in den Bilddaten zuverlässig zu quantifizieren, wird eine Möglichkeit zur zuverlässigen Bestimmung des Verschleißzustandes des KGTs geschaffen. Einschränkend muss jedoch angemerkt werden, dass hierzu die Oberflächenzerrüttung als lebensdauerbegrenzendes Verschleißmerkmal auftreten muss. Dies muss anwendungsindividuell entschieden werden. Die Definition verbietet es aber nicht, eine weitere lebensdauerbestimmende Größe mit hinzuzuziehen.

Im folgenden Abschnitt werden die Ergebnisse des bildbasierten Ansatzes im Kontext bestehender Arbeiten am KGT diskutiert.

6.1.6 Diskussion der Erkenntnisse im Kontext bestehender Arbeiten

Von besonderer praktischer Relevanz für Anwender ist, dass Fehlstellen möglichst eindeutig identifiziert werden können, um rechtzeitig und zuverlässig Wartungsmaßnahmen einzuleiten. Weiterhin muss ein System von den Umgebungs- und Betriebsbedingungen sowie der Bauart der Maschine möglichst unabhängig sein. Ein weiterer relevanter Aspekt ist die benötigte Sensorik. Neben der Fähigkeit eines Systems, Verschleiß rechtzeitig für die Umsetzung von Instandhaltungsmaßnahmen zu identifizieren, sind als letztes Kriterium eine exakte Quantifizierung sowie Prognose des Verschleißverhaltens hilfreich.

Hinsichtlich dieser Kriterien werden in diesem Abschnitt der vorgestellte Ansatz und die Ergebnisse der empirischen Analyse im Kontext bestehender Arbeiten diskutiert.

Zunächst lässt sich festhalten, dass der vorgestellte Ansatz den ersten direkten Ansatz zur Quantifizierung von Verschleiß und speziell Oberflächenzerrüttung auf der KGT-Spindel darstellt und damit genauer als bestehende Arbeiten ist. Alle bestehenden Arbeiten identifizieren Verschleiß über indirekte Sensorgrößen. Besonders die auf Körperschall basierenden Ansätze von (Schopp 2009), (Schmid & Gädeke et al. 2010), (Hennrich 2013), (Helwig 2018) und (Xi & Kehne et al. 2020) zeigen auf, dass Verschleiß im Signal erst bei Überschreiten von ca. 60 % der Komponentenlebensdauer zuverlässig identifiziert werden kann. Darüber hinaus stellt (Hillenbrand 2021) fest, dass hochfrequenter Körperschall nur eingeschränkt zur Bestimmung des Verschleißfortschrittes von Oberflächenzerrüttung genutzt werden kann²², da das Signal sehr stark von den aktuell wirkenden Betriebsbedingungen, der Platzierung des Sensors sowie der Bauform der Maschine abhängig ist und sich damit ein punktueller, ortsaufgelöster Schaden nur schwer identifizieren lässt (Hillenbrand 2021).

Aus den Untersuchungen von (Schopp 2009) ist bekannt, dass der Verschleißverlauf unabhängig von den Lastkollektiven und den Betriebsbedingungen stets den von ihm beschriebenen dreiphasigen Verlauf annimmt. Die Phasen (die Kurve) werden lediglich gestreckt oder gestaucht, was sich auch mit den hier vorliegenden Ergebnissen deckt. Aus diesem Grund ist der präsentierte Ansatz geeignet, da mittels der Bilddaten Fehlstellen unabhängig von den vorliegenden Betriebsbedingungen und Lastkollektiven identifiziert werden können. Das Bildsignal ist im Gegensatz zu indirekten Signalen (bspw. Motorstrom und (Körper)Schall) von den Lastkollektiven und Betriebsbedingungen unabhängig. Es wurde weiterhin gezeigt, dass sich die beobachteten Verläufe zwar mit den Erkenntnissen der bestehenden Arbeiten decken, Fehlstellen jedoch schon weit vor Erreichen der 60 % Lebensdauer zuverlässig identifiziert werden konnten. Die Unsicherheit im Bereich vor Erreichen der 60 % Lebensdauer, in dem (Schopp 2009) eine Unsicherheit in den Messdaten feststellte, kann mit dem vorgestellten Verfahren verrin-

²² Hochfrequenter Körperschall ist nach (Hillenbrand 2021) allerdings besonders gut geeignet, um Abrasivverschleiß zu identifizieren. Dies unterstreicht die Eignung bestimmter Sensorsysteme zur Identifikation bestimmter Verschleißerscheinungsformen und Verschleißmechanismen.

gert werden. Weiterhin nehmen die Unsicherheiten in den Messwerten des hier vorgestellten Systems, anders als bei (Schopp 2009), mit zunehmender Laufleistung nur minimal zu.

Im Gegensatz zu den bisherigen sensorbasierten Ansätzen zeichnet sich das bildbasierte Verfahren durch ein Sensorsignal aus, das die Entwicklung des Schadens direkt beobachten kann. Dies ist auch in Abbildung 6-22 angelehnt an (Walther 2011) qualitativ dargestellt. Die bisherigen sensorbasierten Ansätze sind in der Lage, eine Verschleißfolge in Form einer Auswirkung auf das Betriebsverhalten der Komponente oder der Maschine indirekt zu identifizieren, indem bspw. eine Änderung im Schall- oder Stromsignal festgestellt wird. Dies unterstreicht die Vorteilhaftigkeit hinsichtlich der Genauigkeit des hier vorgestellten bildbasierten Ansatzes.

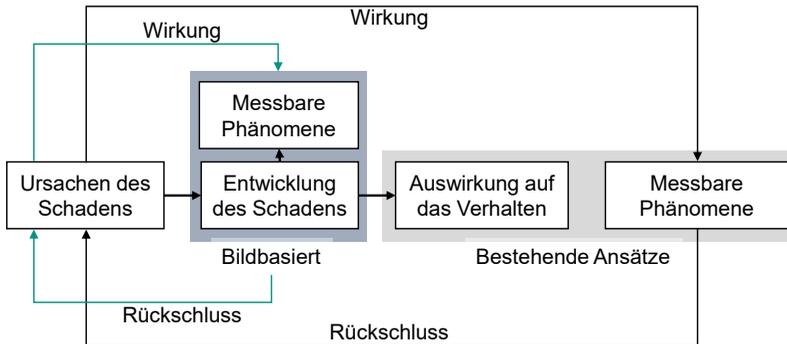


Abbildung 6-22 Qualitative Darstellung der Ursache-Wirkungs-Kette bei der Messung von Oberflächenzerrüttung am KGT.

Die Ansätze von (Imiela 2006), (Veith & Zimmermann et al. 2020), (Möhring & Bertram 2012) sowie (Herder 2013) unterscheiden sich insofern von dem präsentierten Ansatz, als deren primäres Ziel die Quantifizierung des Vorspannungsverlustes an KGT ist. Damit ist deren Ziel nicht die Beobachtung von Verschleiß, sondern die Beobachtung einer Verschleißfolge. Die Messung dieser Verschleißfolge erlaubt es zu entscheiden, ob eine Komponente noch die geforderten technischen (Steifigkeits-)Eigenschaften erfüllt. Dies ist möglich, da in diesem Fall die Verschleißfolge gleichbedeutend mit den technischen Eigenschaften des Systems ist. Zwar muss auch hier definiert werden, welche Vorspannung zulässig ist, bei den sonstigen Ansätzen muss jedoch zusätzlich noch die Verbindung zwischen einem Verschleißmerkmal und den technischen Anforderungen

an das System geschaffen werden. Unter diesem Aspekt ist die Überwachung der Vorspannung zielführend. Limitierend ist dagegen, dass die Systeme durch deren Spezialisierung den Vorspannungsverlust zwar sehr gut quantifizieren können, die Relevanz des Vorspannungsverlustes allerdings anwendungsindividuell bewertet werden muss und die Systeme darüber hinaus nur eingeschränkt für weitere Verschleißgrößen anwendbar sind. In dieser Hinsicht ergeben sich für Systeme zur Überwachung der Vorspannung dieselben Limitationen wie für den vorgestellten bildbasierten Ansatz. Allerdings haben die Ansätze, welche den Vorspannungsverlust im System messen, besonders bei der Quantifizierung der Oberflächenzerrüttung Nachteile gegenüber dem hier vorgestellten Ansatz. Ein interessanter Aspekt und eine mögliche Erweiterung dieser Dissertation sind die Verbindung zwischen der Oberflächenzerrüttung und des Vorspannungsverlustes sowie die Integration in ein Gesamtmodell.

Vielsprechende Ansätze zur Identifikation von Verschleiß sind jene, die Aussagen über den vorliegenden Verschleiß ohne zusätzliche Sensorik treffen können, da sie besonders aufwandsarm implementiert werden können. Dabei bieten sich Ansätze an, die steuerungsinterne Größen wie Motorstrom oder Motormoment verwenden (Walther 2011), (Henrich 2013) (lediglich der belastungsorientierte Teil), (Yagmur 2014), (Verl & Heisel et al. 2009), (Riaz & Shah et al. 2020), (Riaz & Shah et al. 2021). Im Gegensatz zum vorgestellten System besteht der Nachteil der präsentierten Verfahren darin, dass diese den Verschleiß zum einen nur indirekt messen können und dass zum anderen bekannt ist, dass steuerungsinterne Größen stark durch Lastkollektive und Betriebsbedingungen beeinflusst werden, was eine möglichst frühzeitige Abgrenzung von Verschleiß gegenüber dem Rauschen des Signals erschwert.

Weitere sensorlose Ansätze sind jene, die auf Simulationen und Modellen basieren (Imiela 2006), (Stockinger 2011), (Broos 2012), (Münzing 2017), wobei kritisch betrachtet werden muss, bei welchen Ansätzen zusätzliche Sensorik zur Parametrisierung der Modelle notwendig ist. Der Nachteil dieser Ansätze ist, dass sie nur eingeschränkt auf nicht deterministische Größen wie Materialschwankungen, Fertigungs- und Einbaufehler sowie betriebsbedingte Effekte wie Verschleißpartikel und Verschmutzung reagieren können. Müssen die Methoden dann mittels zusätzlicher Sensorik parametrisiert werden, so stellt sich die Frage nach der Effizienz der Ansätze. Die Abhängigkeit von nicht deterministischen Größen kann bei dem vorgestellten Ansatz weitestgehend ausgeschlossen werden – es wird immer der aktuell vorliegende Verschleißzustand angezeigt. Weiterhin parametrisiert sich der hier vorgestellte Ansatz im Betrieb selbst, da die

Prognosefunktion (siehe Kapitel 7) immer auf Basis des aktuell vorliegenden Verschleißzustandes gebildet wird.

Neben der direkten Identifikation von Verschleiß stellt auch die positionsgenaue Auflösung einen Vorteil des gezeigten Systems dar. Im Gegensatz zu den bestehenden Arbeiten, die weitestgehend keine höhere Auflösung als den Gesamtzustand der Spindel erlauben, können mit dem betrachteten Ansatz Fehler individuell betrachtet werden. Damit kann eine höhere Auflösung erzielt und die Entwicklungen einzelner Fehler können als Größe für die Beschreibung des Verschleißzustandes verwendet werden.

Weiterhin erlauben die Bilddaten eine einfache Interpretation des Verschleißzustandes durch die Domänenexpert*innen. Damit ist eine einfache und schnelle Analyse des Verschleißzustandes der Spindel möglich. Durch die Möglichkeit, Fehlstellen direkt in den Bilddaten zu identifizieren, erschließt sich das Potenzial, den Bilddaten direkt Klassen zuzuweisen und diese für das Training von Verfahren des Maschinellen Lernens zu verwenden. Mithilfe der Bilddaten kann also entweder eine manuelle oder automatisierte, auf Maschinellern basierende Inspektion der Spindeln durchgeführt werden. Die Abgrenzung des präsentierten Ansatzes von bestehenden Arbeiten ist in Tabelle 6-2 zusammengefasst. Es kann festgehalten werden, dass nahezu alle bisherigen Ansätze ebenfalls in der Lage sind, das Lebensdauerende des KGTs auf unterschiedliche Arten zu erkennen, und hierbei spezifische Stärken aufweisen. Der präsentierte Ansatz soll also nicht als grundsätzlich substituierender Ansatz für alle bestehenden Systeme gewertet werden. Allerdings hat der hier vorgestellte Ansatz gegenüber den bestehenden Systemen den Vorteil, dass mittels der Bilddaten speziell die Oberflächenzerrüttung direkt, ortsaufgelöst und für die Domänenexpert*innen einfach interpretierbar dargestellt werden kann. Weiterhin kann die Genauigkeit des Ansatzes aufgrund der frühzeitigen Sichtbarkeit von Pittings in den Bilddaten als genauer eingeschätzt werden. Gegenüber bestehenden Ansätzen können erstmals Verfahren des Maschinellen Lernens auf Basis dieser Bilddaten trainiert werden. Grundsätzlich kann beobachtet werden, dass sich ein Trade-off zwischen Genauigkeit und Universalität der Systeme einstellt.

Tabelle 6-2 Gegenüberstellung der Spezifikationen der bisherigen Arbeiten und des bildbasierten Ansatzes.

Kriterien Sensorgrößen	Rechtzeitiger Austausch der Komponente	Sensorlose Integration	Störanfälligkeit gegenüber Umgebungs- und Betriebsbedingungen	Interpretierbarkeit des Signals	Positionsgenaue Auflösung des Signals	Vielseitigkeit des Systems
Bildbasiert (Schlagenhauf & Landwehr 2021) (Schlagenhauf 2021) (Schlagenhauf & Hillenbrand et al. 2019)	1	1	1	1	1	1
Schallbasiert (Xi & Kehne et al. 2020), (Helwig 2018), (Henrich 2013), (Schmid & Gädeke et al. 2010), (Schopp 2009)	-1	1	-1	-1	-1	2
Vorspannungsbasiert (Veith & Zimmermann et al. 2020), (Herder 2013), (Möhring & Bertram 2012), (Imiela 2006)	1	1	-1	1	-2	-1
Steuerungsinterne Größen (Riaz & Shah et al. 2021), (Riaz & Shah et al. 2020), (Spohrer 2019), (Yagmur 2014), (Henrich 2013), (Verl & Heisel et al. 2009)	-1	3	3	-1	-2	3
Modellbasiert	-1	2	3	1	-2	2

(Münzing 2017), (Broos 2012), (Stockinger 2011), (Imiela 2006)						
Temperaturbasiert (A. Mannesmann Maschinenfabrik GmbH, 2022) ²³	-1	1	-1	1	-1	2
-2 Nicht fähig -1 Nachteile ggü. Referenz 1 Referenz, bildbasiert 2 Vorteile ggü. Referenz 3 große Vorteile ggü. Referenz						

Zusammenfassend wurden im vorherigen Abschnitt die Erkenntnisse und Eigenschaften des bildbasierten Systems im Kontext bestehender Arbeiten diskutiert. Es wurden die besonderen Stärken gegenüber bestehenden Ansätzen vor allem mit Blick auf die Genauigkeit und Interpretierbarkeit der Oberflächenzerrüttung aufgezeigt. Werden die Punktzahlen in Tabelle 6-2 aufsummiert, so ergibt sich ein genereller Vorteil des bildbasierten Ansatzes gegenüber den bestehenden Systemen. Dennoch sei hier betont, dass Sensorsysteme maßgeblich nach dem zu erwartenden Schadensmechanismus und Schadensort ausgewählt werden müssen und das vorgestellte System nicht geeignet ist, bestehende Systeme vollständig zu substituieren, sondern vielmehr dazu, diese zu erweitern.

Im folgenden Abschnitt wird auf die visuellen Charakteristika der Verschleißerscheinungsform Pitting eingegangen und aufgezeigt, wie diese in den Bilddaten beobachtet werden können. Es wird damit eine Grundlage für die Analyse und Klassifikation der Oberflächenzerrüttung geschaffen.

6.1.7 Visuelle Charakteristika der Verschleißerscheinungsform Pitting

Die Untersuchung visueller Charakteristika der Pittings in den Bilddaten erhöht das Verständnis über das Verschleißmerkmal Pitting und bildet die Grundlage für die Untersuchung von Systemen zur automatisierten Identifikation von Pittings in den Bilddaten.

In Abschnitt 2.4 wurden Bildmerkmale theoretisch betrachtet und es wurde dargelegt, wie mittels dieser Form-, Farb- und Texturmerkmale ein Bild beschrieben werden kann. Diese Betrachtung soll nun auf die Verschleißerscheinungsform Pitting übertragen und

²³ A. Mannesmann Maschinenfabrik GmbH (2021), *Sensorik für Kugelgewindtriebe*. <https://www.amannesmann.de/produkte/kugelgewindtriebe/sensorik/> [21.01.2021].

erweitert werden (siehe Abbildung 6-23). Auf dieser Basis lässt sich das Pitting basierend auf Domänenwissen anhand von Bildmerkmalen beschreiben, die im Folgenden zur Klassifikation eines Bildes verwendet werden können.



Abbildung 6-23 Qualitative Abbildung der Bildmerkmale am Beispiel des Pittings.

Die Verschleißerscheinungsform Pitting resultiert aus der Oberflächenzerrüttung der Spindel. Hierbei ist die Kinematik des KGTs für den Entstehungsort des Pittings ein relevanter Einflussfaktor. Die Wälzkörper laufen innerhalb der Spindelgänge, wobei die maximale Last aufgrund der axialen Belastung der Kugelgewindetriebmutter an den Flanken der Spindellaufbahnen anliegt. Pittings auf der Spindel treten damit ausschließlich an den Flanken der Laufbahnen auf. Durch fortschreitenden Verschleiß sowie eine mögliche Überlagerung (bspw. mit adhäsiven Verschleißmechanismen) wachsen sie zusammen und es entstehen flächige Pittings (überlagert mit Abschälungen) an den Flanken der Spindellaufbahnen. Dies ist in der Mitte von Abbildung 6-24 dargestellt.



Abbildung 6-24 Markante Zustände der Laufbahn des KGTs (A_Hofmann 2021).

Durch die in KGT herrschenden Betriebsbedingungen sowie den verwendeten Schmierstoff kommt es zu einer Verunreinigung der Spindellaufflächen, die sich auch in den Bilddaten zeigt (siehe Abbildung 6-24 rechts). Es ist ersichtlich, dass sich das

optische Erscheinungsbild je nach Zustand unterscheidet und damit in unterschiedlichen Bildmerkmalen codiert werden kann.

Zur Beschreibung der visuellen Charakteristika werden bei (A_Heinzler 2020) die visuellen Charakteristika von Bildern mit Pitting, Bildern ohne Pitting sowie Bildern mit Verschmutzung und Pitting gegenübergestellt. Die Erkenntnisse werden im Folgenden anhand der Bildausschnitte in Abbildung 6-24 qualitativ diskutiert, bevor diese in Abschnitt 6.2.2 formalisiert und mittels Algorithmen zur Merkmalsextraktion extrahiert werden.

6.1.7.1 Farbmerkmale

Es ist ersichtlich, dass in den Bildern mit Pitting hauptsächlich Braun- und Grautöne vorherrschen, wohingegen in Bilddaten ohne Pitting die Farbtöne hauptsächlich grau sind. Überlagert Verschmutzung ein Pitting oder liegt Verschmutzung ohne Pitting vor (siehe Abbildung 6-24 rechts unten und oben), so überwiegen dunkle Braun- und Grautöne. Gleichzeitig liegt bei einem Bild ohne Pitting sowie bei einem Bild mit starker Verschmutzung eine geringe Anzahl an unterschiedlichen Farbtönen vor. Diese sind bei nicht verschmutzten Oberflächen meist grau, während sie bei stark verschmutzten Oberflächen zunehmend dunklere Farben annehmen. Bei einem Bild mit Pitting bzw. bei der Überlagerung mit Schmierstoff findet sich eine hohe Anzahl an Farbtönen, wobei dies bei nicht vorliegender Verschmutzung deutlicher ist.

6.1.7.2 Texturmerkmale

Auf Bilddaten ohne Pitting und Verschmutzung findet sich eine regelmäßige Oberfläche, wohingegen diese bei Aufnahmen von Pittings unregelmäßig ist und bei verschmutzten Bildern von unregelmäßig bis regelmäßig reicht. Ähnliches zeigt sich bei den Kontrasten der Bilder: Im nicht geschädigten Zustand finden sich wenige Kontrastunterschiede, wohingegen diese bei Vorliegen von Pittings stärker sind. Die Texturen sind bei Bildern ohne Pittings gleichmäßig, während sie bei Bildern mit Pitting innerhalb des Pittings zufällig erscheinen. Bei Bildern mit Verschmutzung richtet sich die Textur parallel zur Laufbahn entlang von rinnenförmigen Verdrängungen aus, was sich durch das Überrollen des Schmierstoffes durch die Kugeln erklären lässt. Dieser Effekt ist bei verschmutzten Bildern mit Pitting durch die zufällige Struktur des Pittings weniger stark ausgeprägt. Die zufällig erscheinende Textur der Pittings lässt sich in ihrer Gestalt als plateauartig beschreiben. Dies lässt sich durch die zwischen den geschädigten Stellen verbleibende nicht geschädigte Spindeloberfläche erklären. Sind Pittings nur vereinzelt

vorhanden bzw. klein, so tritt diese plateauartige Struktur weniger deutlich zum Vorschein.

6.1.7.3 Formmerkmale

Tritt ein Pitting auf, so zeichnet sich dieses durch klare Kanten gegenüber dem nicht geschädigten Hintergrund aus. Diese Kanten sind bei nicht geschädigter Oberfläche nicht vorhanden und verschwimmen mit zunehmendem Verschmutzungsgrad, bis sie schließlich bei vollständiger Überdeckung durch Verschmutzung nicht mehr klar zu erkennen sind (siehe Abbildung 6-24 rechts unten).

6.1.7.4 Lagemerkmal

Als letzte visuelle Charakteristik einer Pitting-Schädigung werden die im Stand der Forschung genannten Form-, Farb- und Texturmerkmale um das sogenannte Lagemerkmal erweitert.

Aufgrund der Kinematik des KGTs treten Schädigungen immer an der Laufbahnflanke auf. Liegt Verschmutzung vor, so ist diese meist über die gesamte Laufbahn der Spindel verteilt. Finden sich Merkmale abseits der Laufbahnflanke, so kann mittels Domänenwissen Verschleiß mit einer hohen Wahrscheinlichkeit ausgeschlossen werden.

In den vorherigen Abschnitten wurden Bildmerkmale von Pittings in Form von Farb-, Textur-, Form- und Lagemerkmalen spezifiziert. Diese werden im Folgenden zusammengefasst und den entsprechenden Kategorien zugewiesen.

6.1.7.5 Zuordnung zu Bildmerkmalen

Die visuellen Charakteristika sind in Tabelle 6-3 zusammengefasst und dem entsprechenden Bildmerkmal zugeordnet.

Hierbei lassen sich alle hauptsächlich auf die Farbe bezogenen Charakteristika den Farbmerkmalen zuschreiben. Charakteristika, die Regelmäßigkeiten sowie Kontraste und Texturen beschreiben, lassen sich am ehesten der Kategorie der Texturmerkmale zuordnen. Formmerkmale beschreiben charakteristische Formen innerhalb eines Bildes. Die klaren Kanten des Pittings lassen sich als einzige für das Pitting charakteristische Form identifizieren. Dem Lagemerkmal wird die Position der Schädigung zugewiesen.

Tabelle 6-3 Zuordnung der visuellen Merkmale zu Bildmerkmalen.

Kein Pitting	Pitting	Verschmutzung	Bildmerkmal
Hellbraune, graue Farbtöne	Dunkelbraune Farbtöne	Schwarze, dunkelgraue Farbtöne	Farbmerkmal
Geringe Anzahl an Farbtönen	Hohe Anzahl an Farbtönen	Geringe Anzahl an Farbtönen	Farbmerkmal
Regelmäßige Oberfläche	Unregelmäßige Oberfläche	Gemischte Regelmäßigkeit	Texturmerkmal
Geringe Kontrastunterschiede	Viele globale und lokale Kontrastunterschiede	Gemischte Kontrastunterschiede	Texturmerkmal
Gleichmäßige Textur	Zufällige Textur	Textur entlang der Laufbahn	Texturmerkmal
Keine plateauartigen Texturen	Plateauartige Texturen	Gemischte plateauartige Texturen	Texturmerkmal
Keine Kanten	Zahlreiche Kanten	Zahlreiche Kanten	Formmerkmal
-	Tritt an den Flanken der Laufbahn auf	Zufälliges Auftreten, meist über gesamte Laufbahn	Lagemerkmale

Zusammenfassend lassen sich charakteristische Merkmale für Pittings sowie für Bereiche, die keine Pittings zeigen, in den Bilddaten identifizieren. Die Definition von Charakteristika aufgrund von Domänenwissen und die Zuweisung zu Bildmerkmalen dienen nach (Richter 2018) als Grundlage für die Klassifikation der Bilddaten.

Die Untersuchung der Eignung und Notwendigkeit der definierten Merkmale zur Klassifikation ist das Ziel des nächsten Abschnitts.

6.2 Klassifikation von Oberflächenzerrüttung in den Bilddaten des Kugelgewindetriebes

In den vorherigen Abschnitten wurde die Entwicklung von Pittings auf den Spindeln eingehend untersucht und die visuellen Eigenschaften von Pittings wurden entsprechenden Farbmerkmalskategorien zugewiesen. Um eine automatisierte Einschätzung des Vorhandenseins von Pittings auf der Kugelgewindetriebsspindel zu realisieren, wird

im folgenden Unterkapitel auf der empirischen Analyse der Bilddaten aufgebaut und die Verfahren zur automatisierten Klassifikation der Bilddaten werden untersucht.

6.2.1 Grundüberlegungen zur Auswahl von Klassifikationsansätzen

Zur Klassifikation von Pittings auf der Kugelgewindetriebspindel werden zwei Ansätze untersucht und deren Ergebnisse diskutiert. Der erste zu untersuchende Ansatz beschäftigt sich mit der Klassifikation von Bilddaten, indem Merkmale zur Klassifikation mittels Domänenwissen definiert und aus den Bilddaten extrahiert werden. Dieser Ansatz wird nachfolgend als merkmalsbasierter Ansatz bezeichnet.

Dieser erste Ansatz baut direkt auf den im vorherigen Abschnitt definierten visuellen Charakteristika von Pittings auf und nutzt diese zur Klassifikation. Zunächst werden mittels der in Abschnitt 6.1.7 definierten visuellen Charakteristika von Pittings und der bei (Schlagenhauf, Heinzler & Fleischer 2020) weiterentwickelten Merkmalsextraktionsalgorithmen geeignete Merkmale aus Bildern extrahiert. Diese Merkmale werden dann einem Modell des Maschinellen Lernens zur Klassifikation übergeben.

Das Ziel dieses ersten Ansatzes ist es, die Hypothese zu überprüfen, dass die semantisch sinnvollen²⁴ Merkmale, die von Domänenexpert*innen zur Differenzierung von Bildern mit Pittings definiert werden können, auch zur automatischen Klassifikation mittels Verfahren des Maschinellen Lernens eingesetzt werden können. Ein weiteres Ziel ist es, die Relevanz der definierten Merkmale für die Klassifikation zu überprüfen.

Nach (Fink & Wang et al. 2020) ist bekannt, dass es schwierig ist, sämtliche Merkmale zu definieren, sodass diese alle Zustände eines Objektes beschreiben, weshalb sich hierfür besonders automatisierte Verfahren zur Merkmalsextraktion eignen. Diese Verfahren basieren nicht auf der Definition von Merkmalen, sondern erlernen diese selbstständig aus den Daten. Daher werden in einem zweiten Ansatz unterschiedliche Varianten von CNN untersucht, die in der Lage sind, Merkmale automatisiert aus den Bilddaten zu extrahieren. Zur Untersuchung der CNN-Architekturen werden diese mittels eines teilfaktoriellen Versuchsplans erstellt und auf die Problemstellung ausgelegt optimiert. Das Ziel des zweiten Teilkapitels ist es, CNN-Modelle mit einer möglichst hohen Klassifikationsgenauigkeit zu untersuchen.

²⁴ Semantisch sinnvoll bedeutet hier, dass die Merkmale einfach von Domänenexper*innen interpretiert werden können, z. B.: Das Pitting zeichnet sich durch eine dunkle Färbung gegenüber dem Hintergrund aus.

Die Motivation für die Betrachtung der beiden Ansätze resultiert aus deren unterschiedlichen Stärken, die jeweils im Kontext der Klassifikation von KGT noch nicht untersucht wurden. Nach (Richter 2018) ist der merkmalsbasierte Ansatz mittels klassischer Methoden des Maschinellen Lernens hoch relevant, da sich durch Domänenwissen relevante „hochdiskriminative, robuste, störungsinvariante und möglichst niedrigdimensionale Merkmale“ (Richter 2018 S.38) extrahieren lassen, die nur die für die Klassifikation notwendigen Informationen enthalten. Lassen sich Objekte bspw. nach einer einfachen Eigenschaft wie der Farbe klassifizieren, genügt ein Abgleich der Farben zur Lösung der Klassifikationsaufgabe. Ein Ansatz, der die Merkmale aus den Daten automatisiert erlernt, ist nicht unbedingt notwendig. Weitere Vorteile der klassischen Definition und Extraktion von Merkmalen in Kombination mit Methoden des Maschinellen Lernens sind die Interpretierbarkeit der Merkmale sowie die Interpretierbarkeit der Modellentscheidungen. Die tiefgehende domänenbezogene Auseinandersetzung mit der Problemstellung und den dafür benötigten Daten erfüllt die im Quasi-Standard-Machine-Learning-Prozessmodell CRISP-DM (Cross Industry Standard Process for Data Mining) (Chapman & Clinton et al. 2000) geforderten drei Phasen des Business Understanding, Data Understanding und Data Preparation und zwingt die Domänenexpert*innen dazu, sich intensiv und inhaltlich mit der Thematik zu befassen. Vor allem in technischen Domänen wie bspw. dem Maschinenbau ist für die Erstellung der Datensätze und das Labeln der Daten Domänenwissen erforderlich. So konnte am wbk Institut für Produktionstechnik des KIT im Rahmen von zahlreichen Projekten und Gesprächen mit Unternehmensvertretern festgestellt werden, dass die Implementierung von Maschinellern im industriellen Kontext je nach Anwendung ein erhebliches Maß an Verständnis für die jeweilige Domäne erfordert. Ohne Domänenwissen kann die Anwendung von Verfahren des Maschinellen Lernens zur Klassifikation als nicht zielführend eingestuft werden. Als weiteren Vorteil von Merkmalen, die mittels Domänenwissen extrahiert wurden, unterstreicht (Richter 2018) die Interpretierbarkeit von Modellen, die mit diesen interpretierbaren Merkmalen trainiert wurden. Die Hilfestellung, die den Domänenexpert*innen durch maschinelle Lernverfahren geboten wird, besteht darin, dass die Klassifikation von Datenpunkten automatisiert durch das maschinelle Lernverfahren vorgenommen wird. Hierbei entscheidet das Modell selbstständig auf Basis der von den Anwender*innen erstellten Merkmale, welche Kombination der Merkmale für eine Klassifikation geeignet ist. Dies ist besonders bei einer großen Anzahl an Merkmalen bzw. in Fällen, in

denen eine Klassifikation nicht durch die Anwender anhand einzelner weniger Merkmale vorgenommen werden kann, hilfreich.

Letztendlich besteht die Stärke der manuellen Definition und Extraktion von Merkmalen auf Basis von Domänenwissen besonders darin, dass eine intensive Auseinandersetzung mit der Problemstellung stattfindet und dass daraus Merkmale hervorgehen können, die eine interpretierbare Klassifikation ermöglichen.

Die Vorteile, die sich durch die Definition von Merkmalen auf Basis von Domänenwissen ergeben, werden von (Richter 2018) einerseits betont, andererseits werden aber auch die Nachteile des Ansatzes diskutiert: Die intensive Auseinandersetzung mit einer Materie ist zeitaufwendig und damit teuer. Liegen genügend Daten für die Modellbildung vor und ist die Problemstellung ausreichend verstanden sowie das Ergebnis der Modellbildung zufriedenstellend, liegt es in der Verantwortung der Anwender, die Sinnhaftigkeit der manuellen Definition von Merkmalen zu bewerten. Eine weitere Herausforderung im Kontext klassischer Ansätze ist es, dass die Merkmale problemspezifisch definiert sind und damit nur schwer auf andere Anwendungsgebiete übertragen werden können (Richter 2018).

Gegenüber den klassischen Ansätzen haben sich in den letzten Jahren mit großem Erfolg Deep-Learning-basierte Ansätze zur Klassifikation etabliert. Nach (Chollet 2018) basiert die Stärke von Deep-Learning-Modellen auf dem Vorhandensein von großen Datenmengen, mit deren Hilfe die große Anzahl an Modellparametern automatisiert angepasst werden kann. Dies resultiert in der Tatsache, dass neben der Datenaufnahme und dem Labeling der Daten kein weiteres Domänenwissen für die Generierung von Merkmalen notwendig ist, sondern diese automatisiert aus den Daten erlernt werden. So konnte ein Team um Alex Krizhevsky (Krizhevsky, Sutskever & Hinton 2012) beim ImageNet-Wettbewerb im Jahr 2012 mittels der sogenannten AlexNet-Deep-Learning-Architektur eine Top-fünf-Genauigkeit²⁵ von 83.6 % erreichen. Damit übertrafen sie die bisherigen auf klassischen maschinellen Lernverfahren basierenden Modelle erstmals deutlich, die bis dahin eine maximale Top-fünf-Genauigkeit von 74.3 % erzielt hatten und seither die Deep-Learning-basierten Ansätze nicht mehr übertreffen konnten. Im Jahr 2015 konnte eine Top-fünf-Genauigkeit von 96.4 % erreicht werden (Russakovsky

²⁵ Unter den ersten fünf vorhergesagten Klassen befindet sich die wahre Klasse.

& Deng et al. 2015). Die menschliche Leistungsfähigkeit in diesem Wettbewerb liegt bei ca. 95 % und wurde damit 2015 erstmals übertroffen.

Diese Beispiele zeigen die Fähigkeit von Deep-Learning-Architekturen bei Vorhandensein von großen Trainingsdatenmengen. Ein weiterer Vorteil von Deep-Learning-Architekturen liegt darin begründet, dass die Verfahren auf unterschiedliche Kontexte angewandt werden können, da die manuelle Definition von Merkmalen entfällt. Dies ist mit klassischen Ansätzen, bei denen die Merkmale domänenspezifisch definiert werden, nicht ohne Weiteres möglich.

Wie im Stand der Technik in Abschnitt 3.1 beschrieben wurde, konnten Deep-Learning-basierte Modelle zwar auch in technischen Domänen bereits überzeugende Ergebnisse erzielen, sie wurden allerdings noch nicht am KGT untersucht. Aus diesem Grund besteht das Ziel im zweiten Ansatz im Erreichen einer möglichst hohen Klassifikationsgenauigkeit. Dagegen grenzt sich das Ziel des ersten Ansatzes ab, bei dem die Möglichkeit der Klassifikation mittels semantisch sinnvoller Merkmale untersucht werden soll. Aufgrund der unterschiedlichen Stärken und Zielsetzungen der Ansätze werden die maximal erreichbaren Genauigkeiten der beiden Ansätze nicht direkt verglichen.

In den folgenden beiden Unterkapiteln werden die beiden Ansätze weiter konkretisiert und die Ergebnisse dargestellt.

6.2.2 Klassifikation mittels extrahierter Merkmale

Ziel dieses Abschnitts ist die Untersuchung von Modellen des Maschinellen Lernens, die mit Merkmalen basierend auf Domänenwissen über die Tribologie und die Verschleißmerkmale des KGTs trainiert werden. Die Hypothese ist hierbei, dass sich die Bilddaten über extrahierte Form-, Farb- und Texturmerkmale und die Integration von weiteren Domänenwissen-spezifischen Charakteristika so beschreiben lassen, dass durch Kombination der Merkmale Bilder mit Schädigung von Bildern ohne Schädigung unterschieden werden können. Es wird also angenommen, dass bestimmte Merkmalskombinationen charakteristisch für Pittings und andere charakteristisch für nicht geschädigte Oberflächen sind. Diese Kombination an semantischen Merkmalen soll mithilfe von Verfahren des Maschinellen Lernens automatisiert vorgenommen werden.

Als Ergebnis liegt der Fokus der Untersuchung hierbei zunächst auf der Klärung der Frage, ob eine automatische Klassifikation mittels dieser Merkmale möglich ist, und

darauf aufbauend auf der Frage, ob die zur Klassifikation verwendeten Merkmale mit der Einschätzung von Domänenexpert*innen übereinstimmen.

Zur Klassifikation werden die in Abschnitt 6.1.7 definierten Charakteristika zunächst mittels zum Stand der Technik erweiterten Merkmalsextraktionsalgorithmen extrahiert. Die so extrahierten Merkmale werden dann in einem für die Bilddaten repräsentativen Merkmalsset gespeichert, das zusammen mit den Labeln (Pitting, kein Pitting) als Grundlage zur Klassifikation dient.

6.2.2.1 Auswahl und Erweiterung von Algorithmen zur Merkmalsextraktion

In Abschnitt 6.1.7 wurden Charakteristika eines Pitting-Schadens (siehe Tabelle 6-3) definiert. Diese Merkmale beschreiben semantische Eigenschaften der Bilder und müssen in einem nächsten Schritt zunächst formalisiert werden, um sie dann zur Klassifikation mittels eines maschinellen Lernverfahrens zu verwenden. Hierzu werden im Folgenden Merkmalsextraktionsalgorithmen zur Extraktion von Farb-, Form- und Texturmerkmalen ausgewählt. Bei der Auswahl soll auf die in Abschnitt 2.4.2 beschriebenen Grundlagen hinsichtlich der Bildmerkmale aufgebaut werden. Die nachfolgend dargelegten Ergebnisse wurden in der Arbeit von (A_Heinzler 2020) erarbeitet.

6.2.2.1.1 Farbmerkmale

Die Oberfläche der nicht verschmutzten und unbeschädigten Spindellaufbahn zeigt sich in den Bilddaten in vornehmlich hellbraunen und grauen Farbtönen, wohingegen sich Bilddaten, in denen ein Pitting vorhanden ist, in dunkelbraunen Farbtönen zeigen.

Diese qualitative Beschreibung lässt sich durch Abbildung 6-25 quantifizieren. Es ist ersichtlich, dass sich Bilder mit Pittings aufgrund der unterschiedlichen Farben sowohl von Bildern der sauberen und unbeschädigten Spindel als auch von Bildern mit Verschmutzung bzw. Verschmutzung und Pitting unterscheiden lassen. Liegt jedoch Verschmutzung zusammen mit einem Pitting vor, so sind die Unterschiede in den Farbhistogrammen nicht eindeutig. Weitere Form- und Texturmerkmale müssen dann für eine Klassifikation hinzugezogen werden.

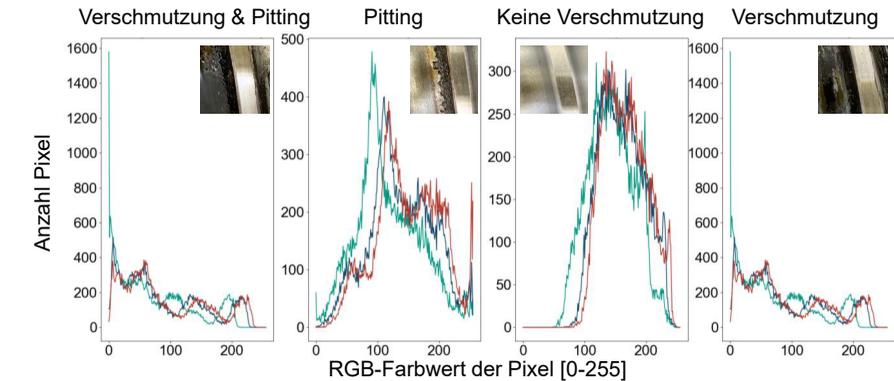


Abbildung 6-25 Farbhistogramme der Bildklassen.

Die Extraktion von Farbmerkmalen aus den Farbhistogrammen erfolgt mittels des sogenannten Clustured Colour Share, der in der Arbeit von (Schlagenhauf, Heinzler & Fleischer 2020) erstmals aufbauend auf einem Ansatz nach (Bahant, 2018)²⁶ entwickelt wurde. Zur Berechnung des Clustured Colour Share werden die Farbwerte zunächst mittels eines K-Means-Algorithmus in 20 Cluster für Bildausschnitte mit Pitting und Bildausschnitte ohne Pitting zusammengefasst. Dargestellt in einem Kuchendiagramm lässt sich damit eine Metrik für das Verhältnis von Farbintensität und Anteil der Farbcluster ermitteln. Zur Generierung von für das Training geeigneten Farbmerkmalen wird der Begriff des Farbanteiles eingeführt.

Hierbei wird der Mittelwert der Farbcluster mit deren Anteil multipliziert und damit in einem Wert zusammengefasst. Damit ergeben sich für 20 Cluster 20 Farbmerkmale zur Klassifikation. Der Mittelwert, die Standardabweichung, das Maximum sowie das Minimum der RGB-Werte werden ebenfalls als Merkmale aus den Bilddaten extrahiert. Der Pseudocode des Ansatzes ist in Abbildung 6-26 dargestellt.

²⁶ Bahant, K. (2018), *Color Identification in Images*. Towards Datascience (Hrsg.). <https://towardsdatascience.com/color-identification-in-images-machine-learning-application-b26e770c4c71> [05.08.2020].

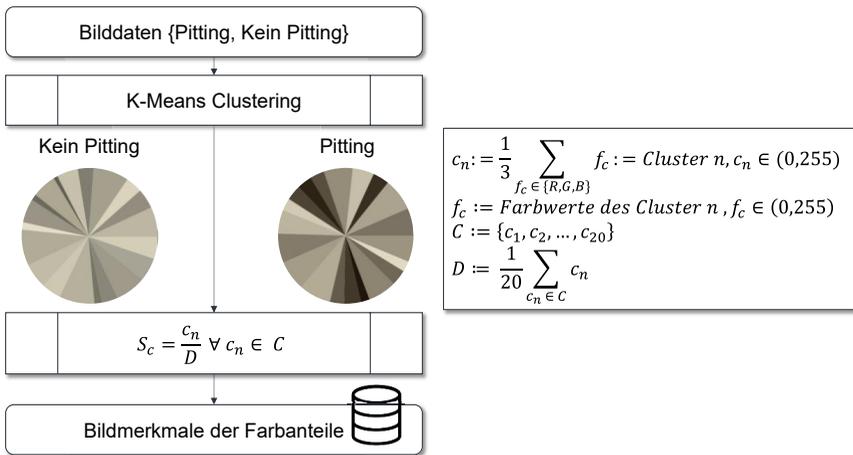


Abbildung 6-26 Vorgehen zur Extraktion von Farbmerkmalen.

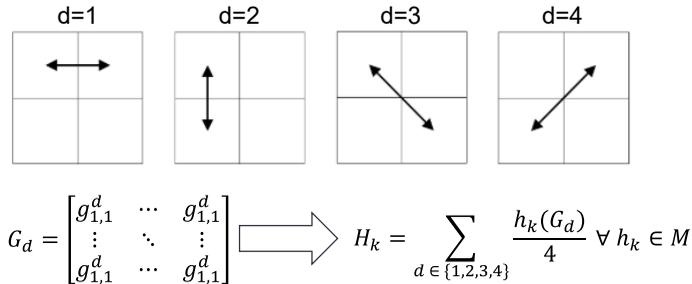
Zusammenfassend lassen sich die Farbmerkmale vor allem durch den Ansatz des Clustered Colour Share extrahieren und damit die semantischen Merkmale in eine für Klassifikationsmodelle verwendbare mathematische Form überführen. Dieser Ansatz ist neuartig und im Stand der Technik bisher so nicht verwendet.

6.2.2.1.2 Texturmerkmale

Anschließend werden Texturmerkmale aus den Bilddaten extrahiert.

Wie in Abschnitt 6.1.7 dargelegt, unterscheiden sich Pittings von nicht geschädigten Spindelabschnitten hauptsächlich durch deren unregelmäßige und teilweise plateauartige Oberflächentextur im Bild. Tritt zusätzlich Verschmutzung auf, so zeigt sich diese durch in Laufbahnrichtung weisende, rinnenförmige Texturen in den Bildern. Zur Extraktion dieser Texturmerkmale werden auf Basis der Grey-Level-Co-Occurrence-Matrix sogenannte Haralick-Merkmale (Haralick 1979) zur Beschreibung der Oberflächentextur berechnet. Die Haralick-Merkmale werden global für ein Bild berechnet und beschreiben damit die Textureigenschaften des Bildes als Ganzes. Zur Berechnung der Haralick-Merkmale werden die in einem Bild auftretenden Grauwerte n geordnet und in einer $n \times n$ Matrix aufgetragen. Die Matrix enthält dabei sowohl entlang der Spalten als auch entlang der Zeilen die n im Bild vorkommenden Grauwerte. Für jede mögliche Nachbarschaftsrichtung zwischen zwei Pixeln wird die Anzahl der nebeneinander (co-

occurrence) auftretenden Pixel aufgetragen. Die betrachteten Richtungen sind hierbei zusammen mit der Berechnungsvorschrift in Abbildung 6-27 dargestellt.



G_d := GLCOO – Matrix für die Richtung d

$g_{1,1}^d$:= Anzahl der zusammen auftretenden benachbarten Grauwerte

H_k := Menge der berechneten Mittelwerte der Haralick Merkmale über alle Richtungen d

h_k := k – tes Haralick Merkmal

M := Menge der Haralick Merkmale

Abbildung 6-27 Grey-Level-Co-Occurrence-Matrix nach (A_Heinzler 2020) und (Haralick 1979).

Für die vier möglichen Richtungen ergeben sich so vier Matrizen, aus denen 13 Haralick-Merkmale als Texturmerkmale zur Beschreibung der Bilddaten berechnet werden. Die Haralick-Merkmale werden für zwei beispielhafte Bilder in Tabelle 0-1 im Anhang gegenübergestellt.

Die Haralick-Merkmale beschreiben globale Texturmerkmale eines Bildes und eignen sich damit bspw. zur Beschreibung der zufälligen Strukturierung der Oberfläche bei Auftreten von Pittings. Zur Extraktion weiterer Texturmerkmale werden sogenannte Local-Binary-Patterns (LBP) nach (Ojala, Pietikainen & Maenpaa 2002) berechnet und aus den Bilddaten extrahiert. Beim LBP-Ansatz werden lokale, rotationsinvariante Texturmerkmale berechnet, die zur Beschreibung von lokalen Texturen wie dem Auftreten der plateauartigen Texturen oder der Textur entlang der Laufbahn verwendet werden können. Beim LBP-Ansatz wird der Grauwert eines Pixels mit den Pixelwerten der umgebenden acht Pixel verglichen. Einem umliegenden Pixel wird der Wert 1 zugewiesen, falls der Pixelwert größer ist als der Wert des zentralen Pixels, andernfalls wird dem Pixel der Wert 0 zugewiesen. Die acht Werte werden in einem 8-Bit-Vektor codiert. Es ergeben sich somit pro Pixel insgesamt 256 mögliche Binary-Patterns. Der Ansatz wird

für alle Pixel des Bildes berechnet und aus der resultierenden LBP-Matrix wird ein Histogramm mit 256 Abschnitten gebildet. Jeder Bereich des Histogramms enthält damit die Summe der auftretenden LBP im Bild, die den Wert des jeweiligen Abschnitts annehmen. Dieser Ansatz ist in Abbildung 6-28 dargestellt. Durch den LBP-Ansatz werden insgesamt 256 Texturmerkmale extrahiert.

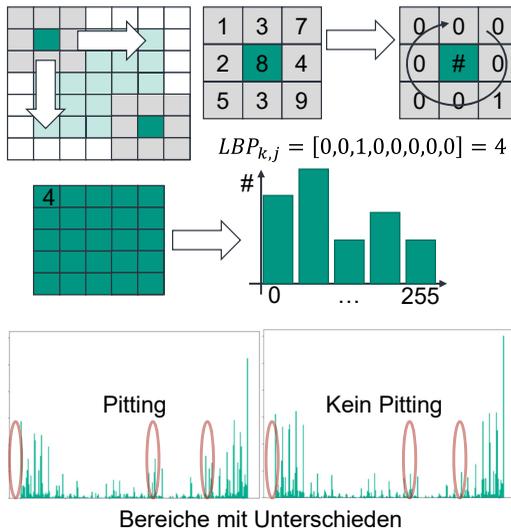


Abbildung 6-28 Ansatz zur Extraktion der verwendeten Locally-Binary-Patterns und Codierung derselben dargestellt in einem Histogramm.

Die in den Bilddaten enthaltenen Texturen, die sich zur Unterscheidung eines Bildes mit Pitting von einem Bild ohne Pitting eignen, wurden mittels LBPs codiert. Diese codierten Merkmale können für die Klassifikation mittels eines maschinellen Lernverfahrens verwendet werden.

6.2.2.1.3 Formmerkmale

Eine weitere charakteristische Eigenschaft von Pitting-Schäden ist der markante Übergang der geschädigten Stelle hin zum nicht geschädigten Material, der sich durch eine Art Kante beschreiben lässt. Diese Charakteristik ist den Formmerkmalen zuzuschreiben. Wird diese Charakteristik global für die Bereiche im Bild berechnet, die diese Charakteristik aufweisen, so ergibt sich ein räumliches Merkmal in den Bilddaten, das die Position der Kanten im Bild codiert. Zur Extraktion der Formmerkmale wird das Scale-

Invariant-Feature-Transform(SIFT)-Verfahren nach (Lowe 2004) angewandt, das zu dem in (Schlagenhauf, Heinzler & Fleischer 2020) beschriebenen Key-Points-Per-Sub-Region-Ansatz weiterentwickelt wird. Durch den Key-Points-Per-Sub-Region-Ansatz ist es möglich, die räumliche Position von Formmerkmalen im Bild zu extrahieren. Dieser Ansatz ist neuartig und im Stand der Technik bisher so nicht beschrieben.

Durch den SIFT-Ansatz werden zunächst rotations- und skalierungsinvariante Merkmale für ein Bild generiert. Durch die Invarianz in der Skalierung und Rotation ist es möglich, eine bestimmte Form in unterschiedlicher Größe und Ausrichtung im Bild zu detektieren. Aus der Kenntnis, dass charakteristische Kanten eines Pittings im Bild rechts oder links an den Flanken auftreten, werden für das Bild zunächst die SIFT-Merkmale berechnet.

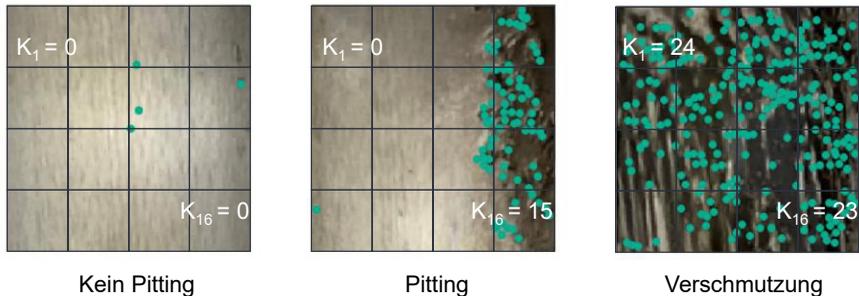


Abbildung 6-29 Verteilung der SIFT-Key-Points in Subregionen nach (Schlagenhauf, Heinzler & Fleischer 2020). Links: kein Defekt; Mitte: Defekt an rechter Flanke; rechts: Verschmutzung.

Darauf aufbauend wird das Bild in insgesamt 16 quadratische Unterbilder (Sub-Regionen) geteilt und jeder Sub-Region wird die Anzahl der innerhalb der Sub-Region auftretenden SIFT-Key-Points zugeordnet. Diese 16 Werte werden dann als Merkmale zur Kodierung der Position der Pittings im Bild in die Merkmalstabelle übernommen. Abbildung 6-29 veranschaulicht, wie mittels dieses Ansatzes die für ein Pitting charakteristischen Formmerkmale extrahiert werden und durch Erweiterung des SIFT-Ansatzes die relevante räumliche Anordnung der Merkmale im Bild in Form von 16 Merkmalsvektoren codiert wird.

