Check for updates

Latest updates: https://dl.acm.org/doi/10.1145/3708359.3712139

PDF Download
3708359.3712139.pdf
21 January 2026
Total Citations: 0
Total Downloads: 1502

Published: 24 March 2025

Citation in BibTeX format

IUI '25: 30th International Conference on Intelligent User Interfaces
March 24 - 27, 2025
Cagliari, Italy

Conference Sponsors:
SIGAI

RESEARCH-ARTICLE

Will Health Experts Adopt a Clinical Decision Support System for Game-Based Digital Biomarkers? Investigating the Impact of Different Explanations on Perceived Ease-of-Use, Perceived Usefulness, and Trust

YU CHEN, KU Leuven, Leuven, Vlaams-Brabant, Belgium

KATRIEN VERBERT, KU Leuven, Leuven, Vlaams-Brabant, Belgium

KATHRIN MARIA GERLING, Karlsruhe Institute of Technology, Karlsruhe, Baden-Wurttemberg, Germany

MARIE ELENA VANDEN ABEELE

VERO A VANDEN ABEELE, KU Leuven, Leuven, Vlaams-Brabant, Belgium

Open Access Support provided by:

KU Leuven

Karlsruhe Institute of Technology

IUI '25: Proceedings of the 30th International Conference on Intelligent User Interfaces (March 2025)
https://doi.org/10.1145/3708359.3712139
ISBN: 9798400713064

Will Health Experts Adopt a Clinical Decision Support System for Game-Based Digital Biomarkers? Investigating the Impact of Different Explanations on Perceived Ease-of-Use, Perceived Usefulness, and Trust

Yu Chen
Department of Computer Sciences
KU Leuven
Leuven, Belgium
yu.chen@kuleuven.be

Katrien Verbert
Department of Computer Sciences
KU Leuven
Leuven, Belgium
katrien.verbert@cs.kuleuven.be

Kathrin Gerling
Human-Computer Interaction and Accessibility
Karlsruhe Institute of Technology
Karlsruhe, Germany
kathrin.gerling@kit.edu

Marie-Elena Vanden Abeele
Department of Geriatrics
Jessa Hospital
Hasselt, Belgium
marie-elena.vandenabeele@jessazh.be

Vero Vanden Abeele
Department of Computer Sciences
KU Leuven
Leuven, Belgium
vero.vandenabeele@kuleuven.be

Abstract

This paper explores the adoption of a clinical decision support system (cDSS) utilizing game-based digital biomarkers for diagnosing mild cognitive impairment (MCI). Specifically, it investigates how different explanation methods, with a focus on data-centric explanations, impact perceived ease-of-use, perceived usefulness, and trust among healthcare professionals (HCPs). Through a qualitative study with 12 HCPs, we assess their interactions with an explainable AI (XAI)-enriched cDSS. The findings indicate that HCPs are open to adopting XAI-enriched cDSS to communicate the outcomes of game-based digital biomarkers. HCPs preferred to receive key diagnostic information in an easily digestible format. Both local explanations of intra-personal evolutionary data and global overview of normative data were found to be valuable for interpreting digital biomarkers. HCPs tended to trust the machine learning algorithms as a black box, but they considered the dataset used for training the model and the outcome prediction to be crucial. Therefore, presenting the uncertainty alongside the prediction was deemed important. These insights underscore the importance of designing cDSS tools that foster trust through clear, actionable explanations, paving the way for improved decision-making in clinical contexts.

CCS Concepts

• **Human-centered computing → User studies**; • **Information systems → Decision support systems**; • **Computing methodologies →** *Artificial intelligence.*

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
IUI '25, Cagliari, Italy
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1306-4/25/03
https://doi.org/10.1145/3708359.3712139

Keywords

Healthcare, Decision support system, Digital biomarkers, Mild cognitive impairment, Explainable AI, XAI, Video games

ACM Reference Format:
Yu Chen, Katrien Verbert, Kathrin Gerling, Marie-Elena Vanden Abeele, and Vero Vanden Abeele. 2025. Will Health Experts Adopt a Clinical Decision Support System for Game-Based Digital Biomarkers? Investigating the Impact of Different Explanations on Perceived Ease-of-Use, Perceived Usefulness, and Trust. In *30th International Conference on Intelligent User Interfaces (IUI '25), March 24–27, 2025, Cagliari, Italy*. ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3708359.3712139

1 Introduction

The increasing recognition of Artificial Intelligence (AI) by the healthcare sector has led to its widespread application, with diagnostic assistance being one of the prominent use cases [12, 51, 77]. While relying solely on algorithmic judgments for medical decisions poses risks, collaborative efforts between healthcare professionals (HCPs) and AI have demonstrated superior performance in supporting case-finding and diagnosis, compared to working independently [20, 83, 99]. A specific and novel application of AI in health for diagnostic support concerns *digital biomarkers*, defined as *"objective, quantifiable physiological and behavioral data, collected and measured by means of digital devices such as portables, wearables, implantables, or ingestibles, and used to explain, influence, and/or predict health-related outcomes."* [28, 58, 79, 82]. Artificial intelligence is warranted to transform raw data captured via sensing devices (heart rate, blood pressure, night agitation, etc.) and the traces that users leave behind via interactive technologies (number of keystrokes, speed of finger tapping, sustained gaze on screen, etc.) into a digital biomarker.

Digital biomarkers can be generated from a variety of sources [68]. Among them, mobile and wearable devices are popular due to their widespread usage, immediate access, and the plurality of included sensors [49]. However, recently more specific and novel interactive devices have also been found useful, such as virtual reality [22] or even videogames [38, 57, 104]. Games elicit user interaction and provide challenges that demand motor performance (hand-eye coordination, reaction speed, etc.), cognitive performance (dual tasks, wayfinding, puzzles), or social behaviors (collaborative or competitive play) [10]. This trail of player data [30], in combination with sophisticated machine learning models, can lead to specific metrics supporting diagnosis and case-finding, and hold potential as *game-based digital biomarkers* [63].

Despite the promise of digital biomarkers, their current uptake by domain experts remains limited [35]. The process of AI-assisted diagnosis and case-finding, drawing on digital biomarkers, requires knowledge from two distinct realms. On the one hand, disease diagnosis demands a high level of medical specialization. On the other hand, interpreting the outcome of AI models necessitates a certain knowledge of data analytics and algorithms [87]. Given that few HCPs in clinical practice possess such an AI background, the challenge remains to integrate medical domain knowledge with the outcome of an AI model, to interpret outcomes and come to a diagnosis. Therefore, *explainable AI* (XAI) has become an important focus of research in the health domain, *"developing methods and frameworks to enhance the interpretability and transparency of AI models, bridging the gap between accuracy and explainability"* [65]. To further promote the adoption of XAI in healthcare, a promising approach involves utilizing visual analytics integrated into interactive dashboards, also known as clinical decision support systems (cDSS) [72]. Such clinical decision support dashboards 1) *visualize the outcomes* of AI models, 2) *provide interactivity* with AI models, e.g., exploring or filtering data [59] and 3) *directly communicate (shepherding)* how they work with explicit explanations.

The main goal of XAI-enriched cDSS is to be perceived as *useful* and to align with clinical use cases. It is important that XAI-enriched dashboards help HCPs gain a comprehensive *clinical understanding* of digital biomarkers to aid in their case findings. In addition to providing usefulness, cDSS also needs to be perceived as *ease to use*, ensuring that HCPs understand how to interact with the dashboard, interpret cDSS outcomes, and utilize designed functionalities. Furthermore, in a health context, experiences of safety and trust play a key role in the adoption of technology [11, 25, 31]. Therefore, examining the extent to which XAI can support HCPs in *trusting* digital biomarkers is crucial for the adoption of cDSS [108]. The need for trustworthiness may particularly apply to the novel, game-based digital biomarkers, which are associated with entertainment and may not be deemed credible or fit within a medical context.