Es ist ersichtlich, dass in einem Bild, das weder Verschmutzung noch eine Schädigung zeigt, wenige charakteristische Punkte vom SIFT-Algorithmus gefunden werden. In einem Bild, das ein Pitting zeigt, werden in den jeweiligen Sub-Regions am Rand vermehrt charakteristische Punkte detektiert. In Bilddaten, in denen Verschmutzung oder eine Schädigung zusammen mit Verschmutzung auftritt, sind die Merkmale zufällig über die Sub-Regions verteilt.

Zusammenfassend lässt sich die charakteristische Position eines Pittings in den Bilddaten über den Key-Points-Per-Sub-Region-Ansatz aus den Bilddaten extrahieren.

Bevor eine Klassifikation der extrahierten Merkmale erfolgt, werden zunächst die Ergebnisse der Analyse und Extraktion von Bildmerkmalen aus den Originaldaten zusammengefasst.

6.2.2.2 Zusammenfassung der Analyse und Extraktion von Bildmerkmalen

Durch die beschriebene Definition von Merkmalen und deren Extraktion mittels geeigneter Merkmalsextraktionsalgorithmen wurden die Bilddaten auf 318 Merkmale codiert, die wie folgt zusammengesetzt sind: 28 Farbmerkmale, die mittels der Clustered-Colour-Share-Methodik extrahiert wurden, 274 Texturmerkmale, die mittels des LBP-Ansatzes (261) sowie der Haralick-Merkmale (13) extrahiert wurden, und 14 Formmerkmale, die mittels des Key-Points-Per-Sub-Region-Ansatzes extrahiert wurden. Die Merkmalskategorien, deren Anzahl sowie die verwendeten Algorithmen sind in Tabelle 0-2 im Anhang dargestellt. Für die Zustände ‚Pitting‘ und ‚kein Pitting‘ wurden die Bilddaten in Merkmalsvektoren transformiert. Zusätzlich zu diesen Merkmalen liegt die Information über die Klasse eines jeweiligen Bildes (Pitting, kein Pitting) vor. Diese Informationen bilden zusammen den Datensatz, aus dem die korrekte Kombination der Merkmale zur Klassifikation erlernt werden soll.

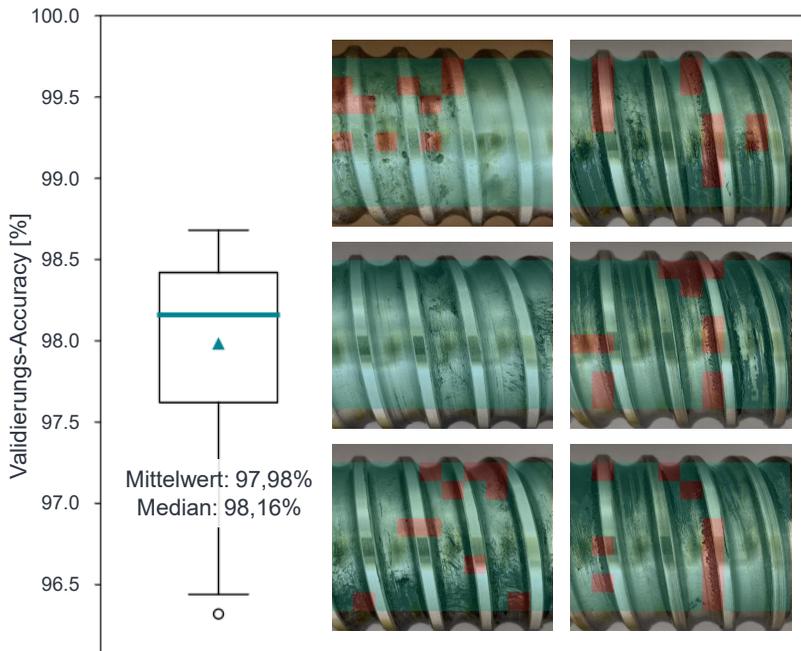
Die Fähigkeit, mittels der extrahierten Merkmale eine Klassifikation durchzuführen, wird im Folgenden untersucht.

6.2.2.3 Klassifikation der Merkmale

Das Ziel dieses Abschnitts ist die Überprüfung, ob sich die definierten Merkmale zur automatisierten Klassifikation eignen.

Die 318 extrahierten Merkmale werden in ein Neuronales Netz zur Klassifikation gegeben. Hierfür ist es notwendig, ein möglichst leistungsfähiges Modell zur Klassifikation

auszuwählen, um auszuschließen, dass eine nicht zielführende Klassifikation auf die Wahl des Klassifikationsalgorithmus zurückzuführen ist. Das Neuronale Netz ist im Stand der Technik als leistungsfähiges Modell zur Klassifikation bekannt. Nach (Chollet 2018) kombiniert und gewichtet das Neuronale Netz die extrahierten Merkmale automatisch in einem mehrstufigen Verfahren und eignet sich damit besonders zur Klassifikation, weshalb zunächst diese Architektur gewählt wird.



Auf Domänenwissen
basierende Merkmale

Abbildung 6-30 Validierungs-Accuracy des Neuronalen Netzes, trainiert mittels Merkmalen, die auf Basis von Domänenwissen extrahiert wurden. Siehe Abbildung 0-9 im Anhang für Vergleichsdaten.

Das Neuronale Netz zur Klassifikation der Merkmale besteht aus fünf Schichten mit zehn Neuronen pro Schicht. Es wird die ReLu-Aktivierungsfunktion verwendet. Die L2-Regularisierung mit einem alpha-Wert von 10^{-4} wird eingesetzt, um *Overfitting* zu ver-

meiden. Das Modell wird für 200 Epochen mit dem Adam-Optimierer trainiert. Das Neuronale Netz wird jeweils fünfmal mit 80 % der Trainingsdaten trainiert und auf 20 % der Daten getestet.

Die Ergebnisse der Klassifikation sind in Abbildung 6-30 dargestellt. Es ist ersichtlich, dass eine mediane Validierungs-*Accuracy* von 98.16 % für die Kombination der Merkmale erreicht werden kann. Wenngleich die Klassifikationsgenauigkeit sehr hoch ist, fällt auf, dass das Modell verschmutzte Pittings teilweise fehlklassifiziert bzw. Verschmutzungsmuster fälschlicherweise als Pittings klassifiziert. Auch für Domänenexpert*innen kann diese Art der Bilder als die herausforderndste Art eingestuft werden. Gleichzeitig wird ein Großteil der Schädigungen auf den Bildern korrekt erkannt (vergleich hierzu Abbildung 0-9 im Anhang).

Als Ergebnis kann die Möglichkeit zur Klassifikation der Bilddaten auf Basis der definierten Merkmale bestätigt werden. Aufgrund dieser Erkenntnis und der bereits hohen Klassifikationsgenauigkeit soll auf einen Vergleich mit weiteren Verfahren des Maschinellen Lernens verzichtet werden.

Neben der Fähigkeit, eine Klassifikation auf Basis der Merkmale vorzunehmen, ist hier besonders die Relevanz der Merkmale zur Klassifikation von Interesse, auf die im folgenden Abschnitt eingegangen werden soll.

6.2.2.4 Untersuchung der Merkmalsrelevanz

Im vorangehenden Abschnitt konnte gezeigt werden, dass sich mittels der definierten und extrahierten Merkmale Bilddaten zuverlässig klassifizieren lassen. Ein Nachteil der Architektur des Neuronalen Netzes besteht darin, dass die Entscheidungsfindung nicht einfach nachvollzogen werden kann. Es ist zwar möglich, durch eine gezielte Permutation der Merkmale und das anschließende Beobachten der Änderung in der Validierungsgenauigkeit auf die Relevanz einzelner Merkmale oder Merkmalskategorien zu schließen (Form-, Farb- oder Texturmerkmale). Allerdings ist hierbei nicht ersichtlich, wie das Modell die vorhandenen Merkmale zur Klassifikation nutzt. Um zu dieser Erkenntnis zu gelangen, wird im Folgenden die Entscheidungsfindung bei der Klassifikation mithilfe eines Entscheidungsbaumes untersucht.

Entscheidungsbäume zeichnen sich nach (Richter 2018) besonders dadurch aus, dass die Merkmalsbewertung schon in die Entscheidungsfindung des Ansatzes integriert ist.

Zunächst soll überprüft werden, ob der Entscheidungsbaum in der Lage ist, eine Bewertung der Merkmalswichtigkeit vorzunehmen. Dies geschieht über die Auswertung seiner Klassifikationsgenauigkeit: Ist der Entscheidungsbaum in der Lage, eine hohe Klassifikationsgenauigkeit zu erzielen, so ist er in der Lage, die Relevanz der Merkmale für die Klassifikation korrekt einzuschätzen.

Abbildung 6-31 zeigt die prinzipielle Struktur eines Entscheidungsbaumes. Die Merkmale zur Klassifikation sind im Baum ihrer Relevanz nach absteigend angeordnet: Wichtige Merkmale finden sich oben, weniger wichtige Merkmale entsprechend tiefer im Entscheidungsbaum. Dies resultiert aus der Logik, mit der eine Entscheidung zustande kommt. Wichtige Merkmale werden zuerst geprüft und daran anschließend weniger wichtige Merkmale weiter unten im Baum geprüft. Somit wird eine Klassifikation durchgeführt und gleichzeitig eine Aussage über die Relevanz der Merkmale zur Klassifikation zurückgegeben.

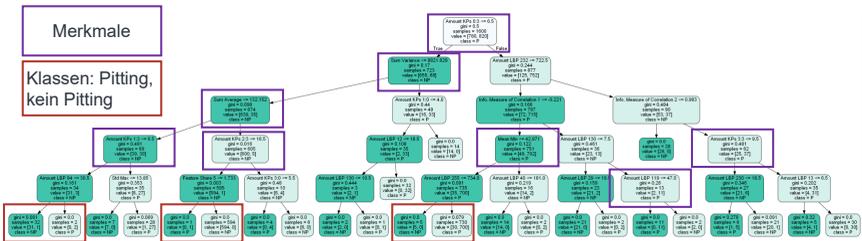


Abbildung 6-31 Struktur eines Entscheidungsbaumes.

Der Entscheidungsbaum wird mit denselben Daten wie das Neuronale Netz trainiert. Um eine statistische Absicherung zu erhalten, wird er 300-mal mit zufällig ausgewählten Trainings- und Validierungsdaten trainiert und eine mittlere Validierungs-Accuracy von 91.43 % wird erzielt. In Abbildung 6-32 wird die Klassifikationsgenauigkeit des Entscheidungsbaumes der Klassifikationsgenauigkeit des Neuronales Netzes gegenübergestellt. Die Klassifikationsgenauigkeit des Neuronales Netzes übersteigt zwar die Klassifikationsgenauigkeit des Entscheidungsbaumes, nichtsdestoweniger kann die Funktionsfähigkeit des Entscheidungsbaumes zur Klassifikation der Bilddaten auf Basis der definierten Merkmale herangezogen werden.

Zusammenfassend kann die Eignung des Entscheidungsbaumes zur Klassifikation der Merkmale belegt werden. Der Entscheidungsbaum eignet sich damit zur Bewertung der

Relevanz der Merkmale. Im Folgenden wird diese Relevanz der Merkmale zur Klassifikation untersucht.

Abbildung 6-33 listet die über die 300 Testläufe gemittelten wichtigsten zehn Merkmale zur Klassifikation der Bilddaten auf. Dabei wird die hohe Relevanz der Formmerkmale (Nummer 1) für die Entscheidungsfindung ersichtlich. Die erklärte Varianz gemittelt über die zehn wichtigsten Merkmale beträgt 83.5 %, wobei die erklärte Varianz der drei wichtigsten Merkmale 58.6 % beträgt. Die verbleibenden 308 Merkmale erklären lediglich 16.5 % der Varianz. Die vier wichtigsten Merkmale sind allesamt den Formmerkmalen zuzuschreiben.

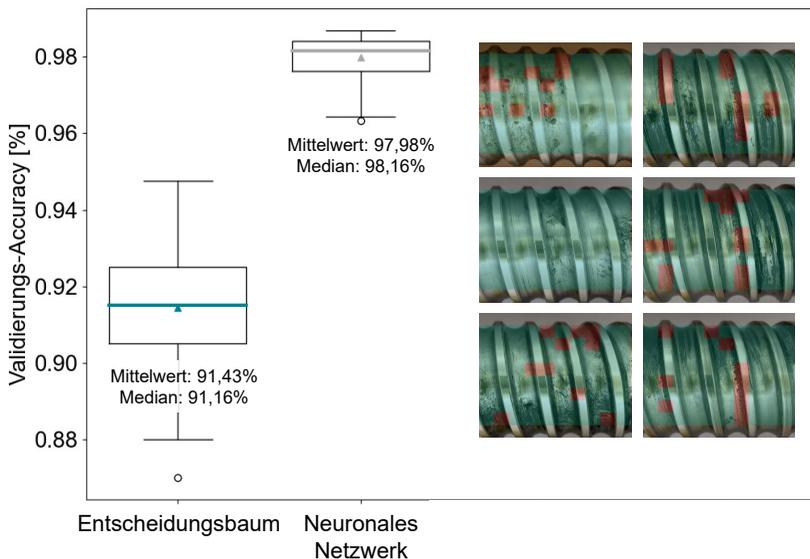


Abbildung 6-32 Klassifikationsgenauigkeit des Entscheidungsbaumes im Vergleich zum Neuronalen Netz.

Konkret handelt es sich bei diesen Merkmalen um die Anzahl an Keypoints in den Subbereichen des Bildes auf der rechten Seite. Dies ist der Bereich, in dem Schädigungen aufgrund der Kinematik und der Tribologie am KGT zu erwarten sind. Damit kann die vermutete Relevanz dieses Merkmals für die Klassifikation bestätigt werden. Abbildung 6-34 stellt die Gesamtwichtigkeit der Bildmerkmale einander normiert gegenüber.

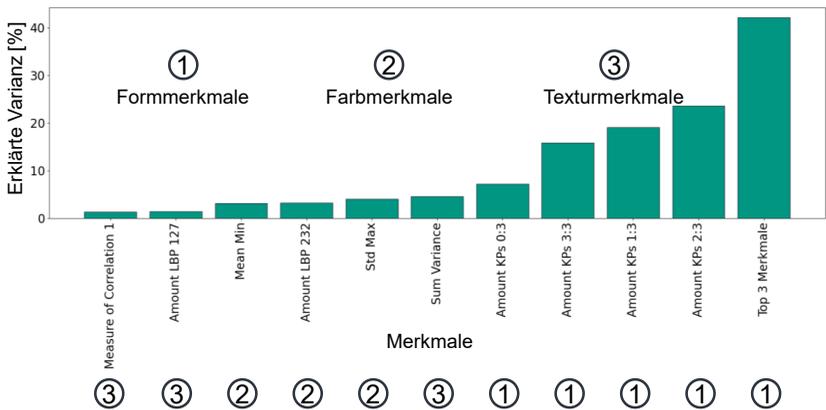


Abbildung 6-33 Merkmale zur Klassifikation von Bilddaten.

Hierbei wurde die Gesamtvarianz durch die Anzahl der Merkmale pro Kategorie geteilt. Es ist ersichtlich, dass die extrahierten Formmerkmale für die vorliegende Klassifikation die höchste absolute Relevanz besitzen, gefolgt von den Farbmerkmalen. Texturmerkmale haben für die vorliegende Klassifikationsaufgabe und den vorliegenden Bilddatensatz die geringste Relevanz. Dies lässt sich damit begründen, dass die Unterschiede in den Farben vor allem bei Bilddaten, die lediglich Verschmutzung sowie Verschmutzung gepaart mit einer Schädigung enthalten, nicht groß sind. Dieses Verhalten unterstreicht die Beobachtung, die auf Basis der Farbhistogramme (Abbildung 6-25) gemacht wurde.

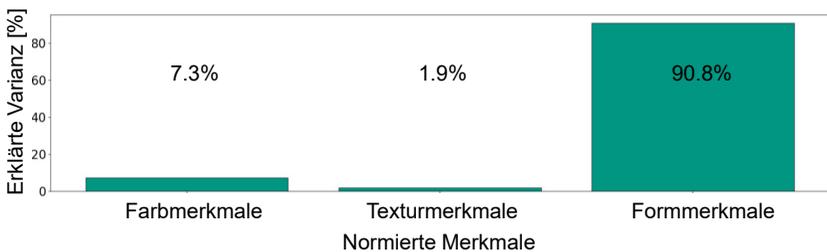


Abbildung 6-34 Normierter Anteil der Bildmerkmalskategorien an der Gesamtvarianz.

Um nachzuvollziehen, wie der Entscheidungsbaum eine Entscheidung trifft, zeigt Abbildung 6-35 die Entscheidungspfade eines Entscheidungsbaumes für vier charakteristische Bilddaten auf. Die Ergebnisse zeigen, dass die Entscheidungen klar nachvollzogen und interpretiert werden können. Die Bilder, die keinen Pitting enthalten, folgen demselben Pfad im Baum – unabhängig davon, ob eine Verschmutzung vorhanden ist. Weiterhin wird als wichtigstes Entscheidungsmerkmal in der ersten Ebene des Baumes geprüft, ob ein Bild im Quadrat 0:3, d. h. im rechten oberen Quadrat in Abbildung 6-29, eine bestimmte Anzahl an Key-Points überschreitet.

Zusammenfassend konnte in den vorherigen Abschnitten gezeigt werden, dass für die Klassifikation von Bilddaten charakteristische Merkmale definiert und mittels geeigneter Algorithmen aus den Bilddaten extrahiert werden können. Dabei wurden die Ansätze des Clustured-Colour-Share sowie der Key-Points-Per-Sub-Region erstmals im Rahmen dieser Arbeit entwickelt und angewandt. Die qualitative Relevanz von Merkmalen, wie sie von Domänenexpert*innen eingeschätzt wird, konnte quantifiziert werden. Weiterhin entsprechen die vom Entscheidungsbaum zur Klassifikation abgeleiteten Regeln den domänenbasierten Heuristiken.

Nachdem gezeigt wurde, dass es möglich ist, mittels der definierten Merkmale eine Klassifikation vorzunehmen, soll im folgenden Abschnitt untersucht werden, welche Klassifikationsgenauigkeit erreicht werden kann, wenn die Merkmale automatisiert mittels Deep-Learning-Verfahren aus den Bilddaten extrahiert werden.

6.2.3 Klassifikation mittels Convolutional-Neural-Network-Architekturen

In den vorausgehenden Abschnitten wurde das Wachstum der Pittings auf dem KGT empirisch analysiert und es wurden charakteristische Merkmale auf Basis von Domänenwissen abgeleitet. Weiterhin wurde gezeigt, dass mittels dieser Merkmale eine zuverlässige Klassifikation der Bilddaten möglich ist, und die Relevanz der definierten Merkmale für die Klassifikation wurde untersucht.

Um eine automatisierte Klassifikation der Bilddaten möglichst zuverlässig zu gestalten, werden in diesem Abschnitt Deep-Learning-basierte Verfahren mit dem Ziel untersucht, eine möglichst hohe Klassifikationsgenauigkeit zu erzielen. Hierzu werden zunächst aus dem Stand der Technik CNN-Architekturen zur Klassifikation ausgewählt und untersucht. Aufbauend darauf werden in einer Parameterstudie insgesamt 486 CNN-Architekturen auf einer speziell für maschinelles Lernen ausgelegten Rechnerarchitektur am wbk Institut für Produktionstechnik des KIT trainiert und die Ergebnisse untersucht. Aufbauend auf der Parameterstudie wird der Einfluss von Data Augmentation auf die Leistung des besten Modelles untersucht. Zur Absicherung der Modellergebnisse werden die Bereiche in den Bilddaten untersucht, die für eine Klassifikation ausschlaggebend sind.

6.2.3.1 Vorauswahl von Klassifikationsmodellen

Der Fokus der zu untersuchenden Modellarchitekturen liegt auf Modellen, die sich der Convolution-Rechenoperation bedienen und sich damit, wie in Abschnitt 2.4.5 dargestellt, der Familie der CNN zuordnen lassen. CNN eignen sich nach (Aggarwal 2019) besonders für die Klassifikation von Bilddaten, da die Merkmale in Bilddaten starke räumliche Abhängigkeiten aufweisen. CNN-Architekturen wurden, wie im Stand der Technik beschrieben, in technischen Domänen bspw. von (Dong, Taylor & Cootes 2018) zur Klassifikation von Defekten auf Schweißstellen, von (Balzategui & Eciolaza et al. 2019) zur Detektion von Fehlstellen auf Solarpanelen sowie von (Bozic, Tabernik & Skocaj 2021) zur Klassifikation von Defekten auf Stahlbrammen basierend auf dem Severstal-Datensatz¹⁷ verwendet. Die grundsätzliche Funktionsfähigkeit von CNN zur Klassifikation von Defekten konnte bereits in Vorarbeiten nachgewiesen werden, weshalb eine Untersuchung der Modelle auf Bilddaten von Pittings als sinnvoll eingeschätzt werden kann.

Im Folgenden werden die Architekturen von performanten Modellen, wie sie auch bei der von (Russakovsky & Deng et al. 2015) begründeten, jährlich durchgeführten Image Net Large Scale Visual Recognition Challenge (ILSVRC) zum Einsatz kommen, auf die vorliegenden Daten angepasst und untersucht. Als Grundarchitektur wird die LeNet-Architektur von Yann LeCun (Lecun & Bottou et al. 1998) untersucht, die eine der ersten performanten CNN-Architekturen darstellt und zur Vergleichbarkeit untersucht werden soll. Als zweite Architektur wird die AlexNet-Architektur untersucht. Die AlexNet-Architektur (Krizhevsky, Sutskever & Hinton 2012) konnte erstmals die Fehlerrate im Image Net Contest von bisher 25 auf 15.4 % senken, weshalb die Fähigkeit dieser Modellarchitektur weiter untersucht werden soll. Als dritte und vierte Architektur werden die VGG16- und VGG19-Architekturen (Simonyan & Zisserman 2015) untersucht, bei denen es sich um Weiterentwicklungen der AlexNet-Architektur handelt, die sich hauptsächlich in ihrer Tiefe (16 und 19 Convolutional Layer) von dieser unterscheiden. Die fünfte zu untersuchende Architektur ist die GoogLeNet-Architektur (Szegedy & Liu et al. 2015), welche die ILSVRC-Top-fünf-Fehlerrate auf 6.7 % reduzieren konnte und damit erstmals der menschlichen Leistungsfähigkeit nahe kam. Die Neuerungen der GoogLeNet-Architektur beruhen hauptsächlich auf der Verwendung von sogenannten Inception-Schichten, die auch als „network within a network“ (Aggarwal 2019, S. 339) beschrieben werden. Die sechste zu untersuchende Architektur ist die ResNet-Architektur (He & Zhang et al. 2016), die sich hauptsächlich durch ihre größere Tiefe sowie das Einführen von Abkürzungen, den sogenannten Skip-Connections, zwischen den Schichten zum Informationstransfer auszeichnet und mit einer Top-fünf-Fehlerrate von 3.6 % erstmals die Marke der menschlichen Leistungsfähigkeit überschritt.

Als Ergänzung zum Stand der Technik wird als siebte Architektur eine CNN-Architektur entwickelt und an den Dantesatz angepasst. Für diese Architektur wird eine ausführliche Parameterstudie durchgeführt. Die Ergebnisse werden den Ergebnissen der Modelle aus dem Stand der Technik gegenübergestellt und diskutiert. Der Einfluss von Data Augmentation auf die Ergebnisse wird untersucht. Die Ergebnisse der Arbeiten werden im Folgenden dargestellt.

6.2.3.2 Voruntersuchungen zur Auswahl von Stand-der-Technik-Architekturen

Zur Untersuchung und Bewertung der Modelle wird in einem dreistufigen Verfahren vorgegangen: Zunächst werden Voruntersuchungen mit den beschriebenen Modellen aus dem Stand der Technik durchgeführt, bei denen die Modelle ohne Anpassung der

Hyper- und *Modellparameter* trainiert werden. Hierdurch soll die grundsätzliche Eignung der Modellarchitekturen für die vorliegende Klassifikationsaufgabe untersucht werden. Im zweiten Schritt werden aus diesen Modellen die performantesten Modelle ausgewählt und eine Optimierung der Parameter wird durchgeführt. In einem dritten Schritt wird eine weitere CNN-Architektur aufgebaut und eine Hyperparameter-Studie mittels eines Versuchsplans durchgeführt. Der Einfluss von Data Augmentation auf die Leistungsfähigkeit des besten Modelles wird im Anschluss untersucht.

Für die Modellbildung wird der in Abschnitt 5.4 beschriebene Datensatz verwendet. Die Ergebnisse der Vorversuche sind in Abbildung 6-36 dargestellt.

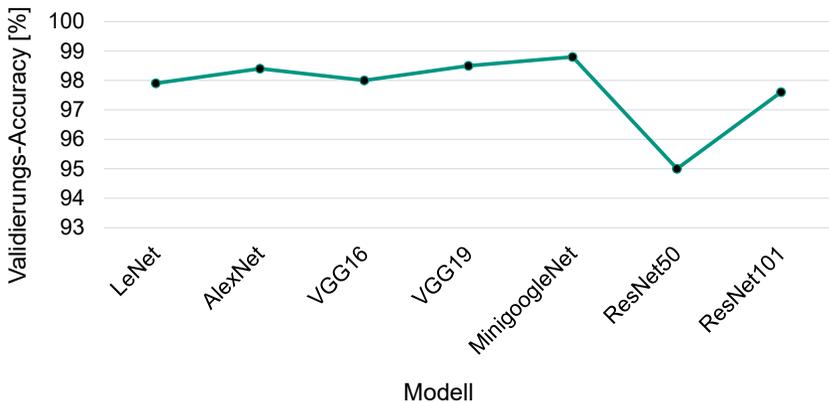


Abbildung 6-36 Gegenüberstellung der Validierungs-Accuracy von verschiedenen Stand-der-Technik-Architekturen.

Als Ergebnis des ersten Schrittes ist ersichtlich, dass alle Modelle eine hohe Validierungs-Accuracy aufweisen. Die LeNet-Architektur liefert im Vergleich mit den übrigen Modellen ein durchschnittliches Ergebnis. Die AlexNet-Architektur sowie die VGG-Architekturen liegen in ihren Ergebnissen eng beisammen. Auch die GoogleNet-Architektur liegt gemessen an der Validierungs-Accuracy zwar nahe bei den vorherigen Modellen, sie ist allerdings in der Klassifikation aufgrund der geringeren Modellgröße um den Faktor 3 schneller als die VGG-Architekturen. Die ResNet-Architekturen liegen um zwei Prozentpunkte auseinander, wobei die ResNet101-Architektur mit den vorherigen Modellen vergleichbare Ergebnisse aufweist. Basierend auf diesen Ergebnissen wird als Vergleichsbasis und aufgrund der hohen Trainings-Accuracy die LeNet-Architektur in

die zweite Testphase übernommen. Aufgrund der hohen Leistungsfähigkeit und der vielversprechenden Inception-Schichten wird als zweites Modell die MinigoogLeNet-Architektur zur Optimierung ausgewählt. Als drittes Modell wird das VGG19-Netzwerk für weitere Untersuchungen ausgewählt. Als vierte Architektur wird die ResNet101-Architektur aufgrund ihrer vielversprechenden Skip-Connections und der Tatsache, dass diese erstmals die menschliche Leistungsfähigkeit bei Klassifikationsaufgaben übertroffen hat, ausgewählt.

Als Ergebnis wurde auf Basis des Standes der Technik eine erste Auswahl an performanten Modellen getroffen. Für diese Modelle wird im folgenden Abschnitt eine Optimierung vorgenommen.

6.2.3.3 Hyperparameter-Optimierung der Stand-der-Technik-Architekturen

Im Anschluss an die erste Versuchsphase wird im zweiten Schritt eine Optimierung der Hyperparameter für die ausgewählten Modelle vorgenommen.

Die grundsätzlichen Strukturen der Modelle werden hierbei nicht verändert, da diese nach dem Stand der Technik als weitestgehend optimiert angesehen werden können. Zur Optimierung der Hyperparameter wird zunächst Dropout eingeführt, da dieses zur Vermeidung von Overfitting, wie es bspw. bei der ResNet50- und LeNet-Architektur in den Voruntersuchungen vermutet wird, eingesetzt werden kann. Der Wertebereich für die Dropout-Werte wird mit $d \in \{0.2, 0.5, 0.8\}$ angesetzt. Als zweiter Hyperparameter wird die *Learning Rate* der Modelle mit $l \in \{0.0001, 0.001, 0.01, 0.1\}$ gewählt. Die Learning Rate hat einen entscheidenden Einfluss auf die Anpassung der Modellgewichte während des Trainings und damit auf die zu erreichende Genauigkeit. Als dritter Hyperparameter wird der zu verwendende Optimierungsalgorithmus untersucht. Der Optimierer wird mit $o \in \{\text{Adam}, \text{Nadam}, \text{Adamax}\}$ gewählt. Er entspricht hierbei dem Algorithmus, mit dem die Gewichte im Modell angepasst werden, und hat damit einen wesentlichen Einfluss auf das Training des Modelles. Für die Hyperparameter-Suche werden alle Modelle für zehn Iterationen pro Parameterwahl trainiert. Die Trainingsdaten werden analog zu den Vorversuchen mit einer Aufteilung von 80 % Trainingsdaten und 20 % Validierungsdaten gewählt. Für die performanteste Modell-Parameter-Kombination wird eine 25-Fold-Cross-Validation durchgeführt, um die Ergebnisse abzusichern. In diesem Versuch wird die Architektur für 50 Iterationen trainiert.

Die ausführlichen Ergebnisse der Modelle sind in Abbildung 0-8 im Anhang dargestellt. Die Boxplots der Validierungs-Accuracy der Modelle, gemittelt über alle Hyperparameter-Werte, sind zur besseren Veranschaulichung in Abbildung 6-37 dargestellt. Der Median der Modelle liegt bei allen Modellen mit Ausnahme des VGG19-Modells über 90 %: 98.60 % für das GoogLeNet, 94.26 % für das LeNet, 50.00 % für das VGG19 sowie 96.21 % für das ResNet. Die Validierungsgenauigkeiten liegen bei allen Modellen eng zusammen – mit nur wenigen Ausreißern nach oben und unten, was an den Quantilen der Boxplots zu erkennen ist. Eine Ausnahme bildet das VGG19-Modell, dessen Ergebnisse sehr stark über die Hyperparameterbereiche schwanken. Anhand der Daten ist erkennbar, dass die Validierungsgenauigkeit des VGG19 bei einer Lernrate > 0.0001 stets bei 0.5 liegt, was darauf schließen lässt, dass die Anpassung der Parameter bei diesen Lernraten nicht korrekt erfolgt. Eine Ausnahme stellt die Kombination Lernrate = 0.001, Dropout = 0.5 und Optimierer = Nadam dar, bei der eine Validierungs-Accuracy von 88.5 % erreicht werden konnte. Weitere Ausnahmen, bei denen teils hohe Validierungsgenauigkeiten von > 96 % erreicht werden konnten, stellen Kombinationen dar, die den Adagrad-Optimierer nutzen. Ein Beispiel hierfür ist die Kombination Lernrate = 0.01, Dropout = 0.2, Optimierer = Adagrad, mit der eine Genauigkeit von 96.1 % erreicht werden konnte. Abbildung 6-37 rechts zeigt zur besseren Veranschaulichung die Ergebnisse der Modelle ohne das VGG19.

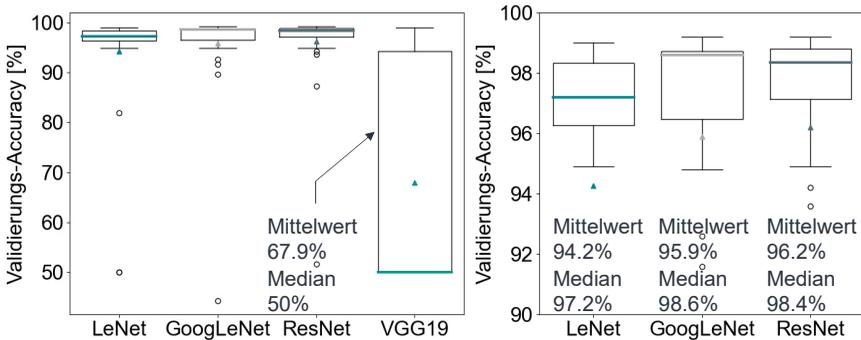


Abbildung 6-37 Boxplots der trainierten CNN-Architekturen.

Den höchsten Median über alle Hyperparameter-Werte hat mit 98.60 % das GoogLeNet. Den höchsten erreichten Wert der Validierungsgenauigkeit besitzt ebenfalls das GoogLeNet zusammen mit dem ResNet mit 99.19 %. Da bei der Optimierung der Hyperparameter nach der Modell-Hyperparameter-Kombination mit der höchsten Validierungsgenauigkeit gesucht wird, ist diese Modell-Hyperparameter-Kombination als die leistungsstärkste Kombination einzustufen und es wird aufgrund des höheren Medians das GoogLeNet als performantestes Modell ausgewählt. Die Validierungsgenauigkeit von 99.19 % wurde beim ResNet mit der Hyperparameterkombination {Learning Rate = 0.001; Dropout = 0.8 Optimierer = Adagrad} erzielt. Das GoogLeNet erzielte diese Validierungsgenauigkeit mit der Hyperparameterkombination {Learning Rate = 0.1; Dropout = 0.2 Optimierer = Adagrad}. Die Analyse der ausführlichen Ergebnisse in Abbildung 0-8 zeigt, dass bei den Werten 0.2 und 0.8 für die Dropout-Rate die besten Ergebnisse und die geringste Streuung erreicht werden. Außerdem zeigt sich, dass sich die Ergebnisse bei einer Learning Rate von 0.1 nahe 98 % konzentrieren, wohingegen sie bei einer Learning Rate von 0.01 stärker streuen. Das GoogLeNet erzielt die höchste Leistungsfähigkeit bei einer Learning Rate von 0.1. Die Streuung der Learning Rate 0.0001 ist nach der Learning Rate 0.1 die geringste, wobei das beste Ergebnis von einem ResNet erzielt wird.

Die Ergebnisse für die Wahl des Optimierungsalgorithmus zeigen, dass die Genauigkeiten bei Verwendung des Adam-Optimierers am wenigsten streuen und die besten Ergebnisse mit dem Adagrad- und Nadam-Optimierer erreicht werden können. Auf Basis der Analyse können die Ergebnisse als nachvollziehbar eingestuft werden.

Zusammenfassend ist ersichtlich, dass die Modelle insgesamt größtenteils > 90 % Validierungsgenauigkeit erreichen und damit als sehr genau eingestuft werden können. Weiterhin führen bestimmte Parameterkombinationen zu einer Häufung von Modellen mit hoher Validierungs-Accuracy, womit diese Parameter als besonders relevant für eine hohe Accuracy eingestuft werden können.

Zur Absicherung der Ergebnisse des GoogLeNet wird für das Modell mit der besten Modell-Hyperparameter-Kombination eine 25-fache Cross Validation durchgeführt. Die Ergebnisse sind in Abbildung 6-38 links dargestellt. Es ist ersichtlich, dass die mittlere Validierungsgenauigkeit von 99.22 % die im vorherigen Versuch ermittelte Genauigkeit bestätigt.

Zusammenfassend lässt sich feststellen, dass mittels der Architekturen aus dem Stand der Technik eine zuverlässige Klassifikation der Bilddaten möglich ist, wobei die besten Ergebnisse mit der GoogLeNet-Architektur erzielt werden können. Da dieses Modell ursprünglich nicht für den Kontext der Klassifikation von Fehlstellen entwickelt wurde, wird im nächsten Abschnitt eine versuchsplanbasierte CNN-Architektur untersucht.

6.2.3.4 Versuchsplanbasierte CNN-Architektur

Zum Vergleich der Leistungsfähigkeit der höchst performanten Modell-Parameter-Kombination des GoogLeNet-Ansatzes werden in einem nächsten Schritt mittels eines Versuchsplanes weitere CNN-Modelle gebildet und deren Leistungsfähigkeit wird im Rahmen einer Parameterstudie untersucht. Das Ziel ist es, eine möglichst hohe Klassifikationsgenauigkeit zu erzielen. Damit wird eine Methode geschaffen, mittels welcher flexibel CNN-Architekturen generiert werden können. Dies erlaubt den Aufbau und das Training von Modellen mit einer unterschiedlichen Anzahl an Convolution- und Pooling-Schichten. Im Rahmen der folgenden Untersuchung wird, anders als bei der Untersuchung der Modelle aus dem Stand der Technik, auch die Grundarchitektur der Modelle verändert. Hiermit ist es möglich, CNN-Modelle automatisiert mit unterschiedlichen Parameterkombinationen aufzubauen und zu trainieren. Es soll gezeigt werden, dass eine neue Architektur zur Klassifikation von Bildern des KGTs geschaffen werden kann, welche die bestehenden Modelle aus dem Stand der Technik übertrifft.

Da die Modelle aus dem Stand der Technik insgesamt sehr gute Ergebnisse erzielen, wird die Wahl der Parameterräume an die oben beschriebenen Stand-der-Technik-Architekturen angelehnt. Die Parameterräume werden hierbei gewählt, wie in Tabelle 6-4 dargestellt ist. Als Optimierer wird der Adam-Optimierer für alle Versuche gewählt, da dieser in den Voruntersuchungen die stabilsten Ergebnisse erzielte. Für den Versuchsplan ergeben sich somit insgesamt 486 CNN-Kombinationen, die im Folgenden aufgebaut und trainiert werden. Für das Training der Modelle werden dieselben Daten wie in den Vorversuchen verwendet.

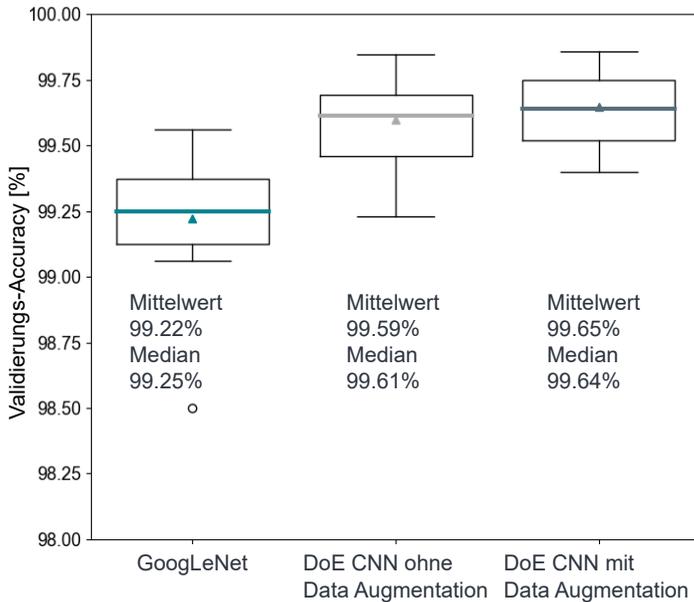


Abbildung 6-38 Validierungs-Accuracy bei 25-facher Cross Validation.

Für das Training und die Validierung der Modelle wird ein 80/20-Split angewandt. Die Anzahl an Iterationen wird mit 50 gewählt und als Klassifikationsmodell wird ein fixes Neuronales Netz mit fünf Schichten und zehn Neuronen pro Schicht an die Convolutional Base angeschlossen. Der Aufbau des Neuronale Netzes entspricht damit dem Aufbau, wie er auch in Abschnitt 6.2.2 zur Klassifikation der Merkmale basierend auf Domänenwissen verwendet wurde. Als Aktivierungsfunktion wird die ReLu-Aktivierungsfunktion verwendet. Nach der dritten Schicht des Klassifikationsnetzes ist ein Dropout Layer implementiert, das die Dropout-Werte aus dem Versuchsplan erhält.

Tabelle 6-4 DoE-Parameterraum für die Erstellung von CNN-Architekturen.

Parameter	Parameterwerte
Anzahl Convolutional Layer	{3, 5, 7}
Filtergröße	{2, 3}
Dropout Rate	{0.3, 0.1, 0}
Pooling-Größe ²⁷	{2, 3, 0}
Filteranzahl	{64, 128, 256}
Lernrate	{0.0001, 0.001, 0.1}

Die ausführlichen Ergebnisse in Abhängigkeit der Modellparameter sind in Abbildung 0-7 im Anhang dargestellt. Es ist ersichtlich, dass für die vorliegende Klassifikationsaufgabe die maximale Validierungsgenauigkeit mit steigender Learning Rate abnimmt und die besten Ergebnisse mit einer Learning Rate von 0.0001 erzielt werden können, wohingegen sich die Ergebnisse bei einer Learning Rate von 0.1 auf zwei Punkte im unteren 90%-Bereich konzentrieren. Ebenfalls steigt die maximale Validierungsgenauigkeit mit steigender Anzahl an Convolutional Layers und die besten Ergebnisse können mit sieben Convolutional Layers erzielt werden. Die Wahl des Pooling-Parameters hat ebenfalls Einfluss auf die maximal zu erreichende Validierungsgenauigkeit, wobei kein Pooling zu niedrigen Validierungsgenauigkeiten führt. Dies kann mit einem Overfitting der Modelle auf die Trainingsdaten erklärt werden, das durch die Integration von Pooling verringert wird. Das beste Modell erreicht damit eine Validierungsgenauigkeit von 99.8 % mit der Parameterkombination {Convolutional Layer = 7, Filtergröße = 3, Dropout Rate = 0.1, Pooling Größe = 2, Filteranzahl = 128, Lernrate = 0.0001}.

Für diese Modell-Parameter-Kombination wird im Folgenden eine 25-fache Cross Validation durchgeführt, um die Ergebnisse abzusichern. Die Ergebnisse sind in Abbildung 6-38 (Mitte) den Ergebnissen des GoogLeNet-Modelles gegenübergestellt. Es ist ersichtlich, dass die mittlere Validierungsgenauigkeit von 99.59 % das vorherige Ergebnis bestätigt und damit das Ergebnis des Stand-der-Technik-Modells übertroffen werden kann.

Als Fazit zur Klassifikation konnte somit gezeigt werden, dass ein mittels eines Versuchsplans generiertes CNN, zugeschnitten auf die vorliegende Klassifikationsaufgabe, die Modelle im Stand der Technik übertreffen kann.

²⁷ Hinter dem dritten und sechsten (sofern vorhanden) Convolutional Layer folgt ein Pooling Layer. Ist die Pooling-Größe im Versuchsplan mit 0 spezifiziert, so erfolgt kein Pooling. Für das Pooling wird max Pooling verwendet.

Für die bisherigen Untersuchungen wurde der unveränderte Bilddatensatz verwendet, ohne diesen künstlich zu vergrößern (Data Augmentation). Im folgenden Abschnitt soll der Einfluss der künstlichen Vergrößerung des Datensatzes mittels Data Augmentation auf das Modellergebnis untersucht werden.

6.2.3.5 Vergrößerung der Datenmenge durch Data Augmentation

Zur Untersuchung der Möglichkeit, die Modellgüte weiter zu steigern, wird zusätzlich der Einfluss von Data Augmentation auf die beste Modell-Parameter-Kombination untersucht. Zur Augmentation werden die Bilddaten X , Bildtransformationen θ einer Menge an Bildtransformationen Θ unterzogen. Die Abbildung lautet damit: $\theta: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$; $X \mapsto \theta(X)$; $\theta \in \Theta$. Die Transformationen werden hierbei so ausgewählt, dass sie Zustände darstellen, die auch in realen Umgebungen durch die natürlichen Varianzen im Prozess auftreten können und damit aus Domänensicht eine sinnvolle Erweiterung der Bilddaten darstellen. Der Datensatz wird damit künstlich um realistische Abwandlungen der Bilddaten erweitert und vergrößert. Durch diese Vergrößerung des Datensatzes soll die Leistungsfähigkeit des Modelles verbessert werden. Domänenrelevante Änderungen in den Bilddaten, die abgedeckt werden sollen, sind zusammen mit den angewandten Techniken zur Augmentierung in Tabelle 6-5 dargestellt.

Tabelle 6-5 Domänenspezifische Varianz in den Bilddaten und angewandte Techniken zur Data Augmentation.

Domänenspezifische Varianz	Technik zur Augmentierung
Änderung der Gewindesteigung	Rotation im Bereich von $[-5^\circ, +5^\circ]$
Änderung des Kamerawinkels	Transformation der Perspektive (0,0.15)
Verschmutzung und Partikel	Laplace Rauschen (0.03*255 pro Kanal)
Unschärfe Bereiche	Hinzufügen von Unschärfe
Schädigungen auf beiden Flankenseiten	Horizontales- und vertikales Spiegeln
Helligkeits-, Farb- und Kontrastunterschiede	Multiplikation und Addition von RGB Werten $\{0.7, \dots, 1.3\}$. CLAHE-Algorithmus.

Zusätzlich wird zur Erhöhung des Kontrastes im Bild der CLAHE-Algorithmus (Contrast Limited Adaptive Histogram Equalization) im Bereich $[1, 2, \dots, 10]$ nach (Zuiderveld

2000) angewandt. Der CLAHE-Algorithmus wird häufig in medizinischen Anwendungen zur Erhöhung des Kontrastes in Gewebeaufnahmen verwendet. Abbildung 6-39 zeigt, dass die Kontrastanpassung in den untersuchten Aufnahmen von Pittings – ähnlich den medizinischen Anwendungen – ebenfalls als zielführend eingeschätzt werden kann.

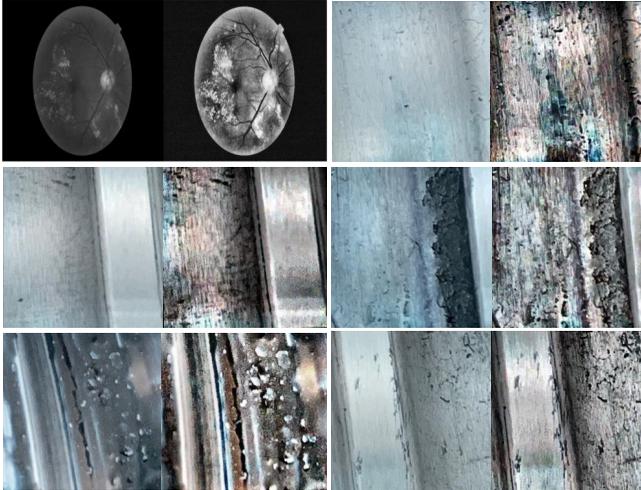


Abbildung 6-39 Anwendung des CLAHE-Algorithmus zur Kontrastanpassung auf Bilddaten. Erste Zeile links Darstellung der menschlichen Retina, sonst KGT-Bilddaten. Links jeweils Originalbild, rechts CLAHE-transformiertes Bild. Bildpaare sind getrennt durch weiße Linien. Abbildung der Retina nach (Bhat & Patil 2014).

Für die Modellbildung werden die Daten jeweils mittels der beschriebenen Techniken zur Augmentierung erweitert und dem Modell für das Training übergeben. Die Validierungsdaten werden nicht augmentiert, um die Leistungsfähigkeit auf Realdaten zu testen. Zur Überprüfung der Wirksamkeit und zur statistischen Absicherung wird für das Modell ebenfalls eine 25-fache Cross Validation durchgeführt. Abbildung 6-38 rechts zeigt die Ergebnisse unter Verwendung von Data Augmentation. Mittels Data Augmentation kann eine maximale Validierungsgenauigkeit von 99.86 % erreicht werden. Die erreichte Validierungsgenauigkeit übersteigt damit die Validierungsgenauigkeit des DoE-CNN ohne Data Augmentation. Es ist ersichtlich, dass die mittlere Validierungsgenauigkeit von 99.65 % das beobachtete maximale Ergebnis bestätigt und über dem Ergebnis des DoE-CNN ohne Data Augmentation (99.59 %) sowie des GoogleNet-Modelles (99.22 %) liegt.

Zusammenfassend kann durch die Integration von Data Augmentation die maximale Validierungsgenauigkeit für die Klassifikation von Pitting-Bildern erhöht werden. Im Folgenden soll eine weitere visuelle Validierung der Ergebnisse erfolgen.

6.2.3.6 Visuelle Validierung des Ansatzes

Zum Test der ausgewählten Modell-Parameter-Kombination wird das Modell abschließend an neuen Bildern eines verschlissenen KGT getestet (Abbildung 0-9 im Anhang).

Die Bilder stellen alle auftretenden Szenarien inklusive starker Verschmutzung und teilweise überdeckten Schädigungen dar und sind auch für Domänenexpert*innen schwer klassifizierbare Bilddaten. Die Klassifikation wird mittels eines Sliding-Window-Ansatzes durchgeführt, bei dem das Bild in 150x150-Pixel-Rastern überlaufen wird. Die Ergebnisse der Klassifikation sind in Abbildung 6-40 dargestellt.

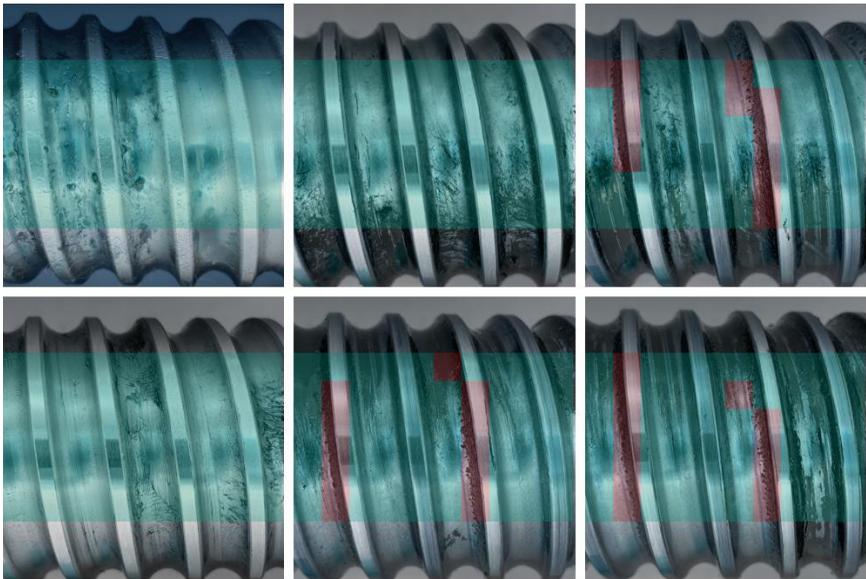


Abbildung 6-40 Klassifikationsergebnisse des Versuchsplanbasierten CNN unter Verwendung von Data Augmentation.

Grün überblendete Stellen wurden vom Modell als nicht fehlerhaft und rot überblendete Stellen wurden als fehlerhaft klassifiziert. Es ist ersichtlich, dass das Modell in der Lage

ist, fehlerhafte Stellen zuverlässig von nicht fehlerhaften Stellen im Bild zu unterscheiden. Das Modell ist robust gegenüber Verschmutzungen und fehlklassifiziert Verschmutzungen in keinem Fall als Schädigung. So ist erkennbar, dass das Modell die ersten beiden Bilder in der oberen Reihe korrekt klassifiziert, obwohl hier auch für die Domänenexpert*innen nicht triviale Bereiche enthalten sind. Im dritten Bild der oberen Reihe wird lediglich im unteren Bereich der ersten Laufbahn eine vom Schmierstoff überdeckte Schädigung nicht erkannt, wobei hier von einem Extremfall gesprochen werden kann. Diese Schädigung in Kombination mit Schmierstoff ähnelt sehr stark einer Stelle, bei der lediglich Verschmutzung vorherrschend ist. Die Verschmutzung kann hierbei als Rauschen in den Bilddaten betrachtet werden. Dennoch findet das Modell auch in stark verschmutzten Bereichen zuverlässig Pittings, wie im zweiten und dritten Bild der unteren Reihe erkennbar ist. Die in Abschnitt 6.1 beschriebenen Ergebnisse der Verschleißversuche zeigen, dass der Verschmutzungsgrad, wie er in den Testdaten vorhanden ist, meist nur kurz vor Ausfall der Komponente auftritt.