Most recently, to improve perceptions of usefulness, ease-of-use, and trust, XAI in healthcare has particularly focused on the use of data-centric explanations and counterfactuals versus more traditional approaches such as feature importance [3, 17, 61, 107]. Whereas certain studies suggest non-expert users found the data-centric explanations helpful and easy to understand, these studies also found that machine learning experts had concerns about over-simplification and perceived limitations [8]. Other scholars suggest that providing multiple counterfactual examples improves non-expert users' understanding of the model's predictions. However, these studies also reported that this increased the complexity of the explanation [18] and thus may lower the overall appreciation by non-AI experts. In sum, preferences for data-centric explanations and counterfactuals vary and may depend on different contexts of use.

Consequently, several researchers [19, 52, 74], have called for further research on how XAI and visual analytics can support HCPs in their practice with the ultimate aim of increasing the adoption of cDSS. Therefore, this paper presents the design of an XAI-enriched clinical dashboard. The purpose of this cDSS is to communicate game-based digital biomarkers to support the diagnosis of Mild Cognitive Impairment (MCI). The study is an early-stage investigation aimed at understanding the adoption readiness of HPCs rather than an evaluation of the clinical efficacy.

This cDSS was evaluated with 12 HCPs using qualitative methods, exploring the willingness to adopt such an XAI-enriched cDSS, and by zooming in on three research questions (RQs):

- **RQ1 (Perceived Ease-of-use):** How easy to use do HCPs perceive the differ visualizations and interaction functionalities in the XAI-enriched cDSS, to communicate game-based digital biomarkers?
- **RQ2 (Perceived Usefulness):** How useful do the HCPs find the different visualizations and interaction functionalities in the XAI-enriched cDSS in supporting their clinical needs in case-finding and diagnosis?
- **RQ3 (Trust):** How and to what extent do HCPs trust predictions based on game-based digital biomarkers for screening MCI with our XAI-enriched cDSS?

This paper outlines three main research contributions. Firstly, we introduce the design of an XAI-enriched cDSS, which facilitates clinical case-finding and diagnosis through game-based digital biomarkers. Secondly, we analyze and report on HCPs' perceptions of the XAI-enriched cDSS in terms of ease-of-use, usefulness, and trust. Finally, we conduct an in-depth exploration of XAI methods (data-centric explanations, counterfactuals, and feature importance) in this cDSS for novel digital biomarkers, and how these offer insights to HCPs.

2 Background

In this section, we first provide more information on game-based digital biomarkers and their use in health. Next, we detail clinical decision support systems and related XAI applications. We end the background section with previous work on game-based digital biomarkers, and how player metrics are used for the screening of MCI.

2.1 Game-based Digital Biomarkers

The term biomarker is shorthand for "biological marker" and is defined as an objective indication of the medical state observed from *outside* the patient. By definition, the characteristic of a biomarker is that it is **objective**, **quantifiable**, and **reproducible** [102]. Biomarkers are widely applied in clinical research and practices [27, 44, 54, 90]. The most common biomarkers are laboratory evaluation markers (e.g., blood glucose concentration, "white dots" on MRI scans, etc.) [9]. With the digitization in medical and healthcare research, digital health technologies have expanded possible sources of biomarkers. Biomarkers generated from physiological and behavioral data obtained via one or more *digital* devices are hence called *digital* biomarkers. Research has shown that digital biomarkers can assist in the diagnosis of various diseases, such as Multiple Sclerosis [26], Major Depression [85][64], Alzheimer's Disease [58], and Parkinson's Disease [88][111].