Als Ergebnis kann somit gezeigt werden, dass mit dem untersuchten Ansatz, in dem ein CNN mittels eines Versuchsplanes gebildet und mit einem durch Data-Augmentation-Techniken vergrößerten Datensatz trainiert wurde, zuverlässig Fehlstellen auf KGT identifiziert werden können. Das Modell schneidet hierbei für die vorliegende Aufgabe besser ab als die untersuchten Stand-der-Technik-Algorithmen.

Um sicherzustellen, dass die untersuchten Modelle zur Klassifikation sinnvolle Merkmale in den Bilddaten verwendet haben, folgt eine abschließende Untersuchung über die Bereiche in den Bilddaten, die für eine Klassifikation besonders relevant sind.

6.2.3.7 Absicherung der erlernten Merkmale

In den vorherigen Abschnitten konnte die zuverlässige Klassifikation mittels Deep-Learning-Architekturen gezeigt werden. Es ist allerdings noch nicht abschließend geklärt, welche Bereiche in den Bilddaten für das Modell für eine Klassifikation besonders relevant sind.

Aus dem Stand der Technik ist bekannt, dass Deep-Learning-Architekturen Merkmale in den Bilddaten zur Klassifikation verwenden können, die nicht mit der eigentlichen Klassifikationsaufgabe zusammenhängen. So konnten (Lapuschkina & Wäldchen et al. 2019) zeigen, dass bei der Klassifikation von bestimmten Pferdebildern nicht die Merkmale der Pferde erlernt wurden, sondern Markierungen im Bild, die von den Fotografen händisch eingefügt wurden. Das Modell erfüllte durch Erlernen dieser Markierungen die

Klassifikationsaufgabe, obwohl es hierzu nicht die den Pferden spezifischen Merkmale erlernte. Wurden die Markierungen entfernt, so konnte das Modell die Bilddaten nicht mehr korrekt klassifizieren.

Die Hypothese (die es zu widerlegen gilt) ist, dass das verwendete CNN Merkmale in den Bilddaten der Pittings erlernt, anhand derer zwar eine Klassifikation vorgenommen werden kann, die aber nicht den Merkmalen der Pittings zuzuordnen sind. Hierzu wird ein sogenannter Heat-Map-Ansatz nach (Selvaraju & Cogswell et al. 2020) und (Chollet 2018) verwendet, bei dem die Bereiche in einem Bild hervorgehoben werden, die für die Klassifikation des Neuronalen Netzes ausschlaggebend waren. Hierbei werden die Gradienten des CNN nach der Klasse Pitting berechnet. Die Gradienten der Bereiche, die für eine Klassifikation besonders relevant sind, werden hoch sein und sich damit in der Heat-Map abzeichnen.



Abbildung 6-41 Darstellung der Relevanz der Bildbereiche zur Klassifikation der Bilder mittels eines Heat-Map-Ansatzes.

Zur Überprüfung wurden für 15 zufällig gewählte Bilddaten mit Pittings Klassifikationen durchgeführt. Die absoluten Werte der Gradienten wurden als Heat-Map über die Bilder gelegt, wobei hohe Gradienten in Rot abgebildet sind. Die Ergebnisse sind in Abbildung 6-41 dargestellt.

Es ist ersichtlich, dass Bereiche, in denen Pittings vorherrschend sind, besonders hervorgehoben werden, wohingegen die Bereiche, in denen keine charakteristischen Merkmale vorhanden sind, weniger stark betrachtet werden. Somit kann belegt werden, dass das Modell für Pittings charakteristische Formen für die Klassifikation verwendet.

Als Fazit lässt sich somit die effektive Klassifikation von Pitting-Bilddaten mittels Deep-Learning-basierten Verfahren bestätigen.

Neben der Effektivität ist grundsätzlich auch die Effizienz in technischen Domänen von Relevanz. Übertragen auf den Kontext der Klassifikation soll im folgenden Abschnitt ein Exkurs zur dateneffizienten Klassifikation in technischen Domänen durchgeführt werden.

6.2.4 Dateneffiziente Klassifikation in technischen Domänen

Dieser Abschnitt steht nicht in direktem Zusammenhang mit den vorangegangenen Abschnitten der empirischen Analyse sowie mit der automatischen Identifikation von Pittings in Bilddaten. Nichtsdestoweniger hat die Dateneffizienz in technischen Domänen und auch für die Identifikation von Pittings eine große Bedeutung und soll darum behandelt werden. Um den inhaltlichen Faden der Dissertation wieder aufzugreifen, folgt im Anschluss an diesen Abschnitt die Untersuchung eines Ansatzes zur Prognose der Entwicklung von Pittings am KGT.

Im Rahmen der versuchsplanbasierten Untersuchung von CNN-Architekturen konnte gezeigt werden, dass es möglich ist, Pittings zuverlässig zu klassifizieren. Die zum Training und zur Validierung der Modelle verwendeten Daten wurden über lange Zeiträume hinweg aufgenommen und lagen in moderaten Mengen vor. Eine bislang wenig betrachtete, aber große Herausforderung in der technischen Domäne sind die oftmals geringen zur Verfügung stehenden Datenmengen (Fink & Wang et al. 2020) zum Training von Machine-Learning-Modellen in der technischen Domäne (Bengio, LeCun & Hinton 2021), (Li, Liu & Bilen 2021), (Wang, Michau & Fink 2019).

Im Bereich der Klassifikation mittels Deep-Learning-Modellen heben (Chollet 2018) und (Aggarwal 2019) die Notwendigkeit von großen Datenmengen für das zuverlässige Training von Deep-Learning-Modellen hervor. Gleichzeitig stellen (He & Song et al. 2019), (Ren, Hung & Tan 2018) und (Koch 2015) die Problematik dar, dass in Domänen, in denen Fehlerfälle entweder selten sind oder die Generierung von Daten teuer ist, Trainingsdaten meist nur in begrenzten Mengen vorhanden sind. Solche Domänen

finden sich bspw. im Bereich der Medizin bei der Klassifikation von Tumorzellen, aber auch in technischen Domänen, bspw. bei der Klassifikation von Defekten auf metallischen Oberflächen.²⁸ Am Beispiel des vorliegenden KGT müssen Trainingsdaten entweder im Feld aufgenommen oder im Labor mittels Lebensdauerexperimenten generiert werden. In beiden Fällen sind die so zu generierenden Datenmengen sowohl aufgrund der Dauer als auch wegen der Kosten zur Generierung der Daten begrenzt. Der vorgestellte Ansatz (das sogenannte Siamese Basis Function Network) adressiert ebendieses Spannungsfeld der Klassifikation unter begrenzten Datenmengen. Der Siamese-Basis-Function-Network-Ansatz ist so im Stand der Technik noch nicht vorhanden und stellt damit eine neuartige Methodik dar, die im Rahmen dieser Arbeit entwickelt wurde. Der Ansatz basiert hierbei auf Grundbausteinen, die bereits im Stand der Technik vorhanden sind und auf welche weiter unten eingegangen wird. Die Wirksamkeit des untersuchten Ansatzes wird sowohl am KGT-Datensatz (Schlagenhauf 2021) als auch anhand mehrerer gängiger Datensätze aus dem Stand der Technik (NEU-Datensatz (Song & Yan 2013), MNIST-Datensatz²⁹, CIFAR-Datensatz³⁰, TEX-Datensatz (Bergmann & Fauser et al. 2019)) aus technischen sowie nichttechnischen Domänen getestet, um seine Übertragbarkeit zu überprüfen.

Der Siamese-Basis-Function-Ansatz (SBF-Net-Ansatz) beruht auf dem Vergleich und der anschließenden Klassifikation von Bilddaten mittels Ähnlichkeitsmaßen, wie sie in den Grundlagen in Abschnitt 2.4.5.5 eingeführt wurden. Die diesem Abschnitt zugrunde liegenden Arbeiten finden sich in (Schlagenhauf, Yildirim & Brückner 2022) und (A_Yildirim 2021). Die Einzelkomponenten des Ansatzes sowie die Ergebnisse werden im Folgenden dargestellt.

6.2.4.1 Übersicht zur Architektur des SBF-Net-Ansatzes

Zunächst soll eine Übersicht zum Ansatz gegeben werden.

Menschen sind in der Lage, aus wenigen Daten eines bestimmten Objektes auf unterschiedliche Erscheinungsbilder desselben Objektes zu schließen und darüber hinaus auf neue Objekte zu generalisieren, die sich mit dem bekannten Objekt Merkmale teilen

²⁸ Sankesara, H. (2019), *N-Shot Learning: Learning More with Less Data*. <https://blog.floydhub.com/n-shot-learning/> [26.01.2021].

²⁹ Tensorflow (2022), *mnist*. <https://www.tensorflow.org/datasets/catalog/mnist> [26.01.2022].

³⁰ Tensorflow (2022), *cifar10*. <https://www.tensorflow.org/datasets/catalog/cifar10> [26.01.2022].

(Koch 2015). So kann ein Mensch bspw. aus einem einzelnen Fehlerbild auf einer bestimmten Komponente dasselbe Fehlerbild auf einer anders gearteten Komponente erkennen sowie über den spezifischen Fehler hinaus auf den Zustand der Komponente schließen. Der Mensch kann somit als sehr effizient in der Extraktion und Zuordnung von Merkmalen angesehen werden. Die klassischen Deep-Learning-Modelle sind zwar generell dazu in der Lage, benötigen aber eine große Anzahl an Trainingsdaten (Koch 2015). Dies stellt vor allem Domänen vor Herausforderungen, in denen wenige Daten vorhanden sind.

Der vorgestellte Ansatz hat zum Ziel, die Merkmalseffizienz von Deep-Learning-Modellen in der technischen Domäne zu erhöhen und dadurch eine dateneffiziente Klassifikation zu erreichen.

Der gewählte Ansatz basiert auf dem Siamese-Network-Ansatz nach (Koch 2015) sowie auf der Repräsentation von Bildern als d -dimensionale Merkmalsvektoren nach (Schroff, Kalenichenko & Philbin 2015), in denen ein $n \times m$ -dimensionales Bild als Vektor $x \in \{0,1, \dots, 255\}^r$ mittels eines Modelles auf einen niederdimensionalen Merkmalsraum abgebildet wird: $f(x): x \mapsto \mathbb{R}^d; d \leq r$. Die so extrahierten Merkmalsvektoren beinhalten dabei die relevanten Informationen (Merkmale) eines Bildes und können für die Klassifikation genutzt werden. Das Ziel ist es, ein Modell dazu zu befähigen, diese Merkmale effizient zu extrahieren.

Die dem Ansatz zugrunde liegenden Modelle folgen der Architektur von CNN-Modellen, wie sie in den vorherigen Abschnitten zur Klassifikation verwendet wurden. Als zusätzliches Element werden diese Modelle mittels der sogenannten Triplet-Loss-Funktion (Schroff, Kalenichenko & Philbin 2015) trainiert, auf die weiter unten eingegangen wird. Um nun auf Basis der Merkmalsvektoren, welche die Bilddaten repräsentieren, eine Klassifikation der Daten vorzunehmen, wird die Ähnlichkeit zwischen den aus den Bildern resultierenden Vektoren mittels einer Gauss-Distanzmetrik angelehnt an (Fort 2017) berechnet. Hierbei wird der Merkmalsvektor eines neuen, unbekanntes Bildes mit den Merkmalsvektoren von bekannten Bildern verglichen und eine Distanz berechnet. Dem unbekanntes Bild wird die Klasse (bspw. Pitting) desjenigen Bildes zugewiesen, zu dem die Distanz am geringsten ist. Abbildung 6-42 stellt die Grundarchitektur des Siamese Network schematisch dar. Hierbei übernehmen die sogenannten Siamese-Kernel die Extraktion der Merkmale (Merkmalsvektoren). Die Distanzberechnung

und Klassifikation der Vektoren wird dann mittels eines an den RBF-Ansatz angelehnten Ansatzes vorgenommen, wie er bei (Aggarwal 2019) beschrieben wird. Die Architektur besteht aus einer Kombination mehrerer der in Abbildung 6-42 gezeigten Module. Auf die einzelnen Komponenten wird im Folgenden detailliert eingegangen.

6.2.4.2 Grundüberlegung

Die zugrundeliegende Überlegung ist, Merkmale von Objekten verschiedener Klassen direkt miteinander zu vergleichen, um zu entscheiden ob diese derselben Klasse angehören. Mathematisch bedeutet dies, die Klasse C_i eines unbekanntes Objektes (Bild x) zu bestimmen, indem es mit mehreren Objekten (Bilder x_j) verglichen wird, deren Klassen C_j^i mit $i \in S$ bekannt sind und S eine Menge an Klassen ist.

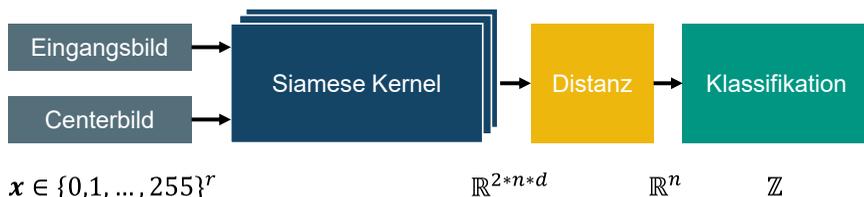


Abbildung 6-42 Grundarchitektur des SBF-Net-Ansatzes. Das Centerbild stellt dabei ein Bild dar, von dem die Klasse bekannt ist.

6.2.4.3 Distanzmaß

Im vorangegangenen Abschnitt wurde die grundsätzliche Aufgabe einer Klassifikation durch einen Vergleich der Merkmalsvektoren eingeführt. Im Folgenden soll hierfür eine Distanzfunktion definiert werden.

Der Vergleich bedarf hierbei eines Distanzmaßes zwischen den Merkmalsvektoren der Bilder und der anschließenden Zuordnung des unbekanntes Objektes x zu derjenigen Klasse C_i , die von der Mehrheit der nächsten Punkte geteilt wird. Diese Idee ist analog zum K-Nearest-Neighbor-(KNN)-Klassifikator, wie er bspw. von (Raschka 2016) beschrieben wird. Die Distanzmetrik zwischen zwei Datenpunkten ergibt sich danach mit:

$$d(x, x_j) = \sqrt[p]{\sum_k |x^k - x_j^k|^p}, \text{ wobei } k \text{ die Dimension der Datenpunkte darstellt. Die Dis-}$$

tanz entspricht der euklidischen Distanz für $p = 2$. Die Distanzen werden geordnet und dem Datenpunkt x , die Mehrheitsklasse der k den nächsten Datenpunkten (Nachbarn)

zugewiesen. In der Arbeit von (Xiao & Jiang et al. 2017) konnte allerdings gezeigt werden, dass der klassische KNN-Klassifikator für die Klassifikation von Fehlstellen auf metallischen Oberflächen nur unzureichende Ergebnisse liefert, wenn die unverarbeiteten Pixelwerte für die Klassifikation verwendet werden.

Als Ergebnis scheidet der klassische KNN-Ansatz zur Klassifikation der Bilddaten aus. Im Folgenden soll die Idee der distanzbasierten Klassifikation mittels des KNN-Ansatzes erweitert werden, um damit eine effektive distanzbasierte Klassifikation durchführen zu können.

6.2.4.4 Distanzbasierte Klassifikation

Ein Ansatz, der die Funktionalität des KNN-Klassifikators implementiert und erweitert, ist das sogenannte RBF-Network, das nach (Aggarwal 2019) eine vergessene Architektur von Neuronalen Netzen darstellt, dessen volles Potenzial wenig erforscht ist. Die Architektur eines RBF-Networks ist in Abbildung 6-43 links qualitativ dargestellt.

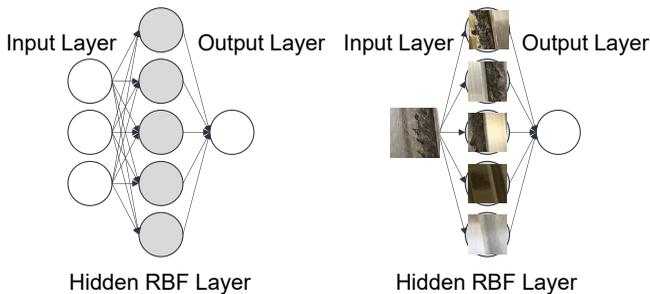


Abbildung 6-43 Grundarchitektur eines RBF-Networks und qualitative Übertragung auf die Klassifikation von Pittings.

Das RBF-Network besteht aus einem Input Layer, das die Informationen an das Hidden-RBF-Layer weitergibt, sowie einem *Output Layer*, in dem dann eine gewichtete Summe der Outputs des RBF-Layers berechnet wird. Im Unterschied zum klassischen Neuronalen Netzwerk, bei dem die Outputs der Hidden-Layer-Knoten mit $o_p = \sigma(\sum_k x_k w_p^k) = \sigma(\mathbf{x}\mathbf{w}_p^T)$ als gewichtete Summe der Inputs berechnet werden (σ ist hierbei die Sigmoid-Funktion zur Implementierung der Nichtlinearität), implementiert das RBF-Network sogenannte RBF-Kernel als Knoten in dem Hidden Layer. Die RBF-Kernel vergleichen hierbei den Input \mathbf{x} gegen einen sogenannten Prototyp-Vektor $\boldsymbol{\mu}_p$, der dieselbe

Dimensionalität wie der Inputvektor x besitzt mit: $\sigma_p = \exp\left(-\frac{\sqrt{\sum_k(x_k - \mu_k^p)^2}}{2\sigma_p^2}\right)$. Der Parameter σ dient hierbei als Gewichtung der RBF-Knoten. Die im Zähler implementierte Distanzfunktion entspricht der euklidischen Distanz, wie sie auch im KNN-Klassifikator verwendet werden kann. Der Output des RBF-Netzwerks berechnet sich damit wie folgt:

$\hat{y} = \sum_p \left(\exp\left(-\frac{\sqrt{\sum_k(x_k - \mu_k^p)^2}}{2\sigma_p^2}\right)\right) w_{out}^p = \mathbf{o} \mathbf{w}_{out}^T$. Es wird also im einfachsten Fall eine lineare Gewichtung der transformierten Abstände durchgeführt. Nach (Zadeh, Hosseini & Sra 2018) kann der RBF-Ansatz damit als eine Erweiterung des KNN-Klassifikators angesehen werden.

Durch den Vergleich des Eingangsdatenpunktes mit dem Prototyp-Vektor erfüllt die RBF-Architektur die Grundintention des Ansatzes, Eingangsdatenpunkte gegen bestehende Datenpunkte mittels einer Distanzfunktion zu vergleichen, um dadurch ein Ähnlichkeitsmaß zu berechnen und die Klasse des Eingangsdatenpunktes zu bestimmen. Übertragen auf die Aufgabe zur Klassifikation von Pittings stellt sich das RBF-Netzwerk qualitativ wie in Abbildung 6-43 rechts dar. Ein neues Bild wird in das Netzwerk gegeben (Input Layer) und die Ähnlichkeit zu Bilddaten im Hidden RBF Layer berechnet.

Als Ergebnis kann festgehalten werden, dass die Klassifikation der Bilder durch eine Gewichtung der Ähnlichkeiten zwischen Merkmalsvektoren vorgenommen werden kann.

Im Folgenden soll die Architektur des RBF-Netzwerks zunächst erweitert werden, um deren Leistungsfähigkeit zu erhöhen.

6.2.4.5 Effektive Klassifikation mittels RBF-Netzwerk

Eine Schwäche des RBF-Ansatzes gegenüber klassischen mehrschichtigen Neuronalen Netzen ist nach (Aggarwal 2019) dessen geringe Tiefe und die damit einhergehende Schwäche, die Strukturen in den Daten zu erlernen. Zur Behebung dieser Schwäche werden der RBF-Architektur zusätzlich zum vorhandenen Output Layer weitere Layer hinzugefügt, um die Leistungsfähigkeit des Klassifikationsmodelles zu erhöhen. Die oben beschriebene lineare Gewichtung der transformierten Abstände wird also erweitert mit $\hat{y} = f(\mathbf{o})$, wobei die Funktion f ein klassisches mehrschichtiges Neuronales Netzwerk ist. Hiermit wird der Kritik von (Aggarwal 2019) an der geringen Tiefe des RBF-Ansatzes direkt begegnet.

Als Ergebnis liegt eine erweiterte RBF-Architektur vor. Die verbleibende Herausforderung besteht in der zielführenden Berechnung der Vektoren x und μ_p , die im RBF-Netzwerk verwendet werden. Auf diesen Aspekt wird im Folgenden eingegangen.

6.2.4.6 Effiziente Merkmalsextraktion

Damit eine Klassifikation der Bilder erfolgen kann, müssen diese zunächst von einem Bild in einen Merkmalsvektor transformiert werden. Die Klassifikation erfolgt dann, indem diese Merkmale einem RBF-Netz zur Klassifikation übergeben werden. Die im folgenden Abschnitt zu beantwortende Fragestellung lautet, wie diese Merkmale (Merkmalsvektoren) effizient generiert werden können.

Im vorgestellten Ansatz sollen Bilddaten verglichen werden es ist allerdings bekannt, dass ein Vergleich der Pixelwerte nicht zielführend ist.

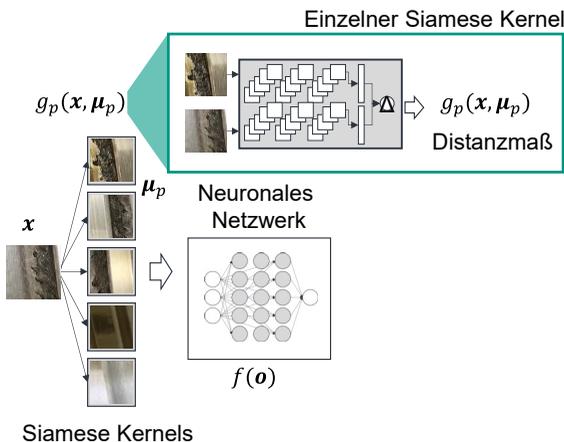


Abbildung 6-44 Erweiterung des RBF-Netzwerk-Ansatzes zum SBF-Net-Ansatz bestehend aus Deep-Learning-Siamese-Kernel, Bild-Prototypen sowie einem Neuronales Netzwerk zur Klassifikation. Das grün umrandete Element entspricht dem vorderen Teil aus Abbildung 6-42.

Um anstelle der Pixelwerte semantisch relevante Merkmale zu vergleichen (Merkmalsvektoren), werden als zweites zentrales Element die Prototyp-Vektoren (μ_p) direkt durch Bilder mit und ohne Schädigung implementiert, aus denen dann relevante Merkmale extrahiert werden.

Um aus diesen Bilddaten effektiv relevante Merkmalsvektoren zu extrahieren und gleichzeitig eine Distanzfunktion zum Vergleich der Daten beizubehalten, werden die RBF-Kernel durch Deep-Learning-Architekturen in Form von Siamese Networks (Siamese-Kernel) ergänzt. Diese Siamese-Kernel überführen damit ein Eingangsbild x in einen Merkmalsvektor μ , der dann als Prototyp-Vektor im RBF-Net verwendet wird. Durch den Tausch des Prototyp-Vektors in Form eines Bildes zu einem Merkmalsvektor des Bildes, der mittels einer Deep-Learning-Architektur (Siamese Network) extrahiert wird, ergibt sich die Erweiterung und Umbenennung des RBF-Nets zum beschriebenen SBF-Net. Die Architektur des SBF-Nets ist in Abbildung 6-44 schematisch dargestellt.

Als Ergebnis werden zur Klassifikation nicht die Pixelwerte eines Bildes, sondern Merkmalsvektoren verwendet, die mittels sogenannter Siamese-Kernel (CNN-Architekturen) aus den Bilddaten extrahiert werden.

Im Folgenden soll der Aufbau der Siamese-Kernel zur Merkmalsextraktion beschrieben werden.

6.2.4.7 Siamese-Kernel

Die Motivation für die Anpassung der Kernel innerhalb des RBF-Netzwerks basiert hauptsächlich auf der Kritik der geringen Tiefe eines klassischen RBF-Netzwerks. Eine weitere Motivation für eine Erweiterung ist die Möglichkeit, hochdimensionale Bilddaten effizient miteinander zu vergleichen. Dies ist mit einer pixelbasierten euklidischen Distanzfunktion, wie sie in den RBF-Knoten verwendet wird, nicht ohne Weiteres möglich, da hierbei lediglich die Werte der Pixel verglichen werden und die in den Bilddaten vorhandenen Strukturen und Merkmale weitestgehend unbeachtet bleiben.

Dies ist in Abbildung 6-45 dargestellt und soll wie folgt verdeutlicht werden: Sei \mathcal{A}_j die Menge der Klassen/Objekte in einem Bild und seien P_1 und P_2 Bilder $\in \mathbb{R}^{r \times r}$, für die $\mathcal{A}_1 \neq \mathcal{A}_2$ ³¹. Es sei angenommen, dass die Sättigung eines Bildes dessen Klassen nicht verändert. Sei $P_3 : \mathcal{A}_1 = \mathcal{A}_3$ ein Bild derselben Dimensionalität wie P_1 , $\exists R \in \mathbb{R}^{r \times r}$: $\mathcal{A}_1 = \mathcal{A}_R$, sodass $\|P_1 - P_2\|^2 \leq \|P_1 - R\|^2$. Es werde R konstruiert mit $R = P_3 + M$, $M = m * \mathbf{1}$. Nach Umformen ergibt sich $\|P_2\|^2 \geq \|R\|^2$. Indem das Bild P_3 in seiner Sättigung

³¹ $\mathcal{A}_1 \neq \mathcal{A}_2$ bedeutet, dass die Bilder semantisch unterschiedlich sind, d. h. keine gemeinsamen Klassen besitzen. Es sei angenommen, dass \mathcal{A}_j vom Betrachter oder von einem Klassifikationsmodell bestimmt wurde.

angepasst wird, kann die euklidische Distanz zwischen zwei semantisch unterschiedlichen Bildern kleiner gleich der euklidischen Distanz zwischen zwei semantisch gleichen Bildern gemacht werden. Dies zeigt, dass die Verwendung der Pixelwerte in Kombination mit einer euklidischen Distanz nicht ausreicht, um die inhaltliche Ähnlichkeit zweier Bilder zu bestimmen.



Abbildung 6-45 Die Bilder P_1 und R enthalten dieselben Objekte und können damit als semantisch identisch angesehen werden. Die Bilder P_1 und P_2 teilen hingegen keine gemeinsamen Klassen. Die euklidische Distanz der Pixel zwischen den Bildern P_1 und R ist größer als die euklidische Distanz zwischen den Bildern P_1 und P_2 . R ist gesättigte Version von P_3 .

Es wird also nach einer Möglichkeit gesucht, ein Bild in seine Bildmerkmale zu zerlegen bzw. diese aus dem Bild zu extrahieren, um dann in einem weiteren Schritt die Merkmale zweier Bilder zu vergleichen und darüber eine Entscheidung über die Gleichheit der Bilder im Sinne der darin vorhandenen Objekte (Pitting, kein Pitting) vorzunehmen.

Als Ergebnis kann festgehalten werden, dass semantisch sinnvolle Merkmale aus Bildern extrahiert werden müssen, um diese hinsichtlich ihres Inhalts zu klassifizieren.

Im folgenden Abschnitt soll auf die Extraktion der Merkmale eingegangen werden.

6.2.4.8 Merkmalsextraktion mittels Siamese Neural Networks

Nachdem im vorherigen Abschnitt gezeigt wurde, dass eine Ähnlichkeitsberechnung auf Basis der Pixelwerte nicht zielführend ist, wird im folgenden Abschnitt das Vorgehen bei der Extraktion von Merkmalen aus den Bilddaten behandelt.

Eine Möglichkeit der Merkmalsextraktion bieten die folgenden klassischen Verfahren: Scale Invariant Feature Transformation (SIFT) (Lowe 1999), Structural Similarity Index (SSI) (Wang & Bovik et al. 2004), Speeded Up Robust Features (SURF) (Bay & Tinne

et al. 2006), Binary Robust Independent Elementary Features (BRIEF) (Calonder & Lepetit et al. 2010) sowie deren schnellere Weiterentwicklung, Oriented FAST and Rotated BRIEF (ORB) (Rublee & Rabaud et al. 2011). Die klassischen Verfahren wurden allerdings weitestgehend von den performanteren Deep-Learning-Verfahren wie den CNN zur Extraktion von Merkmalen abgelöst, wie in Abschnitt 6.2.3 bei der Klassifikation von KGT-Bildern gezeigt werden konnte.

Um die Extraktion von Merkmalen durchzuführen und gleichzeitig ein Ähnlichkeitsmaß zweier Bilddaten zu berechnen, werden als Kernel sogenannte Siamese Neural Networks implementiert. Die Grundarchitektur eines Siamese Network und damit der eingesetzten Siamese-Kernel ist in Abbildung 6-46 dargestellt. Hierbei sind die Grundelemente des Siamese Network mit den Elementen eines klassischen CNN identisch.



Abbildung 6-46 Siamese Network als Grundarchitektur der Siamese-Kernel. Die Distanz von Bildern identischer Klasse soll möglichst gering und jene von Bildern unterschiedlicher Klasse möglichst hoch sein.

Ein CNN kann dabei, wie in den Grundlagen in Abschnitt 2.4.5 dargestellt, als Funktion $f: \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^p, x \mapsto f(x)$ verstanden werden. Die so extrahierten Merkmale können dann in einem nächsten Schritt zum Vergleich von Bilddaten verwendet werden. Dem Vergleich der Bilder liegt die Annahme zugrunde, dass zwei Bilder derselben Klasse

aufgrund ihrer semantischen Ähnlichkeit durch ähnliche Merkmalsvektoren repräsentiert werden. Wird dann eine Distanz bspw. in Form der euklidischen Distanz mit $\delta = \|f(x_i) - f(x_j)\|^2$ berechnet, so ergeben sich bei übereinstimmenden Klassen niedrige und bei nicht übereinstimmenden Klassen hohe Werte. Um dieses Verhalten zu implementieren, besteht das Siamese Network, wie in Abbildung 6-46 dargestellt, aus zwei identischen CNN-Architekturen, die jeweils eine Merkmalsextraktion vornehmen. Entstammen die Bilder x_i, x_j derselben Klasse, so sind die resultierenden Merkmale ähnlich und damit ist die Distanz zwischen deren Merkmalsvektoren gering.

Die verwendete CNN-Architektur basiert auf der Grundarchitektur des VGG16-Netzwerks (Simonyan & Zisserman 2015), wobei die Architektur dahingehend angepasst wird, dass anstelle der beiden 4096-dimensionalen Schichten zwei 100-dimensionale Fully Connected Layer implementiert wurden, um dichtere Merkmalsvektoren zu erhalten.

Basierend auf Abbildung 6-46 besteht das Ziel darin, die Merkmalsdistanz zwischen Bildern derselben Klasse möglichst zu minimieren. Dazu werden während des Trainings die Gewichte des CNN so angepasst, dass das Netzwerk Merkmalsvektoren mit hoher Übereinstimmung bei Bildern derselben Klasse und mit geringer Übereinstimmung bei Bildern unterschiedlicher Klassen generiert. Da die Gewichte identisch sind, ergibt sich für ein identisches Bild x derselbe Merkmalsvektor und damit eine Distanz der Bilddaten von 0. Zur Erlangung dieses Verhaltens wird für das Training des CNN die Triplet-Loss-Verlustfunktion angelehnt an (Schroff, Kalenichenko & Philbin 2015) verwendet mit: $Triplet Loss = \max(0, g(\varphi(a), \varphi(p)) + \alpha - g(\varphi(a), \varphi(n)))$, wobei $g(\cdot)$ die Cosinus-Distanzfunktion und $\varphi(\cdot)$ das CNN darstellt. a entspricht einem sogenannten Anker-Bild, das dieselbe Klasse wie das Positiv-Bild (p) teilt. n wird als Negativ-Bild bezeichnet und ist von einer anderen Klasse als a und p . Ist die Distanz zwischen Anker-Bild und Negativ-Bild im Merkmalsraum also kleiner als die Distanz zwischen Anker-Bild und Positiv-Bild plus einem zusätzlichen Abstand α , so ist das Kriterium, Bilder unterschiedlicher Klassen im Merkmalsraum weit entfernt zu platzieren, erfüllt und der Loss ist 0. Ist dies nicht gegeben, wird ein positiver Wert als Loss zurückgegeben. Die Reduzierung der Verlustfunktion führt damit zu einer korrekten Transformation von Bilddaten in aussagekräftige Merkmalsvektoren. Die Distanzfunktion $g(\cdot)$ innerhalb des Triplet Loss berechnet sich zu $Cosinus Distanz = 1 - \frac{x^T y}{\|x\| \times \|y\|}$. Werte nahe 0 repräsentieren

tieren hierbei eine hohe Ähnlichkeit (oder geringe Distanz). Auf Basis von Voruntersuchungen werden die einzelnen Siamese-Kernel mittels des Adam-Optimierers für 5000 Iterationen mit einer Lernrate von 10^{-5} trainiert, wobei α mit 0.3 gewählt wird. Diese Werte wurden empirisch ermittelt.

Als Ergebnis werden aus den Eingangsbildern die Vektoren x und μ extrahiert. Teilen die Eingangsbilder eine Klasse, so ist die Distanz zwischen den Bildern gering. Das Training der CNNs zur Extraktion der Merkmale wird mittels des Triplet Loss vorgenommen.

Nachdem in den obigen Abschnitten dargestellt wurde, wie die Verarbeitung der Eingangsbilder stattfindet und wie der Vergleich mit neuen Bilddaten umgesetzt wird, soll im Folgenden die Kombination mehrerer Siamese-Kernel zu dem vorgestellten Siamese-Net-Ansatz vorgestellt werden.

6.2.4.9 SBF-Net

Jeder Siamese-Kernel berechnet die Distanz zwischen einem Eingangsbild und einem hinterlegten Center-Bild. Damit nicht nur ein Center-Bild betrachtet wird, sondern das Modell einen Vergleich eines Eingangsbildes mit mehreren Bildern einer Klasse vornehmen kann, werden mehrere Siamese-Kernels verknüpft. Jeder Kernel beinhaltet ein Center-Bild, mit dem das Eingangsbild verglichen wird. Hierbei kann die Identität des Eingangsbildes durch einen Vergleich mit mehreren Bilddaten festgestellt werden. Gleichzeitig bleibt der Ansatz dateneffizient, da die verwendete Menge an Bilddaten gering bleibt.

Im SBF-Net-Ansatz wird nun eine Anzahl an n Siamese-Kernels eingesetzt und die resultierenden Distanzwerte werden mittels eines erweiterten RBF-Netzes zur Klassifikation verwendet. Hierbei lernen die Siamese-Kernels, die Zugehörigkeit des jeweiligen Anker-Bildes genau zu bestimmen.

6.2.4.9.1 Center-Bilddaten

Für das Training führt jeder Siamese-Kernel einen Vergleich von Bilddaten (positive, negative) mit einem entsprechenden Anker-Bild durch. Im vorgestellten Ansatz wird pro Siamese-Kernel ein einzelnes Bild zufällig aus den Trainingsdaten als Anker-Bild gewählt. Es werden m Anker-Bilder pro Klasse zufällig ausgewählt und für jedes der Anker-Bilder wird ein Siamese-Kernel trainiert. Die im Folgenden beschriebene Kombination der Siamese-Kernel in einem Gesamtmodell kann damit als eine Art Kombination

(Ensemble) von Expert*innen angesehen werden, die gemeinsam zu einer Entscheidung kommen. Durch die Wahl von mehreren Anker-Bildern pro Klasse erhält das Gesamtnetz eine Art Klassenbewusstsein, da jedes Eingangsbild mit mehreren Bildern einer Klasse verglichen wird. Als eingängiges Beispiel würde ein Fahrzeug nicht lediglich mit einem bestimmten Fahrrad und einem bestimmten PKW, sondern mit mehreren Bildern dieser Klassen verglichen werden.

Als Ergebnis werden zufällig Center-Bilddaten aus den Datensätzen ausgewählt.

Im Folgenden wird die Kombination der Siamese-Kernels zu einem Gesamtmodell beschrieben.

6.2.4.9.2 Kombination der Siamese-Kernels

Pro Siamese-Kernel ergibt sich aus der Distanz der Merkmalsvektoren ein Ähnlichkeitswert, der zur Klassifikation verwendet werden kann. Eine Klassifikation erfolgt dann, indem pro Klasse m Siamese-Kernels trainiert werden. Für das Gesamtnetz ergibt sich somit eine Anzahl von $n * m$ Siamese-Kernels.

Ein neuer Bilddatenpunkt x_i kann dann klassifiziert werden mit:

$c = \operatorname{argmin}(\sum_m g(\varphi(x_i), \varphi(x_m^{c_1})), \sum_m g(\varphi(x_i), \varphi(x_m^{c_2})), \dots, \sum_m g(\varphi(x_i), \varphi(x_m^{c_n})))$. Dies entspricht einer K-Nächsten-Nachbarn-Klassifikation, wie sie vom KNN-Algorithmus

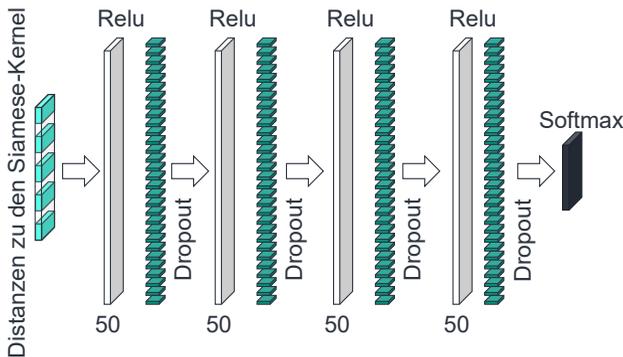


Abbildung 6-47 Architektur des Neuronalen Netzes zur Klassifikation der Bilddaten mit der Rectified-Linear-Unit-Aktivierungsfunktion (Relu-Aktivierungsfunktion) und vier Schichten à 50 Neuronen nach (Schlagenhauf, Yildirim & Brückner 2022).

durchgeführt wird. In der vorliegenden Arbeit werden die pro Siamese-Kernel berechneten Distanzen $g((x_i, x_m^c))$ nicht direkt zur Klassifikation verwendet, sondern wie in Abschnitt 6.2.4.1 beschrieben zunächst durch einen RBF-Kernel transformiert und als Merkmale in ein mehrschichtiges Neuronales Netzwerk gegeben. Das Neuronale Netz besteht hierbei aus vier Schichten mit je 50 Neuronen, wobei die Relu-Aktivierungsfunktion verwendet wird. Jeder Schicht folgt eine Dropout-Schicht mit einem Dropout-Wert von 0.1. Das Modell wird für 1000 Iterationen mittels der Categorical-Cross-Entropy-Verlustfunktion mit einer Lernrate von 10^{-6} trainiert.

Das Neuronale Netzwerk erhält damit pro zu klassifizierendem Bild $n * m$ Merkmale der Form $o(g(\varphi(x_i), \varphi(x_m^c)))$, wobei o die RBF-Kernels darstellen.

Zusammenfassend überführen die Siamese-Kernels die Bilder (Pixel-Matrizen) in semantisch relevante Merkmale, die dann transformiert und durch ein mehrschichtiges Neuronales Netz zur Klassifikation verwendet werden.

Nachdem in den vorherigen Abschnitten der Aufbau der SBF-Net-Architektur beschrieben wurde, folgt im Anschluss die Vorstellung der Datensätze zur Untersuchung des Ansatzes.

6.2.4.9.3 Daten zur Modellbildung und zum Modellvergleich

Der Fokus des Ansatzes liegt auf der dateneffizienten Klassifikation in der technischen Domäne, weshalb neben dem KGT-Datensatz (Schlagenhauf 2021) ebenfalls der NEU-Datensatz (Song & Yunhui 2019) sowie der TEX-Datensatz (Bergmann & Fauser et al. 2019) verwendet werden. Die Übertragbarkeit des Ansatzes soll anhand der nicht-technischen Datensätze MNIST²⁹ und cifar10³⁰ gezeigt werden.

Der NEU-Datensatz repräsentiert die Klassifikation von Fehlstellen auf Bandstahl, wobei die sechs Klassen Eindrücke, Marken, Risse, Pittings, Einschlüsse und Kratzer

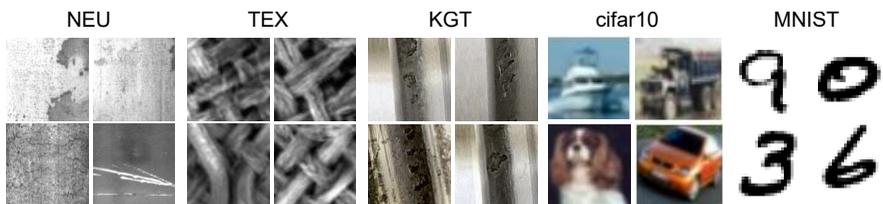


Abbildung 6-48 Beispieldaten aus den verwendeten Datensätzen.

gleichmäßig auf insgesamt 1800 Graustufenbilder der Größe 200 x 200 Pixel verteilt sind. Der Datensatz stellt einen Stand-der-Technik-Datensatz dar, der zum Vergleich von Modellen zur Klassifikation von Fehlstellen auf metallischen Oberflächen verwendet wird. Mittels dieses Datensatzes ist es möglich, die Leistungsfähigkeit des eigenen Ansatzes im Vergleich zu bestehenden Ansätzen zur Fehlerklassifikation auf metallischen Oberflächen einzuordnen.

Der TEX-Datensatz repräsentiert die Klassifikation von Fehlstellen in gewobenen Textilien und besteht aus insgesamt 108 000 Graustufenbildern der Größe 64 x 64 Pixel. Der Datensatz ist gleichmäßig auf die folgenden Klassen aufgeteilt: kein Fehler, Verfärbung, Einschnitt, Loch, Fehlstelle und metallischer Einschluss.

Der MNIST-Datensatz besteht aus den handgeschriebenen Ziffern 0 bis 9 und enthält insgesamt 60 000 Graustufenbilder der Größe 32 x 32 Pixel, die gleichmäßig auf die Klassen aufgeteilt sind.

Der cifar10-Datensatz repräsentiert die zehn Klassen Flugzeug, Automobil, Vogel, Katze, Rentier, Hund, Frosch, Pferd, Schiff und LKW, die gleichmäßig auf insgesamt 60 000 RGB-Bilder der Größe 32 x 32 Pixel verteilt sind. Abbildung 6-48 beinhaltet Beispiele aus den Datensätzen.

Alle Bilder in den Datensätzen werden sowohl in ihrer originalen als auch in augmentierter Form verwendet. Für die Augmentierung der Daten wird jedes Bild der Datensätze viermal einer zufälligen Auswahl $\mathcal{P} \subseteq \mathcal{M}$ an Data-Augmentation-Techniken unterzogen. Die Menge \mathcal{M} enthält dabei die Elemente: {Contrast Limited Histogram Equalization (CLAHE) aller Farbkanäle in Limit von (0,10), zufällige Rotation im Bereich $[-5^\circ, 5^\circ]$, horizontales Spiegeln, vertikales Spiegeln, Laplace Rauschen pro Farbkanal, zufällige Multiplikation der Farbkanäle mit einem Wert im Bereich $[0.7, 1.3]$ }. Weiterhin wurden alle Bilder auf Werte zwischen 0 und 1 normalisiert. Für die Augmentierung der Daten wurde die Python Bibliothek `imgaug`³² verwendet.

Pro Datensatz werden jeweils 25 % der Gesamtdaten zufällig ausgewählt und als Validierungsdatsatz entfernt. Diese Bilder stehen dem Modell für das Training also nicht zur Verfügung und das Modell kann darüber validiert werden. Das Modell wird mit 3, 5,

³² `imgaug` (2020), *imgaug — imgaug 0.4.0 documentation*. <https://imgaug.readthedocs.io/en/latest/index.html> [16.11.2020].

10, 20, 50 und 100 Bildern pro Klasse trainiert. Hierdurch kann die Dateneffizienz des Ansatzes überprüft werden.

Zur Überprüfung der Dateneffizienz wird der Ansatz mit den zwei leistungsfähigen Stand-der-Technik-Ansätzen ResNet50 sowie ResNet101 (He & Zhang et al. 2016) verglichen. Diese werden mit derselben Anzahl an Trainings- und Validierungsdaten trainiert.

Als Ergebnis wurden Datensätze zur Untersuchung der SBF-Net-Architektur ausgewählt.

Da der SBF-Net-Ansatz mit einer begrenzten Anzahl an Trainingsdaten umgesetzt wird, folgt nun die Beschreibung der Auswahl der Anzahl an Center-Bildern.

6.2.4.9.4 Auswahl der Anzahl an Center-Bilder

Sowohl die Auswirkungen durch die Anzahl an Center-Bildern als auch die Dateneffizienz des SBF-Ansatzes werden experimentell untersucht. Die Ergebnisse werden in den folgenden zwei Abschnitten dargestellt.

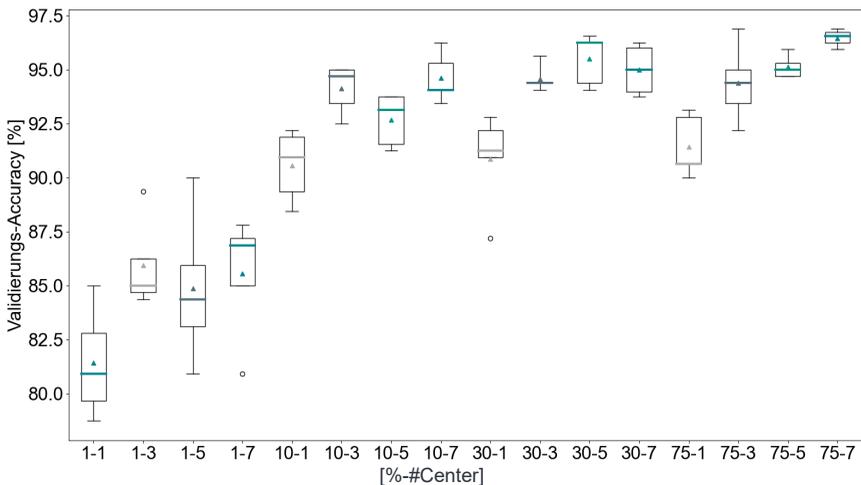


Abbildung 6-49 Modellergebnisse des SBF-Nets (KGT-Datensatz) nach (Schlagenhauf, Yildirim & Brückner 2022) bei Variation der Anzahl an Center-Bildern pro Klasse sowie der Datenmenge zum Training. Die x-Achsen-Beschriftung folgt der Struktur [Datenmenge in % – Anzahl Center pro Klasse].

Integraler Bestandteil der SBF-Architektur ist die Auswahl der Center-Bilder (Anker) pro Klasse. Die Hypothese ist, dass die Leistungsfähigkeit des Modells mit steigender Anzahl an Center-Bildern steigt, da das Modell dann eine größere Vielfalt an möglichen Fällen abdeckt. Eine weitere Hypothese ist die Steigerung der Leistungsfähigkeit mit steigender Anzahl an Trainingsdatenpunkten. Für die Auswahl der Anzahl an Center-Bildern, die für die weiteren Experimente zur Untersuchung der Dateneffizienz verwendet werden, wird das SBF-Net mit 1, 3, 5 und 7 Center-Bildern pro Klasse trainiert und dabei wird die Trainingsdatenmenge mit 1 %, 10 %, 30 % und 75 % gewählt. Es werden jeweils 25 % der Daten als Validierungsdaten verwendet. Als Datensatz wird zunächst der KGT-Datensatz verwendet. Die Ergebnisse für die Klassifikation des KGT-Datensatzes sind in Abbildung 6-49 dargestellt.

Zunächst kann die Hypothese bestätigt werden, dass die Leistungsfähigkeit des Modells mit steigender Anzahl an Trainingsdaten steigt. Es ist weiterhin ersichtlich, dass die Steigerungsraten mit steigender Anzahl an Trainingsdaten sinken. Dieser Effekt ist im Stand der Forschung bekannt und lässt sich damit erklären, dass diejenigen Bilder des Validierungsdatensatzes, die vom Modell bei einer schon hohen Klassifikationsgenauigkeit dennoch nicht korrekt klassifiziert werden, sehr schwer zu klassifizierende Beispiele darstellen. Im Falle der Bilddaten des KGTs sind dies bspw. diejenigen Daten, die kleine Pittings zeigen und gleichzeitig stark verschmutzt sind.

Die zweite Hypothese, dass die Modellgenauigkeit mit steigender Anzahl an Centern pro Klasse steigt, kann ebenfalls grundsätzlich bestätigt werden. Es ist auffällig, dass in einigen Fällen mit fünf oder teilweise drei Center-Bildern pro Klasse die besten Ergebnisse erzielt werden können. Dieser Effekt lässt sich durch zwei Ursachen erklären: Zunächst werden die Center pro Klasse zufällig aus den Trainingsdaten ausgewählt. Dies führt dazu, dass es Center-Bilder gibt, welche die zugehörige Klasse besser repräsentieren als andere. Dies führt zu einer grundsätzlichen Varianz und erklärt, warum teilweise mit einer geringeren Anzahl an Centern bessere Ergebnisse erzielt werden können. Der zweite Aspekt, den es zu beachten gilt ist, dass mit steigender Anzahl an Trainingsdaten die Varianz mit zunehmender Anzahl an Centern pro Klasse sinkt und eine klarere Steigerung der Genauigkeit mit steigender Anzahl an Centern pro Klasse zu beobachten ist. Da die Center bspw. auch im Falle von 75 % Trainingsdaten zufällig gewählt werden, besteht nahezu in gleichem Maße die Möglichkeit, dass diese Center repräsentativ oder nichtrepräsentativ für die jeweilige Klasse sind. Da bei größeren Datenmengen aber mehr Positiv- und Negativ-Datenpunkte zur Verfügung stehen, hat das

Modell mehr Möglichkeiten, charakteristische Eigenschaften der Center-Bilder durch Vergleich mit diesen zu identifizieren, die dann eine korrekte Zuordnung zulassen. Dies ist besonders deutlich, wenn 75 % der Daten für das Training verwendet werden.

Zusammenfassend lässt dies auf die Notwendigkeit einer korrekten Auswahl der Center-Bilder schließen. Dieser Aspekt ist allerdings nur von Relevanz, solange die Datenmengen groß genug sind, um eine entsprechende Auswahl treffen zu können. Bei sehr kleinen Datenmengen sind die zu verwendenden Center-Bilder praktisch vorbestimmt. Für die weiteren Untersuchungen werden fünf Center pro Klasse gewählt.

Nachdem die Auswahl und Relevanz der Center-Daten untersucht wurden, soll im folgenden Abschnitt die Dateneffizienz des SBF-Ansatzes untersucht werden.

6.2.4.9.5 Dateneffizienz

Unter Verwendung von fünf Centern pro Klasse werden Experimente zur Untersuchung der Dateneffizienz des SBF-Net-Ansatzes mit den KGT-, NEU-, TEX-, MNIST- und cifar10-Datensätzen durchgeführt, wobei das Modell jeweils mit 3, 5, 10, 20, 50 und 100 Datenpunkten pro Klasse trainiert wird. Die Ergebnisse für die Datensätze sind in Abbildung 6-50 und Tabelle 6-6 dargestellt. Im Fall der KGT- und NEU-Datensätze ist ersichtlich, dass das SBF-Net für alle Datenmengen klare Vorteile gegenüber den Architekturen aus dem Stand der Technik hat und die größte Differenz bei fünf und zehn Datenpunkten zu verzeichnen ist. Mit nur zehn Datenpunkten wird bereits eine Validierungsgenauigkeit zwischen 70 und 80 % erreicht. Auch ist der Effekt zu beobachten, dass die Wachstumsraten mit steigender Anzahl an Trainingsdatenpunkten in der Genauigkeit abflachen. In beiden Fällen ist die maximal erreichte Accuracy höher als bei den Modellen aus dem Stand der Technik.

Im Falle der MNIST-Daten liefert das SBF-Net für alle Versuche die besten Ergebnisse. Bei drei, fünf und zehn Datenpunkten ist dies auch für den TEX-Datensatz der Fall. Bei 20 Datenpunkten liefert das ResNet101 leicht bessere Ergebnisse, allerdings ist der Unterschied gering. Im Falle von 50 sowie 100 Datenpunkten liefert das SBF-Net wiederum die besten Ergebnisse, weshalb das Ergebnis mit 20 Datenpunkten als Ausreißer-Ergebnis gewertet werden kann.

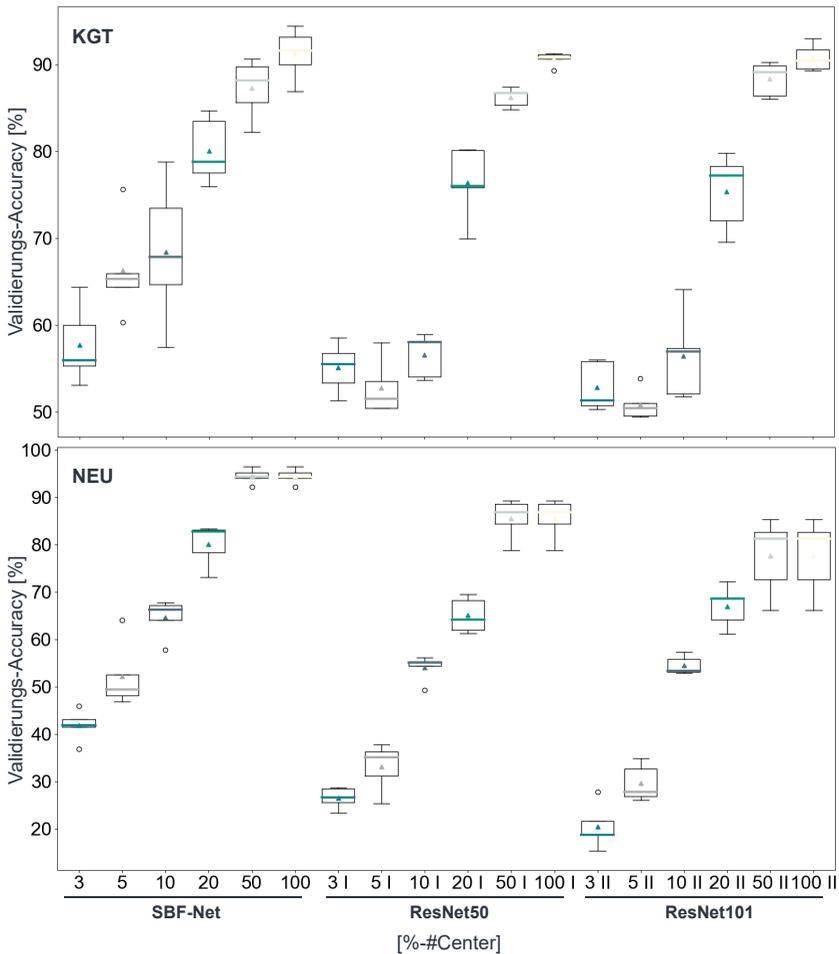


Abbildung 6-50 Validierungs-Accuracy des SBF-Net bei unterschiedlichen Anzahlen an Trainingsdaten im Vergleich zu den Stand-der-Technik-Modellen nach (Schlagenhaut, Yildirim & Brückner 2022).

Für den cifar10-Datensatz ergibt sich ein etwas differenzierteres Bild: Hier ist die Klassifikationsgenauigkeit des SBF-Nets für drei und fünf Datenpunkte am höchsten, bleibt dann aber hinter der Klassifikationsgenauigkeit der ResNet. Dieser Effekt lässt sich dadurch erklären, dass die Bilder im cifar10-Datensatz strukturell unterschiedlich zu

den übrigen Datensätzen sind, die einige charakteristische Bildmerkmale teilen. Da die ResNet-Architekturen eine größere Kapazität aufweisen als die einzelnen Siamese-Kernels, scheinen sie besser in der Lage zu sein, die unterschiedlichen Merkmale des cifar10-Datensatzes zu codieren. Weiterhin wurden die ResNet-Architekturen für den ImageNet Contest (Russakovsky & Deng et al. 2015) entwickelt. Die Klassen des cifar10-Datensatzes bilden eine Teilmenge der Klassen des ImageNet-Datensatzes.

Tabelle 6-6 Gegenüberstellung der erreichten Validierungs-Accuracy des SBF-Nets für die TEX-, MNIST- und cifar10-Datensätze im Vergleich zu Stand-der-Technik-Ansätzen nach (Schlagenhauf, Yildirim & Brückner 2022). Ungefärbte Zellen zeigen durchgeführte Versuche. Gefärbte Zellen zeigen den Durchschnitt der durchgeführten Versuche. Die Farben Grün, Gelb und Rot ordnen die Ergebnisse nach deren Platzierung hinsichtlich der Leistungsfähigkeit.

SBF-NET						ResNet50						ResNet101					
3	5	10	20	50	100	3	5	10	20	50	100	3	5	10	20	50	100
TEX																	
0.209	0.269	0.334	0.294	0.466	0.491	0.249	0.210	0.259	0.321	0.357	0.490	0.177	0.198	0.276	0.305	0.346	0.487
0.191	0.247	0.231	0.275	0.387	0.512	0.233	0.234	0.257	0.293	0.368	0.451	0.235	0.176	0.263	0.320	0.378	0.470
0.275	0.269	0.319	0.316	0.416	0.469	0.177	0.236	0.276	0.318	0.361	0.464	0.234	0.233	0.291	0.311	0.361	0.443
0.262	0.237	0.278	0.303	0.369	0.484	0.247	0.216	0.270	0.303	0.368	0.440	0.178	0.203	0.279	0.295	0.371	0.417
0.247	0.284	0.319	0.303	0.434	0.509	0.191	0.252	0.251	0.281	0.400	0.477	0.180	0.207	0.310	0.287	0.368	0.447
0.237	0.261	0.296	0.298	0.414	0.493	0.219	0.230	0.263	0.303	0.371	0.464	0.201	0.204	0.284	0.304	0.365	0.453
MNIST																	
0.691	0.775	0.881	0.950	0.959	0.981	0.172	0.486	0.766	0.872	0.926	0.954	0.112	0.272	0.507	0.782	0.876	0.936
0.753	0.791	0.903	0.913	0.959	0.972	0.252	0.307	0.832	0.880	0.916	0.954	0.198	0.499	0.675	0.824	0.879	0.900
0.806	0.719	0.875	0.934	0.956	0.981	0.176	0.466	0.744	0.859	0.933	0.952	0.160	0.382	0.642	0.850	0.916	0.943
0.688	0.806	0.894	0.928	0.966	0.969	0.190	0.475	0.735	0.898	0.930	0.958	0.190	0.276	0.690	0.853	0.885	0.916
0.625	0.747	0.875	0.922	0.950	0.972	0.175	0.448	0.718	0.887	0.929	0.959	0.121	0.286	0.673	0.831	0.893	0.908
0.713	0.767	0.886	0.929	0.958	0.975	0.193	0.436	0.759	0.879	0.927	0.955	0.156	0.343	0.637	0.828	0.890	0.921
CIFAR10																	
0.206	0.219	0.206	0.213	0.356	0.259	0.146	0.188	0.263	0.269	0.328	0.398	0.134	0.197	0.235	0.253	0.281	0.275
0.169	0.200	0.259	0.259	0.297	0.131	0.138	0.170	0.254	0.270	0.329	0.360	0.114	0.150	0.215	0.221	0.313	0.318
0.216	0.203	0.262	0.269	0.281	0.259	0.132	0.173	0.238	0.277	0.355	0.385	0.118	0.152	0.233	0.262	0.331	0.243
0.188	0.206	0.253	0.228	0.275	0.309	0.162	0.180	0.274	0.269	0.339	0.394	0.132	0.162	0.222	0.271	0.316	0.262
0.178	0.172	0.219	0.247	0.188	0.228	0.121	0.189	0.243	0.281	0.315	0.335	0.122	0.170	0.220	0.267	0.287	0.297
0.191	0.200	0.240	0.243	0.279	0.238	0.140	0.180	0.254	0.273	0.333	0.375	0.124	0.166	0.225	0.255	0.306	0.279

Um ein klareres Bild der Leistungsfähigkeit des SBF-Nets gegenüber den Ansätzen aus dem Stand der Technik zu zeichnen, ist in Abbildung 6-51 die durchschnittliche Validierungs-Accuracy aus allen Datensätzen dargestellt.

Der Vorteil des SBF-Net-Ansatzes ist erkennbar. Die beiden ResNet-Architekturen erzielen ähnliche Ergebnisse, mit einem leichten Vorteil für das ResNet50. Es ist in den

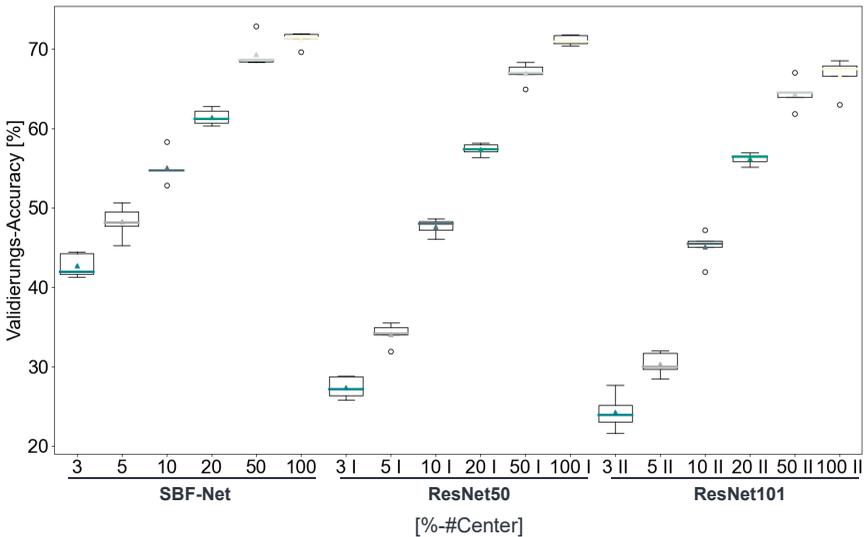


Abbildung 6-51 Vergleich der durchschnittlichen Validierungs-Accuracy über alle Datensätze bei gegebener Anzahl an Trainingsdatenpunkten nach (Schlagenhauf, Yildirim & Brückner 2022).

vorherigen Versuchen erkennbar, dass das SBF-Net auch bei größeren Datenmengen on-par ist mit den ResNet-Architekturen oder sogar bessere Ergebnisse erzielt. Diese Aussage wird auch durch Abbildung 6-52 bekräftigt, in der die SBF-Net-Architektur mit 75 % des NEU-Datensatzes trainiert wurde und die Ergebnisse in den internationalen Stand der Technik nach (Song & Yunhui 2019) eingeordnet sind. Die erreichten 99.69 % liegen über dem aktuellen Benchmark, wobei zusätzlich einige der dargestellten Architekturen mit einem 80/20-Train-Test-Split trainiert wurden und damit 5 % mehr Trainingsdaten zur Verfügung standen. Die maximale Leistungsfähigkeit des verwendeten ResNet50 ist ebenfalls eingeordnet.

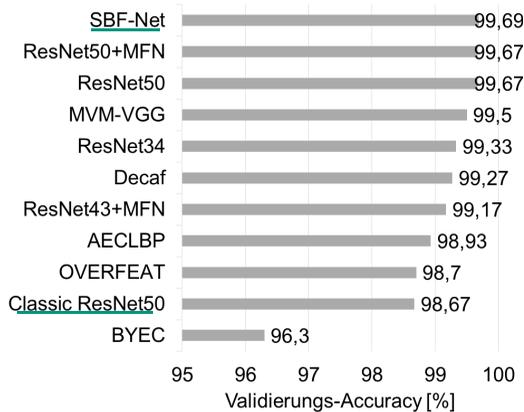


Abbildung 6-52 Vergleich der Leistungsfähigkeit des SBF-Net-Ansatzes gegen internationalen Stand der Forschung nach (Schlagenhauf, Yildirim & Brückner 2022) und (Song & Yunhui 2019) anhand des NEU-Datensatzes.

Zusammenfassend konnten sowohl die Dateneffizienz als auch die Übertragbarkeit des SBF-Nets auf andere Datensätze (besonders für kleine Datensätze), gezeigt werden. Es konnte weiterhin gezeigt werden, dass die Architektur auch bei größeren Datenmengen eine vergleichbare Leistungsfähigkeit wie die Modelle aus dem Stand der Technik aufweist. Die dateneffiziente Klassifikation im Kontext der KGT-Bilder sei besonders betont, wodurch eine Grundlage für die Entwicklung von dateneffizienten Verfahren zur Klassifikation von Oberflächendefekten auf metallischen Oberflächen bereitgestellt werden kann.

Nachfolgend wird ein Fazit zum Ansatz gezogen, bevor zur Identifikation von Pittings in Bilddaten des KGTs und zur Untersuchung eines Prognosemodelles für die Oberflächenzerrüttung am KGT zurückgekehrt wird.

6.2.4.10 Siamese Basis Function Network – Fazit

Die SBF-Net-Architektur extrahiert mittels der eingebetteten Siamese-Kernel effektiv die Merkmale der zugewiesenen Center-Bilder und führt anschließend eine ähnlichkeitsbasierte Klassifikation durch. Es wurde gezeigt, dass der Ansatz sowohl für die Klassifikation von Fehlstellen auf dem KGT als auch zur Klassifikation in unterschiedlichen technischen und nicht technischen Domänen geeignet ist. Die Untersuchung der SBF-Net-Architektur zeigt, dass diese leistungsfähige Architekturen aus dem Stand der

Technik hinsichtlich Dateneffizienz bei der Klassifikation übertrifft und gleichzeitig auch bei großen Datenmengen gute Ergebnisse erzielt werden können. Es konnte gezeigt werden, dass der aktuelle Benchmark bei der Klassifikation des NEU-Datensatzes unter Verwendung von 75 % der Daten leicht überschritten werden konnte.

Ein Aspekt, der weiter untersucht werden sollte, ist die optimale Wahl und Zusammenstellung der Center-Bilder. Weiterhin müssen Strategien entwickelt werden, wie die Lücke in der Leistungsfähigkeit der Modelle bei unterschiedlichen Trainingsdatenmengen verkleinert und damit die absolute Validierungs-Accuracy des Ansatzes auch bei geringen Datenmengen weiter gesteigert werden kann.

Nachdem die dateneffiziente Klassifikation im Kontext technischer Domänen in den vorherigen Abschnitten untersucht wurde, soll im folgenden Abschnitt zur Identifikation von Pittings in Bilddaten zurückgekehrt und die Untersuchungen sollen zusammengefasst werden.

6.2.5 Fazit zu der Klassifikation von Oberflächenzerrüttung

Der Vergleich des Ansatzes basierend auf Merkmalen, die durch Domänenwissen extrahiert wurden, mit dem Deep-Learning-basierten Ansatz zeigt, dass in beiden Fällen zuverlässig Pittings in Bilddaten klassifiziert werden können.

Die Untersuchungen zeigen weiterhin, dass Merkmale, die mittels Domänenwissen definiert wurden, zur Klassifikation von Pittings in Bildern des KGTs verwendet werden können. Die Wichtigkeit der Merkmale zur Klassifikation konnte gezeigt werden. Ein interessanter Aspekt hierbei ist, dass die Wichtigkeit der zur Klassifikation verwendeten Merkmale mit der qualitativen Wichtigkeit der Merkmale für die Domänenexpert*innen übereinstimmt.

Ein entscheidender Vorteil des Ansatzes, zunächst Merkmale händisch auf Basis von Domänenwissen zu definieren, ist das damit einhergehende Problemverständnis, das bei der Generierung der Merkmale aufgebaut wird. Die Analyse der Bilddaten und die Suche nach diskriminativen Merkmalen erlaubt es den Prozessexpert*innen, ein tiefes Verständnis für die Problemstellung zu entwickeln und damit der ersten und zweiten Phase des CRISP-DM-Prozesses (Chapman & Clinton et al. 2000), dem Business Understanding und Data Understanding, Rechnung zu tragen.

Basierend auf den Erkenntnissen in dieser Dissertation sowie den Vorteilen der untersuchten Ansätze ist in Abbildung 6-53 ein Prozess als Vorgehensempfehlung zur Implementierung von Machine-Learning-Modellen in der technischen Domäne dargestellt. Angelehnt an den CRISP-DM-Prozess (Chapman & Clinton et al. 2000) steht die Bildung von Domänenwissen zur effizienten Lösung eines Problems im Fokus.³³

Im Zuge des Prozesses der Problemanalyse und Merkmalsdefinition ist es möglich, ein Merkmal zu identifizieren, das als diskriminatives Merkmal zur Klassifikation genutzt werden kann. Ist ein solches Merkmal identifiziert, kann ein Klassifikationsalgorithmus mit einer nur geringen Menge an Daten trainiert werden oder es kann gänzlich auf Maschinelles Lernen verzichtet werden. Ein Einsatz von Deep-Learning-Architekturen, die große Datenmengen benötigen, ist damit nicht weiter notwendig.

In einem Setup, das alleine auf Maschinellern Lernen bzw. spezifischer auf Deep Learning basiert, müssen Merkmale zunächst aus den Daten erlernt werden. Es lässt sich somit schließen, dass es von Vorteil ist, einen Ansatz zur Klassifikation basierend auf definierten Merkmalen als Grundarchitektur zu implementieren, um hierüber Prozessverständnis zu generieren und einen Vergleichswert für die Klassifikation mittels Deep-Learning-Architekturen zu schaffen.

Werden mittels dieses ersten Schrittes zufriedenstellende Ergebnisse erzielt, so kann auf weitere Untersuchungen verzichtet werden. Sind die Ergebnisse nicht zielführend, sollten die entwickelten Merkmale überdacht und ggf. neue Merkmale implementiert werden.

Parallel dazu ist zu empfehlen, ein Deep-Learning-Modell auf Basis der vorhandenen Datenmenge zu implementieren. Sind die Ergebnisse des Deep-Learning-Ansatzes zielführend, kann dieses Modell implementiert werden. Sind die Ergebnisse nicht zielführend, aber vielversprechend, können auf Basis dieses Ergebnisses weitere Daten aufgenommen und der Deep-Learning-Ansatz kann weiter entwickelt werden.

Führt die Optimierung des Deep-Learning-Ansatzes zu zufriedenstellenden Ergebnissen, so kann die Architektur implementiert werden. Sind die Ergebnisse hingegen unzureichend und weiterhin nicht vielversprechend, so sollten die Datengrundlage sowie die Problemstellung kritisch hinterfragt werden. Beispielsweise sollte das Labeling der Daten überprüft werden, um sicherzustellen, dass die Zuordnung der Label korrekt ist.

³³ Die Inhalte stellen eine qualitative Empfehlung des Autors dar.

Weiterhin sollte kritisch hinterfragt werden, ob die aufgenommene Information für die Repräsentation der Problemstellung geeignet ist. So lassen sich etwa Defekte im Gefüge von Metallen höchstwahrscheinlich nicht auf Basis von Standard-Kameraaufnahmen erkennen und es muss ein geeignetes Sensorsystem implementiert werden.

Nach diesem Vorgehen wird sichergestellt, dass in allen Fällen eine kritische Auseinandersetzung mit dem Prozess stattfindet und gleichzeitig die Vorteile der Generierung von Merkmalen sowie des Deep Learnings optimal ausgenutzt werden.³⁴ Dieses Vorgehen steht in Einklang mit dem in der Literatur beschriebenen Vorgehen zur Durchführung von Machine-Learning-Projekten.

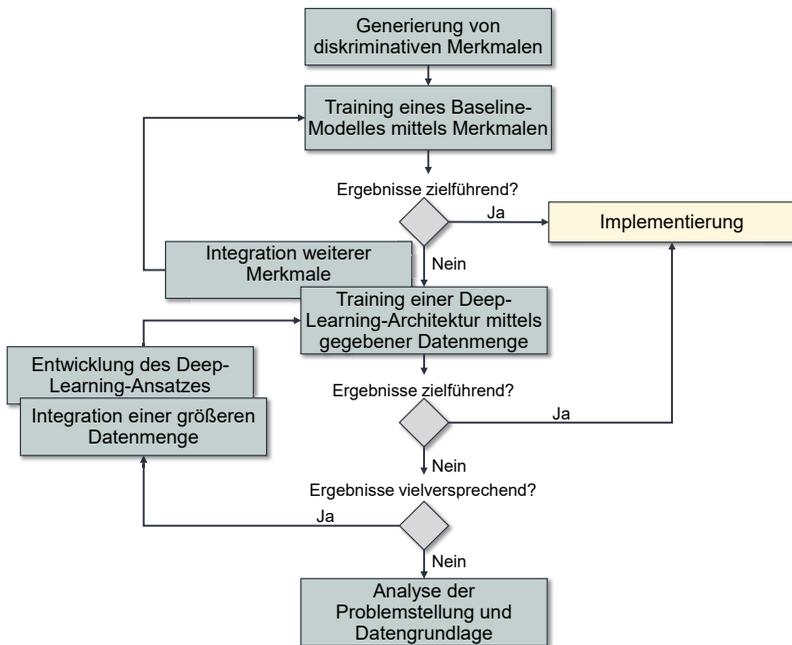


Abbildung 6-53 Handlungsempfehlung zum Einsatz von Maschinellem Lernen unter Verwendung von klassischen merkmalsbasierten Ansätzen und Deep-Learning-basierten Ansätzen.

³⁴ Diese Betrachtung beschreibt den Prozess der Modellbildung unter Beachtung der erreichten Genauigkeit. Weitere Zielgrößen, bspw. die erforderliche Rechenkapazität für das Training und die Inferenz oder die Kosten für die Generierung von Trainingsdaten, können den Prozess beeinflussen und sind hier nicht dargestellt.

Als Ergebnis wurde eine Zusammenfassung der Erkenntnisse der vorherigen Abschnitte gegeben und daraus wurde ein qualitatives Vorgehensmodell zum Einsatz von Maschinellern in der technischen Domäne abgeleitet.

In den vorherigen Abschnitten wurde die Entwicklung in den Bilddaten empirisch analysiert und es wurden unterschiedliche Ansätze zur automatischen Identifikation der Fehlstellen in Bilddaten untersucht. Mit Blick auf die autonome Produktionsmaschine sowie eine vorausschauende Instandhaltung ist es notwendig, dass Fehlstellen nicht nur identifiziert und quantifiziert, sondern ebenfalls in ihrer Entwicklung prognostiziert werden können. Dies ist notwendig, um einen zukünftigen Schädigungszustand möglichst frühzeitig abzuschätzen. Die Untersuchung eines solchen Prognosesystems ist das Ziel der folgenden Kapitel.

7 Verschleißprognose

In den vorherigen Kapiteln wurde die Entwicklung des Verschleißerscheinungsbildes Pitting untersucht und dabei ein charakteristisches Wachstum beobachtet. Aufbauend auf diesen Erkenntnissen soll in diesem Kapitel die Möglichkeit untersucht werden, automatisiert aus den Bilddaten das Wachstum von Pittings zu prognostizieren und damit den Verschleißzustand des KGTs einzuordnen. Ziel ist es hierbei, eine Prognosefunktion zu schaffen, die sich auf Basis der aktuellen Daten selbständig an das Wachstum der Schädigungen anpasst. Hierbei sei zunächst erwähnt, dass die Lebensdauer eine Größe ist, die anwendungsabhängig zu spezifizieren ist. Das primäre Ziel des Kapitels ist es damit zu zeigen, dass eine geeignete Funktion und Methodik zur Prognose der Schädigungsentwicklungen auf Basis der Bilddaten grundsätzlich aufgestellt werden kann. Eine Funktion zur Prognose der Entwicklung von Pittings am Kugelgewindtrieb ist im Stand der Forschung bisher nicht betrachtet. Der Ansatz zeichnet sich gegenüber schon bestehenden Ansätzen besonders dadurch aus, dass erstmals eine direkte Information über das Schadensmerkmal verarbeitet wird. Weiterhin wird diese Information kontinuierlich verwendet, um die Prognosefunktion zu aktualisieren. Die gezeigten Ergebnisse wurden im Rahmen von (Schlagenhauf & Burghardt 2021), (Schlagenhauf & Landwehr 2021) sowie (A_Burkhardt 2021) erarbeitet und werden im Folgenden zusammengefasst dargestellt.

7.1 Grundüberlegungen

Zunächst werden Grundüberlegungen zum Einsatz von Verfahren des Maschinellen Lernens sowie zur Integration von Expertenwissen zur Umsetzung eines Prognosesystems angestellt.

Verfahren des Maschinellen Lernens sind in der Lage, zuverlässig aus aktuellen Daten zu lernen. Darüber hinaus können mittels Domänenwissen über das Verschleißverhalten des KGTs wertvolle Information integriert werden. Hieraus lässt sich die Hypothese ableiten, dass durch eine Kombination ein performantes Tandem zur Prognose geschaffen werden kann.

Aus den vorhergehenden Untersuchungen der Entwicklung von Pittings in den Bilddaten aus Kapitel 6 ist bekannt, dass ein Pitting, das zu einem bestimmten Zeitpunkt entsteht, auch zu jedem späteren Zeitpunkt an derselben Position auf der Spindel vorhanden ist, d. h. $p(x^t) = p(x^{t+1})$ mit $p(\cdot)$ als Funktion, welche die Position der Schädigung

auf der Spindel zurückgibt. Ebenfalls wird ein Pitting in seiner Größe nicht kleiner, sondern lediglich größer werden, d. h. $a(x^t) \leq a(x^{t+1})$ mit $a(\cdot)$ einer Funktion, welche die Fläche eines Pittings zurückgibt.

Kritische Größe bei der Verschleißprognose ist die Genauigkeit pro Zeitschritt, mit der ein Modell die Entwicklung prognostizieren kann. Je früher ein Modell die Entwicklung korrekt vorhersagen kann, d. h. je effizienter es die Vergangenheitswerte zur Prognose nutzen kann, desto frühzeitiger kann eine bevorstehende Wartung geplant werden. Die vorhergesagte Größe eines Fehlers (\hat{a}_x^{t+j}) an der Position x zu einem j Zeitschritte in der Zukunft liegenden Zeitpunkt wird also durch eine Funktion f einer Menge an Funktionen F bestimmt. Die Aufgabe ist es, diejenige Funktion f zu finden, welche die zukünftige Entwicklung von Schädigungen auf Basis von historischen Entwicklungen am besten beschreibt, d. h. diejenige Funktion f zu finden, die eine gewählte Kostenfunktion K am ehesten minimiert. Formal lässt sich diese Aufgabe mit $\operatorname{argmin}_f(\sum_{j=1}^k K_f(\hat{a}_x^{t+j}, a_x^{t+j}))$ ausdrücken.

Zur Prognose eines Fehlers in den Bilddaten wird mittels eines dreistufigen Prozesses vorgegangen, wie er in Abbildung 7-1 dargestellt ist. Zunächst wird der Fehler im Bild lokalisiert. In einem zweiten Schritt folgt die Quantifizierung des Fehlers. Zu diesem Zeitpunkt liegt ein klassisches System zur Zustandsüberwachung vor, das den aktuellen Zustand der Komponente beschreibt.

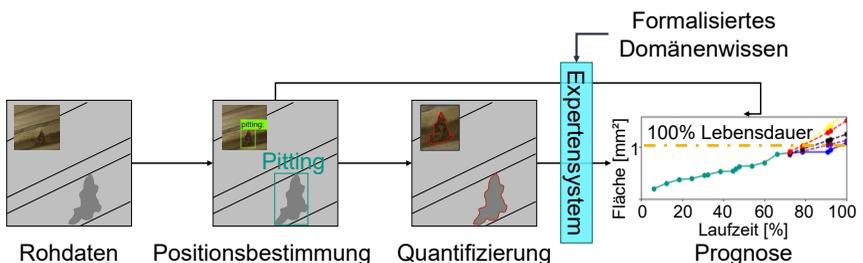


Abbildung 7-1 Prozess zur Prognose der Größenentwicklung von Fehlstellen nach (Schlagenhauf & Burghardt 2021).

Um das System zu befähigen, Aussagen über die zukünftige Entwicklung eines Fehlers machen zu können, bedarf es zusätzlich eines Prognosemoduls sowie der Einordnung der prognostizierten Größe relativ zur Definition der 100 % Lebensdauer. Ein relevanter Aspekt ist hierbei die Integration von Domänenwissen zur Integration von bestehendem

Wissen über die Schädigungsentwicklungen. Durch eine Formalisierung des Domänenwissens wird ein Expertensystem abgeleitet, das regulierend auf die Vorhersagen des Quantifizierungsmoduls einwirken kann.

Als Ergebnisse wurden die Kombination von Verfahren des Maschinellen Lernens sowie die Integration von Domänenwissen durchgeführt und der Prozess zur Prognose der Schädigungsentwicklung wurde definiert.

In den folgenden Abschnitten wird auf die Einzelelemente zur Prognose der Schädigungsentwicklung eingegangen.

7.2 Positionsbestimmung und Quantifizierung

Basis für die Prognose des Fehlers sind die Bestimmung seiner Position im Bild sowie die anschließende Berechnung seiner Größe.

Die für die Modellbildung verwendeten Bilddaten entsprechen den in Abschnitt 6.1 analysierten sowie den von (Schlagenhauf & Landwehr 2021) vorgestellten Daten. Die Bilder zeigen keine, kleine oder wachsende Fehler in unterschiedlichen Größen, wie sie in Abbildung 7-9 dargestellt sind. Zur Bestimmung der Position eines Fehlers in einem Bild wird das Modell EfficientDet D0 512x512³⁵ aus der Tensorflow Object Detection API verwendet. Gegenüber dem in Abschnitt 6.2.3 vorgestellten Modell besitzt dieses den Vorteil, dass es die Position einer Schädigung im Bild direkt grob bestimmen kann, wenngleich die Genauigkeit bei der Klassifikation geringer ist (siehe unten). Zur Anpassung des Modelles auf die vorliegende Aufgabe wird das letzte Layer des Modelles mittels 170 RGB-Bildern (190 x 190 Pixel) angepasst, wobei 70 % der Bilder für das Training und 30 % für den Test des Modelles verwendet werden. Abbildung 7-2 zeigt die Ergebnisse des Detektionsmodelles für 60 neue Bilder, die im Training des Modelles nicht vorhanden waren. Das Modell erreicht hierbei eine Validierungsgenauigkeit von 92 %.

³⁵ TensorFlow (2022), *efficientdet/d0*. <https://tfhub.dev/tensorflow/efficientdet/d0/1> [26.01.2022]

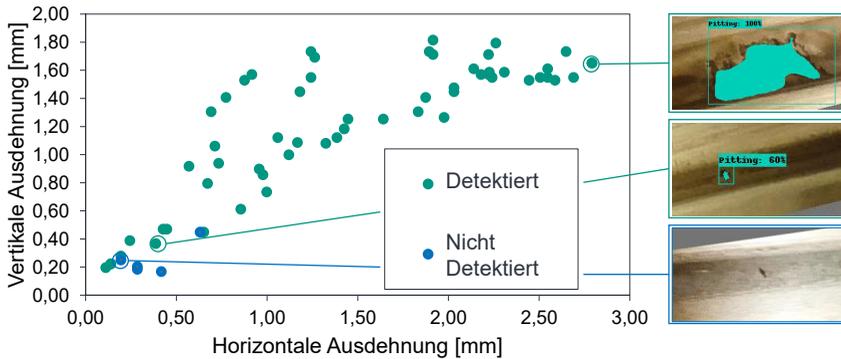


Abbildung 7-2 Validierung des Modelles zur Detektion von Fehlstellen nach (Schlagenhaut & Landwehr 2021).

In Grün sind jene Bilddaten dargestellt, die korrekt detektiert werden konnten, in Blau diejenigen, die das Modell nicht korrekt lokalisieren konnte. Es ist ersichtlich, dass das Modell lediglich sehr frühe Fehlstellen übersieht. Eine Korrelation der Genauigkeit mit der Größe der Fehler ist ersichtlich.

Die vom Modell zurückgegebene, den Defekt einschließende Box dient als erste Quantifizierung seiner Ausdehnungen. Der zurückgegebene Wert kann bei Bedarf direkt verwendet werden.

Zur Steigerung der Genauigkeit wird der Bereich innerhalb der Box mittels eines Schwellwert-Algorithmus (Bradski, Kaehler & Bradski 2017) weiter verarbeitet, der die Fläche des Fehlers von der umliegenden Fläche segmentiert. Dabei werden die besten Ergebnisse erzielt, wenn das Bild als binäres Bild vorliegt, da dann Kanten markant sichtbar sind. Die Binarisierung eines Bildes erfolgt mittels der Schwellwertfunktion $f: \{0, \dots, 255\}^{n \times m \times p} \rightarrow \{0, 1\}^{n \times m \times p}, f(x) = \begin{cases} 1: x > \gamma \\ 0: x \leq \gamma \end{cases}$, wobei n, m, p die Dimensionen (Länge, Breite, Farbkanäle) des Bildes darstellen und γ einem Schwellwert entspricht, bei dessen Überschreiten der Wert des Pixels x auf 1 (schwarz) gesetzt wird. Das Resultat ist ein Schwarz-Weiß-Bild, wobei die Fehlerfläche schwarz und die umliegende Fläche weiß eingefärbt ist.

Da die Bilder der Fehlstellen in unterschiedlichen Ausprägungen sowie mit unterschiedlichen Verschmutzungsgraden vorliegen, existiert kein einheitlicher Schwellwert, der es erlaubt, alle Bilder so zu binarisieren, dass die Schadensfläche klar extrahiert werden

kann. Um eine zuverlässige Extraktion der Fehlerfläche zu gewährleisten, wird ein Meta-Algorithmus angewandt, der als Vorstufe für den Schwellwertalgorithmus dient. Der vorgestellte Meta-Algorithmus zur Vorhersage von optimalen Schwellwerten ist so im Stand der Technik nicht vorhanden. Dieser Meta-Algorithmus besteht aus einem CNN, das optimale Schwellwert-Parameter auf Basis eines Eingangsbildes vorschlägt. Es werden sechs Schwellwerte gewählt, die einen möglichst großen Bereich abdecken: $m: x \rightarrow \{35, 40, 45, 52, 62, 72\}$. Die Auswahl der Schwellwerte erfolgte auf Basis einer Schwellwertanalyse, wobei festgestellt wurde, dass eine Erhöhung der Anzahl an Schwellwerten im vorliegenden Fall zu keiner Verbesserung der Vorhersage führt.

Zur Prognose der Schwellwerte wird das Metamodell zunächst mit 600 Bildern nach dem in Abbildung 7-3 dargestellten Schema trainiert. 70 % der Bilddaten wurden für das Training und 30 % der Bilddaten für die Validierung des Modelles verwendet. Zur Vorbereitung für das Training wird zunächst für jedes Bild mit allen Schwellwerten eine Kantenextraktion durchgeführt und die Ergebnisse werden analysiert. Diese Bilder werden begutachtet und derjenige Schwellwert, der nach Experteneinschätzung zum besten Ergebnis führt, wird als Label für das Bild übernommen.

Es ist ersichtlich, dass für den Großteil der Bilder ein spezifischer Schwellwert zu einer optimalen Kantenextraktion führt. Weiterhin ist zu beobachten, dass sich die Schwellwerte unterscheiden und dass die Wahl eines nicht optimalen Schwellwertes nicht zwangsläufig zu einer Nicht-Funktionsfähigkeit des Algorithmus, sondern lediglich zu suboptimalen Ergebnissen führt. Die mit den Schwellwerten versehenen (mit Labels versehenen) Bilddaten werden für das Training des CNN verwendet. Das CNN besteht hierbei aus vier Convolutional-Schichten, die jeweils von einer 2x2-max-Pooling-Operation gefolgt werden. Ein linearer Klassifikator wird zur Klassifikation der Bilder verwendet.

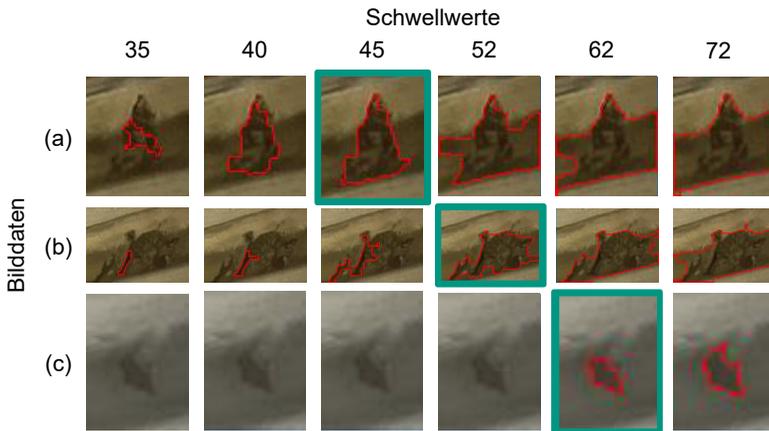


Abbildung 7-3 Prozess zur Auswahl des optimalen Schwellwertes nach (Schlagenhauf & Burghardt 2021). Optimaler Schwellwert ist grün umrandet.

Abbildung 7-4 zeigt den Vergleich mit einem Standardansatz (Otsu 1979) zur Schwellwertfindung. Es ist ersichtlich, dass das klassische Verfahren nicht in der Lage ist, den notwendigen Schwellwert vorherzusagen, wohingegen das konzipierte Metamodell gute Ergebnisse erzielt.

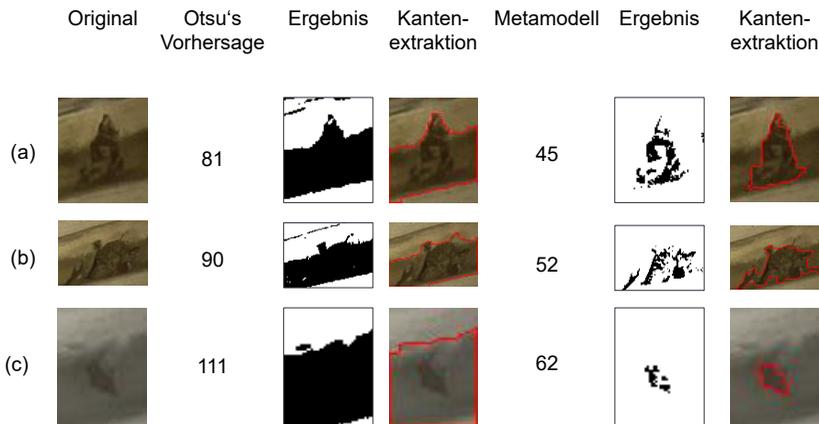


Abbildung 7-4 Initialer Vergleich eines klassischen Ansatzes zur Kantenextraktion mit dem verwendeten Metamodell zur Vorhersage der optimalen Schwellwerte nach (Schlagenhauf & Burghardt 2021).

Die Pipeline zur Positionsbestimmung und Quantifizierung kann folgendermaßen zusammengefasst werden: Zunächst wird durch das EfficientDet-Modell ein Bildausschnitt extrahiert, der ein Pitting enthält. Durch das Metamodell wird anschließend ein geeigneter Schwellwert zur Binarisierung dieses Bildausschnittes ausgewählt, sodass die anschließende Kantendetektion und Segmentierung der Fehlerfläche optimale Ergebnisse liefert.

Als Ergebnis liegt ein Metamodell vor, das den optimalen Schwellwert für die Kantenextraktion mit einer Genauigkeit von 92 % vorhersagen kann. Das Modell zur Bestimmung des optimalen Schwellwertes kann damit als validiert angesehen und im nächsten Schritt zur Annäherung der realen Schadensfläche verwendet werden.

Im folgenden Abschnitt werden die Vorhersagen des Quantifizierungsmoduls (Detektion, Metamodell, Kantenextraktion) mittels eines Expertensystems verfeinert.

7.3 Integration von Domänenwissen

In Abbildung 7-1 wird Domänenwissen an der Schnittstelle zwischen der Größenextraktion und der Prognose der Größenentwicklung mittels eines Expertensystems integriert. Dies kann an dieser Stelle als zielführend eingeschätzt werden, da durch die Analyse der Bilddaten Wissen über die reale Entwicklung von Pittings vorliegt, das an dieser Stelle sinnvoll integriert werden kann.

Dabei werden die Größen, die vom Quantifizierungsmodell zurückgegeben werden, geprüft und die Werte ggf. hin zu realistischeren Werten korrigiert. Die Korrektur der Werte basiert auf der Annahme der zeitlichen Kontinuität von Schädigungen. Ist eine Schädigung einmal vorhanden, so wird sie nicht mehr verschwinden, sondern lediglich über die Zeit wachsen. Dies bedeutet, dass eine Schädigung, die zum Zeitpunkt t an Position x detektiert wurde, auch zum Zeitpunkt $t + n$ an Position x vorhanden sein muss. Wahrgenommene Größenunterschiede werden dann bspw. durch Verschmutzungen verursacht, welche die Schädigung überdecken. Dies lässt sich mit folgender Formel beschreiben: $c_x^t = \begin{cases} 1 & \text{für } c_x^{t-1} \vee B_x^t \\ \text{sonst } 0 & \end{cases}$. Eine bestimmte Stelle im Bild, spezifiziert durch vier

Koordinaten (abgekürzt mit x), wird zu einem Zeitpunkt t als 1, d. h. ein Pitting enthaltend, klassifiziert, falls für diese Position vom Detektionsmodell eine Bounding-Box B_x gefunden wurde. Ansonsten wird der Variablen der Wert 0 zugeschrieben. Damit wird eine Logik implementiert, die prüft, ob zu einem früheren Zeitpunkt an einer bestimmten Stelle schon ein Fehler vorhanden war.

Aus den Analysen in Kapitel 6 ist bekannt, dass sich die Fläche eines Pittings durch eine Ellipse annähern lässt. Auf dieser Basis wird die Fläche der umschließenden Bounding-Box in eine Ellipse transferiert, indem die Seiten des Rechtecks als Halbachsen der Ellipse verwendet werden. Die Funktion $g(t) = A_t$ stellt die Methode zur Extraktion der Größe der Schädigung dar. Ist $g(t+1) < g(t) * 0.9$, so widerspricht dies der Tatsache, dass ein Fehler in seinen Dimensionen nicht kleiner werden kann, und wird als falsch angenommen. Es wird $g(t) * 1.01$ zurückgegeben. Die Werte wurden gewählt, um ein kontinuierliches leichtes Wachstum sicherzustellen. Weiterhin haben die Versuche gezeigt, dass die Situation eintreten kann, dass ein Fehler bspw. aufgrund von Verschmutzung vom Quantifizierungssystem überschätzt wird. Unter der Annahme, dass ein Fehler in seinen Dimensionen nicht kleiner werden darf, resultiert dies in einer dauerhaften Überschätzung des Fehlers. Durch den Faktor 0.9 ist ein langsames Anpassen der Größe auch in negativer Richtung zulässig.

Weicht die vom Segmentierungsmodul für den Zeitschritt t zurückgegebene Größe innerhalb der vorgegebenen Grenzen von der zurückgegebenen Größe zum Zeitschritt $t - 1$ ab, so muss dies im Modell berücksichtigt und eine Anpassung der Werte vorgenommen werden. Hierzu wird eine prognostizierte Größe A_{t*} mittels der Berechnungsvorschrift $A_t = A_{t*} * (1 - (A_{t-1} - S_t) * \beta * e^{-\frac{1}{2}(A_{t-1} - S_t)})$ verarbeitet. S_t steht dabei für die Segmentierungsfläche des Fehlers. Der Faktor β dient der Gewichtung und wird mit einem niedrigen Wert von 0.2 gewählt, um lediglich moderate Anpassungen vorzunehmen. Mittels dieser Berechnungsvorschrift wird der aktuell vom Expertensystem zurückgegebene Wert mit der Abweichung zwischen der Fehlerfläche zum Zeitpunkt $t - 1$ und der Segmentierungsfläche gewichtet.

Aus der Versuchsanalyse ist weiterhin bekannt, dass Fehler zwar stetig wachsen, große Zuwächse zu vorherigen Fehlerflächen allerdings unwahrscheinlich sind. Aus diesem Grund werden starke Änderungen mit einem sinkenden Faktor gewichtet. Starke Flächenzuwächse sind damit zwar möglich, allerdings sinkt deren Wahrscheinlichkeit. Das Vorgehen zur Anpassung ist in Abbildung 7-5 dargestellt. Die gelbe Kurve zeigt die Kurve zur Regulierung des aktuellen Differenzwertes und gibt damit die Wahrscheinlichkeit für eine große Abweichung vor. Die Multiplikation mit dem absoluten Wert der Abweichung resultiert in der grün dargestellten Kurve. Es ist ersichtlich, dass große Korrekturwerte zunehmend geringer gewichtet werden.

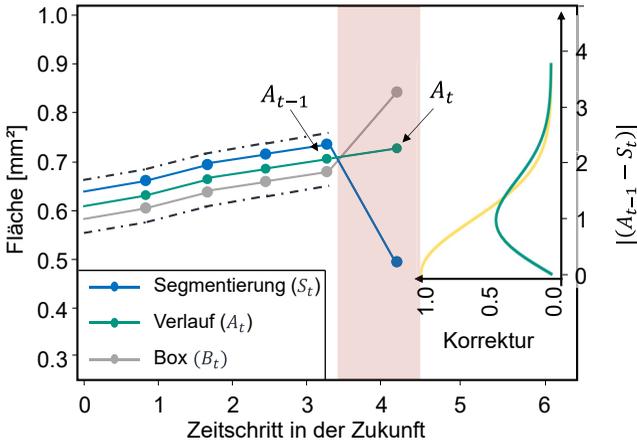


Abbildung 7-5 Anpassung der Fehlerfläche mittels eines gewichteten Abstandsmaßes.

Abbildung 7-6 stellt die vom Quantifizierungsmodul zurückgegebene Größe einer Schädigung zu unterschiedlichen Zeitpunkten dar. Dargestellt sind ebenfalls die vom Expertensystem verarbeiteten Werte des Quantifizierungsmoduls sowie die Realdaten. Hierbei entspricht 0 % der Zeit, zu der erstmals eine Schädigung identifiziert wurde. Die Laufleistung bezieht sich damit auf den Zeitraum zwischen Identifikation eines Fehlers und Ausfall der Komponente. Es ist ersichtlich, dass das Modell zur Quantifizierung die Größe der Schädigung approximiert, diese zu Beginn jedoch leicht unterschätzt. Zu späteren Zeitpunkten neigt das Modell hingegen dazu, die Größe der Schädigung eher zu überschätzen.

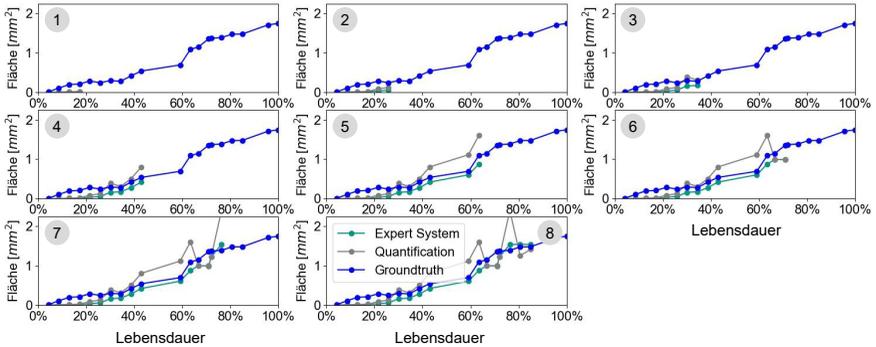


Abbildung 7-6 Vergleich der vom System berechneten Fehlergröße mit Experten-system (Expert System) und ohne Expertensystem (Quantification) im Vergleich zu den Realdaten (Groundtruth) nach (Schlagenhauf & Burghardt 2021).

Als Ergebnis wird die zurückgegebene Größe durch Integration des Expertensystems deutlich näher an den Realverlauf angenähert und Ausreißer werden reduziert. Zu betonen ist, dass das Expertensystem lediglich die vom Quantifizierungsmodul segmentierten Größen erhält, d. h. die Groundtruth-Daten dienen lediglich der Überprüfung der Genauigkeit des Ansatzes.

Im Folgenden wird eine passende Funktion zur Prognose der vom Expertensystem zurückgegebenen Werte untersucht.

7.4 Prognose

Auf Basis der mittels Domänenwissen korrigierten Größen soll im Folgenden ein Modul zur Prognose dieser Größen konzipiert werden. Hierzu werden zunächst Anforderungen an eine Prognosefunktion aufgestellt und die Erfüllung dieser Anforderungen wird durch verschiedene Prognosefunktionen überprüft. Im Anschluss erfolgt die Integration und Validierung der Prognosefunktion.

7.4.1 Auswahl der Prognosefunktion

Im Folgenden werden auf Basis der Erkenntnisse aus Kapitel 6 Funktionen zur Prognose der Entwicklung von Schädigungen untersucht. Aufgrund des Verlaufes der Schädigungen in den Bilddaten werden Funktionen ersten Grades, Funktionen zweiten Grades, Funktionen dritten Grades sowie Exponentialfunktionen als in Frage kommende Funktionen betrachtet. Als Maß für die Güte des Systems wird zunächst der *Root Mean*

Squared Error (RMSE) zwischen den prognostizierten Werten und den Realdaten implementiert: $E = \frac{1}{j} \sum_{j=1}^j \sqrt{(\hat{a}_x^{t+j} - a_x^{t+j})^2}$, mit \hat{a}_x^{t+j} gleich der prognostizierten Größe j Zeitschritte nach dem Zeitschritt t und a_x^{t+j} dem zugehörigen Groundtruth-Wert. Hierbei ist relevant, dass die Funktion für die Zeitschritte 0 bis t gebildet wird, die Prognosegüte aber lediglich für die Zeitschritte $t + j$ berechnet wird. Dies ist relevant, da eine Funktion gewünscht ist, die in der Lage ist, neue, für das Training nicht vorhandene Datenpunkte zu prognostizieren.

Neben der Anforderung, dass der Fehler der Prognosefunktion lediglich für neue Daten berechnet werden darf, ergeben sich nach (Nesges 2005) und (Wieser 2009) bei dem Betrieb einer Maschine oder Komponente eine mittlere Zeit zwischen Fehlern (MTBF) sowie eine Zeit zur Planung und Umsetzung von Instandhaltungsmaßnahmen. Aus praktischer Sicht ist die Prognose von Punkten, die weit über den notwendigen Planungshorizont hinaus in der Zukunft liegen, nicht von großem Vorteil. Ebenso ist eine Prognose, die kurzfristig sehr genau ist, von geringem Vorteil, da dann der notwendige Planungshorizont nicht mehr eingehalten werden kann. Es ergibt sich also ein individuell zu wählendes Optimum, bei dem die Prognose möglichst exakt sein soll. Dieses Optimum muss anwendungsspezifisch auf Basis des gewünschten Planungshorizontes gewählt werden.

Hinsichtlich dieses Aspektes wird der Fehlerterm wie folgt erweitert:

$E = \frac{1}{j} \sum_{j=1}^j f(j) \sqrt{(\hat{a}_x^{t+j} - a_x^{t+j})^2}$ – und damit wird die Möglichkeit implementiert, Fehlerwerte zeitabhängig zu gewichten. Die Funktion $f(j)$ wird hierbei als Glockenkurve mit $f(j) = e^{-0.15 * (\text{ceil}(\frac{\alpha}{2}) - j)^2}$ gewählt.³⁶ Damit werden in der Zukunft liegende Punkte entsprechend gewichtet. Dies ist auch in Abbildung 7-7 dargestellt. α stellt einen Term dar, der die in der Zukunft liegenden Zeitpunkte gewichtet und die Glockenkurve entlang der Zeitachse verschiebt. Mittels α kann eingestellt werden, wie weit in der Zukunft die höchste Genauigkeit liegen soll und wird mit dem empirisch für diesen Datensatz ermittelten Wert von 7 gewählt. Der zu einem Zeitpunkt berechnete Fehler wird also gewichtet, wobei aufgrund der Symmetrie der Funktion die maximale Gewichtung von 1 am

³⁶ $\text{ceil}(\cdot)$ rundet die Gleitkommazahl auf die nächste \mathbb{Z} auf.

Hochpunkt der Glockenkurve (4 in diesem Beispiel) liegt. Werte davor ($j < \text{ceil}(\frac{\alpha}{2})$) sowie Werte danach ($j > \text{ceil}(\frac{\alpha}{2})$) werden geringer gewichtet.

Einen weiteren praktischen Aspekt für die Prognose von Fehlern stellt die Dateneffizienz dar, mit der das Modell korrekte Prognosen geben kann. Je weniger Datenpunkte für das Training des Modelles verwendet werden müssen, desto dateneffizienter ist das Modell und desto frühzeitiger können korrekte Aussagen getroffen werden.

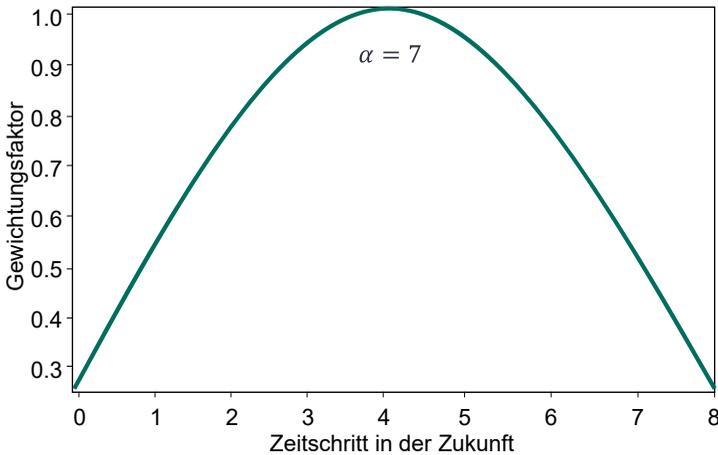


Abbildung 7-7 Qualitative Darstellung der Gewichtungsfunktion.

Zur Implementierung dieses Verhaltens wird der Fehlerterm reziprok zur verwendeten Anzahl an Datenpunkten gewichtet und der summierte Fehler über eine Anzahl an N Datenpunkten berechnet. Der erweiterte Fehlerterm ergibt sich mit:

$$E = \sum_{p=\beta}^N p^{-1} \left(\frac{1}{J} \sum_{j=1}^J f(j) \sqrt{(\hat{a}_x^{t+j} - a_x^{t+j})^2} \right) = \sum_{\beta}^N p^{-1} RMSE_{\beta}^{\alpha}. \text{ Die Anzahl der für das}$$

Training verwendeten Datenpunkte ist mit β dargestellt.

Dieser Fehlerterm wird für den Vergleich der Funktionen ersten, zweiten und dritten Grades sowie der Exponentialfunktion verwendet. Insgesamt wurden für fünf zufällig ausgewählte Fehlerverläufe Prognosen durch die Funktionen durchgeführt (siehe Tabelle 7-1) und der Fehler E berechnet. Es ist ersichtlich, dass die Regression ersten Grades für alle Schädigungsverläufe in dem niedrigsten Fehler resultiert und damit als

die bestgeeignete Funktion angesehen werden kann. Abbildung 7-8 stellt die Prognose eines Fehlers zusammen mit den resultierenden Fehlerwerten dar.

Tabelle 7-1 Prognosegüte ausgewählter Funktionen zur Vorhersage der Schädigungsentwicklung.

Schädigungs- verlauf	Funktion 1. Grades	Funktion 2. Grades	Funktion 3. Grades	Exponentialfunktion
1	0.018	0.079	0.188	0.317
2	1.286	4.972	25.973	160.951
3	0.664	0.736	4.116	1.635
4	0.042	0.151	0.287	0.134
5	0.156	0.627	1.117	373.91

In den Fehlerwerten ist der Einfluss der erweiterten Fehlerdefinition ersichtlich, die mittelfristige Prognosen stärker gewichtet und die Dateneffizienz implementiert. Fehlerwerte sind dann gering, wenn der Abstand der Kurven mittelfristig gering ist, und besonders dann, wenn dies mittels weniger Datenpunkte erreicht werden kann. Langfristige Abweichungen haben dagegen einen weniger starken Einfluss.

Als Ergebnis wurden Anforderungen an eine Prognosefunktion definiert und Untersuchungen mit Funktionen ersten, zweiten und dritten Grades sowie mit Exponentialfunktionen durchgeführt. Die lineare Funktion konnte als bestgeeignete Funktion zur Prognose identifiziert werden.

Im folgenden Abschnitt wird die Prognosefunktion validiert.

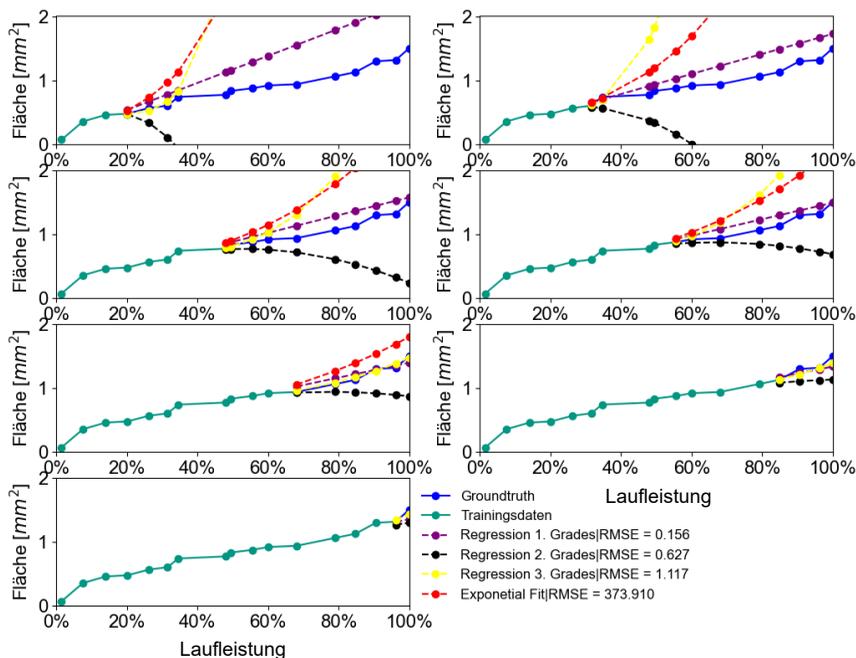


Abbildung 7-8 Prognose einer Schädigung, zeilenweise von links nach rechts mit zunehmender Anzahl an Trainingsdaten nach (Schlagenhau & Burghardt 2021).

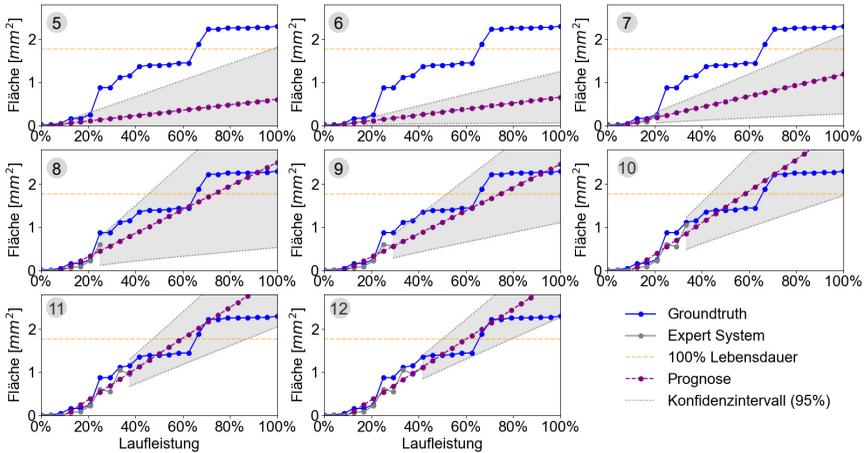
7.4.2 Validierung des Prognosesystems

Im vorherigen Schritt wurde die lineare Funktion auf Basis von historischen Daten gebildet und die Prognosegüte für zukünftige, nicht während des Trainings vorhandene Daten geprüft. Damit hat eine erste Validierung des Prognosesystems stattgefunden.

Zur weiteren Validierung des Prognosesystems werden für sechs neue Schädigungen die Entwicklungen prognostiziert. Hierbei werden möglichst frühzeitige Schädigungen ausgewählt, da diese als erste beobachtet werden und damit die höchste Relevanz für die Bestimmung der Lebensdauer aufweisen. Zur Prognose der Schädigungsverläufe werden die Schritte aus Abbildung 7-1 sequentiell durchgeführt. Zunächst werden die Defekte mittels des Detektionsalgorithmus detektiert, um dann in einem nächsten Schritt die Größe der Schädigungen zu extrahieren. Daraufaufgehend werden die Werte

mittels des Expertensystems verarbeitet. Anschließend folgt die Prognose der Schädigungsentwicklungen.

Zur Validierung des Prognosesystems hinsichtlich eines rechtzeitigen Tausches der Komponente vor einem mechanischen Ausfall wird zusätzlich die 100-%-Lebensdauermarkante angelehnt an die erweiterte Definition der Lebensdauer des KGTs ($100\% \text{ Lebensdauer} \triangleq 0.3 * \varnothing_{Kugel}$), umgerechnet in Fehlerfläche, integriert (siehe hierzu auch Abbildung 0-6 im Anhang). Damit berechnet sich die maximal zulässige Schadensfläche mit: $A_{max} = \pi * \left(1.5 * 0.3 * \frac{3.969}{2}\right)^2 = 1.77 \text{ mm}^2$, wenn $\alpha = 1.5$ gewählt wird. Bei den vorliegenden Versuchen erhöht ein niedrigerer Wert zwar den Spielraum für den rechtzeitigen Tausch der Komponente, dies geht allerdings zu Lasten der Einsatzdauer der Komponente. Wie in den folgenden Ergebnissen gezeigt werden wird, verbleibt bei Wahl des Faktors 1.5 noch eine Resteinsatzzeit von über 50 % bei Überschreiten der definierten 100 % Lebensdauer. Dies ist auch in Abbildung 7-9 ersichtlich, in der die Prognose eines neuen Fehlers zusammen mit Ausschnitten aus der Entwicklungshistorie eines Fehlers und mit dem angewandten Lokalisierungs- und Quantifizierungsmodul dargestellt ist.



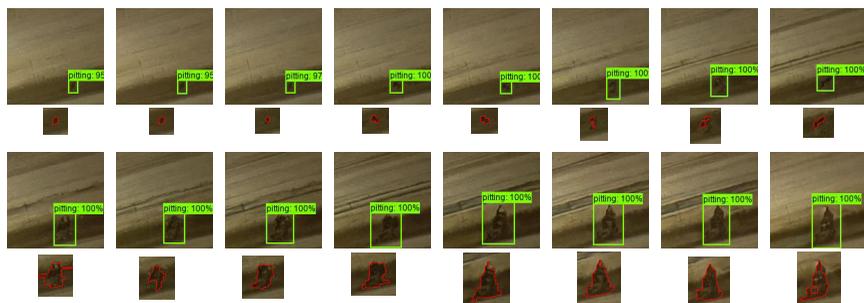


Abbildung 7-9 Validierung des Prognosesystems zusammen mit Ausschnitten des Quantifizierungsmodules, angelehnt an (Schlagenhauf & Burghardt 2021). Die Zahlen in den grau eingefärbten Kreisen stellen die für das Training des Prognosemodelles verwendeten Datenpunkte dar.

Die Prognosefunktion wurde hierbei mittels der Datenpunkte trainiert, die vom Expertensystem zurückgegeben werden. Die Mengen der dafür zur Verfügung stehenden Datenpunkte sind als Zahlen in den grauen Kreisen dargestellt. Analog zum Realeinsatz wird das Modell also mit zunehmend mehr Vergangenheitswerten trainiert, um sich damit an den realen Verlauf der Schädigungen anzupassen.

Es ist ersichtlich, dass sich die Prognosefunktion zügig an den Verlauf der wahren Entwicklung anpasst und in der Lage ist, diese zu prognostizieren. Allerdings wird zu Beginn der Prognose die Fehlergröße unterschätzt und damit die Prognosefunktion als zu flach gewählt, was dazu führt, dass das Erreichen der 100 % Lebensdauer nach dem tatsächlichen Erreichen der 100 % Lebensdauer prognostiziert wird. Dieses Verhalten ist dadurch zu erklären, dass dem Modell zu Beginn lediglich wenige, nämlich fünf Datenpunkte für das Training zur Verfügung stehen. In diesen fünf Datenpunkten ist die reale Entwicklung der Schädigung noch nicht ausreichend codiert, weshalb die Prognosefunktion ungenau ist. Das Modell passt sich mit zunehmender Anzahl an Trainingsdatenpunkten zunehmend besser an den realen Verlauf an.

Im Falle, dass Fehlstellen vom Quantifizierungssystem leicht unterschätzt werden, kann es zu einer für den praktischen Betrieb negativen Abweichung vom realen Verlauf der Schädigung kommen. Aus diesem Grund wurde die Prognosefunktion mit einem 95%-Konfidenzintervall versehen. Dieses Konfidenzintervall beinhaltet den wahren Wert der Prognosefunktion zu 95 %. Da für den praktischen Betrieb besonders ein Überschätzen der Lebensdauer zu einem ungeplanten Stillstand von Maschinen führen

kann, ist vor allem die obere Schranke des Konfidenzintervalls von Relevanz. Diese zeigt das Erreichen der Lebensdauer früher an als die Prognosefunktion und implementiert damit eine Art Sicherheitsfaktor für den Betrieb.

Ein besonders praxisrelevanter Aspekt ist die automatische Anpassung des Konfidenzintervalls. Aus dem Verlauf des Konfidenzintervalls ist ersichtlich, dass die Breite des Intervalls mit der Abweichung der Prognosefunktion korreliert. Das Konfidenzintervall codiert damit die Ungenauigkeit des Modelles und passt sich selbstständig an. Sind die Residuen zwischen Prognosefunktion und den vom Expertensystem zurückgegebenen Werten groß, so codiert das Konfidenzintervall eine hohe Sicherheit und wird damit breiter. Dies führt in der Konsequenz zwar zu einer ungenaueren Prognose, verhindert damit aber einen ungeplanten Ausfall der Komponente. Sind die Residuen gering, so ist auch das Konfidenzintervall schmaler und damit die Sicherheit der Prognose höher.

Zur Validierung der Prognosefunktion werden mittels der definierten, maximal erlaubten Schadensfläche für weitere sechs unterschiedliche Schädigungen Prognosen durchgeführt.

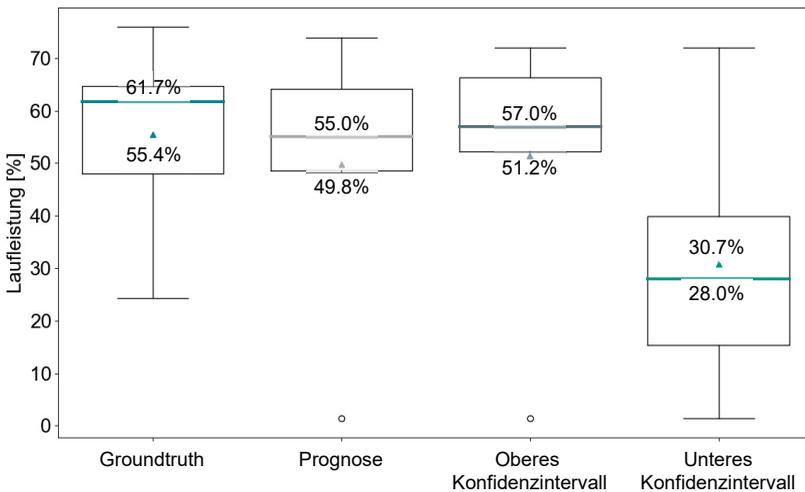


Abbildung 7-10 Abnutzungsvorrat (synonym verwendet zu Remaining Useful Lifetime [RUL]) bei Erreichen der 100 % Lebensdauer nach Definition durch die

Prognosefunktionen. Das Konfidenzintervall erlaubt einen sichereren Betrieb der Komponenten.

Abbildung 7-10 zeigt den verbleibenden Abnutzungsvorrat zum mechanischen Ausfall bei Erreichen der 100 % Lebensdauer.

Es ist ersichtlich, dass die verbleibende Lebensdauer der Komponenten mittels des Prognosesystems zuverlässig abgeschätzt werden kann, d. h. die prognostizierten Werte nahe an den Realdaten (Groundtruth-Daten) liegen. Weiterhin ist ersichtlich, dass nach Erreichen der definierten 100 % Lebensdauer im Mittel über 50 % des Abnutzungsvorrates bis zum mechanischen Ausfall des Systems verbleiben. Die Wahl des Parameters α mit 1.5 erlaubt damit eine zuverlässige Prognose anhand der definierten 100 % Lebensdauer und stellt weiterhin sicher, dass das System auch bei Überschreiten der 100 % Lebensdauer weiterhin mechanisch intakt bleibt. Ein unvorhergesehener Ausfall aufgrund von Oberflächenzerrüttung ist damit nahezu ausgeschlossen.

Die Genauigkeit der Prognosefunktion ist auch in Abbildung 7-11 ersichtlich, in der für die sechs prognostizierten Fehler die Differenz zwischen Prognosesystem und tatsächlichem Schadensverlauf bei Überschreiten der 100%-Lebensdauermarke dargestellt ist.

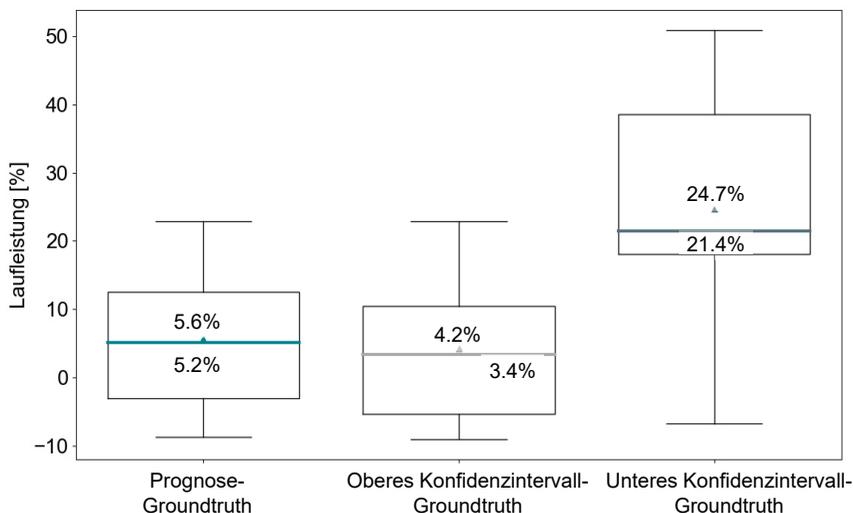


Abbildung 7-11 Abweichung der Prognose zu den Realdaten bei Überschreiten der definierten 100%-Lebensdauermarke.

Bei Verwendung des Prognosesystems prognostiziert dieses das Erreichen der 100 % Lebensdauer im Vergleich zu den Realdaten im Mittel 5.6 % später. Zwar bleibt in allen Fällen genügend Zeit zwischen der Prognose der 100 % Lebensdauer und dem mechanischen Ausfall des Systems, allerdings wird die Verwendung der oberen Schranke des Konfidenzintervalls empfohlen. Es ist ersichtlich, dass damit die Ungenauigkeit der Prognosefunktion reduziert und die Abweichung im Mittel auf 4.2 % reduziert werden kann.

Im Resultat kann gezeigt werden, dass sich die Prognosefunktion eignet, den Schädigungsverlauf der Oberflächenzerrüttung bei Definition einer 100%-Lebensdauermarke vorherzusagen. Gleichzeitig ist mit hoher Wahrscheinlichkeit ein sicherer Tausch der Komponenten vor dem mechanischen Ausfall des Systems gewährleistet. Dies lässt sich darauf zurückführen, dass in allen Fällen das Erreichen der definierten Lebensdauermarke zuverlässig und rechtzeitig prognostiziert werden konnte.

Damit kann das System grundsätzlich auf drei Arten eingesetzt werden: Im ersten Fall wird die Komponente als Defekt deklariert, sobald ein Fehler vom Quantifizierungssystem identifiziert wird. In diesem Fall spielt die Prognosefunktion keine Rolle, da bei erstmaligem Detektieren eines Fehlers Instandhaltungsmaßnahmen angestoßen werden. Dies entspricht der klassischen Definition der Lebensdauer des KGTs nach Norm. Im zweiten Fall wird lediglich das Modul zur Quantifizierung der Fehlergröße inklusive des Expertensystems verwendet, um den aktuellen Schädigungszustand abzuleiten. Die Instandhaltung wird angestoßen, sobald die zulässige Schädigung überschritten wird. Eine vorausschauende Instandhaltung ist damit zwar nicht direkt möglich, dafür aber die Nutzung der verschleißabhängigen 100 % Lebensdauer. Es muss anwendungsabhängig entschieden werden, ob eine Prognose zielführend ist. Ist dies der Fall, so kann weiterhin die Prognosefunktion verwendet werden, um eine vorausschauende Instandhaltung anzustoßen. Alle drei Anwendungsfälle konnten mittels der obigen Versuche validiert werden.

Zusammenfassend ist das System damit in der Lage, die Entwicklung von Pittings zuverlässig für die Zukunft zu prognostizieren. Dies ermöglicht eine genauere Planung der Instandhaltung und eine Beurteilung der Komponenten basierend auf deren tat-

sächlichem Verschleißzustand, ohne hierzu direkt eine manuelle Inspektion durchführen zu müssen. Die Gefahr eines ungeplanten Stillstandes von Maschinen aufgrund eines unvorhergesehenen Ausfalls eines KGT durch Oberflächenzerrüttung kann damit nahezu ausgeschlossen werden. Großer Vorteil ist die unmittelbare Sichtbarkeit von Schäden in den Bilddaten.

Im folgenden Abschnitt wird das Prognosesystem im Kontext bestehender Arbeiten diskutiert und ein Fazit gezogen.

7.5 Fazit zu der Verschleißprognose im Kontext bestehender Arbeiten

Die folgenden Diskussionen ergänzen die Diskussionen in Abschnitt 6.1.6.

Das vorgestellte bildbasierte Prognosesystem zur Zustandsüberwachung soll bestehende Systeme und Ansätze zur Überwachung und Prognose der Verschleißentwicklung am KGT um einen weiteren Ansatz erweitern. Vorangegangene Arbeiten am KGT, die explizit die Prognose von Schädigungen zur Zustandsabschätzung beschreiben, sind die Arbeiten von (Imiela 2006), (Schopp 2009) und (Henrich 2013).

Das von (Imiela 2006) vorgestellte System nutzt den Vorspannungsverlust des KGTs, der im Kontext von Verschleißvorratskurven zur Prognose der Restlebensdauer des KGTs herangezogen werden kann. Hierbei müssen die Verschleißvorratskurven individuell für KGT aus Verschleißuntersuchungen abgeleitet werden, was mit dem hier vorliegenden Ansatz nicht notwendig ist. Weiterhin bietet die in dieser Dissertation vorgeschlagene Definition zur Bestimmung der 100 % Lebensdauer in Kombination mit der optischen Aufnahme der Schädigungen eine von der Komponente und den Betriebsbedingungen weitestgehend unabhängige Möglichkeit zur Erfassung und Prognose von Verschleiß. Dies ist mit dem Ansatz von (Imiela 2006) so nicht ohne Weiteres möglich.

(Schopp 2009) verwendet die letzten drei aufgezeichneten Messpunkte zur bereichsweisen linearen Prognose des Verschleißmerkmals. Hiermit kann die Resteinsatzdauer des KGTs abgeschätzt werden. Der zulässige Verschleiß muss betreiberspezifisch definiert werden. Die Fähigkeit des Prognosesystems von (Schopp 2009) kann allerdings nicht beziffert werden, da (Schopp 2009) den Prognoseansatz zwar vorgeschlagen, aber nicht validiert hat. Anders als bei (Schopp 2009) werden in dem hier präsentierten Ansatz alle bisherigen Verschleißmesspunkte für die Prognose verwendet, was eine insgesamt genauere Verschleißprognose möglich macht. Weiterhin kann speziell die

Oberflächenzerrüttung exakt in ihrer zeitlichen Entwicklung beobachtet werden, was den Ansatz genauer als den von (Schopp 2009) vorgeschlagenen Prognoseansatz macht.

Der von (Hennrich 2013) präsentierte Ansatz ist anwenderfreundlich, da er zunächst steuerungsinterne Größen in Kombination mit der Formel zur nominellen Lebensdauer des KGTs nutzt. Nachteilig hinsichtlich der Prognosefunktion ist, dass exponentielle Diagnosefunktionen genutzt werden, die allerdings zunächst bauartspezifisch durch Lebensdauerversuche parametrisiert werden müssen. Darüber hinaus muss (Hennrich 2013) körperschallbasierte Messungen zum Abgleich durchführen, um das Prognose-system zu re-parametrisieren. Dadurch wird der Aufwand zur Prognose der Restein-satzdauer erhöht. Dies ist vergleichbar mit der Integration von Domänenexpert*innen, die im vorgestellten Ansatz zu festen Zeitpunkten die tatsächliche Schadensgröße aus den Bilddaten vermessen und diese in die Prognosefunktion geben. Die Genauigkeit wird hierdurch zwar erhöht, allerdings steigt der (manuelle) Aufwand zur Prognose. Damit zeichnet sich der in dieser Dissertation vorgestellte Prognoseansatz besonders dadurch aus, dass hierbei die direkten bildlichen Verschleißinformationen verwendet werden.

Als Ergebnis hebt sich der vorgestellte Ansatz besonders durch die folgenden Eigenschaften von bestehenden Arbeiten ab.

Aussagekräftiges Verschleißmerkmal

Unabhängig von der erreichten nominellen Lebensdauer der Komponente liefert das bildbasierte System ein klares Anzeichen für Verschleiß in Form von Oberflächenzerrüttung. Hierbei ist besonders zu betonen, dass es Anwender*innen damit möglich ist, den KGT auszutauschen, nachdem ein tatsächliches Merkmal für Verschleiß identifiziert wurde, und somit den Abnutzungsvorrat der Komponente maximal auszunutzen. Dies hat Vorteile hinsichtlich der Anlagenverfügbarkeit, der Instandhaltungskosten für neue KGT sowie der Ressourceneffizienz in der Produktion.

Frühzeitige Identifikation von Fehlstellen

In den Versuchen konnte gezeigt werden, dass ein Fehler frühzeitig auf der Spindel zu erkennen ist. Bisherige Arbeiten beschreiben, dass ein Defekt teilweise erst bei fortgeschrittenem Stadium zweifelsfrei nachgewiesen werden konnte. In den Versuchen wurde weiterhin gezeigt, dass der KGT auch bei einer recht hohen anliegenden Kraft, wie sie in den Versuchen gewählt wurde, nach Auftreten von Pittings weiterbetrieben

werden kann, ohne einen mechanischen Ausfall zu riskieren. Damit hat das vorgeschlagene System Vorteile hinsichtlich der Genauigkeit bei der Prognose der Oberflächenzerrüttung.

Bestätigung bestehender Arbeiten

In den Versuchen konnte gezeigt werden, dass ein dreiphasiges bzw. anders interpretiert ein exponentielles Wachstum des Gesamtverschleißes zu beobachten ist und dass einem initialen Fehler eine längere Zeit ohne auffälliges Verhalten folgt. Damit konnten die Beobachtungen der bestehenden Arbeiten überprüft und erstmals visuell bestätigt werden.

Kontinuierliche Erfassung und Prognose

Es konnte gezeigt werden, dass das Wachstum eines Fehlers auf der Spindel mittels einer Prognosefunktion prognostiziert werden kann. Dies war so in vorherigen Arbeiten nicht beschrieben worden. Durch die Definition eines Abbruchkriteriums kann somit eine zuverlässige Abschätzung des aktuell und zukünftig vorliegenden Verschleißes umgesetzt werden. Innovativ gegenüber bestehenden Arbeiten ist dabei besonders das sich kontinuierlich auf Basis der aktuellen Daten selbst anpassende sogenannte Konfidenzintervall. Die so implementierte Sicherheit für Anwender*innen geht über bestehende Arbeiten hinaus.

Integration von Verfahren des Maschinellen Lernens

Erstmals wurden in dieser Arbeit wirkungsvoll Verfahren des Maschinellen Lernens zur Klassifikation von Oberflächenzerrüttung in Bilddaten untersucht und angewandt. Die erreichten Genauigkeiten sind dabei sehr hoch und unterstreichen den Vorteil bei der Genauigkeit des Systems. Es wurde damit außerdem die Anwendung des Maschinellen Lernens im Kontext der Zustandsüberwachung und Zustandsprognose auf Basis von Bilddaten von Kugelgewindetrieben initiiert.

Datensatz

Erstmals wurde ein Datensatz über die Verschleißentwicklung am KGT veröffentlicht (Schlagenhauf 2021) und künftigen Wissenschaftler*innen zugänglich gemacht. Der Datensatz ist so ausgelegt, dass dieser direkt für das Training von Verfahren des Maschinellen Lernens verwendet werden kann. Eine Kombination des bildbasierten Ansatzes mit weiteren Sensorgrößen ist damit möglich und bietet sich an.

Dateneffizienz

Gegenüber bestehenden Arbeiten wurde erstmals der Aspekt der Dateneffizienz bei der Klassifikation in technischen Domänen untersucht und ein neuartiger Ansatz vorgestellt.

Im folgenden Abschnitt soll die Erweiterung des Standes der Forschung und Technik durch das vorgestellte System diskutiert und die Dissertation damit inhaltlich abgeschlossen werden.

7.6 Erweiterung der Zustandsüberwachung am KGT

Die Vorteilhaftigkeit des bildbasierten Ansatzes, speziell zur Quantifizierung von Oberflächenzerrüttung, wurde in den vorherigen Kapiteln dargestellt.

Abschließend soll eine Empfehlung ausgesprochen werden, welche Überwachungssysteme für unterschiedliche Betriebs- und Verschleißszenarien besonders geeignet sind, um ungeplante Maschinenstillstände zu verhindern. Diese ist in Tabelle 7-2 dargestellt.

Hierbei ist erkennbar, dass das bildbasierte Verfahren eine sinnvolle Erweiterung der bestehenden Systeme zur Überwachung des KGTs darstellt. Es gibt Betriebsszenarien, bspw. die Anwendung an einem Transport-KGT, bei denen sich der Einsatz des bildbasierten Systems besonders eignet. Gleichzeitig gibt es Betriebsszenarien, bspw. Kurzhubanwendungen, bei denen körperschallbasierte Systeme als zielführender angesehen werden können. Als Empfehlung für weitere Untersuchungen bietet sich die Kombination des bildbasierten Systems mit weiteren Sensorsystemen an, wobei als Ziel die sensorlose Überwachung der Komponente angestrebt werden sollte.

Tabelle 7-2 Betriebsszenarien, vorherrschende Ausfallmechanismen und empfohlene Prinzipien zur Überwachung.³⁷

Betriebsszenario	Vorherrschende Ausfallmechanismen	Monitoring-Prinzip
Kurzhub	Mangelschmierung, Passungsrost, lokale Ermüdung	
Transport-KGT*	Ermüdung	
Hohe Lasten	Ermüdung	
Hohe Geschwindigkeit	Vorspannungsverlust, Mangelschmierung	 
Lange Hübe	Ermüdung in Mutter/Kugeln	 
Hohe Verschmutzungsbelastung	Partikel in Mutter, Vorspannungsverlust	  
Präzisionsanwendung**	Vorspannungsverlust	
Normalbetrieb***	Verkleben, Ermüdung, Brüche	  

	Bildbasierte Überwachung		Motorstrommessung		Acoustic Emission	* Kugelgewindetrieb ohne spezielle Genauigkeitsanforderung
	Vibrations-sensorik		Guard-Plus Steinmeyer		Temperatur-sensorik	** Vorspannungsverlust unerwünscht
						*** Keine spezielle Anforderung, aber Maschinenstillstand ist kritisch

Als Ergebnis kann festgehalten werden, dass der präsentierte Ansatz den bisherigen Stand der Forschung und Technik um ein System erweitert, das speziell die Oberflächenzerrüttung als lebensdauerbegrenzendes Verschleißmerkmal in den Fokus nimmt und sich dabei gegenüber bestehenden Ansätzen abhebt.

³⁷ Die dargestellten Prinzipien zur Überwachung stellen lediglich Vorschläge des Autors dar und erheben keinen Anspruch auf Vollständigkeit.

8 Bewertung

Das Kapitel der Bewertung gliedert sich in eine Zusammenfassung der Arbeiten und einen Ausblick auf vielversprechende Forschungsrichtungen.

8.1 Zusammenfassung

Die vorliegende Dissertation hatte zum Ziel, die Entwicklung der Oberflächenzerrüttung an der Kugelgewindetriebspindel zu analysieren. Darauf basierend war es das Ziel, eine Erweiterung des bestehenden Standes der Forschung und Technik zur Zustandsüberwachung und -prognose am KGT auf Basis von Bilddaten zu untersuchen.

Das betrachtete Themenfeld ist hierbei von hoher wirtschaftlicher und wissenschaftlicher Relevanz. Die Komponente KGT ist ein häufig in Maschinen eingesetztes Maschinenelement und maßgeblich für ungeplante Maschinenstillstände in Werkzeugmaschinen verantwortlich. Aufgrund der Tatsache, dass ungeplante Maschinenstillstände hohe Kosten nach sich ziehen, werden KGT häufig vor Erreichen des nutzbaren Abnutzungsvorrates ausgetauscht. Dies wirkt sich wiederum negativ auf die Verfügbarkeit von Maschinen aus, zieht erhöhte Instandhaltungskosten nach sich und ist nicht nachhaltig. Eine zuverlässige Abschätzung des tatsächlich vorliegenden Komponentenzustandes hat daher hohe Relevanz.

Wissenschaftliches Potenzial ergibt sich einerseits aus der Tatsache, dass der Verschleißfortschritt am KGT im bestehenden Stand der Forschung bisher nicht auf Bilddaten abgebildet wurde. Von Relevanz ist außerdem die Integration von Verfahren des Maschinellen Lernens in technische Prozesse, um eine Grundlage für die Entwicklung von weiteren intelligenten Systemen zur Überwachung von Maschinenelementen zu schaffen. Die Arbeit leistet damit sowohl einen Beitrag zum klassischen Maschinenbau als auch zur Umsetzung von intelligenten und autonom agierenden Produktionsmaschinen.

Eine Herausforderung im Kontext der Zustandsüberwachung des KGTs besteht darin, dass die Komponentenlebensdauer abhängig vom Prozess stark schwankt. Dies macht es unmöglich, eine für alle Anwendungen universelle, fixe und sinnvolle Lebensdauer sowie lebensdauerbestimmende Verschleißmechanismen zu bestimmen. Da ohne Überwachung nicht zweifelsfrei klar ist, wann ein KGT ausgetauscht werden muss, kann es zu der oben angesprochenen (zu frühen) Instandsetzung des KGTs kommen.

Im bestehenden Stand der Forschung finden sich zahlreiche Ansätze, die sich zum Ziel gesetzt haben, den aktuell am KGT vorliegenden Verschleiß festzustellen. Die Hauptherausforderung bei den bestehenden Ansätzen ist, dass diese Verschleißmerkmale lediglich indirekt messen und damit den aktuell vorliegenden Verschleißzustand nicht zweifelsfrei bestimmen können. Dazu zählt auch, dass das Vorliegen von Verschleiß und speziell von Oberflächenzerrüttung erst bei fortgeschrittenem Verschleiß zweifelsfrei identifiziert werden kann, was den Handlungszeitraum für Anwender einschränkt. Als dritte Herausforderung bestehender Ansätze ergibt sich, dass das Sensorsignal eines Systems zur Zustandsüberwachung oftmals stark von der Komponente, dem Prozess und den Betriebsbedingungen beeinflusst wird. Die Signale sind darüber hinaus häufig für Domänenexpert*innen nur schwer interpretierbar. Insgesamt führt dies zu einer teilweise recht aufwändigen und dennoch mit Unsicherheit behafteten Identifikation von Verschleiß, wobei die Oberflächenzerrüttung im aktuellen Stand der Forschung bisher nicht explizit untersucht wurde.

Es bedarf also eines Systems, das Verschleiß im Allgemeinen und Oberflächenzerrüttung im Speziellen zuverlässig identifizieren und quantifizieren kann, um damit die Sicherheit für Unternehmen hinsichtlich der Instandhaltungsplanung zu erhöhen und die Verfügbarkeit von Maschinen zu maximieren. Darüber hinaus muss der Schädigungszustand für die Domänenexpert*innen klar aus dem Sensorsignal ersichtlich sein, um eine Verschleißeinschätzung treffen bzw. die Verschleißeinschätzung des Systems überprüfen zu können. Die Prognose des Verschleißzustandes auf Basis des aktuell vorliegenden Fehlerbildes, wie sie im Stand der Technik nur eingeschränkt untersucht wurde, erhöht weiterhin die Sicherheit für Unternehmen.

Der hier verfolgte Ansatz adressiert die obigen Punkte, indem mittels eines Kamerasystems, das an der Mutter des KGTs angebracht ist, Bilddaten der Oberfläche der Kugelgewindetriebsspindel aufgenommen werden. In den aufgenommenen Bilddaten sind Defekte direkt identifizierbar, sobald diese auf der Spindel auftreten. Die Größe der Defekte und der Gesamtzustand der Spindel lassen sich daraus ableiten und das Signal ist für die Domänenexpert*innen einfach interpretierbar. Als Folge können die Bilddaten mit Labeln versehen und zum Training von Verfahren des Maschinellen Lernens verwendet werden.

Das System wird realisiert, indem zunächst ein Kamerasystem konzipiert wird, das nahe der KGT-Mutter angebracht ist. Mittels dieses Kamerasystems werden insgesamt zwölf KGT-Teilsysteme während Lebensdauerversuchen aufgenommen und darüber

Bilddaten generiert, welche die gesamte Historie der Oberflächenzerrüttung auf den Spindeln zeigen. Weitere Verschleißmerkmale sowie miteinander wechselwirkende Verschleißmechanismen können in den Bilddaten ebenfalls erkannt werden. Die Bilddaten werden in einem nächsten Schritt ausführlich hinsichtlich der Schädigungsentwicklung analysiert. Weiterhin werden die Bilddaten der Forschungsgemeinschaft zur weiteren Untersuchung sowie zur Entwicklung von Condition-Monitoring-Systemen zur Verfügung gestellt.

Der Analyse der Bilddaten folgte die Untersuchung der automatischen Klassifikation von Pittings in Bilddaten. Dazu wurden zunächst charakteristische Merkmale der Schädigungen auf Basis von Domänenwissen definiert und mittels Merkmalsextraktionsalgorithmen aus den Bilddaten extrahiert. Diese Merkmale wurden zum Training von Klassifikationsmodellen verwendet, um die Fähigkeit der Klassifikation mittels dieser Merkmale zu überprüfen. In einem weiteren Schritt wurden Deep-Learning-basierte Verfahren zur automatischen Merkmalsextraktion und Klassifikation untersucht. Dabei hat sich gezeigt, dass Pittings in den Bilddaten mittels des konzipierten CNN zuverlässig identifiziert werden können.

Auf Basis dieser Betrachtungen wurde weiterhin der in technischen Domänen kritische Aspekt der geringen vorliegenden Datenmengen berücksichtigt, indem ein Ansatz zur dateneffizienten Klassifikation von Bilddaten untersucht und dessen Wirksamkeit gegenüber dem bestehenden Stand der Technik aufgezeigt wurde.

Auf Basis der Klassifikation der Bilddaten wurde in einem nächsten Schritt ein System untersucht, das es erlaubt, den zukünftigen Schädigungszustand der Spindel auf Basis der aufgenommenen Bilddaten zu prognostizieren. Hierzu wurde ein regressionsbasiertes Prognosesystem untersucht. Zur Optimierung der Prognosegüte wurde das System um ein Expertensystem ergänzt, das die Prognosen mit auf Domänenwissen basierenden Charakteristika abgleicht. Die Prognosefähigkeit des Ansatzes sowie die Wirksamkeit des Tandems aus Prognosesystem und Expertensystem konnten dargelegt werden. Durch die Definition einer Lebensdauermarke, die an eine bestehende Norm angelehnt ist, wird eine Möglichkeit zur Einschätzung des aktuell vorliegenden Verschleißzustandes geschaffen. Ein sich automatisch anpassendes Konfidenzintervall verleiht dem System zusätzliche Sicherheit. Die Untersuchungen haben gezeigt, dass es mittels des Prognosesystems möglich ist, den Verschleißzustand für künftige Zeitpunkte

zu prognostizieren und Instandhaltungsmaßnahmen vor dem mechanischen Ausfall der Komponenten einzuleiten.

Zusammenfassend kann festgehalten werden, dass alle im Rahmen der Arbeit gesetzten Ziele erreicht wurden.

Mit den im Rahmen dieser Arbeit untersuchten Bilddaten sowie den vorgestellten Modellen des Maschinellen Lernens wurden Bausteine zur weiteren Untersuchung und Entwicklung von industriellen Condition-Monitoring-Systemen sowie zur Untersuchung von performanten Machine-Learning-Verfahren zur Identifikation von Verschleiß auf metallischen Oberflächen geschaffen.

8.2 Ausblick

Zur Steuerung der Kamerafahrt wird das aktuelle System manuell mit der Maschinensteuerung verbunden. Zur Realisierung einer direkt industriell einsetzbaren Plug-and-Play-Lösung muss dieser Prozess automatisiert werden. Weiteres direktes Optimierungspotenzial liegt in der Reduzierung des notwendigen Bauraums des Systems. Dies ist relevant, da durch das Kamerasystem der Fahrweg einer Achse eingeschränkt wird. Hierbei ist anzunehmen, dass sich ein Trade-off zwischen den Kosten für das Sensorsystem und dem Mehrwert durch die Abmessungsoptimierung einstellt. Gleiches gilt für eine optimierte Kameratechnik oder das Beleuchtungssystem.

Ein optimales System zur Zustandsüberwachung am KGT sollte in der Lage sein, alle Verschleißerscheinungsformen am KGT zuverlässig in ihrem Typ zu identifizieren und zu verorten. Dabei sollte das System in der Lage sein, ohne zusätzliche Sensorik zu arbeiten. Das bildbasierte Verfahren zur Zustandsüberwachung am KGT hat die Einschränkung, dass es zum einen lediglich die Spindeloberfläche beobachten kann und zum anderen sensorbasiert ist. Der nächste Schritt sollte die Kombination von bestehenden Verfahren und Signalen, bspw. des Motorstroms mit dem bildbasierten Verfahren sein. Die Bilddaten können hierbei als Meta-Informationen zum Labeling der weiteren Signale verwendet werden. Eine Herangehensweise könnte darin bestehen, ein Modell zu trainieren, das zwar mittels aller zur Verfügung stehenden Daten trainiert wird, allerdings zum Ziel hat, Fehler lediglich auf Basis des Motorstroms zu identifizieren. Dabei werden die in den Bild- und Schalldaten vorhandenen Informationen genutzt, um dem Signal des Motorstromes wertvolle Informationen hinzuzufügen bzw. in den Motorstromdaten Muster zu erkennen, die ohne diese zusätzlichen Daten nicht erkannt

werden könnten. Im Ergebnis hat das System zwar aus Sensordaten gelernt, der Betrieb erfolgt jedoch sensorlos. Ebenfalls von Interesse ist die Untersuchung eines Gesamtansatzes, der sowohl die Oberflächenzerrüttung (langsamer Spätausfall) als auch den langsamen Vorspannungsverlust in einem System vereint.

Eine offene Fragestellung, die sich im Kontext der Prognosefunktion ergibt, ist die Integration von sich in der Zukunft ändernden Lastkollektiven in die Prognose. Da die Prognosefunktion mit Vergangenheitswerten trainiert wurde, bildet sie diese ab. Zwar passt sie sich iterativ an hinzukommende Daten an, sie kann jedoch niemals sich in der Zukunft ändernde Lastkollektive berücksichtigen. Dieser Aspekt sollte zusammen mit der Kombination von Sensorsystemen untersucht werden, um damit zu einem Gesamtansatz zu gelangen.

Zur Umsetzung von autonomen Instandhaltungssystemen müssen die Modelle über die oben genannten Kriterien hinaus in der Lage sein, maschinenübergreifend ohne erneute Parametrisierung bzw. ohne Nachtraining Fehler zuverlässig zu identifizieren und neue Fehlerbilder selbstständig zu integrieren. Dies ist mit bestehenden Systemen bisher so nicht möglich. Diese Eigenschaften sollten im Supervised-Learning-Fall entweder sehr dateneffizient erreicht werden können oder das System sollte in der Lage sein, ohne gelabelte Daten, nicht überwacht (Unsupervised) und dateneffizient zu arbeiten. Daraus ergeben sich die drei folgenden relevanten Forschungsfelder: 1. generalisierende und resiliente Modelle, 2. dateneffiziente Modelle und 3. kontinuierlich lernende Systeme.

Publikationsliste des Autors

Schlagenhauf & Hillenbrand et al. 2019

Schlagenhauf, T.; Hillenbrand, J.; Klee, B. & Fleischer, J. (2019), „Integration von Machine Vision in Kugelgewindespindeln“, *wt Werkstatttechnik online*, 109 (7/8), S. 605–610.

Schlagenhauf & Feuring et al. 2019

Schlagenhauf, T.; Feuring, C.-P.; Hillenbrand, J. & Fleischer, J. (2019), „Camera Based Ball Screw Spindle Defect Classification System. System zur kamerabasierten Defekterkennung auf Kugelgewindetribspindeln“. *Production at the Leading Edge of Technology. Proceedings of the 9th Congress of the German Academic Association for Production Technology (WGP), September 30th - October 2nd, Hamburg 2019*. Hamburg, Hrsg. J. P. Wulfsberg, Springer Berlin / Heidelberg, Berlin, Heidelberg, S. 503–512. ISBN: 9783662604175.

Schlagenhauf, Heinzler & Fleischer 2020

Schlagenhauf, T.; Heinzler, M. & Fleischer, J. (2020), „Extraction of surface image features for wear detection on ball screw drive spindles“. *Forum Bildverarbeitung 2020*. Karlsruhe, Hrsg. T. Längle & M. Heizmann, KIT Scientific Publishing, Karlsruhe, S. 305–314. ISBN: 9783731510536.

Schlagenhauf, Ruppelt & Fleischer 2020

Schlagenhauf, T.; Ruppelt, P. & Fleischer, J. (2020), „Frühzeitige Detektion von Oberflächenzerrüttungen“, *wt Werkstattstechnik online*, 110 (07-08), S. 501–506.

Schlagenhauf, Ammann & Fleischer 2021

Schlagenhauf, T.; Ammann, N. & Fleischer, J. (2021), „Online Learning für die präventive Verschleißdetektion. Online Learning for preventive wear detection - Online Retraining of Deep Learning models for unknown wear patterns“, *wt Werkstattstechnik online*, 111 (07-08), S. 475–480.

Schlagenhauf & Landwehr 2021

Schlagenhauf, T. & Landwehr, M. (2021), „Industrial machine tool component surface defect dataset“, *Data in Brief*, 39 (1), S. 107643.

Schlagenhauf & Burghardt 2021

Schlagenhauf, T. & Burghardt, N. (2021), „Intelligent vision based wear forecasting

on surfaces of machine tool elements“, *Springer Nature Applied Sciences*, 3 (858), S. 1–13.

Schlagenhauf, Brander & Fleischer 2021

Schlagenhauf, T.; Brander, T. & Fleischer, J. (2021), „A stitching algorithm for automated surface inspection of rotationally symmetric components“, *CIRP Journal of Manufacturing Science and Technology*, 35 (1), S. 169–177.

Albers & Fleischer et al. 2021

Albers, A.; Fleischer, J.; Gönzheimer, P.; Mühlbeier, E.; Schlagenhauf, T.; Wäschle, M.; Anderl, R.; Giese, T.; Wang, Y.; Aurich, J.; Glatt, M. & Litsche, S. & Steglich, S. (2021), *5G in der Industrie. Wege in die Technologieführerschaft in Produktentwicklung und Produktion*, Acatech, München. ISBN: 2702-7627.10.48669/aca_2021-2.

Schlagenhauf, Netzer & Hillinger 2022

Schlagenhauf, T.; Netzer, M. & Hillinger, J. (2022), „ Text Detection on Technical Drawings for the Digitization of Brown-field Processes “, *16TH CIRP CONFERENCE ON INTELLIGENT COMPUTATION IN MANUFACTURING ENGINEERING 2022. Italy, accepted for Publication.*

Literaturverzeichnis

A_Burkhardt 2021

A_Burkhardt, N. (2021), *Algorithm for Quantifying the Damage Progression on Ball Screw Drives*. Masterarbeit, Albstadt-Sigmaringen University, Albstadt, Faculty of Industrial Engineering.

A_Hansjosten 2020

A_Hansjosten, M. (2020), *Entwicklung eines intelligenten Kamerasystems zum Condition Monitoring von Kugelgewindetriebspindeln*. Masterarbeit, Karlsruher Institut für Technologie, Karlsruhe, wbk Institut für Produktionstechnik.

A_Heinzler 2020

A_Heinzler, M. (2020), *Examination of image filter methods for feature extraction and the application of machine learning classifiers in the context of ball screw drive surface images*. Masterarbeit, Albstadt-Sigmaringen University, Faculty of Industrial Engineering.

A_Hofmann 2021

A_Hofmann, D. (2021), *Analyse des bildlich detektierbaren Verschleißfortschrittes auf Kugelgewindetrieben*. Masterarbeit, Karlsruher Institut für Technologie, Karlsruhe, wbk Institut für Produktionstechnik.

A_Krasnikow 2021

A_Krasnikow, O. (2021), *Verschleißprognose am Kugelgewindetrieb*. Masterarbeit, Karlsruher Institut für Technologie, Karlsruhe, wbk Institut für Produktionstechnik.

A_Schlund 2019

A_Schlund, J. (2019), *Konzeption und Umsetzung eines integrierten Kamerasystems zur Überwachung des Verschleißzustands von Kugelgewindetrieben*. Bachelorarbeit, Hochschule Karlsruhe Technik und Wirtschaft, Fakultät für Maschinenbau und Mechatronik.

A_Sun 2019

A_Sun, C. (2019), *GAN based Ball Screw Drive picture database enlargement for Failure classification*. Masterarbeit, Karlsruher Institut für Technologie, Karlsruhe, wbk Institut für Produktionstechnik.

A_Yildirim 2021

A_Yildirim, F. (2021), *Klassifikation unter Verwendung von Deep Learning extrahierten Ähnlichkeitsmaßen*. Bachelorarbeit, Karlsruher Institut für Technologie, Karlsruhe, wbk Institut für Produktionstechnik.

Abdusamad, Gao & Muljadi 2013

Abdusamad, K. B.; Gao, D. W. & Muljadi, E. (2013), „A condition monitoring system for wind turbine generator temperature by applying multiple linear regression model“. *North American Power Symposium (NAPS), 2013*. Manhattan, KS, USA, Hrsg. Institute of Electrical and Electronics Engineers (IEEE), IEEE, Piscataway, NJ, S. 1–8. ISBN: 9781479912537.

Achard 1980

Achard, J. F. (1980), „Wear Theory and Mechanisms“ in *Wear Control Handbook*, Hrsg. M. Peterson & W. Winer, ASME, New York, S. 35–80.

Aggarwal 2019

Aggarwal, C. C. (2019), *Neural Networks and Deep Learning // Neural networks and deep learning. A textbook*, Springer; Springer International Publishing, Heidelberg. ISBN: 9783319944630.

Al-Arbi 2012

Al-Arbi, S. (2012), *Condition Monitoring of Gear Systems using Vibration Analysis*. Dissertation, University of Huddersfield, Huddersfield.

Albers, Dickerhof & Burger 2009

Albers, A.; Dickerhof, M. & Burger, W. (2009), „Condition-Monitoring Based on Structure-Borne Ultrasound Analysis“. *Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference - 2008*. Brooklyn, New York, USA, ASME, New York, NY, S. 465–470. ISBN: 9780791843260.

Alegre & Alaíz-Rodríguez et al. 2009

Alegre, E.; Alaíz-Rodríguez, R.; Barreiro, J.; Ruiz, J.; Alegre, E.; Alaíz-Rodríguez, R.; Barreiro, J. & Ruiz, J. (2009), „Use of contour signatures and classification methods to optimize the tool life in metal machining“, *Estonian Journal of Engineering*, 15 (1), S. 3.

Alqatawneh & Rabeyee et al. 2021

Alqatawneh, I.; Rabeyee, K.; Zhang, C.; Feng, G.; Gu, F.; Ball, A.; Gu, F. & Ball, A. D. (2021), „A Modified Activation Function for Deep Convolutional Neural Network and Its Application to Condition Monitoring“ in *Proceedings of IncoME-V & CEPE Net-2020*, Hrsg. D. Zhen, D. Wang, T. Wang, H. Wang, B. Huang, J. K. Sinha & A. D. Ball, Springer International Publishing, Heidelberg, S. 895–909.

Balzategui & Eciolaza et al. 2019

Balzategui, J.; Eciolaza, L.; Arana-Arexolaleiba, N.; Altube, J.; Aguerre, J.-P.; Legarda-Ereno, I. & Apraiz, A. (2019), „Semi-automatic quality inspection of solar cell based on Convolutional Neural Networks“. *Proceedings 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. Zaragoza, Spain, Hrsg. A. Jaunasse & M. David, IEEE Press; INSEE; IEEE, Limoges, S. 529–535. ISBN: 9781728103037.

Bay & Tinne et al. 2006

Bay, H.; Tinne, T.; Van Gool, L. & Tuytelaars, T. (2006), „SURF: Speeded Up Robust Features“. *Computer vision - ECCV 2006*. Graz, Austria, 2006, Hrsg. A. Leonardis, H. Bischof & A. Pinz, Springer, Berlin, S. 404–4017. ISBN: 3540338365.

Bender & Göhlich 2020

Bender, B. & Göhlich, D. (2020), *Dubbel Taschenbuch für den Maschinenbau 2. Taschenbuch für den Maschinenbau Anwendungen*, Springer Vieweg, Berlin. ISBN: 9783662597125.

Bengio, LeCun & Hinton 2021

Bengio, Y.; LeCun, Y. & Hinton, G. (2021), „Deep learning for AI“, *Communications of the ACM*, 64 (7), S. 58–65.

Benker, Kleinwort & Zäh 2019

Benker, M.; Kleinwort, R. & Zäh, M. F. (2019), „Estimating remaining useful life of machine tool ball screws via probabilistic classification“. *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*. San Francisco, CA, USA, Hrsg. IEEE, IEEE, Piscataway, NJ, S. 1–7. ISBN: 9781538683569.

Bergmann & Fauser et al. 2019

Bergmann, P.; Fauser, M.; Sattlegger, D. & Steger, C. (2019), „MVTec AD — A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection“. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long

Beach, CA, USA, Hrsg. A. Gupta, D. Hoiem, G. Hua, Z. Tu & L. Davis, IEEE, Piscataway, NJ, S. 9584–9592. ISBN: 9781728132938.

Beyerer & Puente León et al. 2016

Beyerer, J.; Puente León, F.; Frese, C. & Meyer, J. (2016), *Automatische Sichtprüfung. Grundlagen, Methoden und Praxis der Bildgewinnung und Bildauswertung*, Springer Vieweg; Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN: 9783662477861.

Beyerer, Richter & Nagel 2018

Beyerer, J.; Richter, M. & Nagel, M. (2018), *Pattern recognition. Introduction, features, classifiers and principles*, De Gruyter, Berlin, Boston. ISBN: 9783110537932.

Bhat & Patil 2014

Bhat, M. & Patil, T. (2014), „Adaptive clip limit for contrast limited adaptive histogram equalization (CLAHE) of medical images using least mean square algorithm“. *International Conference on Advanced Communication, Control and Computing Technologies (ICACCCT)*. Ramanathapuram, India, Hrsg. Institute of Electrical and Electronics Engineers. Madras Section, Syed Ammal Engineering College & Institute of Electrical and Electronics Engineers, IEEE, Piscataway, NJ, S. 1259–1263. ISBN: 9781479939145.

Bousdekis & Apostolou et al. 2020

Bousdekis, A.; Apostolou, D.; Mentzas, G.; Apostolou, D. & Mentzas, G. (2020), „Predictive Maintenance in the 4th Industrial Revolution: Benefits, Business Opportunities, and Managerial Implications“, *IEEE Engineering Management Review*, 48 (1), S. 57–62.

Bozic, Tabernik & Skocaj 2021

Bozic, J.; Tabernik, D. & Skocaj, D. (2021), „End-to-end training of a two-stage neural network for defect detection“. *Proceedings of ICPR 2020. 25th International Conference on Pattern Recognition*. Milan, Italy, Hrsg. R. Vezzani, IEEE, Piscataway, NJ, S. 5619–5626. ISBN: 9781728188089.

Bradley & Ebrahimi et al. 2014

Bradley, W.; Ebrahimi, M.; Ehsani, M.; Bradley, W. J. & Ebrahimi, M. K. (2014), „A General Approach for Current-Based Condition Monitoring of Induction Motors“,

Journal of Dynamic Systems Measurement and Control-transactions of The Asme, 136 (4), S. 41024.

Bradski, Kaehler & Bradski 2017

Bradski, G.; Kaehler, A. & Bradski, G. R. (2017), *The OpenCV Library // Learning OpenCV 3. Computer vision in C++ with the OpenCV library*, O'Reilly Media, Beijing, Boston, Farnham, Sebastopol, Tokyo. ISBN: 1491937998.

Brecher, Klein & Paepenmüller 2006

Brecher, C.; Klein, W. & Paepenmüller, F. (2006), „Ansätze zur Schadensdiagnose an Linearführungssystemen“, *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 101 (7-8), S. 408–411.

Bredies & Lorenz 2018

Bredies, K. & Lorenz, D. (2018), *Mathematical image processing*, Birkhäuser; Springer Nature, Basel. ISBN: 3030014576.

Broos 2012

Broos, A. (2012), *Simulationsgestützte Ermittlung der Komponentenbelastung für die Lebensdauerprognose an Werkzeugmaschinen*. Dissertation, Karlsruher Institut für Technologie, Shaker, Düren. ISBN: 9783844013436.

Caggiano & Caggiano 2018

Caggiano, A. & Caggiano, A. (2018), „Tool Wear Prediction in Ti-6Al-4V Machining through Multiple Sensor Monitoring and PCA Features Pattern Recognition“, *Sensors*, 18 (3), S. 823.

Calonder & Lepetit et al. 2010

Calonder, M.; Lepetit, V.; Strecha, C. & Fua, P. (2010), „BRIEF: Binary Robust Independent Elementary Features“. *Computer vision - ECCV 2010*. Heraklion, Crete, Greece, Hrsg. K. Daniilidis, P. Maragos & N. Paragios, Springer, Berlin, S. 778–792. ISBN: 9783642155604.

Cao, Zhang & Tang 2018

Cao, P.; Zhang, S. & Tang, J. (2018), „Preprocessing-Free Gear Fault Diagnosis Using Small Datasets With Deep Convolutional Neural Network-Based Transfer Learning“, *IEEE Access*, 1 (6), S. 26241–26253.

Carnero 2006

Carnero, M. (2006), „An evaluation system of the setting up of predictive maintenance programmes“, *Reliability Engineering & System Safety*, 91 (8), S. 945–963.

Cha & Choi et al. 2018

Cha, Y.-J.; Choi, W.; Suh, G.; Mahmoudkhani, S. & Büyüköztürk, O. (2018), „Autonomous Structural Visual Inspection Using Region-Based Deep Learning for Detecting Multiple Damage Types“, *Computer-Aided Civil and Infrastructure Engineering*, 33 (9), S. 731–747.

Cha, Choi & Büyüköztürk 2017

Cha, Y.-J.; Choi, W. & Büyüköztürk, O. (2017), „Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks“, *Computer-Aided Civil and Infrastructure Engineering*, 32 (5), S. 361–378.

Chapman & Clinton et al. 2000

Chapman, P.; Clinton, J.; Kerber, R.; Khabaza, T.; Reinartz, T.; Shearer, C. & Wirth, R. (2000), *CRISP-DM 1.0: Step-by-step data mining guide*.

<https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPPWP-0800.pdf> [22.07.2022].

Chen, Tsai & Hsu 2016

Chen, Y.-J.; Tsai, J.-C. & Hsu, Y.-C. (2016), „A real-time surface inspection system for precision steel balls based on machine vision“, *Measurement Science and Technology*, 27 (7), S. 74010.

Cheng, Wang & Zhang 2019

Cheng, Y.; Wang, Z. & Zhang, W. (2019), „A Novel Condition-Monitoring Method for Axle-Box Bearings of High-Speed Trains Using Temperature Sensor Signals“, *IEEE Sensors Journal*, 19 (1), S. 205–213.

Chollet 2018

Chollet, F. (2018), *Deep learning with Python*, Manning, Shelter Island, NY. ISBN: 9781617294433.

Cipollini & Oneto et al. 2019

Cipollini, F.; Oneto, L.; Coraddu, A. & Savio, S. (2019), „Unsupervised Deep Learning for Induction Motor Bearings Monitoring“, *Data-Enabled Discovery and Applications*, 3 (1), S. 1–13.

Czichos & Habig 2015

Czichos, H. & Habig, K.-H. (2015), *Tribologie-Handbuch. Tribometrie, Tribomaterialien, Tribotechnik*, Springer Vieweg, Wiesbaden. ISBN: 9783834822369.

Czichos & Habig 2020

Czichos, H. & Habig, K.-H. (2020), *Tribologie-Handbuch. Tribometrie, Tribomaterialien, Tribotechnik*, Springer Fachmedien Wiesbaden; Springer Vieweg, Wiesbaden. ISBN: 9783658294830.

Czimmermann & Ciuti et al. 2020

Czimmermann, T.; Ciuti, G.; Milazzo, M.; Chiurazzi, M.; Roccella, S.; Oddo, C. M. & Dario, P. (2020), „Visual-Based Defect Detection and Classification Approaches for Industrial Applications-A SURVEY“, *Sensors*, 20 (5), S. 1459.

Dalvand, Kalantar & Safizadeh 2016

Dalvand, F.; Kalantar, A. & Safizadeh, M. S. (2016), „A Novel Bearing Condition Monitoring Method in Induction Motors Based on Instantaneous Frequency of Motor Voltage“, *IEEE Transactions on Industrial Electronics*, 63 (1), S. 364–376.

Dambacher 2005

Dambacher, U. (2005), *Kugelgewindetriebe mit hohem Druckwinkel*. Dissertation, Universität Karlsruhe, Universität Karlsruhe, Karlsruhe.

Davies 2012

Davies, A. (2012), *Handbook of Condition Monitoring. Techniques and Methodology*, Springer Netherlands, Dordrecht. ISBN: 9789401149242.

deBotton, Ben-Ari & Sher 2000

deBotton, G.; Ben-Ari, J. & Sher, E. (2000), „Vibration monitoring as a predictive maintenance tool for reciprocating engines“, *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 214 (8), S. 895–903.

Dehli 2019

Dehli, M. (2019), *Energieeffizienz in Industrie, Dienstleistung und Gewerbe. Energietechnische Optimierungskonzepte für Unternehmen*, Springer Vieweg, Wiesbaden Germany. ISBN: 9783658232030.

Deshpande & Minai et al. 2020

Deshpande, A. M.; Minai, A.; Kumar, M.; Minai, A. A. & Kumar, M. (2020), „One-

Shot Recognition of Manufacturing Defects in Steel Surfaces“, *Procedia Manufacturing*, 48 (1), S. 1064–1071.

Deters 2004

Deters, L. (2004), „Grundsätzliches zu Reibung und Verschleiß in der technischen Anwendung“. *Magdeburger Wissenschaftsjournal*, *Magdeburger Wissenschaftsjournal*, 1 (1), S. 16–26.

Deutsches Institut für Normung 1979

Deutsches Institut für Normung (1979), *Verschleiß-Messgrößen*, DIN 50321 (1979-12-00), Beuth Verlag GmbH, Berlin.

Deutsches Institut für Normung 2010a

Deutsches Institut für Normung (2010), *Wälzlager - Prüfbedingungen zur versuchstechnischen Verifikation der dynamischen Tragzahl von Profilschienenführungen mit kompakten Kugel- oder Rollenumlaufwagen*, DIN 631, Beuth Verlag GmbH, Berlin.

Deutsches Institut für Normung 2010b

Deutsches Institut für Normung (2010), *Zuverlässigkeitsmanagement - Teil 3-11: Anwendungsleitfaden - Auf die Funktionsfähigkeit bezogene Instandhaltung (IEC 60300-3-11:2009)*, DIN EN 60300-3-11, Beuth Verlag GmbH, Berlin.

Deutsches Institut für Normung 2010c

Deutsches Institut für Normung (2010-10), *Wälzlager - Dynamische Tragzahlen und nominelle Lebensdauer*, DIN ISO 281 (2010-10-00), Beuth Verlag GmbH, Berlin.

Deutsches Institut für Normung 2011a

Deutsches Institut für Normung (2011), *Kugelgewindetriebe - Teil 1: Begriffe und Bezeichnungssystem*, DIN ISO 3408-1, Beuth Verlag, Berlin.

Deutsches Institut für Normung 2011b

Deutsches Institut für Normung (2011), *Kugelgewindetriebe - Teil 5: Statische und dynamische axiale Tragzahl und Lebensdauer*, DIN ISO 3408-5 (2011-04-00), Beuth Verlag GmbH, Berlin.

Deutsches Institut für Normung 2018a

Deutsches Institut für Normung (2018), *Instandhaltung - Begriffe der Instandhaltung; Dreisprachige Fassung EN 13306:2017*, DIN EN 13306 (2018-02-00), Beuth Verlag GmbH, Berlin.

Deutsches Institut für Normung 2018b

Deutsches Institut für Normung (2018), *Wälzlager - Linear-Wälzlager - Teil 1: Dynamische Tragzahlen und nominelle Lebensdauer*, DIN ISO 14728-1 (2018-10-00), Beuth Verlag GmbH, Berlin.

Deutsches Institut für Normung 2018c

Deutsches Institut für Normung (2018), *Zustandsüberwachung und -diagnostik von Maschinen - Allgemeine Anleitungen (ISO 17359:2018)*, DIN ISO 17359 (2018-05-00), Beuth Verlag GmbH, Berlin.

Deutsches Institut für Normung 2019

Deutsches Institut für Normung (2019), *Grundlagen der Instandhaltung*, DIN 31051 (2019-06-00), Beuth Verlag GmbH, Berlin.

Di He, Xu & Zhou 2019

Di He, Xu, K. & Zhou, P. (2019), „Defect detection of hot rolled steels with a new object detection framework called classification priority network“, *Computers & Industrial Engineering*, 128 (1), S. 290–297.

Djenouri & Srivastava et al. 2021

Djenouri, Y.; Srivastava, G.; Lin, J.; Djenouri, Y. & Lin, J. C.-W. (2021), „Fast and Accurate Convolution Neural Network for Detecting Manufacturing Data“, *IEEE Transactions on Industrial Informatics*, 17 (4), S. 2947–2955.

Dong, Taylor & Cootes 2018

Dong, X.; Taylor, C. J. & Cootes, T. F. (2018), „Small Defect Detection Using Convolutional Neural Network Features and Random Forests“. *Lecture Notes in Computer Science*. Munich, Hrsg. V. Ferrari, H. Martial, C. Sminchisescu, W. Yair, M. Hebert & Y. Weiss, Springer; Springer International Publishing, Cham, S. 398–412. ISBN: 9783030012182.

Ehrmann 2019

Ehrmann, C. (2019), *Ausfallfrüherkennung von Ritzel-Zahnstangen-Trieben mittels*

Acoustic Emission. Dissertation, Karlsruher Institut für Technologie, Shaker, Düren. ISBN: 9783844077032.

Ehrmann, Isabey & Fleischer 2016

Ehrmann, C.; Isabey, P. & Fleischer, J. (2016), „Condition Monitoring of Rack and Pinion Drive Systems: Necessity and Challenges in Production Environments“, *Procedia CIRP*, 40 (1), S. 197–201.

Ellinger & Benker et al. 2019

Ellinger, J.; Benker, M.; Rowe, J.; Kleinwort, R.; Zäh, M. F. & Unsleber, S. (2019), „Zustandsüberwachung von Vorschubantrieben mithilfe eingebetteter Sensoren“, *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 114 (3), S. 115–118.

Faghih-Roohi & Hajizadeh et al. 2016

Faghih-Roohi, S.; Hajizadeh, S.; Nunez, A.; Babuska, R. & Schutter, B. de (2016), „Deep convolutional neural networks for detection of rail surface defects“. *2016 International Joint Conference on Neural Networks (IJCNN). 24-29 July 2016, Vancouver, Canada*. Vancouver, BC, Canada, Hrsg. IEEE, IEEE, Piscataway, NJ, S. 2584–2589. ISBN: 9781509006199.

Fink & Wang et al. 2020

Fink, O.; Wang, Q.; Svensén, M.; Dersin, P.; Lee, W.-J. & Ducoffe, M. (2020), „Potential, challenges and future directions for deep learning in prognostics and health management applications“, *Engineering Applications of Artificial Intelligence*, 92 (1), S. 103678.

Fleischer & Broos et al. 2009

Fleischer, J.; Broos, A.; Schopp, M.; Wieser, J.; Hennrich, H.; Fleischer, J.; Broos, A.; Schopp, M.; Wieser, J. & Hennrich, H. (2009), „Lifecycle-oriented component selection for machine tools based on multibody simulation and component life prediction“, *CIRP Journal of Manufacturing Science and Technology*, 1 (3), S. 179–184.

Fleischer & Lanza et al. 2013

Fleischer, J.; Lanza, G.; Appel, D.; Stricker, N.; Hennrich, H. & Herder, S. (2013), „Life Cycle Performance 4.0: Strategische und technische Lösungen für den intelligenten Betrieb von Maschinen und Anlagen“, *wt Werkstattstechnik online*, 103 (2), S. 124–129.

Fleischer & Schopp et al. 2007

Fleischer, J.; Schopp, M.; Broos, A. & Wieser, J. (2007), „Datenbasis für lastabhängige Prozesseingriffe - Modularisierung und Analyse von Ausfallursachen zur Erhöhung der Verfügbarkeit von Werkzeugmaschinen“, *wt Werkstattstechnik online*, 97 (7/8), S. 491–497.

Forstmann 2010

Forstmann, J. (2010), *Kugelgewindetriebe im Einsatz an Kunststoffspritzgießmaschinen – Lebensdauerprognose und Optimierung*. Dissertation, Universität Duisburg-Essen, Universität Duisburg-Essen, Duisburg-Essen.

Fort 2017

Fort, S. (2017), *Gaussian Prototypical Networks for Few-Shot Learning on Omniglot*. Arxiv Preprint. <https://arxiv.org/pdf/1708.02735> [22.07.2022].

Fu & Chu et al. 2019

Fu, J.; Chu, J.; Guo, P. & Chen, Z. (2019), „Condition Monitoring of Wind Turbine Gearbox Bearing Based on Deep Learning Model“, *IEEE Access*, 7 (7), S. 57078–57087.

Galan & Orta et al. 2018

Galan, U.; Orta, P.; Kurfess, T. & Ahuett-Garza, H. (2018), „Surface defect identification and measurement for metal castings by vision system“, *Manufacturing Letters*, 15 (1), S. 5–8.

Geister, Hohwieler & Fraunhofer-Institut für Produktionsanlagen und Konstruktionstechnik 2019

Geister, C.; Hohwieler, E. & Fraunhofer-Institut für Produktionsanlagen und Konstruktionstechnik (2019), *Verbundprojekt AMELI4.0. Mikro-elektromechanisches Elektroniksystem zur Zustandsüberwachung in der Industrie 4.0 : Abschlussbericht zum Teilvorhaben AMELI4.0 - Data : Verschleiß-, Anlagen- sowie Prozessmonitoring durch intelligente, autonome Multi-Sensorsysteme : Zeitraum: 1. Dezember 2015 bis 30. November 2018*, Berlin.

Geogi & Umarali et al. 2020

Geogi, S.; Umarali, K.; Lalu, P. & Sudeesh, R. S. (2020), „A neural network approach for detecting surface defects in hot rolling process“, *Journal of Telecommunications System & Management*, 9 (5), 57.

Ghorai & Mukherjee et al. 2013

Ghorai, S.; Mukherjee, A.; Gangadaran, M. & Dutta, P. K. (2013), „Automatic Defect Detection on Hot-Rolled Flat Steel Products“, *IEEE Transactions on Instrumentation and Measurement*, 62 (3), S. 612–621.

Golz 1990

Golz, H. U. (1990), *Analyse, Modellbildung und Optimierung des Betriebsverhaltens von Kugelgewindetrieben*. Dissertation, Universität Karlsruhe, Universität Karlsruhe, Karlsruhe.

Goodfellow, Bengio & Courville 2016

Goodfellow, I.; Bengio, Y. & Courville, A. (2016), *Deep Learning*, The MIT Press, Cambridge, Massachusetts, London, England. ISBN: 9780262035613.

Goodfellow & Pouget-Abadie et al. 2014

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A. & Bengio, Y. (2014), „Generative Adversarial Nets“. *Advances in neural information processing systems* 27. Montreal, Canada, Hrsg. M. Welling, Curran Associates Inc, Red Hook, NY, S. 139–144. ISBN: 9781510800410.

Gronau 2008

Gronau, N. (2008), *Wettbewerbsfähigkeit durch Arbeits- und Betriebsorganisation. 20. Forschungsseminar der Hochschulgruppe Arbeits- und Betriebsorganisation (HAB)*, Gito-Verlag; Gito-Verl., Berlin. ISBN: 9783940019448.

Guo & Infield et al. 2012

Guo, P.; Infield, D.; Yang, X. & Infield, D. (2012), „Wind Turbine Generator Condition-Monitoring Using Temperature Trend Analysis“, *IEEE Transactions on Sustainable Energy*, 3 (1), S. 124–133.

Gupta & Peroulis 2013

Gupta, L. A. & Peroulis, D. (2013), „Wireless Temperature Sensor for Condition Monitoring of Bearings Operating Through Thick Metal Plates“, *IEEE Sensors Journal*, 13 (6), S. 2292–2298.

Haberhauer & Kaczmarek 2014

Haberhauer, H. & Kaczmarek, M. (2014), *Taschenbuch der Antriebstechnik. Mit 7 Tabellen*, Fachbuchverlag Leipzig im Carl Hanser Verlag, München. ISBN: 9783446427709.

Haber Kern 1998

Haber Kern, A. (1998), *Leistungsfähigere Kugelgewindetriebe durch Beschichtung*. Dissertation, Universität Karlsruhe, Universität Karlsruhe, Karlsruhe.

Haralick 1979

Haralick, R. M. (1979), „Statistical and structural approaches to texture“, *Proceedings of the IEEE*, 67 (5), S. 786–804.

Haralick, Shanmugam & Dinstein 1973

Haralick, R. M.; Shanmugam, K. & Dinstein, I. (1973), „Textural Features for Image Classification“, *IEEE Transactions on Systems, Man, and Cybernetics*, 3 (6), S. 610–621.

Hatab, Malekmohamadi & Amira 2020

Hatab, M.; Malekmohamadi, H. & Amira, A. (2020), „Surface Defect Detection Using YOLO Network“, *Intelligent Systems and Applications*, 1250, S. 505–515.

He & Liu et al. 2019

He, T.; Liu, Y.; Xu, C.; Zhou, X.; Hu, Z. & Fan, J. (2019), „A Fully Convolutional Neural Network for Wood Defect Location and Identification“, *IEEE Access*, 7 (1), S. 123453–123462.

He & Song et al. 2019

He, Y.; Song, K.; Dong, H. & Yan, Y. (2019), „Semi-supervised defect classification of steel surface based on multi-training and generative adversarial network“, *Optics and Lasers in Engineering*, 122 (1), S. 294–302.

He & Song et al. 2020

He, Y.; Song, K.; Meng, Q. & Yan, Y. (2020), „An End-to-End Steel Surface Defect Detection Approach via Fusing Multiple Hierarchical Features“, *IEEE Transactions on Instrumentation and Measurement*, 69 (4), S. 1493–1504.

He & Zhang et al. 2016

He, K.; Zhang, X.; Ren, S. & Sun, J. (2016), „Deep Residual Learning for Image Recognition“. *CVPR 2016 : proceedings*. Las Vegas, NV, USA, Hrsg. R. Bajcsy, F.-F. Li & T. Tuytelaars, IEEE, Piscataway, NJ, S. 770–778. ISBN: 9781467388528.

Heise 2002

Heise, W. (2002), *Praxisbuch Zuverlässigkeit und Wartungsfreundlichkeit. R-&M-*

Programm für Automobilzulieferer und den Maschinen- und Anlagenbau, Hanser, München. ISBN: 9783446219304.

Helwig 2018

Helwig, N. J. (2018), *Zustandsbewertung industrieller Prozesse mittels multivariater Sensordatenanalyse am Beispiel hydraulischer und elektromechanischer Antriebssysteme*. Dissertation, Universität des Saarlandes, Shaker, Düren. ISBN: 9783844064940.

Henrich 2013

Henrich, H. (2013), *Aufbau eines kombinierten belastungs- und zustandsorientierten Diagnose- und Prognosesystems für Kugelgewindetriebe*. Dissertation, Karlsruher Institut für Technologie, Shaker, Düren. ISBN: 9783844023121.

Herder 2013

Herder, S. (2013), *Piezoelektrischer Self-Sensing-Aktor zur Vorspannungsregelung in adaptronischen Kugelgewindetriebe*. Dissertation, Karlsruher Institut für Technologie, Shaker, Düren. ISBN: 9783844023428.

Hillenbrand 2021

Hillenbrand, J. (2021), *Erweiterung der Lebensdauertheorie durch Korrelation des Verschleißes mit Acoustic Emission (AE) zu den mechanischen Kennwerten des Kugelgewindetriebes. Vortrag zu Projektergebnissen*, Karlsruhe.

Hillenbrand & Fleischer 2020

Hillenbrand, J. & Fleischer, J. (2020), „Autoconfiguration of a Vibration-Based Anomaly Detection System with Sparse a-priori Knowledge Using Autoencoder Networks“. *Proceedings of the 10th Congress of the German Academic Association for Production Technology (WGP)*. Dresden, Hrsg. B.-A. Behrens, A. Brosius, W. Hintze, S. Ihlenfeldt & J. P. Wulfsberg, Springer, Berlin, Heidelberg, S. 518–527. ISBN: 9783662621387.

Hillenbrand, Spohrer & Fleischer 2018

Hillenbrand, J.; Spohrer, A. & Fleischer, J. (2018), „Zustandsüberwachung bei Kugelgewindetriebe. Integration von DMS-Sensorik in Kugelgewindetriebemuttern“, *wt Werkstatttechnik online*, 7/8 (108), S. 493–498.

Hilmer 1978

Hilmer, H. (1978), *Rechnergestützte Auslegung und Berechnung von Kugelgewindespindeln*. Dissertation, Universität Karlsruhe, Universität Karlsruhe, Karlsruhe.

Hu & Long et al. 2021

Hu, Y.; Long, Z.; Sundaresan, A.; Alfarraj, M.; AlRegib, G.; Park, S.; Jayaraman, S.; Long, Z.; Alfarraj, M.; AlRegib, G. & Jayaraman, S. (2021), „Fabric surface characterization: assessment of deep learning-based texture representations using a challenging dataset“, *The Journal of The Textile Institute*, 112 (2), S. 293–305.

Hubel & Wiesel 1959

Hubel, D. H. & Wiesel, T. N. (1959), „Receptive fields of single neurones in the cat's striate cortex“, *The Journal of Physiology*, 148 (3), S. 574–591.

Huf 2012

Huf, A. (2012), *Kumulative Lastermittlung aus Antriebsdaten zur Bewertung des Zustands von Werkzeugmaschinenkomponenten*. Dissertation, Universität Stuttgart, Universität Stuttgart, Stuttgart. ISBN: 9783939890966.

Imiela 2006

Imiela, J. (2006), *Verfügbarkeitssicherung von Werkzeugmaschinenachsen mit Kugelgewindetrieb durch modellbasierte Verschleißüberwachung*. Dissertation, Universität Hannover, Produktionstechnisches Zentrum, Hannover. ISBN: 3939026042.

Jarosch 2008

Jarosch, P. (2008), *Zur Lebensdauerprognose zyklisch hoch belasteter Kugelgewindetribe*. Dissertation, Universität Duisburg-Essen, Shaker, Düren. ISBN: 9783832272838.

Kang & Liu 2005

Kang, G.-W. & Liu, H.-B. (2005), „Surface defects inspection of cold rolled strips based on neural network“. *Proceedings of 2005 International Conference on Machine Learning and Cybernetics*. Guangzhou, China, Hrsg. IEEE Systems, Man, and Cybernetics Society, IEEE Operations Center, Piscataway, NJ, S. 5034–5037.

Kar & Mohanty 2006

Kar, C. & Mohanty, A. R. (2006), „Multistage gearbox condition monitoring using

motor current signature analysis and Kolmogorov–Smirnov test“, *Journal of Sound and Vibration*, 290 (1-2), S. 337–368.

Kattelus, Miettinen & Lehtovaara 2018

Kattelus, J.; Miettinen, J. & Lehtovaara, A. (2018), „Detection of gear pitting failure progression with on-line particle monitoring“, *Tribology International*, 118 (1), S. 458–464.

Kern 2009

Kern, J. (2009), *Ishikawa Diagramme - Ursache-Wirkungs-Diagramme*, GRIN Verlag GmbH, München. ISBN: 9783640284146.

Kim & Park et al. 2019

Kim, M. S.; Park, T.; Park, P.; Soukup, D. & Huber-Mörk, R. (2019), „Classification of Steel Surface Defect Using Convolutional Neural Network with Few Images“. *Proceedings 12th Asian Control Conference (ASCC)*. Fukuoka, Japan, Hrsg. IEEE Control Systems Society, IEEE, Piscataway, NJ, S. 1398–1401. ISBN: 9781728102634.

Klein 2008

Klein, U. (2008), *Schwingungsdiagnostische Beurteilung von Maschinen und Anlagen*, Stahleisen, Düsseldorf. ISBN: 9783514006874.

Klein 2011

Klein, W. H. (2011), *Zustandsüberwachung von Rollen-Profileschiene Führungen und Kugelgewindetrieben*. Dissertation, RWTH Aachen, Apprimus Verlag, Aachen. ISBN: 3863590295.

Koch 2015

Koch, G. (2015), *Siamese Neural Networks for One-Shot Image Recognition*. Masterthesis, University of Toronto, Department of Computer Science.

Koch & Georgieva et al. 2015

Koch, C.; Georgieva, K.; Kasireddy, V.; Akinci, B. & Fieguth, P. (2015), „A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure“, *Advanced Engineering Informatics*, 29 (2), S. 196–210.

Kodituwakku & Selvarajah 2010

Kodituwakku, S. R. & Selvarajah, S. (2010), „Comparison of Color Features for

Image Retrieval“, *Indian Journal of Computer Science and Engineering*, 1 (3), S. 207–211.

Koizumi & Kawaguchi et al. 2020

Koizumi, Y.; Kawaguchi, Y.; Imoto, K.; Nakamura, T.; Nikaido, Y.; Tanabe, R.; Purohit, H.; Suefusa, K.; Endo, T.; Yasuda, M. & Harada, N. (2020), „Description and Discussion on DCASE2020 Challenge Task2: Unsupervised Anomalous Sound Detection for Machine Condition Monitoring“. *Proceedings of the 5th Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE 2020)*. Tokyo, Japan, Hrsg. Nobutaka Ono, Noboru Harada, Yohei Kawaguchi, Anamaria Mesaros, Keisuke Imoto, Yuma Koizumi & Tatsuya Komatsu, Zenodo, Genève, S. 81–85. ISBN: 9784600005665.

Kolerus & Wassermann 2008

Kolerus, J. & Wassermann, J. (2008), *Zustandsüberwachung von Maschinen. Das Lehr- und Arbeitsbuch für den Praktiker*, expert-Verlag, Renningen. ISBN: 9783816925972.

Kothuru, Nooka & Liu 2019

Kothuru, A.; Nooka, S. P. & Liu, R. (2019), „Application of deep visualization in CNN-based tool condition monitoring for end milling“. *47th North American Manufacturing Research Conference 2019 (NAMRC 47)*. Erie, Pennsylvania, USA, Hrsg. L. Fratini, I. Ragai & L. Wang, Curran Associates Inc, Red Hook, NY, S. 995–1004. ISBN: 9781713801009.

Krishnakumar, Rameshkumar & Ramachandran 2018

Krishnakumar, P.; Rameshkumar, K. & Ramachandran, K. I. (2018), „Feature level fusion of vibration and acoustic emission signals in tool condition monitoring using machine learning classifiers“, *International Journal of Prognostics and Health Management*, 9 (1), S. 1–15.

Krizhevsky, Sutskever & Hinton 2012

Krizhevsky, A.; Sutskever, I. & Hinton, G. E. (2012), „ImageNet Classification with Deep Convolutional Neural Networks“. *Advances in Neural Information Processing Systems*. Lake Tahoe, Nevada, USA, Hrsg. F. Pereira, C. J. C. Burges, L. Bottou & K. Q. Weinberger, Curran Associates Inc, Red Hook, N.Y., S. 1097–1105. ISBN: 9781627480031.

Kumar & Vashishtha et al. 2021

Kumar, A.; Vashishtha, G.; Gandhi, C. P.; Zhou, Y.; Glowacz, A. & Xiang, J. (2021), „Novel Convolutional Neural Network (NCNN) for the Diagnosis of Bearing Defects in Rotary Machinery“, *IEEE Transactions on Instrumentation and Measurement*, 70 (1), S. 1–10.

Kuntoğlu & Sağlam 2021

Kuntoğlu, M. & Sağlam, H. (2021), „Investigation of signal behaviors for sensor fusion with tool condition monitoring system in turning“, *Measurement*, 173 (1), S. 108582.

Lapuschkin & Wäldchen et al. 2019

Lapuschkin, S.; Wäldchen, S.; Binder, A.; Montavon, G.; Samek, W. & Müller, K.-R. (2019), „Unmasking Clever Hans predictors and assessing what machines really learn“, *Nature Communications*, 10 (1), S. 1096.

Lecun & Boser et al. 1989

Lecun, Y.; Boser, B.; Denker, J. S.; Henderson, D.; Howard, R. E.; Hubbard, W. & Jackel, L. D. (1989), „Backpropagation Applied to Handwritten Zip Code Recognition“, *Neural Computation*, 1 (4), S. 541–551.

Lecun & Bottou et al. 1998

Lecun, Y.; Bottou, L.; Bengio, Y. & Haffner, P. (1998), „Gradient-based learning applied to document recognition“, *Proceedings of the IEEE*, 86 (11), S. 2278–2324.

LeCun, Bengio & Hinton 2015

LeCun, Y.; Bengio, Y. & Hinton, G. (2015), „Deep learning“, *Nature*, 521 (7553), S. 436–444.

Lee & Huang et al. 2018

Lee, K.-M.; Huang, Y.; Ji, J. & Lin, C.-Y. (2018), „An Online Tool Temperature Monitoring Method Based on Physics-Guided Infrared Image Features and Artificial Neural Network for Dry Cutting“, *IEEE Transactions on Automation Science and Engineering*, 15 (4), S. 1665–1676.

Lee & Jwo et al. 2020

Lee, C.-H.; Jwo, J.-S.; Hsieh, H.-Y. & Lin, C.-S. (2020), „An Intelligent System for

Grinding Wheel Condition Monitoring Based on Machining Sound and Deep Learning“, *2169-3536*, 8 (1), S. 58279–58289.

Lee & Yu et al. 2021

Lee, S.; Yu, H.; Yang, H.; Song, I.; Choi, J.; Yang, J.; Lim, G.; Kim, K.-S.; Choi, B. & Kwon, J. (2021), „A Study on Deep Learning Application of Vibration Data and Visualization of Defects for Predictive Maintenance of Gravity Acceleration Equipment“, *Applied Sciences*, 11 (4), S. 1564.

Lessmeier 2017

Lessmeier, C. (2017), *Datenbasierte Zustandsüberwachung von Wälzlagerschäden in elektromechanischen Antriebssystemen*. Dissertation, Universität Paderborn, Shaker, Düren. ISBN: 3844054901.

Li & An 2016

Li, L. & An, Q. (2016), „An in-depth study of tool wear monitoring technique based on image segmentation and texture analysis“, *Measurement*, 79 (1), S. 44–52.

Li, Liu & Bilen 2021

Li, W.-H.; Liu, X. & Bilen, H. (2021), „Universal Representation Learning from Multiple Domains for Few-shot Classification“. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, Canada, Hrsg. T. Berg, J. Clark, Y. Matsushita & C. Taylor, Curran Associates Inc, Red Hook, NY, S. 9526–9535.

Li & Su et al. 2018

Li, J.; Su, Z.; Geng, J. & Yin, Y. (2018), „Real-time Detection of Steel Strip Surface Defects Based on Improved YOLO Detection Network“, *IFAC-PapersOnLine*, 51 (21), S. 76–81.

Liu & Wang et al. 2017

Liu, K.; Wang, H.; Chen, H.; Qu, E.; Tian, Y. & Sun, H. (2017), „Steel Surface Defect Detection Using a New Haar–Weibull-Variance Model in Unsupervised Manner“, *IEEE Transactions on Instrumentation and Measurement*, 66 (10), S. 2585–2596.

Liu, Wang & Chen 2019

Liu, Z.; Wang, X. & Chen, X. (2019), „Inception Dual Network for steel strip defect

detection“. *Proceedings of the 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC)*. Banff, Alberta, Canada, Hrsg. H. Zhu, J. Wang & M. Zhou, IEEE, Piscataway, NJ, S. 409–414. ISBN: 9781728100845.

Liu, Xu & Xu 2019

Liu, Y.; Xu, K. & Xu, J. (2019), „Periodic Surface Defect Detection in Steel Plates Based on Deep Learning“, *Applied Sciences*, 9 (15), S. 3127.

Lowe 1999

Lowe, D. G. (1999), „Object recognition from local scale-invariant features“. *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Kerkyra, Greece, Hrsg. IEEE, IEEE Computer Society, Los Alamitos, California, 1150-1157. ISBN: 0769501648.

Lowe 2004

Lowe, D. G. (2004), „Distinctive Image Features from Scale-Invariant Keypoints“, *International Journal of Computer Vision*, 60 (2), S. 91–110.

Lughofer & Sayed-Mouchaweh 2019

Lughofer, E. & Sayed-Mouchaweh, M. (2019), *Predictive Maintenance in Dynamic Systems. Advanced Methods, Decision Support Tools and Real-World Applications*, Springer International Publishing, Cham. ISBN: 9783030056452.

Lundberg & Palmgren 1952

Lundberg, G. & Palmgren, A. (1952), *Dynamic capacity of rolling bearings*, Stockholm.

Luo & Fang et al. 2020

Luo, Q.; Fang, X.; Liu, L.; Yang, C. & Sun, Y. (2020), „Automated Visual Defect Detection for Flat Steel Surface: A Survey“, *IEEE Transactions on Instrumentation and Measurement*, 69 (3), S. 626–644.

Luo & He 2016

Luo, Q. & He, Y. (2016), „A cost-effective and automatic surface defect inspection system for hot-rolled flat steel“, *Robotics and Computer-Integrated Manufacturing*, 38 (1), S. 16–30.

Lv & Duan et al. 2020

Lv, X.; Duan, F.; Jiang, J.-J.; Fu, X. & Gan, L. (2020), „Deep Metallic Surface Defect Detection: The New Benchmark and Detection Network“, *Sensors*, 20 (6), S. 1562.

Macherauch & Zoch 2011

Macherauch, E. & Zoch, H.-W. (2011), *Praktikum in Werkstoffkunde: 91 ausführliche Versuche aus wichtigen Gebieten der Werkstofftechnik*, Vieweg+Teubner Verlag, Wiesbaden. ISBN: 9783834898845.

Maier 2015

Maier, D. (2015), *Sensorlose online Zustandserfassung von Vorschubantriebskomponenten in Werkzeugmaschinen*. Dissertation, Universität Stuttgart, Universität Stuttgart, Stuttgart. ISBN: 9783936100587.

Masci & Meier et al. 2012

Masci, J.; Meier, U.; Ciresan, D.; Schmidhuber, J. & Fricout, G. (2012), „Steel defect classification with Max-Pooling Convolutional Neural Networks“. *2012 International Joint Conference on Neural Networks (IJCNN 2012)*. Brisbane, Australia, Hrsg. IEEE Computational Intelligence Society, IEEE, Piscataway, NJ, S. 1–6. ISBN: 9781467314909.

Medved, Breton & Irving 2004

Medved, J.; Breton, M. & Irving, P. (2004), „Corrosion pit size distributions and fatigue lives—a study of the EIFS technique for fatigue design in the presence of corrosion“, *International Journal of Fatigue*, 26 (1), S. 71–80.

Metzele 2008

Metzele, M. (2008), *Zustandsorientierte Instandhaltung von schnelllaufenden Werkzeugmaschinen-Hauptspindeln*. Dissertation, RWTH Aachen, Shaker, Düren. ISBN: 3832268782.

Mobley 2002

Mobley, R. K. (2002), *An introduction to predictive maintenance*, Butterworth-Heinemann, Amsterdam, New York. ISBN: 9780750675314.

Möhring & Bertram 2012

Möhring, H.-C. & Bertram, O. (2012), „Integrated Autonomous monitoring of ball screw drives“, *CIRP Annals*, 61 (1), S. 355–358.

Münzing 2017

Münzing, T. (2017), *Auslegung von Kugelgewindetrieben bei oszillierenden Bewegungen und dynamischer Belastung*. Dissertation, Universität Stuttgart, Universität Stuttgart, Stuttgart. ISBN: 3946924026.

Munzinger, Schopp & Broos 2009

Munzinger, C.; Schopp, M. & Broos, A. (2009), „Optimizing Life Cycle Performance of Machine Components Using Time-Discrete Condition Monitoring and Prognosis“. *Proceedings of the 16th CIRP International Conference on Life Cycle Engineering*. Cairo, Egypt, Hrsg. CIRP, o.V., o.O., S. 418–424. ISBN: 9780978318727.

Munzinger, Schopp & Hennrich 2010

Munzinger, C.; Schopp, M. & Hennrich, H. (2010), „Increasing maintenance efficiency - A strategy regarding condition and load-based approaches for diagnosis and prognosis“. *CIRP Process Machine Interactions Conference Proceedings*. Vancouver, Kanada, Hrsg. CIRP, o.V., o.O., S. 19–28.

Murrenhoff 2010

Murrenhoff, H. (2010), *Umweltverträgliche Tribosysteme. Die Vision einer umweltfreundlichen Werkzeugmaschine*, Springer, Berlin, Heidelberg. ISBN: 9783642049972.

Neogi, Mohanta & Dutta 2014

Neogi, N.; Mohanta, D. K. & Dutta, P. K. (2014), „Review of vision-based steel surface inspection systems“, *EURASIP Journal on Image and Video Processing*, 2014 (1), S. 50.

Nesges 2005

Nesges, D. (2005), *Prognose operationeller Verfügbarkeiten von Werkzeugmaschinen unter Berücksichtigung von Serviceleistungen*. Dissertation, Universität Karlsruhe, Universität Karlsruhe, Karlsruhe.

Niu & Li et al. 2020

Niu, S.; Li, B.; Wang, X. & Lin, H. (2020), „Defect Image Sample Generation With GAN for Improving Defect Recognition“, *IEEE Transactions on Automation Science and Engineering*, 17 (3), S. 1–12.

Ojala, Pietikainen & Maenpaa 2002

Ojala, T.; Pietikainen, M. & Maenpaa, T. (2002), „Multiresolution gray-scale and rotation invariant texture classification with local binary patterns“, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24 (7), S. 971–987.

Otsu 1979

Otsu, N. (1979), „A Threshold Selection Method from Gray-Level Histograms“, *IEEE Transactions on Systems, Man, and Cybernetics*, 9 (1), S. 62–66.

Park & Oh et al. 2019

Park, K.; Oh, S.; Lee, S. H. & Cho, N. I. (2019), „Blind detection of fabric defects using multiple image features“. *International Workshop on Advanced Image Technology (IWAIT) 2019*. Singapore, Malaysia, Hrsg. Q. Kemao, SPIE, Bellingham, Washington, USA, S. 104–108. ISBN: 9781510627734.

Rabinowicz 1995

Rabinowicz, E. (1995), *Friction and wear of materials*, Wiley, New York, NY. ISBN: 9780471830849.

Rafezi, Akbari & Behzad 2012

Rafezi, H.; Akbari, J. & Behzad, M. (2012), „Tool Condition Monitoring based on sound and vibration analysis and wavelet packet decomposition“. *8th International Symposium on Mechatronics and Its Applications (ISMA)*. Sharjah, United Arab Emirates, Hrsg. IEEE, IEEE, Piscataway, NJ, S. 1–4. ISBN: 9781467308601.

Ran & Zhou et al. 2019

Ran, Y.; Zhou, X.; Lin, P.; Wen, Y. & Deng, R. (2019), *A Survey of Predictive Maintenance: Systems, Purposes and Approaches*. Arxiv Preprint. <https://arxiv.org/pdf/1912.07383> [01.02.2022].

Raschka 2016

Raschka, S. (2016), *Python machine learning. Unlock deeper insights into machine learning with this vital guide to cutting-edge predictive analytics*, Packt Publishing open source, Birmingham, Mumbai. ISBN: 9781783555130.

Reichel, Müller & Haeffs 2018

Reichel, J.; Müller, G. & Haeffs, J. (Hrsg.) (2018), *Betriebliche Instandhaltung*, Springer Vieweg, Berlin, Heidelberg. ISBN: 9783662531341.

Ren & He et al. 2016

Ren, S.; He, K.; Girshick, R. B. & Sun, J. (2016), „Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks“. *Advances in neural information processing systems*. Montreal, Canada, Hrsg. C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama & R. Garnett, Curran Associates Inc, Red Hook, NY, S. 1137–1149. ISBN: 9781510825024.

Ren & Huang et al. 2020

Ren, Y.; Huang, J.; Hong, Z.; Lu, W.; Yin, J.; Zou, L. & Shen, X. (2020), „Image-based concrete crack detection in tunnels using deep fully convolutional networks“, *Construction and Building Materials*, 234 (1), S. 117367.

Ren, Hung & Tan 2018

Ren, R.; Hung, T. & Tan, K. C. (2018), „A Generic Deep-Learning-Based Approach for Automated Surface Inspection“, *IEEE transactions on cybernetics*, 48 (3), S. 929–940.

Riaz & Shah et al. 2020

Riaz, N.; Shah, S. I. A.; Rehman, F.; Gilani, S. O. & Udin, E. (2020), „A Novel 2-D Current Signal-Based Residual Learning With Optimized Softmax to Identify Faults in Ball Screw Actuators“, 8 (1), S. 115299–115313.

Riaz & Shah et al. 2021

Riaz, N.; Shah, S. I. A.; Rehman, F. & Khan, M. J. (2021), „An Intelligent Hybrid Scheme for Identification of Faults in Industrial Ball Screw Linear Motion Systems“, *IEEE Access*, 9 (1), S. 35136–35150.

Richter 2018

Richter, M. (2018), *Über lernende optische Inspektion am Beispiel der Schüttgutsortierung*. Dissertation, Karlsruher Institut für Technologie, KIT Scientific Publishing, Karlsruhe.

Roda, Macchi & Fumagalli 2018

Roda, I.; Macchi, M. & Fumagalli, L. (2018), „The Future of Maintenance Within Industry 4.0: An Empirical Research in Manufacturing“. *Smart manufacturing for industry 4.0*. Seoul, Korea, Hrsg. I. Moon, G. M. Lee, J. Park, D. Kiritsis & G. von Cieminski, Springer, Cham, S. 39–46. ISBN: 9783319997063.

Rojas 1996

Rojas, R. (1996), „The Backpropagation Algorithm“ in *Neural networks. A systematic introduction*, Hrsg. R. Rojas & J. Feldman, Springer, Berlin, Heidelberg, S. 149–182.

Ronneberger, Fischer & Brox 2015

Ronneberger, O.; Fischer, P. & Brox, T. (2015), „U-Net: Convolutional Networks for Biomedical Image Segmentation“. *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich, Hrsg. J. Hornegger, N. Navab, W. M. Wells & A. F. Frangi, Springer, Cham, S. 234–241. ISBN: 9783319245522.

Rosenblatt 1958

Rosenblatt, F. (1958), „The perceptron: A probabilistic model for information storage and organization in the brain“, *Psychological Review*, 65 (6), S. 386–408.

Rublee & Rabaud et al. 2011

Rublee, E.; Rabaud, V.; Konolige, K. & Bradski, G. (2011), „ORB: An efficient alternative to SIFT or SURF“. *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops 2011)*. Barcelona, Spain, Hrsg. IEEE, IEEE, Piscataway, NJ, S. 2564–2571. ISBN: 9781467300629.

Russakovsky & Deng et al. 2015

Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; Berg, A. C. & Fei-Fei, L. (2015), „ImageNet Large Scale Visual Recognition Challenge“, *International Journal of Computer Vision*, 115 (3), S. 211–252.

Samy & Foong et al. 2016

Samy, M. P.; Foong, S.; Soh, G. S. & Yeo, K. S. (2016), „Automatic optical & laser-based defect detection and classification in brick masonry walls“. *Proceedings of the 2016 IEEE Region 10 Conference (TENCON)*. Marina Bay Sands, Singapore, Hrsg. IEEE, IEEE, Piscataway, NJ, S. 3521–3524. ISBN: 9781509025961.

Sauer & Steinhilper 2012

Sauer, B. & Steinhilper, W. (Hrsg.) (2012), *Konstruktionselemente des Maschinenbaus 2. Grundlagen von Maschinenelementen für Antriebsaufgaben*, Springer Vieweg, Berlin. ISBN: 9783642243028.

Schadow 2016

Schadow, C. (2016), *Stillstehende fettgeschmierte Wälzlager unter dynamischer Beanspruchung*. Dissertation, Otto-von-Guericke-Universität Magdeburg, Shaker, Düren. ISBN: 9783844048582.

Schlagenhauf 2021

Schlagenhauf, T. (2021), *Ball Screw Drive Surface Defect Dataset for Classification*. <https://publikationen.bibliothek.kit.edu/1000133819> [25.07.2022].

Schlagenhauf & Burghardt 2021

Schlagenhauf, T. & Burghardt, N. (2021), „Intelligent vision based wear forecasting on surfaces of machine tool elements“, *Springer Nature Applied Sciences*, 3 (858), S. 1–13.

Schlagenhauf, Heinzler & Fleischer 2020

Schlagenhauf, T.; Heinzler, M. & Fleischer, J. (2020), „Extraction of surface image features for wear detection on ball screw drive spindles“. *Forum Bildverarbeitung 2020*. Karlsruhe, Hrsg. T. Längle & M. Heizmann, KIT Scientific Publishing, Karlsruhe, S. 305–314. ISBN: 9783731510536.

Schlagenhauf & Hillenbrand et al. 2019

Schlagenhauf, T.; Hillenbrand, J.; Klee, B. & Fleischer, J. (2019), „Integration von Machine Vision in Kugelgewindespindeln“, *wt Werkstatttechnik online*, 109 (7/8), S. 605–610.

Schlagenhauf & Landwehr 2021

Schlagenhauf, T. & Landwehr, M. (2021), „Industrial machine tool component surface defect dataset“, *Data in Brief*, 39 (1), S. 107643.

Schlagenhauf, Ruppelt & Fleischer 2020

Schlagenhauf, T.; Ruppelt, P. & Fleischer, J. (2020), „Frühzeitige Detektion von Oberflächenzerrüttungen“, *wt Werkstattstechnik online*, 110 (07-08), S. 501–506.

Schlagenhauf & Scheurenbrand et al. 2022

Schlagenhauf, T.; Scheurenbrand, T.; Hofmann, D. & Krasnikow, O. (2022), *Analysis of the Visually Detectable Wear Progress on Ball Screws*. Arxiv Preprint. Under Review in the CIRP Journal of Manufacturing Science and Technology. <https://arxiv.org/abs/2205.01149> [25.07.2022].

Schlagenhauf, Yildirim & Brückner 2022

Schlagenhauf, T.; Yildirim, F. & Brückner, B. (2022), *Siamese Basis Function Networks for Data-efficient Defect Classification in Technical Domains*. Arxiv Preprint. Accepted at the Berlin Workshop on Artificial Intelligence for Engineering Applications (AI4EA). <https://arxiv.org/abs/2012.01338> [25.07.2022].

Schmid & Gädeke et al. 2010

Schmid, J.; Gädeke, T.; Stork, W.; Hennrich, H. & Blank, T. (2010), „A wireless MEMS-sensor network concept for the condition monitoring of ball screw drives in industrial plants“. *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*. Zürich, Switzerland, Hrsg. J. Beutel, ACM, New York, NY, S. 425. ISBN: 9781450303446.

Schopp 2009

Schopp, M. (2009), *Sensorbasierte Zustandsdiagnose und -prognose von Kugelgewindetrieben*. Dissertation, Universität Karlsruhe, Shaker, Düren. ISBN: 9783832287337.

Schroff, Kalenichenko & Philbin 2015

Schroff, F.; Kalenichenko, D. & Philbin, J. (2015), „FaceNet: A unified embedding for face recognition and clustering“. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA, Hrsg. IEEE, IEEE, Piscataway, NJ, S. 815–823. ISBN: 9781467369640.

Selvaraju & Cogswell et al. 2020

Selvaraju, R. R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D. & Batra, D. (2020), „Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization“, *International Journal of Computer Vision*, 128 (2), S. 336–359.

Serin & Sener et al. 2020

Serin, G.; Sener, B.; Ozbayoglu, A. M. & Unver, H. O. (2020), „Review of tool condition monitoring in machining and opportunities for deep learning“, *The International Journal of Advanced Manufacturing Technology*, 109 (3-4), S. 953–974.

Simonyan & Zisserman 2015

Simonyan, K. & Zisserman, A. (2015), *Very Deep Convolutional Networks for Large-Scale Image Recognition*. Arxiv Preprint. <https://arxiv.org/abs/1409.1556> [25.07.2022].

Sommer, Heinz & Schöfer 2018

Sommer, K.; Heinz, R. & Schöfer, J. (2018), *Verschleiß metallischer Werkstoffe. Erscheinungsformen sicher beurteilen*, Springer Vieweg, Wiesbaden, Heidelberg. ISBN: 9783658178505.

Song & Yan 2013

Song, K. & Yan, Y. (2013), „A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects“, *Applied Surface Science*, 285 (2), S. 858–864.

Song & Yunhui 2019

Song, K. & Yunhui, Y. (2019), „Semi-supervised Defect Classification of Steel Surface Based on Multi-training and Generative Adversarial Network“, *Optics and Lasers in Engineering*, 122 (1), S. 294–302.

Soukup & Huber-Mörk 2014

Soukup, D. & Huber-Mörk, R. (2014), „Convolutional Neural Networks for Steel Surface Defect Detection from Photometric Stereo Images“. *Advances in Visual Computing. 10th International Symposium*. Las Vegas, NV, USA, Hrsg. G. Bebis, Springer International Publishing, Cham, S. 668–677. ISBN: 9783319142494.

Spohrer 2019

Spohrer, A. (2019), *Steigerung der Ressourceneffizienz und Verfügbarkeit von Kugelgewindetrieben durch adaptive Schmierung*. Dissertation, Karlsruher Institut für Technologie, Shaker, Düren. ISBN: 9783844067989.

Stamboliska, Rusiński & Moczko 2015

Stamboliska, Z.; Rusiński, E. & Moczko, P. (2015), *Proactive Condition Monitoring of Low-Speed Machines*, Springer, Cham. ISBN: 9783319104942.

Stockinger 2011

Stockinger, M. (2011), *Untersuchung von Methoden zur Zustandsüberwachung von Werkzeugmaschinenachsen mit Kugelgewindtrieb*. Dissertation, Universität Erlangen-Nürnberg, Shaker, Düren. ISBN: 9783844000535.

Szegedy & Liu et al. 2015

Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V. & Rabinovich, A. (2015), „Going deeper with convolutions“. 2015

IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA, Hrsg. IEEE, IEEE, Piscataway, NJ, S. 1–9. ISBN: 9781467369640.

Tao & Zhang et al. 2018

Tao, X.; Zhang, D.; Ma, W.; Liu, X. & Xu, D. (2018), „Automatic Metallic Surface Defect Detection and Recognition with Convolutional Neural Networks“, *Applied Sciences*, 8 (9), S. 1575.

Toh & Park 2020

Toh, G. & Park, J. (2020), „Review of Vibration-Based Structural Health Monitoring Using Deep Learning“, *Applied Sciences*, 10 (5), S. 1680.

Touret & Changenet et al. 2018

Touret, T.; Changenet, C.; Ville, F.; Lalmi, M. & Becquerelle, S. (2018), „On the use of temperature for online condition monitoring of geared systems – A review“, *Mechanical Systems and Signal Processing*, 101 (1), S. 197–210.

Triyono & Prasetyo et al. 2021

Triyono, B.; Prasetyo, Y.; Winarno, B. & Wicaksono, H. H. (2021), „Electrical Motor Interference Monitoring Based On Current Characteristics“, *Journal of Physics: Conference Series*, 1845 (1), S. 12044.

Valenti & Squartini et al. 2017

Valenti, M.; Squartini, S.; Diment, A.; Parascandolo, G. & Virtanen, T. (2017), „A convolutional neural network approach for acoustic scene classification“. *IJCNN 2017. The International Joint Conference on Neural Networks*. Anchorage, AK, USA, Hrsg. IEEE, IEEE, Piscataway, NJ, S. 1547–1554. ISBN: 9781509061822.

van Truong & Chi Hieu et al. 2021

van Truong, H.; Chi Hieu, N.; Ngoc Giao, P. & Xuan Phong, N. (2021), „Unsupervised Detection of Anomalous Sound for Machine Condition Monitoring using Fully Connected U-Net“, *Journal of ICT Research and Applications*, 15 (1), S. 41–55.

Vanderwalde & Hillberry 2008

Vanderwalde, K. & Hillberry, B. (2008), „Characterization of pitting damage and prediction of remaining fatigue life“, *International Journal of Fatigue*, 30 (1), S. 106–118.

Vanoverberghe 2015

Vanoverberghe, M. (2015), *Vision based wear monitoring of complex surfaces*.

Masterthesis, University of Gent, Gent, Department of Mechanical Construction and Production.

Veith & Zimmermann et al. 2020

Veith, M.; Zimmermann, A.; Hillenbrand, J. & Fleischer, J. (2020), „Detektion des Vorspannungsverlusts in Kugelgewindetrieben“, *wt Werkstattstechnik online*, 110 (07-08), S. 485–490.

Verein Deutscher Ingenieure 2017

Verein Deutscher Ingenieure (2017), *Schadensanalyse - Schäden an Metallprodukten durch tribologische Beanspruchungen*, VDI 3822 (2017-06-00), Beuth Verlag GmbH, Berlin.

Verein Deutscher Ingenieure & Verband der Elektrotechnik Elektronik Informations-
technik 2020

Verein Deutscher Ingenieure & Verband der Elektrotechnik Elektronik Informati-
onstechnik (2020), *Entwicklung cyber-physischer mechatronischer Systeme (CPMS)*, VDI/VDE 2206 (2020-09-00), Beuth Verlag GmbH, Berlin.

Verl & Frey 2010

Verl, A. & Frey, S. (2010), „Correlation between feed velocity and preloading in ball screw drives“, *CIRP Annals*, 59 (1), S. 429–432.

Verl & Heisel et al. 2009

Verl, A.; Heisel, U.; Walther, M. & Maier, D. (2009), „Sensorless automated condition monitoring for the control of the predictive maintenance of machine tools“, *CIRP Annals*, 58 (1), S. 375–378.

Walther 2011

Walther, M. (2011), *Antriebsbasierte Zustandsdiagnose von Vorschubantrieben*. Dissertation, Universität Stuttgart, Jost-Jetter, Heimsheim. ISBN: 9783939890775.

Wang & Bovik et al. 2004

Wang, Z.; Bovik, A. C.; Sheikh, H. R. & Simoncelli, E. P. (2004), „Image quality assessment: from error visibility to structural similarity“, *IEEE transactions on image processing*, 13 (4), S. 600–612.

Wang & Han et al. 2019

Wang, T.; Han, Q.; Chu, F. & Feng, Z. (2019), „Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review“, *Mechanical Systems and Signal Processing*, 126 (1), S. 662–685.

Wang, Michau & Fink 2019

Wang, Q.; Michau, G. & Fink, O. (2019), „Domain Adaptive Transfer Learning for Fault Diagnosis“. *2019 Prognostics and System Health Management Conference*. Paris, France, Hrsg. C. Li, IEEE, Piscataway, NJ, S. 279–285. ISBN: 9781728103297.

Wang & Wang 2018

Wang, L. & Wang, X. V. (2018), *Cloud-based cyber-physical systems in manufacturing*, Springer, Cham. ISBN: 9783319676920.

Wieser 2009

Wieser, J. (2009), *Intelligente Instandhaltung zur Verfügbarkeitssteigerung von Werkzeugmaschinen*. Dissertation, Universität Karlsruhe, Shaker, Düren. ISBN: 9783832278823.

Xi & Kehne et al. 2020

Xi, T.; Kehne, S.; Fujita, T.; Epple, A. & Brecher, C. (2020), „Condition Monitoring of Ball-Screw Drives Based on Frequency Shift“, *IEEE/ASME Transactions on Mechatronics*, 25 (3), S. 1211–1219.

Xiao & Jiang et al. 2017

Xiao, M.; Jiang, M.; Li, G.; Xie, L. & Yi, L. (2017), „An evolutionary classifier for steel surface defects with small sample set“, *EURASIP Journal on Image and Video Processing*, 48 (1), S. 1–13.

Yagmur 2014

Yagmur, T. (2014), *Analyse, Verbesserung und Beschreibung des Verschleißverhaltens von Kugelgewindetrieben für Werkzeugmaschinen. Ergebnisse aus der Produktionstechnik*. Dissertation, RWTH Aachen, Apprimus Verlag, Aachen. ISBN: 9783863591915.

Yang & Li et al. 2020

Yang, Q.; Li, X.; Wang, Y.; Ainapure, A. & Lee, J. (2020), „Fault Diagnosis of Ball

Screw in Industrial Robots Using Non-Stationary Motor Current Signals“, *Procedia Manufacturing*, 48 (1), S. 1102–1108.

Youkachen & Ruchanurucks et al. 2018

Youkachen, S.; Ruchanurucks, M.; Phatrapomnant, T. & Kaneko, H. (2018), „Defect Segmentation of Hot-rolled Steel Strip Surface by using Convolutional Auto-Encoder and Conventional Image processing“. *2018 International Conference on Embedded Systems and Intelligent Technology*. Bangkok, Thailand, Hrsg. IEEE, IEEE, Piscataway, NJ, S. 1–5. ISBN: 9781538670637.

Yun & Kim et al. 2017

Yun, J. P.; Kim, D.; Kim, K.; Lee, S. J.; Park, C. H. & Kim, S. W. (2017), „Vision-based surface defect inspection for thick steel plates“, *Optical Engineering*, 56 (5), S. 53108.

Zadeh, Hosseini & Sra 2018

Zadeh, P. H.; Hosseini, R. & Sra, S. (2018), *Deep-RBF Networks Revisited: Robust Classification with Rejection*. Arxiv Preprint. <https://arxiv.org/pdf/1812.03190>.

Zamber & Hillberry 1999

Zamber, J. E. & Hillberry, B. M. (1999), „Probabilistic Approach to Predicting Fatigue Lives of Corroded 2024-T3“, *AIAA Journal*, 37 (10), S. 1311–1317.

Zhang & Liu et al. 2015

Zhang, Y.; Liu, S.; Tu, W.; Yu, H.; Li, C. & Li, C. (2015), „Using computer vision and compressed sensing for wood plate surface detection“, *Optical Engineering*, 54 (10), S. 103102.

Zhang & Lu 2004

Zhang, D. & Lu, G. (2004), „Review of shape representation and description techniques“, *Pattern Recognition*, 37 (1), S. 1–19.

Zheng, Kong & Nahavandi 2002

Zheng, H.; Kong, L. & Nahavandi, S. (2002), „Automatic inspection of metallic surface defects using genetic algorithms“, *Journal of Materials Processing Technology*, 125-126 (1), S. 427–433.

Zuiderveld 2000

Zuiderveld, K. (2000), „Contrast Limited Adaptive Histogram Equalization“ in *Graphics gems*, Hrsg. P. S. Heckbert, Academic Press Inc, Cambridge, Massachusetts, USA, S. 474–484.

Abbildungsverzeichnis

Abbildung 1-1 Aufbau der Dissertation.	5
Abbildung 2-1 Technische Funktion des Tribosystems nach (Sauer & Steinhilper 2012).	8
Abbildung 2-2 Verschleißerscheinungsformen der Adhäsion nach (Czichos & Habig 2020) und (Macherauch & Zoch 2011).	10
Abbildung 2-3 Qualitative Wechselwirkungen zwischen technischen Oberflächen und dem Schmieren nach (Achar 1980).	11
Abbildung 2-4 Verschleißerscheinungsformen des Verschleißmechanismus der Oberflächenzerrüttung nach (Czichos & Habig 2020) und (Sommer, Heinz & Schöfer 2018).	12
Abbildung 2-5 Verschleißerscheinungsformen des abrasiven Verschleiß nach (Macherauch & Zoch 2011) und (Czichos & Habig 2020).	12
Abbildung 2-6 Verschleißerscheinungsform der tribochemischen Reaktion nach (Macherauch & Zoch 2011) und (Czichos & Habig 2020).	13
Abbildung 2-7 Komponenten des tribologischen Systems KGT nach (Neff Gewindetrieb GmbH 2018).	14
Abbildung 2-8 Verschleißmechanismen und Verschleißerscheinungsformen am Kugel-Laufbahn-Kontakt des KGT nach (Czichos & Habig 2020), (Münzing 2017), (NSK 2020) und (Sommer, Heinz & Schöfer 2018).	17
Abbildung 2-9 Sich gegenseitig beeinflussende Verschleißmechanismen. Oben links: Innenring eines Kugellagers (Schaeffler 2000), oben rechts: Pendelrollenlager (Sommer, Heinz & Schöfer 2018), unten: Kugelgewindetriebspindel (Münzing 2017).	18
Abbildung 2-10 Qualitativer Prozess der sich gegenseitig beeinflussenden Verschleißmechanismen angelehnt an (Forstmann 2010).	20
Abbildung 2-11 Badewannenkurve nach (Heise 2002).	21
Abbildung 2-12 Abnutzungsvorrat und Abnutzungsgrenze (A_Hofmann 2021).	24
Abbildung 2-13 Qualitative Definition des Abnutzungsvorrates für verschiedene Anwendungen nach (A_Hofmann 2021).	29

Abbildung 2-14 Prozess zur Durchführung industrieller Zustandsüberwachung nach (Deutsches Institut für Normung 2018c).	33
Abbildung 2-15 Qualitativer Ablauf der Bildverarbeitung anhand von extrahierten Bildmerkmalen nach (Richter 2018).	34
Abbildung 2-16 Qualitative Abbildung eines künstlichen Neuronalen Netzwerkes nach (Zhou 2019).	41
Abbildung 2-17 Klassifikation auf Basis von Pixelwerten oder vorverarbeiteten Merkmalen.	42
Abbildung 2-18 Kernelemente eines einzelnen Perzeptrons.	43
Abbildung 2-19 Bestandteile eines CNN.	47
Abbildung 2-20 Grundelemente eines Convolutional Neural Network.	48
Abbildung 2-21 Funktionsweise der Kernel-Operation. Die Eingangswerte wurden hier auf Werte zwischen 0 und 1 normalisiert. Die farbig hervorgehobenen Zahlen in den Feature-Maps werden aus den gleichfarbigen Bereichen des Inputs durch Multiplikation der Elemente und anschließender Addition berechnet. Zwei Feature-Maps resultieren, da zwei Filter (lila und grün) über den Input gelegt werden.	50
Abbildung 2-22 Anwendung einer nichtlinearen Funktion auf die Elemente der Feature-Maps.	51
Abbildung 2-23 Qualitative Darstellung der Kombination von extrahierten Merkmalen zu zunehmend komplexeren Merkmalen.	52
Abbildung 2-24 Filterelemente, welche die Input-Zellen mit den Feature-Map-Zellen verbinden. Die Farbe der Filterzelle verbindet die Elemente aus Input- und Feature-Map. Zahlen ordnen die Input-Zellen den Feature-Map-Zellen zu.	54
Abbildung 3-1 Dreiphasige Verschleißkurve nach (Schopp, 2009).	69
Abbildung 4-1 Vorgehen zur Zielerfüllung. Blau: Markante Inhalte und Ergebnisse der Kapitel.	79
Abbildung 5-1 Hauptkomponenten des 5-Achs-Versuchsstandes angelehnt an (Hillenbrand, Spohrer & Fleischer 2018).	82

Abbildung 5-2 Links: Metallpartikel auf der Kugelgewindetriebspindel (rot), rechts: defekte Abstreifsysteme bei Versuchsende.	84
Abbildung 5-3 Einflussgrößen auf den Bilddatensatz.	85
Abbildung 5-4 Komponenten des Sensorsystems nach (Schlagenhauf, Ruppelt & Fleischer 2020).	88
Abbildung 5-5 Versuchsergebnisse unterschiedlicher Diffusor-Geometrien nach (A_Schlund 2019).	90
Abbildung 5-6 Sensorsystem im eingebauten Zustand.	91
Abbildung 5-7 Integrierte Sensorsysteme am 5-Achs-Versuchsstand.	92
Abbildung 5-8 Beispielhafte Aufnahmen des Kamerasystems. Links: Oberfläche mit Schädigungen; rechts: nicht geschädigte Oberfläche mit betriebsbedingten Verunreinigungen.	93
Abbildung 5-9 Absicherung der Eigenschaften des Sensorsystems.	94
Abbildung 5-10 Verschleißfortschritt eines Spindelabschnittes. Oben: Spindelabschnitt zu Versuchsbeginn und Versuchsende; unten: Entwicklung einer Schädigung.	95
Abbildung 5-11 Beispiele der aufgenommenen Bilddaten. a) Keine Schädigung, keine Verschmutzung, b) Geringe Schädigung, keine Verschmutzung, c) Geringe Schädigung, Verschmutzung, d) Keine Schädigung, Verschmutzung, e) Große Schädigung, keine Verschmutzung, f) Große Schädigung, Verschmutzung. (Schlagenhauf & Landwehr 2021).	96
Abbildung 6-1 Sich überlagernde Verschleißmechanismen auf der Kugelgewindetriebspindel. Oben: Adhäsion; unten: Abrasion.	98
Abbildung 6-2 Späne im Schmierstoff des KGTs als Indikator für vorliegenden Verschleiß im System.	100
Abbildung 6-3 Fingertest mit kleinen Metallpartikeln.	100
Abbildung 6-4 Korrelation des Schmierstoffzustandes mit dem Verschleißzustand des KGTs. Markierungen stellen Verschleißpartikel dar.	101
Abbildung 6-5 Entwicklung der Schmierstofffärbung über die Lebensdauer der KGT hinweg.	101

Abbildung 6-6 Oben: Pitting ausgehend von adhäsiver Vorschädigung der Oberfläche durch Materialeinschluss (Sommer, Heinz & Schöfer 2018); unten: Ausbrechen von Materialpartikeln (Verein Deutscher Ingenieure 2017).	103
Abbildung 6-7 Mechanismus der Entstehung eines Pittings. Links: Ausgehend von der Oberfläche; rechts: Ausgehend von unterhalb der Oberfläche nach (Czichos & Habig 2015).	104
Abbildung 6-8 Beispiel für die Entstehung eines Pitting ausgehend von einem eingedrückten Abrasivpartikel in der Oberfläche nach (A_Hofmann 2021).	105
Abbildung 6-9 Beispiele von Vorschädigungen auf der Spindeloberfläche, aus denen auch bei fortgeschrittener Lebensdauer keine Pittings entstanden sind.	106
Abbildung 6-10 Beispiel für die Entstehung eines Pittings ausgehend von unterhalb der Oberfläche nach (A_Hofmann 2021).	107
Abbildung 6-11 Beispiele für die Entwicklung unterschiedlicher Pittings (P1–P7) über die Lebensdauer von Komponenten hinweg nach (Schlagenhauf & Scheurenbrand et al. 2022). Erste Spalte: Bilder zu Versuchsbeginn; letzte Spalte: Bilder bei Lebensdauerende. Dazwischen wurden Beispiele gleichmäßig aus den Bilddaten gezogen. P2, P3 und P5 zeigen Partikel auf der Spindel, bevor ein Pitting entsteht.	109
Abbildung 6-12 Anzahl an Pittings auf den Oberflächen der Spindeln über die Lebensdauer hinweg.	111
Abbildung 6-13 Entwicklung der Gesamtfläche der Pittings auf den Spindeloberflächen über die Lebensdauer der Komponenten.	112
Abbildung 6-14 Entwicklung der Einzelflächen von Pittings auf den Spindeloberflächen.	113
Abbildung 6-15 Tangentiale Gesamtausdehnung der Schädigungen auf den Komponenten.	114
Abbildung 6-16 Axiale Gesamtausdehnung der Schädigungen auf den Komponenten.	115
Abbildung 6-17 Entwicklung der tangentialen Ausdehnungen von Pittings auf den Spindeloberflächen.	116

Abbildung 6-18 Entwicklung der axialen Ausdehnungen von Pittings auf den Spindeloberflächen.	116
Abbildung 6-19 Qualitative Darstellung des wirkenden Drucks auf die nicht geschädigten Spindelbereiche.	117
Abbildung 6-20 Gegenüberstellung der Fläche von vier zufällig ausgewählten Pittings mit deren Annäherung als Ellipse. Pitting n^* ist hierbei durch die Formel $TangentialPitting2 * AxialPitting2 * \pi$ angenähert.	118
Abbildung 6-21 Vorgeschlagenes Ausfallkriterium für einen KGT angelehnt an das Ausfallkriterium einer Linearschienenführung nach (Deutsches Institut für Normung 2010a, S. 13).	122
Abbildung 6-22 Qualitative Darstellung der Ursache-Wirkungs-Kette bei der Messung von Oberflächenzerrüttung am KGT.	124
Abbildung 6-23 Qualitative Abbildung der Bildmerkmale am Beispiel des Pittings.	129
Abbildung 6-24 Markante Zustände der Laufbahn des KGTs (A_Hofmann 2021).	129
Abbildung 6-25 Farbhistogramme der Bildklassen.	138
Abbildung 6-26 Vorgehen zur Extraktion von Farbmerkmalen.	139
Abbildung 6-27 Grey-Level-Co-Occurrence-Matrix nach (A_Heinzler 2020) und (Haralick 1979).	140
Abbildung 6-28 Ansatz zur Extraktion der verwendeten Locally-Binary-Patterns und Codierung derselben dargestellt in einem Histogramm.	141
Abbildung 6-29 Verteilung der SIFT-Key-Points in Subregionen nach (Schlagenhauf, Heinzler & Fleischer 2020). Links: kein Defekt; Mitte: Defekt an rechter Flanke; rechts: Verschmutzung.	142
Abbildung 6-30 Validierungs-Accuracy des Neuronalen Netzes, trainiert mittels Merkmalen, die auf Basis von Domänenwissen extrahiert wurden. Siehe Abbildung 0-9 im Anhang für Vergleichsdaten.	144
Abbildung 6-31 Struktur eines Entscheidungsbaumes.	146

Abbildung 6-32 Klassifikationsgenauigkeit des Entscheidungsbaumes im Vergleich zum Neuronalen Netz.	147
Abbildung 6-33 Merkmale zur Klassifikation von Bilddaten.	148
Abbildung 6-34 Normierter Anteil der Bildmerkmalskategorien an der Gesamtvarianz.	148
Abbildung 6-35 Entscheidungspfade für charakteristische Bilddaten.	150
Abbildung 6-36 Gegenüberstellung der Validierungs-Accuracy von verschiedenen Stand-der-Technik-Architekturen.	153
Abbildung 6-37 Boxplots der trainierten CNN-Architekturen.	155
Abbildung 6-38 Validierungs-Accuracy bei 25-facher Cross Validation.	158
Abbildung 6-39 Anwendung des CLAHE-Algorithmus zur Kontrastanpassung auf Bilddaten. Erste Zeile links Darstellung der menschlichen Retina, sonst KGT-Bilddaten. Links jeweils Originalbild, rechts CLAHE-transformiertes Bild. Bildpaare sind getrennt durch weiße Linien. Abbildung der Retina nach (Bhat & Patil 2014).	161
Abbildung 6-40 Klassifikationsergebnisse des Versuchsplanbasierten CNN unter Verwendung von Data Augmentation.	162
Abbildung 6-41 Darstellung der Relevanz der Bildbereiche zur Klassifikation der Bilder mittels eines Heat-Map-Ansatzes.	164
Abbildung 6-42 Grundarchitektur des SBF-Net-Ansatzes. Das Centerbild stellt dabei ein Bild dar, von dem die Klasse bekannt ist.	168
Abbildung 6-43 Grundarchitektur eines RBF-Networks und qualitative Übertragung auf die Klassifikation von Pittings.	169
Abbildung 6-44 Erweiterung des RBF-Network-Ansatzes zum SBF-Net-Ansatz bestehend aus Deep-Learning-Siamese-Kernel, Bild-Prototypen sowie einem Neuronalen Netzwerk zur Klassifikation. Das grün umrandete Element entspricht dem vorderen Teil aus Abbildung 6-42.	171
Abbildung 6-45 Die Bilder $P1$ und R enthalten dieselben Objekte und können damit als semantisch identisch angesehen werden. Die Bilder $P1$ und $P2$ teilen hingegen keine gemeinsamen Klassen. Die euklidische Distanz der Pixel	

zwischen den Bildern $P1$ und R ist größer als die euklidische Distanz zwischen den Bildern $P1$ und $P2$. R ist gesättigte Version von $P3$.	173
Abbildung 6-46 Siamese Network als Grundarchitektur der Siamese-Kernel. Die Distanz von Bildern identischer Klasse soll möglichst gering und jene von Bildern unterschiedlicher Klasse möglichst hoch sein.	174
Abbildung 6-47 Architektur des Neuronalen Netzes zur Klassifikation der Bilddaten mit der Rectified-Linear-Unit-Aktivierungsfunktion (Relu-Aktivierungsfunktion) und vier Schichten á 50 Neuronen nach (Schlagenhauf, Yildirim & Brückner 2022).	177
Abbildung 6-48 Beispielbilder aus den verwendeten Datensätzen.	178
Abbildung 6-49 Modellergebnisse des SBF-Nets (KGT-Datensatz) nach (Schlagenhauf, Yildirim & Brückner 2022) bei Variation der Anzahl an Center-Bildern pro Klasse sowie der Datenmenge zum Training. Die x-Achsen-Beschriftung folgt der Struktur [Datenmenge in % – Anzahl Center pro Klasse].	180
Abbildung 6-50 Validierungs-Accuracy des SBF-Net bei unterschiedlichen Anzahlen an Trainingsdaten im Vergleich zu den Stand-der-Technik-Modellen nach (Schlagenhauf, Yildirim & Brückner 2022).	183
Abbildung 6-51 Vergleich der durchschnittlichen Validierungs-Accuracy über alle Datensätze bei gegebener Anzahl an Trainingsdatenpunkten nach (Schlagenhauf, Yildirim & Brückner 2022).	185
Abbildung 6-52 Vergleich der Leistungsfähigkeit des SBF-Net-Ansatzes gegen internationalen Stand der Forschung nach (Schlagenhauf, Yildirim & Brückner 2022) und (Song & Yunhui 2019) anhand des NEU-Datensatzes.	186
Abbildung 6-53 Handlungsempfehlung zum Einsatz von Maschinellem Lernen unter Verwendung von klassischen merkmalsbasierten Ansätzen und Deep-Learning-basierten Ansätzen.	189
Abbildung 7-1 Prozess zur Prognose der Größenentwicklung von Fehlstellen nach (Schlagenhauf & Burghardt 2021).	192
Abbildung 7-2 Validierung des Modelles zur Detektion von Fehlstellen nach (Schlagenhauf & Landwehr 2021).	194

Abbildung 7-3 Prozess zur Auswahl des optimalen Schwellwertes nach (Schlagenhauf & Burghardt 2021). Optimaler Schwellwert ist grün umrandet.	196
Abbildung 7-4 Initialer Vergleich eines klassischen Ansatzes zur Kantenextraktion mit dem verwendeten Metamodell zur Vorhersage der optimalen Schwellwerte nach (Schlagenhauf & Burghardt 2021).	196
Abbildung 7-5 Anpassung der Fehlerfläche mittels eines gewichteten Abstandsmaßes.	199
Abbildung 7-6 Vergleich der vom System berechneten Fehlergröße mit Expertensystem (Expert System) und ohne Expertensystem (Quantification) im Vergleich zu den Realdaten (Groundtruth) nach (Schlagenhauf & Burghardt 2021).	200
Abbildung 7-7 Qualitative Darstellung der Gewichtungsfunktion.	202
Abbildung 7-8 Prognose einer Schädigung, zeilenweise von links nach rechts mit zunehmender Anzahl an Trainingsdaten nach (Schlagenhauf & Burghardt 2021).	204
Abbildung 7-9 Validierung des Prognosesystems zusammen mit Ausschnitten des Quantifizierungsmodules, angelehnt an (Schlagenhauf & Burghardt 2021). Die Zahlen in den grau eingefärbten Kreisen stellen die für das Training des Prognosemodelles verwendeten Datenpunkte dar.	206
Abbildung 7-10 Abnutzungsvorrat (synonym verwendet zu Remaining Useful Lifetime [RUL]) bei Erreichen der 100 % Lebensdauer nach Definition durch die Prognosefunktionen. Das Konfidenzintervall erlaubt einen sichereren Betrieb der Komponenten.	207
Abbildung 7-11 Abweichung der Prognose zu den Realdaten bei Überschreiten der definierten 100%-Lebensdauermarke.	209
Abbildung 0-1 Backpropagation des Fehlers durch das Neuronale Netzwerk.	XLVI
Abbildung 0-2 Kategorien von sichtbaren Defekten (Czimmermann & Ciuti et al. 2020, S. 3).	L
Abbildung 0-3 Ausschnitt aus der Menge der Bilddaten ohne Pitting.	LI
Abbildung 0-4 Ausschnitt aus der Menge der Bilddaten mit Pitting.	LII
Abbildung 0-5 Technische Zeichnung des verwendeten Kugelgewindetriebes.	LIII

Abbildung 0-6 Beispiele von Pitting-Verläufen, zusammengefasst dargestellt in Abbildung 6-12. Lebensdauer zeilenweise zunehmend. Erstes Bild entspricht 0 % Lebensdauer. Zwischen dem ersten und zweiten Bild sind Bilder zwecks der Übersichtlichkeit nicht dargestellt, da keine Veränderung im Bild zu erkennen ist. Zweites bis letztes Bild zeigen die vollständige Entwicklung des jeweiligen Pittings.

LVI

Abbildung 0-7 Entwicklung der Validierungsgenauigkeit in Abhängigkeit der gewählten Modellparameter.

LVII

Abbildung 0-8 Ergebnisse der Hyperparameterstudie der ausgewählten Standard-Technik-Modelle.

LVIII

Abbildung 0-9 Ausgewählte Testbilder zum Test der Klassifikationsfähigkeit. Grün: Bereiche ohne Pitting, rot: Bereiche mit Pitting.

LIX

Tabellenverzeichnis

Tabelle 5-1 Zuordnung von Anforderungen zu Funktionsträgern.	87
Tabelle 6-1 Erreichte Lebensdauer der KGT, normiert auf die L_{10} Lebensdauer.	109
Tabelle 6-2 Gegenüberstellung der Spezifikationen der bisherigen Arbeiten und des bildbasierten Ansatzes.	127
Tabelle 6-3 Zuordnung der visuellen Merkmale zu Bildmerkmalen.	132
Tabelle 6-4 DoE-Parameterraum für die Erstellung von CNN-Architekturen.	159
Tabelle 6-5 Domänenspezifische Varianz in den Bilddaten und angewandte Techniken zur Data Augmentation.	160
Tabelle 6-6 Gegenüberstellung der erreichten Validierungs-Accuracy des SBF-Nets für die TEX-, MNIST- und cifar10-Datensätze im Vergleich zu Stand-der-Technik-Ansätzen nach (Schlagenhauf, Yildirim & Brückner 2022). Ungefärbte Zellen zeigen durchgeführte Versuche. Gefärbte Zellen zeigen den Durchschnitt der durchgeführten Versuche. Die Farben Grün, Gelb und Rot ordnen die Ergebnisse nach deren Platzierung hinsichtlich der Leistungsfähigkeit.	184
Tabelle 7-1 Prognosegüte ausgewählter Funktionen zur Vorhersage der Schädigungsentwicklung.	203
Tabelle 7-2 Betriebsszenarien, vorherrschende Ausfallmechanismen und empfohlene Prinzipien zur Überwachung.	214
Tabelle 0-1 Haralick-Merkmale für zwei beispielhafte Pittings.	LIX
Tabelle 0-2 Zuordnung von definierten Merkmalen zu Algorithmen zur Merkmalsextraktion.	LX

Anhang

A1 Backpropagation im Neuronales Netzwerk

Abbildung 0-1 zeigt ein Perzeptron im vorfinalen Layer, das mit dem Output-Perzeptron verbunden ist. Da hier ebenfalls die Gewichte angepasst werden sollen, müssen auch für diese die partiellen Ableitungen des Fehlers berechnet werden. Am Beispiel des Gewichtes w_1^{n-1} ³⁸ ergibt sich damit: $\frac{\partial L}{\partial w_i} = \frac{\partial L}{\partial o} * \frac{\partial o}{\partial v} * \frac{\partial v}{\partial w_i} * \frac{\partial o^{n-1}}{\partial v^{n-1}} * \frac{\partial v^{n-1}}{\partial w^{n-1}}$. Hier ist analog zu oben $\frac{\partial o^{n-1}}{\partial v^{n-1}} = o^{n-1}(1 - o^{n-1})$ und $\frac{\partial v^{n-1}}{\partial w^{n-1}} = x_1^{n-1}$. Der Term $\frac{\partial L}{\partial o} * \frac{\partial o}{\partial v} * \frac{\partial v}{\partial w_i} = p$ wurde im Schritt zuvor bereits berechnet. Damit ergibt sich $\frac{\partial L}{\partial w_i} = p * o^{n-1}(1 - o^{n-1}) * x_1^{n-1}$ und die Anpassung des Gewichtes $w_i^{n-1\text{new}}$ mit: $w_i^{n-1\text{new}} = w_i^{n-1\text{alt}} - \alpha * p * o^{n-1}(1 - o^{n-1}) * x_1^{n-1}$. Somit kann der Fehler sukzessive durch das Netzwerk zurückpropagiert werden, indem die jeweiligen Ableitungen für die Gewichte in den tieferen Schichten gespeichert und dann lediglich für Gewichte in früheren Schichten des Neuronales Netzwerkes eingesetzt werden. Dieser Prozess wird so lange iterativ wiederholt, bis die Verlustfunktion minimiert oder eine maximale Anzahl an Iterationen erreicht wurde.

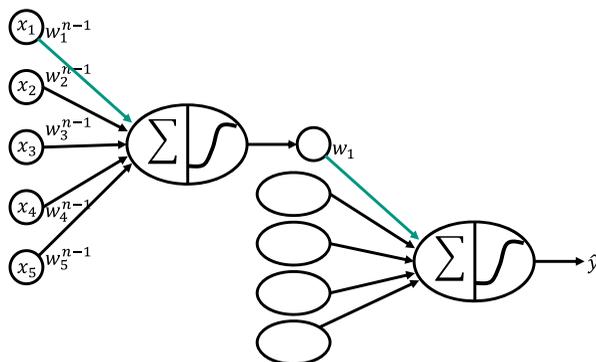


Abbildung 0-1 Backpropagation des Fehlers durch das Neuronales Netzwerk.

³⁸ Die Nomenklatur $(\cdot)^{n-1}$ wurde gewählt, um zu veranschaulichen, dass sich die Gewichte im vorletzten Layer des Netzwerkes befinden.

A2 Backpropagation in der Convolutional Base

Für das in Abbildung 2-24 gezeigte Beispiel berechnet sich die partielle Ableitung $\frac{\partial L}{\partial w_{0,0}}$ konkret wie folgt:

$$\begin{aligned}\frac{\partial L}{\partial w_{0,0}} &= \sum_{S_{i,j} \in S} \frac{\partial L}{\partial o_{i,j}} * \frac{\partial o_{i,j}}{\partial s_{i,j}} * \frac{\partial s_{i,j}}{\partial w_{0,0}} = \frac{\partial L}{\partial o_{0,0}} * 0 * I_{0,0} + \frac{\partial L}{\partial o_{0,1}} * 1 * I_{0,1} + \frac{\partial L}{\partial o_{1,0}} * 1 * I_{1,0} + \frac{\partial L}{\partial o_{1,1}} * 1 * \\ I_{1,1} &= \frac{\partial L}{\partial o_{0,1}} * 0.1 + \frac{\partial L}{\partial o_{1,0}} * 0.3 + \frac{\partial L}{\partial o_{1,1}} * 0.4.\end{aligned}$$

Die Anpassung der Gewichte erfolgt nun analog zum klassischen Neuronalen Netzwerk mit: $w_{0,0}^{neu} = w_{0,0}^{alt} - \alpha * \frac{\partial L}{\partial w_{0,0}^{alt}}$.

Für die Anpassung der Gewichte in tieferliegenden Schichten werden außerdem die partiellen Ableitungen nach den Input-Zellen $\frac{\partial L}{\partial I_{v,w}}$ gefordert, da diese für die Berechnung der partiellen Ableitungen nach den Gewichten erforderlich sind. Hierfür ist zunächst zu beachten, dass analog zu den Gewichten auch die Input-Zellen in unterschiedliche Feature-Map-Zellen eingehen. Die Menge der Feature-Map-Zellen, die durch die Input-Zelle $I_{v,w}$ beeinflusst werden, sei mit $S_c^{v,w}$ bezeichnet. Diese Mengen enthalten unterschiedlich viele Elemente. Beispielsweise ist $S_c^{0,0} = \{S_{0,0}\}$, da der Input in der ersten Zeile der ersten Spalte lediglich in die Feature-Map-Zelle in der ersten Zeile der ersten Spalte eingeht. Dahingegen ist $S_c^{1,1} = \{S_{0,0}, S_{0,1}, S_{1,0}, S_{1,1}\}$, da der Wert in alle Feature-Map-Zellen eingeht. Die Ableitung des Fehlers nach den Input-Zellen

$\frac{\partial L}{\partial I_{v,w}} = \frac{\partial L}{\partial o_{i,j}} * \frac{\partial o_{i,j}}{\partial s_{i,j}} * \frac{\partial s_{i,j}}{\partial I_{v,w}}$ ist hiermit wiederum eine Summe unterschiedlicher Länge:

$\frac{\partial L}{\partial I_{v,w}} = \sum_{S_{i,j} \in S_c^{v,w}} \frac{\partial L}{\partial o_{i,j}} * \frac{\partial o_{i,j}}{\partial s_{i,j}} * \frac{\partial s_{i,j}}{\partial I_{v,w}}$. Für die beiden Beispiele $\frac{\partial L}{\partial I_{0,0}}$ und $\frac{\partial L}{\partial I_{1,1}}$ ergibt sich damit:

$$\frac{\partial L}{\partial I_{0,0}} = \frac{\partial L}{\partial o_{0,0}} * \frac{\partial o_{0,0}}{\partial s_{0,0}} * \frac{\partial s_{0,0}}{\partial I_{0,0}} = \frac{\partial L}{\partial o_{0,0}} * 0 * w_{0,0} = 0$$

und

$$\begin{aligned}\frac{\partial L}{\partial I_{1,1}} &= \frac{\partial L}{\partial o_{0,0}} * \frac{\partial o_{0,0}}{\partial s_{0,0}} * \frac{\partial s_{0,0}}{\partial I_{1,1}} + \frac{\partial L}{\partial o_{0,1}} * \frac{\partial o_{0,1}}{\partial s_{0,1}} * \frac{\partial s_{0,1}}{\partial I_{1,1}} + \frac{\partial L}{\partial o_{1,0}} * \frac{\partial o_{1,0}}{\partial s_{1,0}} * \frac{\partial s_{1,0}}{\partial I_{1,1}} + \frac{\partial L}{\partial o_{1,1}} * \frac{\partial o_{1,1}}{\partial s_{1,1}} * \frac{\partial s_{1,1}}{\partial I_{1,1}} \\ &= \frac{\partial L}{\partial o_{0,0}} * 0 * w_{1,1} + \frac{\partial L}{\partial o_{0,1}} * 1 * w_{1,0} + \frac{\partial L}{\partial o_{1,0}} * 1 * w_{0,1} + \frac{\partial L}{\partial o_{1,1}} * 1 * w_{0,0} \\ &= \frac{\partial L}{\partial o_{0,0}} * 0 * 0 + \frac{\partial L}{\partial o_{0,1}} * 1 * 0 + \frac{\partial L}{\partial o_{1,0}} * 1 * 1 + \frac{\partial L}{\partial o_{1,1}} * 1 * -0.5 = \frac{\partial L}{\partial o_{1,0}} - 0.5 \frac{\partial L}{\partial o_{1,1}}.\end{aligned}$$

Dieser Prozess wird für alle Schichten iterativ durchgeführt, bis der Fehler der Verlustfunktion minimiert oder eine bestimmte Anzahl an Iterationen erreicht wurde.

A3 Ergänzende Materialien

		TYPE	DESCRIPTION	SAMPLE
VISIBLE DEFECTS	Line shape	<i>scratch</i>	<ul style="list-style-type: none"> line shaped defect on the surface (high length/width ratio) hard to detect with tactile sensation significantly longer than deeper and wider usually it occurs on often affected regions 	
		<i>scratches</i>	<ul style="list-style-type: none"> includes some scratch close to each other and heading to the same direction hard to detect with tactile sensation 	
	Patch/spot shape	<i>corrosion</i>	<ul style="list-style-type: none"> patchy shaped defect (low length/width ratio) change in the color dimension of the surface hard to detect with tactile sensation the material turns into its oxide, hydroxide, or sulfide form occurs due to the chemical/electrochemical reaction with the environment 	
		<i>contamination</i>	<ul style="list-style-type: none"> patchy shaped defect (low length/width ratio) change in the color dimension of the surface hard to detect with tactile sensation contains other material(s) than the object 	
		<i>galling</i>	<ul style="list-style-type: none"> patchy shaped defect (low length/width ratio) includes numerous scratches, heading to other directions hard to detect with tactile sensation 	
	Pattern-dependent	<i>shape error</i>	<ul style="list-style-type: none"> abnormality in the shape of the pattern not possible to detect with tactile sensation mostly occurs due to production error 	
		<i>color error</i>	<ul style="list-style-type: none"> abnormality in the color of the pattern hard to detect with tactile sensation mostly occurs due to production error 	

PALPABLE DEFECTS		Line shape		
		<i>crease</i>	<ul style="list-style-type: none"> line shaped defect on the surface (high length/width ratio) possible to detect with vision and tactile significantly longer than wider protusion from the surface mostly occurs due to friction an object on the surface significant height compared to the object's thickness 	
<i>crack tear</i>	<ul style="list-style-type: none"> line shaped defect on the surface (high length/width ratio) possible to detect with vision and tactile immediate transition between the defect and object (a gap with sharp edges) the defect is deep occurs due to fold-trust forces on the object 			
Patch/spot shape		<i>dimple</i>	<ul style="list-style-type: none"> patchy shaped defect (low length/width ratio) possible to detect with vision and tactile significant depth compared to the object's thickness smooth transition between the defect and the object (no edges) occurs due to distributed pressure on the surface 	
		<i>bump</i>	<ul style="list-style-type: none"> patchy shaped defect (low length/width ratio) possible to detect with vision and tactile significant height compared to the object's thickness smooth transition between the defect and the object (no edges) opposite direction as the dimple 	
		<i>pit</i>	<ul style="list-style-type: none"> patchy shaped defect (low length/width ratio) possible to detect with vision and tactile significant depth compared to the object's thickness immediate transition between the defect and object (a gap with sharp edges) occurs due to concentrated pressure on the surface 	
		<i>hole</i>	<ul style="list-style-type: none"> patchy shaped defect (low length/width ratio) possible to detect with vision and tactile the defect reaches through to the other side of the object immediate transition between the defect and object (a gap with sharp edges) occurs due to concentrated pressure on the surface 	

Abbildung 0-2 Kategorien von sichtbaren Defekten (Czimmermann & Ciuti et al. 2020, S. 3).



Abbildung 0-3 Ausschnitt aus der Menge der Bilddaten ohne Pitting.



Abbildung 0-4 Ausschnitt aus der Menge der Bilddaten mit Pitting.

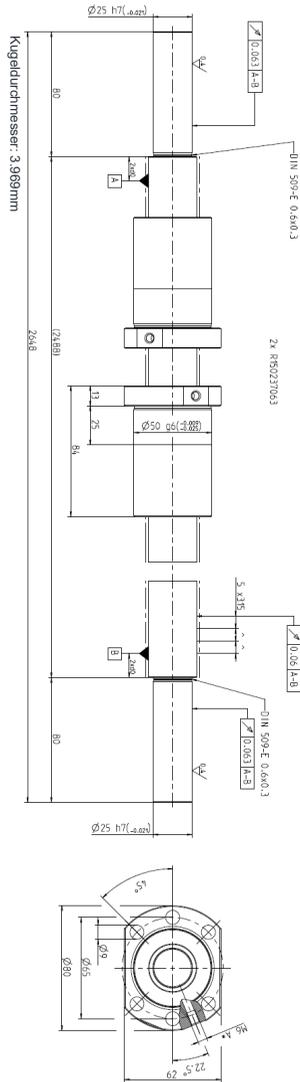
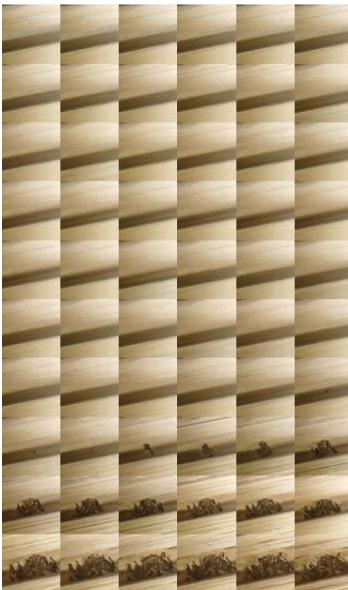
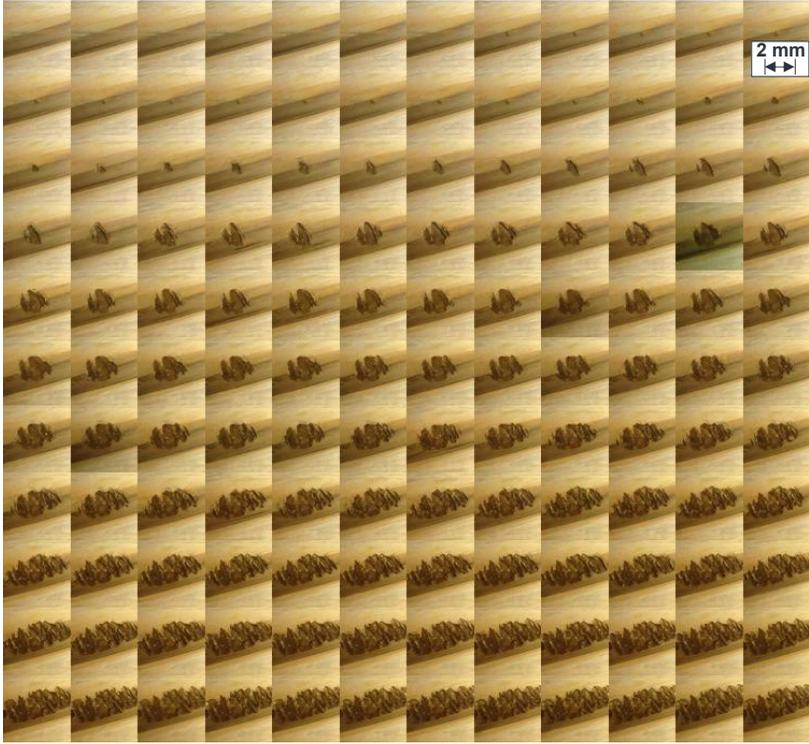


Abbildung 0-5 Technische Zeichnung des verwendeten Kugelgewindetriebes.







*Abbildung 0-6 Beispiele von Pitting-Verläufen, zusammengefasst dargestellt in
Abbildung 6-12. Lebensdauer zeilenweise zunehmend. Erstes Bild entspricht 0 %
Lebensdauer. Zwischen dem ersten und zweiten Bild sind Bilder zwecks der
Übersichtlichkeit nicht dargestellt, da keine Veränderung im Bild zu erkennen ist.
Zweites bis letztes Bild zeigen die vollständige Entwicklung des jeweiligen Pit-
tings.*

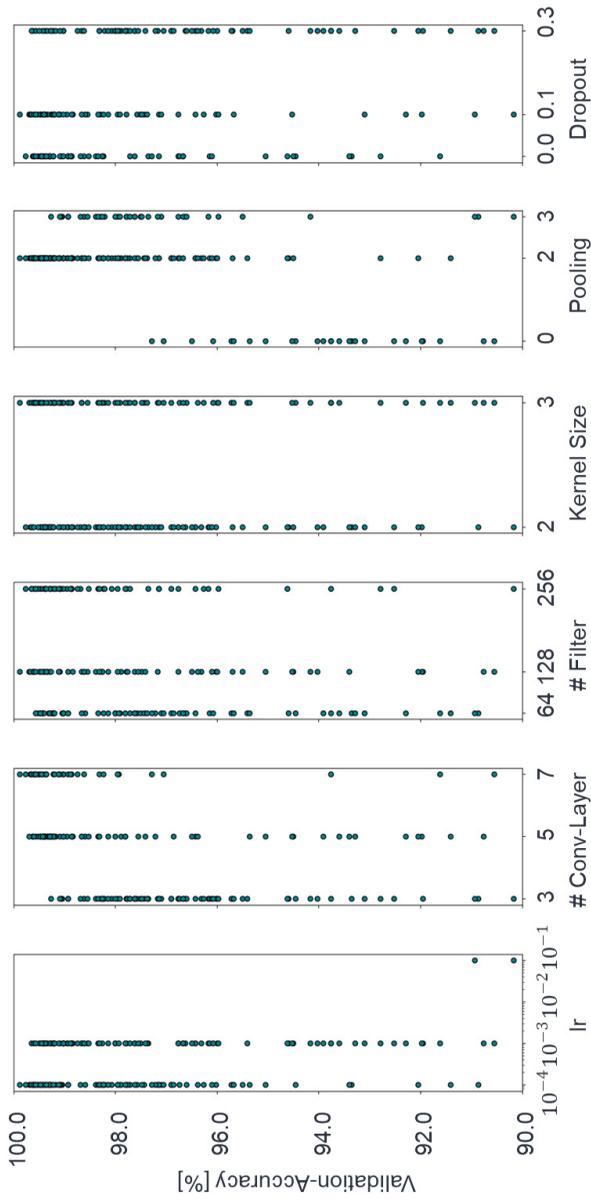


Abbildung 0-7 Entwicklung der Validierungsgenauigkeit in Abhängigkeit der gewählten Modellparameter.

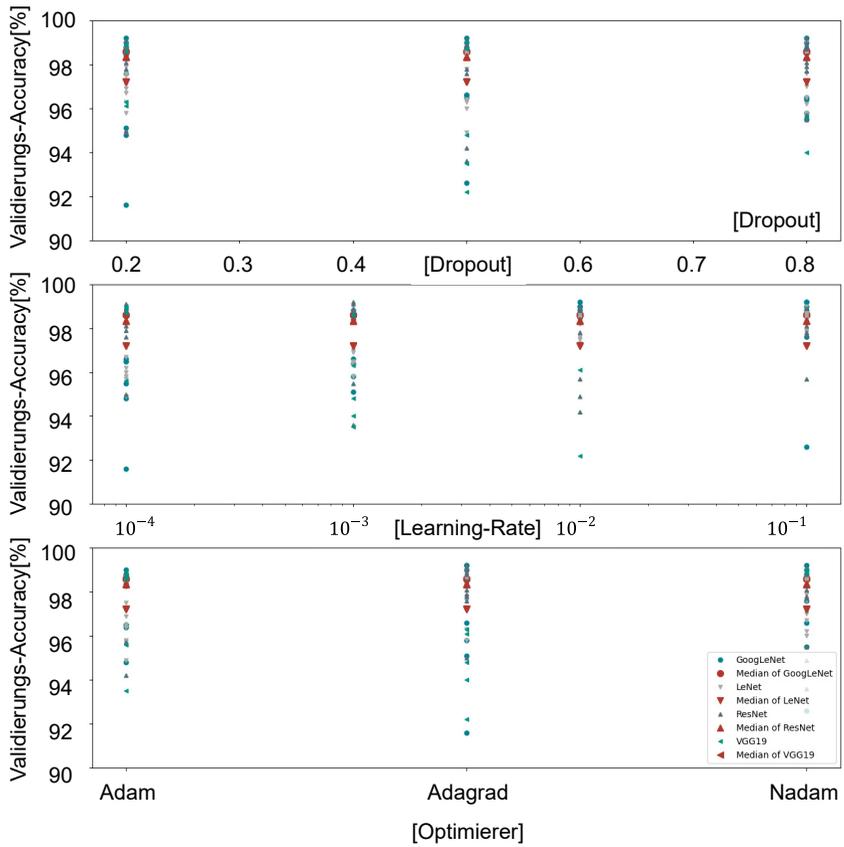


Abbildung 0-8 Ergebnisse der Hyperparameterstudie der ausgewählten Standard-Technik-Modelle.

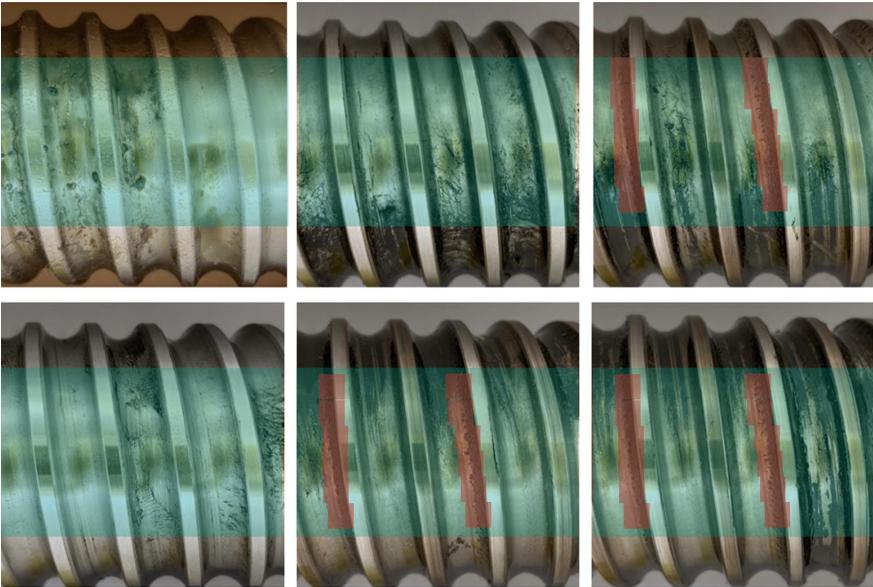


Abbildung 0-9 Ausgewählte Testbilder zum Test der Klassifikationsfähigkeit. Grün: Bereiche ohne Pitting, rot: Bereiche mit Pitting.

Tabelle 0-1 Haralick-Merkmale für zwei beispielhafte Pittings.

	CoO-Matrix	Angular Second Moment	Contrast	Correlation	Sum of Squares: Variance	Inverse Difference Moment	Sum Average	Sum Variance
Kein Pitting	1	0,0005	52,89	0,96	660,78	0,18	338,01	2590,22
	2	0,0005	55,89	0,96	659,98	0,17	338,32	2584,02
	3	0,0010	12,49	0,99	667,97	0,31	337,74	2659,37
	4	0,0005	60,66	0,95	660,02	0,17	338,32	2579,42
	∅	0,0006	45,4842	0,9656	662,1860	0,2064	338,0959	2603,2596
Pitting	1	0,0003	113,15	0,98	2301,80	0,17	262,48	9094,04
	2	0,0003	182,73	0,96	2296,01	0,15	262,57	9001,29
	3	0,0004	70,00	0,98	2304,59	0,25	261,91	9148,37
	4	0,0003	128,25	0,97	2295,96	0,16	262,57	9055,58
	∅	0,0003	123,5324	0,9731	2299,5879	0,1836	262,3825	9074,8191
	CoO-Matrix	Sum Entropy	Entropy	Difference Variance	Difference Entropy	Info. Measure of Correlation 1	Info. Measure of Correlation 2	

Kein Pitting	1	7,65	11,37	0,0003	3,95	-0,30	0,99
	2	7,65	11,41	0,0003	3,99	-0,29	0,99
	3	7,68	10,43	0,0006	2,97	-0,44	1,00
	4	7,65	11,46	0,0003	4,05	-0,29	0,99
	∅	7,6564	11,1651	0,0004	3,7386	-0,3297	0,9921
Pitting	1	8,44	12,24	0,0003	4,29	-0,36	1,00
	2	8,44	12,40	0,0002	4,54	-0,34	1,00
	3	8,46	11,76	0,0004	3,80	-0,43	1,00
	4	8,44	12,33	0,0003	4,38	-0,35	1,00
	∅	8,4442	12,1844	0,0003	4,2511	-0,3685	0,9977

Tabelle 0-2 Zuordnung von definierten Merkmalen zu Algorithmen zur Merkmalsextraktion.

Farbmerkmale			Texturmerkmale			Formmerkmale		
B ³⁹	M ⁴⁰	# ⁴¹	B	M	#	B	M	#
Farbanteil	Clustered Colour Share	20	Local Binary Patterns	Locally Binary Patterns	256	Key Points Per Sub Region	Key Points Per Sub Regions / SIFT-Feature-Extractor	16
Mittelwert		1	Mittelwert		1			
Mittelwert		1	Median		1			
Maximaler Mittelwert		1	Maximum		1			
Minimaler Mittelwert		1	Minimum		1			
Mittelwert Standardabweichung		1	Standardabweichung		1			
Standardabweichung Mittelwert		1	Haralick-Merkmale	Haralick / Gray-Level-Co-Occurrence-Matrix	13			

³⁹ B:= Bezeichnung

⁴⁰ M:= Methode

⁴¹ # := Anzahl

Standardabweichung maximaler Wert		1						
Standardabweichung minimaler Wert		1						

Forschungsberichte aus dem wbk
Institut für Produktionstechnik
Karlsruher Institut für Technologie (KIT)

Bisher erschienene Bände:

Band 0

Dr.-Ing. Wu Hong-qi

Adaptive Volumenstromregelung mit Hilfe von drehzahleregelten Elektroantrieben

Band 1

Dr.-Ing. Heinrich Weiß

**Fräsen mit Schneidkeramik - Verhalten des System
Werkzeugmaschine-Werkzeug-Werkstück und Prozessanalyse**

Band 2

Dr.-Ing. Hans-Jürgen Stierle

Entwicklung und Untersuchung hydrostatischer Lager für die Axialkolbenmaschine

Band 3

Dr.-Ing. Herbert Hörner

Untersuchung des Geräuschverhaltens druckgeregelter Axialkolbenpumpen

Band 4

Dr.-Ing. Rolf-Dieter Brückbauer

Digitale Drehzahlregelung unter der besonderen Berücksichtigung von Quantisierungseffekten

Band 5

Dr.-Ing. Gerhard Staiger

Graphisch interaktive NC-Programmierung von Drehteilen im Werkstattbereich

Band 6

Dr.-Ing. Karl Peters

Ein Beitrag zur Berechnung und Kompensation von Positionierfehlern an Industrierobotern

Band 7

Dr.-Ing. Paul Stauss

Automatisierte Inbetriebnahme und Sicherung der Zuverlässigkeit und Verfügbarkeit numerisch gesteuerter Fertigungseinrichtungen

Band 8

Dr.-Ing. Günter Möckesch

Konzeption und Realisierung eines strategischen, integrierten Gesamtplanungs- und -bearbeitungssystems zur Optimierung der Drehteilorganisation für auftragsbezogene Drehereien

Band 9

Dr.-Ing. Thomas Oestreicher

Rechnergestützte Projektierung von Steuerungen

Band 10

Dr.-Ing. Thomas Selinger

Teilautomatisierte werkstattnahe NC-Programmerstellung im Umfeld einer integrierten Informationsverarbeitung

Band 11

Dr.-Ing. Thomas Buchholz

Prozessmodell Fräsen, Rechnerunterstützte Analyse, Optimierung und Überwachung

Band 12

Dr.-Ing. Bernhard Reichling

Lasergestützte Positions- und Bahnvermessung von Industrierobotern

Band 13

Dr.-Ing. Hans-Jürgen Lesser

Rechnergestützte Methoden zur Auswahl anforderungsgerechter Verbindungselemente

Band 14

Dr.-Ing. Hans-Jürgen Lauffer

Einsatz von Prozessmodellen zur rechnerunterstützten Auslegung von Räumwerkzeugen

Band 15

Dr.-Ing. Michael C. Wilhelm

Rechnergestützte Prüfplanung im Informationsverbund moderner Produktionssysteme

Band 16

Dr.-Ing. Martin Ochs

Entwurf eines Programmsystems zur wissensbasierten Planung und Konfigurierung

Band 17

Dr.-Ing. Heinz-Joachim Schneider

Erhöhung der Verfügbarkeit von hochautomatisierten Produktionseinrichtungen mit Hilfe der Fertigungsleittechnik

Band 18

Dr.-Ing. Hans-Reiner Ludwig

Beanspruchungsanalyse der Werkzeugschneiden beim Stirnplanfräsen

Band 19

Dr.-Ing. Rudolf Wieser

Methoden zur rechnergestützten Konfigurierung von Fertigungsanlagen

Band 20

Dr.-Ing. Edgar Schmitt

Werkstattsteuerung bei wechselnder Auftragsstruktur

Band 21

Dr.-Ing. Wilhelm Enderle

Verfügbarkeitssteigerung automatisierter Montagesysteme durch selbsttätige Behebung prozessbedingter Störungen

Band 22

Dr.-Ing. Dieter Buchberger

Rechnergestützte Strukturplanung von Produktionssystemen

Band 23

Prof. Dr.-Ing. Jürgen Fleischer

Rechnerunterstützte Technologieplanung für die flexibel automatisierte Fertigung von Abkantteilen

Band 24

Dr.-Ing. Lukas Loeffler

Adaptierbare und adaptive Benutzerschnittstellen

Band 25

Dr.-Ing. Thomas Friedmann

Integration von Produktentwicklung und Montageplanung durch neue rechnergestützte Verfahren

Band 26

Dr.-Ing. Robert Zurrin

Variables Formhonen durch rechnergestützte Hornprozesssteuerung

Band 27

Dr.-Ing. Karl-Heinz Bergen

Langhub-Innenrundhonen von Grauguss und Stahl mit einem elektromechanischem Vorschubsystem

Band 28

Dr.-Ing. Andreas Liebisch

Einflüsse des Festwalzens auf die Eigenspannungsverteilung und die Dauerfestigkeit einsatzgehärteter Zahnräder

Band 29

Dr.-Ing. Rolf Ziegler

Auslegung und Optimierung schneller Servopumpen

Band 30

Dr.-Ing. Rainer Bartl

Datenmodellgestützte Wissensverarbeitung zur Diagnose und Informationsunterstützung in technischen Systemen

Band 31

Dr.-Ing. Ulrich Golz

Analyse, Modellbildung und Optimierung des Betriebsverhaltens von Kugelgewindetrieben

Band 32

Dr.-Ing. Stephan Timmermann

Automatisierung der Feinbearbeitung in der Fertigung von Hohlformwerkzeugen

Band 33

Dr.-Ing. Thomas Noe

Rechnergestützter Wissenserwerb zur Erstellung von Überwachungs- und Diagnoseexpertensystemen für hydraulische Anlagen

Band 34

Dr.-Ing. Ralf Lenschow

Rechnerintegrierte Erstellung und Verifikation von Steuerungsprogrammen als Komponente einer durchgängigen Planungsmethodik

Band 35

Dr.-Ing. Matthias Kallabis

Räumen gehärteter Werkstoffe mit kristallinen Hartstoffen

Band 36

Dr.-Ing. Heiner-Michael Honeck

Rückführung von Fertigungsdaten zur Unterstützung einer fertigungsgerechten Konstruktion

Band 37

Dr.-Ing. Manfred Rohr

Automatisierte Technologieplanung am Beispiel der Komplettbearbeitung auf Dreh-/Fräszellen

Band 38

Dr.-Ing. Martin Steuer

Entwicklung von Softwarewerkzeugen zur wissensbasierten Inbetriebnahme von komplexen Serienmaschinen

Band 39

Dr.-Ing. Siegfried Beichter

Rechnergestützte technische Problemlösung bei der Angebotserstellung von flexiblen Drehzellen

Band 40

Dr.-Ing. Thomas Steitz

Methodik zur marktorientierten Entwicklung von Werkzeugmaschinen mit Integration von funktionsbasierter Strukturierung und Kostenschätzung

Band 41

Dr.-Ing. Michael Richter

Wissensbasierte Projektierung elektrohydraulischer Regelungen

Band 42

Dr.-Ing. Roman Kuhn

Technologieplanungssystem Fräsen. Wissensbasierte Auswahl von Werkzeugen, Schneidkörpern und Schnittbedingungen für das Fertigungsverfahren Fräsen

Band 43

Dr.-Ing. Hubert Klein

Rechnerunterstützte Qualitätssicherung bei der Produktion von Bauteilen mit frei geformten Oberflächen

Band 44

Dr.-Ing. Christian Hoffmann

Konzeption und Realisierung eines fertigungsintegrierten Koordinatenmessgerätes

Band 45

Dr.-Ing. Volker Frey

Planung der Leittechnik für flexible Fertigungsanlagen

Band 46

Dr.-Ing. Achim Feller

Kalkulation in der Angebotsphase mit dem selbsttätig abgeleiteten Erfahrungswissen der Arbeitsplanung

Band 47

Dr.-Ing. Markus Klaiber

Produktivitätssteigerung durch rechnerunterstütztes Einfahren von NC-Programmen

Band 48

Dr.-Ing. Roland Minges

Verbesserung der Genauigkeit beim fünffachsigen Fräsen von Freiformflächen

Band 49

Dr.-Ing. Wolfgang Bernhart

Beitrag zur Bewertung von Montagevarianten: Rechnergestützte Hilfsmittel zur kostenorientierten, parallelen Entwicklung von Produkt und Montagesystem

Band 50

Dr.-Ing. Peter Ganghoff

Wissensbasierte Unterstützung der Planung technischer Systeme: Konzeption eines Planungswerkzeuges und exemplarische Anwendung im Bereich der Montagesystemplanung

Band 51

Dr.-Ing. Frank Maier

Rechnergestützte Prozessregelung beim flexiblen Gesenkbiegen durch Rückführung von Qualitätsinformationen

Band 52

Dr.-Ing. Frank Debus

Ansatz eines rechnerunterstützten Planungsmanagements für die Planung in verteilten Strukturen

Band 53

Dr.-Ing. Joachim Weinbrecht

Ein Verfahren zur zielorientierten Reaktion auf Planabweichungen in der Werkstattregelung

Band 54

Dr.-Ing. Gerd Herrmann

Reduzierung des Entwicklungsaufwandes für anwendungsspezifische Zellenrechnersoftware durch Rechnerunterstützung

Band 55

Dr.-Ing. Robert Wassmer

Verschleissentwicklung im tribologischen System Fräsen: Beiträge zur Methodik der Prozessmodellierung auf der Basis tribologischer Untersuchungen beim Fräsen

Band 56

Dr.-Ing. Peter Uebelhoer

Inprocess-Geometriemessung beim Honen

Band 57

Dr.-Ing. Hans-Joachim Schelberg

Objektorientierte Projektierung von SPS-Software

Band 58

Dr.-Ing. Klaus Boes

Integration der Qualitätsentwicklung in featurebasierte CAD/CAM-Prozessketten

Band 59

Dr.-Ing. Martin Schreiber

Wirtschaftliche Investitionsbewertung komplexer Produktionssysteme unter Berücksichtigung von Unsicherheit

Band 60

Dr.-Ing. Ralf Steuernagel

Offenes adaptives Engineering-Werkzeug zur automatisierten Erstellung von entscheidungsunterstützenden Informationssystemen

Band 62

Dr.-Ing. Uwe Schauer

Qualitätsorientierte Feinbearbeitung mit Industrierobotern: Regelungsansatz für die Freiformflächenfertigung des Werkzeug- und Formenbaus

Band 63

Dr.-Ing. Simone Loeper

Kennzahlengestütztes Beratungssystem zur Verbesserung der Logistikleistung in der Werkstattfertigung

Band 64

Dr.-Ing. Achim Raab

Räumen mit hartstoffbeschichteten HSS-Werkzeugen

Band 65,

Dr.-Ing. Jan Erik Burghardt

Unterstützung der NC-Verfahrenskette durch ein bearbeitungselementorientiertes, lernfähiges Technologieplanungssystem

Band 66

Dr.-Ing. Christian Tritsch

Flexible Demontage technischer Gebrauchsgüter: Ansatz zur Planung und (teil-)automatisierten Durchführung industrieller Demontageprozesse

Band 67

Dr.-Ing. Oliver Eitrich

Prozessorientiertes Kostenmodell für die entwicklungsbegleitende Vorkalkulation

Band 68

Dr.-Ing. Oliver Wilke

Optimierte Antriebskonzepte für Räummaschinen - Potentiale zur Leistungssteigerung

Band 69

Dr.-Ing. Thilo Sieth

Rechnergestützte Modellierungsmethodik zerspantechnologischer Prozesse

Band 70

Dr.-Ing. Jan Linnenbuerger

Entwicklung neuer Verfahren zur automatisierten Erfassung der geometrischen Abweichungen an Linearachsen und Drehschwenkköpfen

Band 71

Dr.-Ing. Mathias Klimmek

Fraktionierung technischer Produkte mittels eines frei beweglichen Wasserstrahlwerkzeuges

Band 72

Dr.-Ing. Marko Hartel

Kennzahlenbasiertes Bewertungssystem zur Beurteilung der Demontage- und Recyclingeignung von Produkten

Band 73

Dr.-Ing. Jörg Schaupp

Wechselwirkung zwischen der Maschinen- und Hauptspindelantriebsdynamik und dem Zerspanprozess beim Fräsen

Band 74

Dr.-Ing. Bernhard Neisius

Konzeption und Realisierung eines experimentellen Telemanipulators für die Laparoskopie

Band 75

Dr.-Ing. Wolfgang Walter

Erfolgsversprechende Muster für betriebliche Ideenfindungsprozesse. Ein Beitrag zur Steigerung der Innovationsfähigkeit

Band 76

Dr.-Ing. Julian Weber

Ein Ansatz zur Bewertung von Entwicklungsergebnissen in virtuellen Szenarien

Band 77

Dr.-Ing. Dipl. Wirtsch.-Ing. Markus Posur

Unterstützung der Auftragsdurchsetzung in der Fertigung durch Kommunikation über mobile Rechner

Band 78

Dr.-Ing. Frank Fleissner

Prozessorientierte Prüfplanung auf Basis von Bearbeitungsobjekten für die Kleinserienfertigung am Beispiel der Bohr- und Fräsbearbeitung

Band 79

Dr.-Ing. Anton Haberkern

Leistungsfähigere Kugelgewindetriebe durch Beschichtung

Band 80

Dr.-Ing. Dominik Matt

Objektorientierte Prozess- und Strukturinnovation (OPUS)

Band 81

Dr.-Ing. Jürgen Andres

Robotersysteme für den Wohnungsbau: Beitrag zur Automatisierung des Mauerwerkbaus und der Elektroinstallation auf Baustellen

Band 82

Dr.-Ing. Dipl.Wirtschaftsing. Simone Riedmiller

Der Prozesskalender - Eine Methodik zur marktorientierten Entwicklung von Prozessen

Band 83

Dr.-Ing. Dietmar Tilch

Analyse der Geometrieparameter von Präzisionsgewinden auf der Basis einer Least-Squares-Estimation

Band 84

Dr.-Ing. Dipl.-Kfm. Oliver Stiefbold

Konzeption eines reaktionsschnellen Planungssystems für Logistikketten auf Basis von Software-Agenten

Band 85

Dr.-Ing. Ulrich Walter

Einfluss von Kühlschmierstoff auf den Zerspansprozess beim Fräsen: Beitrag zum Prozessverständnis auf Basis von zerspantechnischen Untersuchungen

Band 86

Dr.-Ing. Bernd Werner

Konzeption von teilautonomer Gruppenarbeit unter Berücksichtigung kultureller Einflüsse

Band 87

Dr.-Ing. Ulf Osmer

Projektieren Speicherprogrammierbarer Steuerungen mit Virtual Reality

Band 88

Dr.-Ing. Oliver Doerfel

Optimierung der Zerspantechnik beim Fertigungsverfahren Wälzstossen: Analyse des Potentials zur Trockenbearbeitung

Band 89

Dr.-Ing. Peter Baumgartner

Stufenmethode zur Schnittstellengestaltung in der internationalen Produktion

Band 90

Dr.-Ing. Dirk Vossmann

Wissensmanagement in der Produktentwicklung durch Qualitätsmethodenverbund und Qualitätsmethodenintegration

Band 91

Dr.-Ing. Martin Plass

Beitrag zur Optimierung des Honprozesses durch den Aufbau einer Honprozessregelung

Band 92

Dr.-Ing. Titus Konold

Optimierung der Fünffachsfräsbearbeitung durch eine kennzahlenunterstützte CAM-Umgebung

Band 93

Dr.-Ing. Jürgen Brath

Unterstützung der Produktionsplanung in der Halbleiterfertigung durch risikoberücksichtigende Betriebskennlinien

Band 94

Dr.-Ing. Dirk Geisinger

Ein Konzept zur marktorientierten Produktentwicklung

Band 95

Dr.-Ing. Marco Lanza

Entwurf der Systemunterstützung des verteilten Engineering mit Axiomatic Design

Band 96

Dr.-Ing. Volker Hüntrup

Untersuchungen zur Mikrostrukturierbarkeit von Stählen durch das Fertigungsverfahren Fräsen

Band 97

Dr.-Ing. Frank Reinboth

Interne Stützung zur Genauigkeitsverbesserung in der Inertialmesstechnik: Beitrag zur Senkung der Anforderungen an Inertialsensoren

Band 98

Dr.-Ing. Lutz Trender

Entwicklungsintegrierte Kalkulation von Produktlebenszykluskosten auf Basis der ressourcenorientierten Prozesskostenrechnung

Band 99

Dr.-Ing. Cornelia Kafka

Konzeption und Umsetzung eines Leitfadens zum industriellen Einsatz von Data-Mining

Band 100

Dr.-Ing. Gebhard Selinger

Rechnerunterstützung der informellen Kommunikation in verteilten Unternehmensstrukturen

Band 101

Dr.-Ing. Thomas Windmüller

Verbesserung bestehender Geschäftsprozesse durch eine mitarbeiterorientierte Informationsversorgung

Band 102

Dr.-Ing. Knud Lembke

Theoretische und experimentelle Untersuchung eines bistabilen elektrohydraulischen Linearantriebs

Band 103

Dr.-Ing. Ulrich Thies

Methode zur Unterstützung der variantengerechten Konstruktion von industriell eingesetzten Kleingeräten

Band 104

Dr.-Ing. Andreas Schmälzle

Bewertungssystem für die Generalüberholung von Montageanlagen –Ein Beitrag zur wirtschaftlichen Gestaltung geschlossener Facility- Management-Systeme im Anlagenbau

Band 105

Dr.-Ing. Thorsten Frank

Vergleichende Untersuchungen schneller elektromechanischer Vorschubachsen mit Kugelgewindetrieb

Band 106

Dr.-Ing. Achim Agostini

Reihenfolgeplanung unter Berücksichtigung von Interaktionen: Beitrag zur ganzheitlichen Strukturierung und Verarbeitung von Interaktionen von Bearbeitungsobjekten

Band 107

Dr.-Ing. Thomas Barrho

Flexible, zeitfenstergesteuerte Auftragseinplanung in segmentierten Fertigungsstrukturen

Band 108

Dr.-Ing. Michael Scharer

Quality Gate-Ansatz mit integriertem Risikomanagement

Band 109

Dr.-Ing. Ulrich Suchy

Entwicklung und Untersuchung eines neuartigen Mischkopfes für das Wasser Abrasivstrahlschneiden

Band 110

Dr.-Ing. Sellal Mussa

Aktive Korrektur von Verlagerungsfehlern in Werkzeugmaschinen

Band 111

Dr.-Ing. Andreas Hühsam

Modellbildung und experimentelle Untersuchung des Wälzschälprozesses

Band 112

Dr.-Ing. Axel Plutowsky

Charakterisierung eines optischen Messsystems und den Bedingungen des Arbeitsraums einer Werkzeugmaschine

Band 113

Dr.-Ing. Robert Landwehr

Konsequent dezentralisierte Steuerung mit Industrial Ethernet und offenen Applikationsprotokollen

Band 114

Dr.-Ing. Christoph Dill

Turbulenzreaktionsprozesse

Band 115

Dr.-Ing. Michael Baumeister

Fabrikplanung im turbulenten Umfeld

Band 116

Dr.-Ing. Christoph Gönzheimer

Konzept zur Verbesserung der Elektromagnetischen Verträglichkeit (EMV) in Produktionssystemen durch intelligente Sensor/Aktor-Anbindung

Band 117

Dr.-Ing. Lutz Demuß

Ein Reifemodell für die Bewertung und Entwicklung von Dienstleistungsorganisationen: Das Service Management Maturity Modell (SMMM)

Band 118

Dr.-Ing. Jörg Söhner

Beitrag zur Simulation zerspanungstechnologischer Vorgänge mit Hilfe der Finite-Element-Methode

Band 119

Dr.-Ing. Judith Elsner

Informationsmanagement für mehrstufige Mikro-Fertigungsprozesse

Band 120

Dr.-Ing. Lijing Xie

Estimation Of Two-dimension Tool Wear Based On Finite Element Method

Band 121

Dr.-Ing. Ansgar Blessing

Geometrischer Entwurf mikromechatronischer Systeme

Band 122

Dr.-Ing. Rainer Ebner

Steigerung der Effizienz mehrachsiger Fräsprozesse durch neue Planungsmethoden mit hoher Benutzerunterstützung

Band 123

Dr.-Ing. Silja Klinkel

Multikriterielle Feinplanung in teilautonomen Produktionsbereichen – Ein Beitrag zur produkt- und prozessorientierten Planung und Steuerung

Band 124

Dr.-Ing. Wolfgang Neithardt

Methodik zur Simulation und Optimierung von Werkzeugmaschinen in der Konzept- und Entwurfsphase auf Basis der Mehrkörpersimulation

Band 125

Dr.-Ing. Andreas Mehr

Hartfeinbearbeitung von Verzahnungen mit kristallinen diamantbeschichteten Werkzeugen beim Fertigungsverfahren Wälzstoßen

Band 126

Dr.-Ing. Martin Gutmann

Entwicklung einer methodischen Vorgehensweise zur Diagnose von hydraulischen Produktionsmaschinen

Band 127

Dr.-Ing. Gisela Lanza

Simulative Anlaufunterstützung auf Basis der Qualitätsfähigkeiten von Produktionsprozessen

Band 128

Dr.-Ing. Ulf Dambacher

Kugelgewindetrieb mit hohem Druckwinkel

Band 129

Dr.-Ing. Carsten Buchholz

Systematische Konzeption und Aufbau einer automatisierten Produktionszelle für pulverspritzgegossene Mikrobauteile

Band 130

Dr.-Ing. Heiner Lang

Trocken-Räumen mit hohen Schnittgeschwindigkeiten

Band 131

Dr.-Ing. Daniel Nesges

Prognose operationeller Verfügbarkeiten von Werkzeugmaschinen unter Berücksichtigung von Serviceleistungen

Im Shaker Verlag erschienene Bände:

Band 132

Dr.-Ing. Andreas Bechle

Beitrag zur prozesssicheren Bearbeitung beim Hochleistungsfertigungsverfahren Wälzschäl

Band 133

Dr.-Ing. Markus Herm

Konfiguration globaler Wertschöpfungsnetzwerke auf Basis von Business Capabilities

Band 134

Dr.-Ing. Hanno Tritschler

**Werkzeug- und Zerspanprozessoptimierung beim Hartfräsen
von Mikrostrukturen in Stahl**

Band 135

Dr.-Ing. Christian Munzinger

**Adaptronische Strebe zur Steifigkeitssteigerung
von Werkzeugmaschinen**

Band 136

Dr.-Ing. Andreas Stepping

**Fabrikplanung im Umfeld von Wertschöpfungsnetzwerken und
ganzheitlichen Produktionssystemen**

Band 137

Dr.-Ing. Martin Dyck

**Beitrag zur Analyse thermische bedingter Werkstückdeformationen
in Trockenbearbeitungsprozessen**

Band 138

Dr.-Ing. Siegfried Schmalzried

**Dreidimensionales optisches Messsystem für eine effizientere
geometrische Maschinenbeurteilung**

Band 139

Dr.-Ing. Marc Wawerla

Risikomanagement von Garantieleistungen

Band 140

Dr.-Ing. Ivesa Buchholz

**Strategien zur Qualitätssicherung mikromechanischer Bauteile
mittels multisensorieller Koordinatenmesstechnik**

Band 141

Dr.-Ing. Jan Kotschenreuther

**Empirische Erweiterung von Modellen der Makrozerspannung
auf den Bereich der Mikrobearbeitung**

Band 142

Dr.-Ing. Andreas Knödel

Adaptronische hydrostatische Drucktascheneinheit

Band 143

Dr.-Ing. Gregor Stengel

Fliegendes Abtrennen räumlich gekrümmter Strangpressprofile mittels Industrierobotern

Band 144

Dr.-Ing. Udo Weismann

Lebenszyklusorientiertes interorganisationelles Anlagencontrolling

Band 145

Dr.-Ing. Rüdiger Pabst

Mathematische Modellierung der Wärmestromdichte zur Simulation des thermischen Bauteilverhaltens bei der Trockenbearbeitung

Band 146

Dr.-Ing. Jan Wieser

Intelligente Instandhaltung zur Verfügbarkeitssteigerung von Werkzeugmaschinen

Band 147

Dr.-Ing. Sebastian Haupt

Effiziente und kostenoptimale Herstellung von Mikrostrukturen durch eine Verfahrenskombination von Bahnerosion und Laserablation

Band 148

Dr.-Ing. Matthias Schlipf

Statistische Prozessregelung von Fertigungs- und Messprozess zur Erreichung einer variabilitätsarmen Produktion mikromechanischer Bauteile

Band 149

Dr.-Ing. Jan Philipp Schmidt-Ewig

Methodische Erarbeitung und Umsetzung eines neuartigen Maschinenkonzeptes zur produktflexiblen Bearbeitung räumlich gekrümmter Strangpressprofile

Band 150

Dr.-Ing. Thomas Ender

Prognose von Personalbedarfen im Produktionsanlauf unter Berücksichtigung dynamischer Planungsgrößen

Band 151

Dr.-Ing. Kathrin Peter

**Bewertung und Optimierung der Effektivität von Lean Methoden
in der Kleinserienproduktion**

Band 152

Dr.-Ing. Matthias Schopp

Sensorbasierte Zustandsdiagnose und -prognose von Kugelgewindetrieben

Band 153

Dr.-Ing. Martin Kipfmüller

Aufwandsoptimierte Simulation von Werkzeugmaschinen

Band 154

Dr.-Ing. Carsten Schmidt

**Development of a database to consider multi wear mechanisms
within chip forming simulation**

Band 155

Dr.-Ing. Stephan Niggeschmidt

**Ausfallgerechte Ersatzteilbereitstellung im Maschinen- und Anlagenbau
mittels lastabhängiger Lebensdauerprognose**

Band 156

Dr.-Ing. Jochen Conrad Peters

**Bewertung des Einflusses von Formabweichungen in der
Mikro-Koordinatenmesstechnik**

Band 157

Dr.-Ing. Jörg Ude

**Entscheidungsunterstützung für die Konfiguration
globaler Wertschöpfungsnetzwerke**

Band 158

Dr.-Ing. Stefan Weiler

Strategien zur wirtschaftlichen Gestaltung der globalen Beschaffung

Band 159

Dr.-Ing. Jan Rühl

Monetäre Flexibilitäts- und Risikobewertung

Band 160

Dr.-Ing. Daniel Ruch

Positions- und Konturerfassung räumlich gekrümmter Profile auf Basis bauteilimmanenter Markierungen

Band 161

Dr.-Ing. Manuel Tröndle

Flexible Zuführung von Mikrobauteilen mit piezoelektrischen Schwingförderern

Band 162

Dr.-Ing. Benjamin Viering

Mikroverzahnungsnormal

Band 163

Dr.-Ing. Chris Becke

Prozesskrafttrichtungsangepasste Frässtrategien zur schädigungsarmen Bohrungsbearbeitung an faserverstärkten Kunststoffen

Band 164

Dr.-Ing. Patrick Werner

Dynamische Optimierung und Unsicherheitsbewertung der lastabhängigen präventiven Instandhaltung von Maschinenkomponenten

Band 165

Dr.-Ing. Martin Weis

Kompensation systematischer Fehler bei Werkzeugmaschinen durch self-sensing Aktoren

Band 166

Dr.-Ing. Markus Schneider

Kompensation von Konturabweichungen bei gerundeten Strangpressprofilen durch robotergestützte Führungswerkzeuge

Band 167

Dr.-Ing. Ester M. R. Ruprecht

Prozesskette zur Herstellung schichtbasierter Systeme mit integrierten Kavitäten

Band 168

Dr.-Ing. Alexander Broos

Simulationsgestützte Ermittlung der Komponentenbelastung für die Lebensdauerprognose an Werkzeugmaschinen

Band 169

Dr.-Ing. Frederik Zanger

Segmentspanbildung, Werkzeugverschleiß, Randschichtzustand und Bauteileigenschaften: Numerische Analysen zur Optimierung des Zerspanungsprozesses am Beispiel von Ti-6Al-4V

Band 170

Dr.-Ing. Benjamin Behmann

Servicefähigkeit

Band 171

Dr.-Ing. Annabel Gabriele Jondral

Simulationsgestützte Optimierung und Wirtschaftlichkeitsbewertung des Lean-Methodeneinsatzes

Band 172

Dr.-Ing. Christoph Ruhs

Automatisierte Prozessabfolge zur qualitätssicheren Herstellung von Kavitäten mittels Mikrobahnerosion

Band 173

Dr.-Ing. Steven Peters

Markoffsche Entscheidungsprozesse zur Kapazitäts- und Investitionsplanung von Produktionssystemen

Band 174

Dr.-Ing. Christoph Kühlewein

Untersuchung und Optimierung des Wälzschälverfahrens mit Hilfe von 3D-FEM-Simulation – 3D-FEM Kinematik- und Spanbildungssimulation

Band 175

Dr.-Ing. Adam-Mwanga Dieckmann

Auslegung und Fertigungsprozessgestaltung sintergefügter Verbindungen für μ MIM-Bauteile

Band 176

Dr.-Ing. Heiko Hennrich

Aufbau eines kombinierten belastungs- und zustandsorientierten Diagnose- und Prognosesystems für Kugelgewindetriebe

Band 177

Dr.-Ing. Stefan Herder

Piezoelektrischer Self-Sensing-Aktor zur Vorspannungsregelung in adaptronischen Kugelgewindetrieben

Band 178

Dr.-Ing. Alexander Ochs

Ultraschall-Strömungsgreifer für die Handhabung textiler Halbzeuge bei der automatisierten Fertigung von RTM-Bauteilen

Band 179

Dr.-Ing. Jürgen Michna

Numerische und experimentelle Untersuchung zerspanungsbedingter Gefügeumwandlungen und Modellierung des thermo-mechanischen Lastkollektivs beim Bohren von 42CrMo4

Band 180

Dr.-Ing. Jörg Elser

Vorrichtungsfreie räumliche Anordnung von Fügepartnern auf Basis von Bauteilmarkierungen

Band 181

Dr.-Ing. Katharina Klimscha

Einfluss des Fügspalts auf die erreichbare Verbindungsqualität beim Sinterfügen

Band 182

Dr.-Ing. Patricia Weber

Steigerung der Prozesswiederholbarkeit mittels Analyse akustischer Emissionen bei der Mikrolaserablation mit UV-Pikosekundenlasern

Band 183

Dr.-Ing. Jochen Schädel

Automatisiertes Fügen von Tragprofilen mittels Faserwickeln

Band 184

Dr.-Ing. Martin Krauß

Aufwandsoptimierte Simulation von Produktionsanlagen durch Vergrößerung der Geltungsbereiche von Teilmodellen

Band 185

Dr.-Ing. Raphael Moser

Strategische Planung globaler Produktionsnetzwerke

Bestimmung von Wandlungsbedarf und Wandlungszeitpunkt mittels multikriterieller Optimierung

Band 186

Dr.-Ing. Martin Otter

Methode zur Kompensation fertigungsbedingter Gestaltabweichungen für die Montage von Aluminium Space-Frame-Strukturen

Band 187

Dr.-Ing. Urs Leberle

Produktive und flexible Gleitförderung kleiner Bauteile auf phasenflexiblen Schwingförderern mit piezoelektrischen 2D-Antriebselementen

Band 188

Dr.-Ing. Johannes Book

Modellierung und Bewertung von Qualitätsmanagementstrategien in globalen Wertschöpfungsnetzwerken

Band 189

Dr.-Ing. Florian Ambrosy

Optimierung von Zerspanungsprozessen zur prozesssicheren Fertigung nanokristalliner Randschichten am Beispiel von 42CrMo4

Band 190

Dr.-Ing. Adrian Kölmel

Integrierte Messtechnik für Prozessketten unreifer Technologien am Beispiel der Batterieproduktion für Elektrofahrzeuge

Band 191

Dr.-Ing. Henning Wagner

Featurebasierte Technologieplanung zum Preforming von textilen Halbzeugen

Band 192

Dr.-Ing. Johannes Gebhardt

**Strukturoptimierung von in FVK eingebetteten metallischen
Lasteinleitungselementen**

Band 193

Dr.-Ing. Jörg Bauer

**Hochintegriertes hydraulisches Vorschubsystem für die Bearbeitung kleiner
Werkstücke mit hohen Fertigungsanforderungen**

Band 194

Dr.-Ing. Nicole Stricker

Robustheit verketteter Produktionssysteme

Robustheitsevaluation und Selektion des Kennzahlensystems der Robustheit

Band 195

Dr.-Ing. Anna Sauer

**Konfiguration von Montagelinien unreifer Produkttechnologien am Beispiel der
Batteriemontage für Elektrofahrzeuge**

Band 196

Dr.-Ing. Florian Sell-Le Blanc

Prozessmodell für das Linearwickeln unrunder Zahnspulen

Ein Beitrag zur orthozyklischen Spulenwickeltechnik

Band 197

Dr.-Ing. Frederic Förster

**Geregeltes Handhabungssystem zum zuverlässigen und energieeffizienten
Handling textiler Kohlenstofffaserzuschnitte**

Band 198

Dr.-Ing. Nikolay Boev

**Numerische Beschreibung von Wechselwirkungen zwischen Zerspanprozess und
Maschine am Beispiel Räumen**

Band 199

Dr.-Ing. Sebastian Greinacher

**Simulationsgestützte Mehrzieloptimierung schlanker und ressourceneffizienter
Produktionssysteme**

Band 200

Dr.-Ing. Benjamin Häfner

Lebensdauerprognose in Abhängigkeit der Fertigungsabweichungen bei Mikroverzahnungen

Band 201

Dr.-Ing. Stefan Klotz

Dynamische Parameteranpassung bei der Bohrungsherstellung in faserverstärkten Kunststoffen unter zusätzlicher Berücksichtigung der Einspannsituation

Band 202

Dr.-Ing. Johannes Stoll

Bewertung konkurrierender Fertigungsfolgen mittels Kostensimulation und stochastischer Mehrzieloptimierung

Anwendung am Beispiel der Blechpaketfertigung für automobiler Elektromotoren

Band 203

Dr.-Ing. Simon-Frederik Koch

Fügen von Metall-Faserverbund-Hybridwellen im Schleuderverfahren ein Beitrag zur fertigungsgerechten intrinsischen Hybridisierung

Band 204

Dr.-Ing. Julius Ficht

Numerische Untersuchung der Eigenspannungsentwicklung für sequenzielle Zerspanungsprozesse

Band 205

Dr.-Ing. Manuel Baumeister

Automatisierte Fertigung von Einzelblattstapeln in der Lithium-Ionen-Zellproduktion

Band 206

Dr.-Ing. Daniel Bertsch

Optimierung der Werkzeug- und Prozessauslegung für das Wälzschälen von Innenverzahnungen

Band 207

Dr.-Ing. Kyle James Kippenbrock

Deconvolution of Industrial Measurement and Manufacturing Processes for Improved Process Capability Assessments

Band 208

Dr.-Ing. Farboud Bejnoud

Experimentelle Prozesskettenbetrachtung für Räumbauteile am Beispiel einer einsatzgehärteten PKW-Schiebemuffe

Band 209

Dr.-Ing. Steffen Dosch

Herstellungsübergreifende Informationsübertragung zur effizienten Produktion von Werkzeugmaschinen am Beispiel von Kugelgewindetrieben

Band 210

Dr.-Ing. Emanuel Moser

Migrationsplanung globaler Produktionsnetzwerke

Bestimmung robuster Migrationspfade und risiko-effizienter Wandlungsbefähiger

Band 211

Dr.-Ing. Jan Hochdörffer

Integrierte Produktallokationsstrategie und Konfigurationssequenz in globalen Produktionsnetzwerken

Band 212

Dr.-Ing. Tobias Arndt

Bewertung und Steigerung der Prozessqualität in globalen Produktionsnetzwerken

Band 213

Dr.-Ing. Manuel Peter

Unwuchtminimale Montage von Permanentmagnetrotoren durch modellbasierte Online-Optimierung

Band 214

Dr.-Ing. Robin Kopf

Kostenorientierte Planung von Fertigungsfolgen additiver Technologien

Band 215

Dr.-Ing. Harald Meier

**Einfluss des Räumens auf den Bauteilzustand in der Prozesskette
Weichbearbeitung – Wärmebehandlung – Hartbearbeitung**

Band 216

Dr.-Ing. Daniel Brabandt

**Qualitätssicherung von textilen Kohlenstofffaser-Preforms mittels
optischer Messtechnik**

Band 217

Dr.-Ing. Alexandra Schabunow

**Einstellung von Aufnahmeparametern mittels projektionsbasierter Qualitäts-
kenngrößen in der industriellen Röntgen-Computertomographie**

Band 218

Dr.-Ing. Jens Bürgin

Robuste Auftragsplanung in Produktionsnetzwerken

Mittelfristige Planung der variantenreichen Serienproduktion unter Unsicherheit
der Kundenauftragskonfigurationen

Band 219

Dr.-Ing. Michael Gerstenmeyer

**Entwicklung und Analyse eines mechanischen Oberflächenbehandlungs-
verfahrens unter Verwendung des Zerspanungswerkzeuges**

Band 220

Dr.-Ing. Jacques Burtcher

**Erhöhung der Bearbeitungsstabilität von Werkzeugmaschinen durch
semi-passive masseneinstellbare Dämpfungssysteme**

Band 221

Dr.-Ing. Dietrich Berger

**Qualitätssicherung von textilen Kohlenstofffaser-Preforms mittels prozess-
integrierter Wirbelstromsensor-Arrays**

Band 222

Dr.-Ing. Fabian Johannes Ballier

Systematic gripper arrangement for a handling device in lightweight production processes

Band 223

Dr.-Ing. Marielouise Schäferling, geb. Zaiß

Development of a Data Fusion-Based Multi-Sensor System for Hybrid Sheet Molding Compound

Band 224

Dr.-Ing. Quirin Spiller

Additive Herstellung von Metallbauteilen mit dem ARBURG Kunststoff-Freiformen

Band 225

Dr.-Ing. Andreas Spohrer

Steigerung der Ressourceneffizienz und Verfügbarkeit von Kugelgewindetrieben durch adaptive Schmierung

Band 226

Dr.-Ing. Johannes Fisel

Veränderungsfähigkeit getakteter Fließmontagesysteme

Planung der Fließbandabstimmung am Beispiel der Automobilmontage

Band 227

Dr.-Ing. Patrick Bollig

Numerische Entwicklung von Strategien zur Kompensation thermisch bedingter Verzüge beim Bohren von 42CrMo4

Band 228

Dr.-Ing. Ramona Pfeiffer, geb. Singer

Untersuchung der prozessbestimmenden Größen für die anforderungsgerechte Gestaltung von Pouchzellen-Verpackungen

Band 229

Dr.-Ing. Florian Baumann

Additive Fertigung von endlosfaserverstärkten Kunststoffen mit dem ARBURG Kunststoff-Freiform Verfahren

Band 230

Dr.-Ing. Tom Stähr

Methodik zur Planung und Konfigurationsauswahl skalierbarer Montagesysteme – Ein Beitrag zur skalierbaren Automatisierung

Band 231

Dr.-Ing. Jan Schwennen

Einbringung und Gestaltung von Lasteinleitungsstrukturen für im RTM-Verfahren hergestellte FVK-Sandwichbauteile

Band 232

Dr.-Ing. Sven Coutandin

Prozessstrategien für das automatisierte Preforming von bebinderten textilen Halbzeugen mit einem segmentierten Werkzeugsystem

Band 233

Dr.-Ing. Christoph Liebrecht

Entscheidungsunterstützung für den Industrie 4.0-Methodeneinsatz
Strukturierung, Bewertung und Ableitung von Implementierungsreihenfolgen

Band 234

Dr.-Ing. Stefan Treber

Transparenzsteigerung in Produktionsnetzwerken
Verbesserung des Störungsmanagements durch verstärkten Informationsaustausch

Band 235

Dr.-Ing. Marius Dackweiler

Modellierung des Fügewickelprozesses zur Herstellung von leichten Fachwerkstrukturen

Band 236

Dr.-Ing. Fabio Echsler Minguillon

Prädiktiv-reaktives Scheduling zur Steigerung der Robustheit in der Matrix-Produktion

Band 237

Dr.-Ing. Sebastian Haag

Entwicklung eines Verfahrensablaufes zur Herstellung von Batteriezellstapeln mit großformatigem, rechteckigem Stapelformat und kontinuierlichen Materialbahnen

Band 238

Dr.-Ing. Raphael Wagner

Strategien zur funktionsorientierten Qualitätsregelung in der Serienproduktion

Band 239

Dr.-Ing. Christopher Ehrmann

Ausfallfrüherkennung von Ritzel-Zahnstangen- Trieben mittels Acoustic Emission

Band 240

Dr.-Ing. Janna Hofmann

Prozessmodellierung des Fünf-Achs-Nadelwickelns zur Implementierung einer trajektoriebasierten Drahtzugkraftregelung

Band 241

Dr.-Ing. Andreas Kuhnle

Adaptive Order Dispatching based on Reinforcement Learning
Application in a Complex Job Shop in the Semiconductor Industry

Band 242

Dr.-Ing. Andreas Greiber

Fertigung optimierter technischer Oberflächen durch eine Verfahrenskombination aus Fliehkraft-Tauchgleitschleifen und Laserablation
Prozesseinflüsse und Prozessauslegung

Band 243

Dr.-Ing. Jan Niclas Eschner

Entwicklung einer akustischen Prozessüberwachung zur Porenbestimmung im Laserstrahlschmelzen

Band 244

Dr.-Ing. Sven Roth

Schädigungsfreie Anbindung von hybriden FVK/Metall-Bauteilen an metallische Tragstrukturen durch Widerstandspunktschweißen

Band 245

Dr.-Ing. Sina Kathrin Peukert

Robustheitssteigerung in Produktionsnetzwerken mithilfe eines integrierten Störungsmanagements

Band 246

Dr.-Ing. Alexander Jacob

Hochiterative Technologieplanung

Rekursive Optimierung produkt- und fertigungsbezogener Freiheitsgrade am Beispiel der hybrid-additiven Fertigung

Band 247

Dr.-Ing. Patrick Moll

Ressourceneffiziente Herstellung von Langfaser-Preforms im Faserblasverfahren

Band 248

Dr.-Ing. Eric Thore Segebadé

Erhöhung der Verschleißbeständigkeit von Bauteilen aus Ti-6Al-4V mittels simulationsgestützter Zerspanung und mechanischer Mikrotexturierung

Band 249

Dr.-Ing. Shun Yang

Regionalized implementation strategy of smart automation within assembly systems in China

Band 250

Dr.-Ing. Constantin Carl Hofmann

Vorausschauende und reaktive Mehrzieloptimierung für die Produktionssteuerung einer Matrixproduktion

Band 251

Dr.-Ing. Paul Ruhland

Prozesskette zur Herstellung von hybriden Faser-Metall-Preforms

Modellbildung und Optimierung des Binderauftrags und der Drapierung für stabförmige Bauteile

Band 252

Dr.-Ing. Leonard Schild

Erzeugung und Verwendung von Anwendungswissen in der industriellen Computertomographie

Band 253

Dr.-Ing. Benedikt Klee

Analyse von Phaseninformationen in Videodaten zur Identifikation von Schwingungen in Werkzeugmaschinen

Band 254

Dr.-Ing. Bruno Vargas

Wälzschalen mit kleinen Achskreuzwinkeln

Prozessgrenzen und Umsetzbarkeit

Band 255

Dr.-Ing. Lucas Bretz

Function-oriented in-line quality assurance of hybrid sheet molding compound

Band 256

Dr.-Ing. Bastian Rothaupt

Dämpfung von Bauteilschwingungen durch einstellbare Werkstückdirektspannung mit Hydrodehnspanntechnik

Band 257

Dr.-Ing. Daniel Kupzik

Robotic Swing Folding of three-dimensional UD-tape-based Reinforcement Structures

Band 258

Dr.-Ing. Bastian Verhaelen

(De-)Zentralisierung von Entscheidungen in globalen Produktionsnetzwerken

Strategie- und komplexitätsorientierte Gestaltung der Entscheidungsautonomie

Band 259

Dr.-Ing. Hannes Wilhelm Weinmann

Integration des Einzelungs- und Stapelbildungsprozesses in ein flexibel und kontinuierlich arbeitendes Anlagenmodul für die Li-Ionen Batteriezellfertigung

Band 260

Dr.-Ing. Florian Stamer

Dynamische Lieferzeit-Preisgestaltung in variantenreicher Produktion

Ein adaptiver Ansatz mithilfe von Reinforcement Learning

Band 261

Dr.-Ing. Patrick Neuenfeldt

Modellbildung des Tauchgleitschleifens zur Abtrag- und Topografievorhersage an komplexen Geometrien

Band 262

Dr.-Ing. Boris Matuschka

Energieeffizienz in Prozessketten: Analyse und Optimierung von Energieflüssen bei der Herstellung eines PKW-Getriebebauteils aus 16MnCr5

Band 263

Dr.-Ing. Tobias Schlagenhauf

Bildbasierte Quantifizierung und Prognose des Verschleißes an Kugelgewindetriebspindeln

Ein Beitrag zur Zustandsüberwachung von Kugelgewindetrieben mittels Methoden des maschinellen Lernens