Recently, it has been argued that player metrics from video games can generate digital biomarkers as well [63]. Player data, generated naturally during gameplay (such as win/loss data, velocity or acceleration of a movement [105], reaction speed [7] and pressure exerted on a screen in a touch-based game [66]), is reflective of cognitive control, motor processing, memory performance, selective attention, etc. Hence, player metrics can be particularly useful to assess cognitive functioning and model mental health [63]. Recent studies have successfully explored and demonstrated the potential of game-based digital biomarkers in the assessment of mental diseases [37, 39, 56, 97, 104].

For example, Vourvopoulos et al. [104] designed a game for the evaluation of MCI named RehabCity. The game was designed as a city block containing more than 200 buildings, several parks, and moving vehicles. In the simulated environment, the game scores, the number of tasks completed, score progression, time needed for wayfinding, and time needed for completing tasks were adopted. These were evaluated as potential game-based digital biomarkers. The result showed that among all player metrics, the game score had the highest correlation with MMSE ($r=0.81$, $p<0.05$). Konstantinidis et al. [56] adapted an exergame to assess MCI. The system comprises sessions of physical training (aerobic, resistance, flexibility, and balance) as well as five games that provide 'light cognitive requirements.' The in-game scores of different levels (game difficulties) and physiological features during the game playing (heart rate, body flexibility, and Borg Scale) were considered as potential digital biomarkers. Using this data, the overall classification accuracy for MCI players was 0.734, and the area under the curve was also 0.734. Tarnanas et al. [97] modified a museum touring simulation game to assess MCI. The system was originally built for virtual museum exhibitions, simulating a museum environment with exhibits. Players in this game should remember five archeological artifacts and locate them in the virtual museum. The reaction time and error numbers were evaluated as game-based digital biomarkers. The analysis of covariance revealed a strong effect ($F(1,50)=79.9$, $p<.001$), indicating that amnestic MCI patients' recalls were impaired compared to healthy control participants.

With respect to digital card games, McKanna et al. [5] designed a game called WarCAT, which is based on the classic card game War. This game was created to assess executive functions and memory. Convolutional neural network classifiers were trained to differentiate between two groups of data, achieving an accuracy of 86.4%. However, it is important to note that the data used in this study was not collected from human participants; instead, it was generated by artificial intelligence bots trained through reinforcement learning. Sirály et al. [91] conducted a study on the card game Find-the-Pair. The researchers collected data on the playing time and the number of clicks needed to complete the game in order to evaluate cognitive functions. They employed logistic regression for classification. Both game metrics demonstrated the ability to differentiate participants with MCI. The number of clicks achieved 83% sensitivity and 62% specificity, while the playing time resulted in 82% sensitivity and 67% specificity. Finally, and most recently, Gielis et al. [36, 39] adapted Klondike Solitaire, a popular digital card game among elders, to capture player metrics. The authors demonstrated that several of these player actions are indicative of MCI. In this paper, we build our cDSS further on the studies of Gielis et al. In section 3.1, we detail the game and the game-based biomarkers further.

It is worth noting that while several studies in the literature have explored the use of card games for assessing MCI, none (including our approach) have been clinically adopted. Further research is required before these methods can achieve widespread clinical adoption.

2.2 XAI for clinical decision support systems

A clinical decision support system (cDSS) integrates targeted clinical knowledge, patient information, and other health information [92], such as (digital) biomarkers [75] to provide diagnostic support for HCPs [92] and clinical case-finding. Consequently, cDSS have been widely applied in many disease case findings, such as diabetes [100][14], cancers [81][69], cardiovascular diseases [101] and mental diseases [29]. However, no prior research has explored the use of cDSS for communicating game-based digital biomarkers.

Nevertheless, algorithm-generated findings and recommendations to support HCPs require careful design in explaining the findings and how the recommendation is made [48, 55]. In this respect, XAI can be a helpful method to present the internal decision-making process of machine learning. XAI is an umbrella term that includes concepts such as transparency[33], interpretability [62], explainability [34], etc. [1].

In the context of technology adoption [103], perceived usefulness and perceived ease-of-use have been found critical. First, the tool should be perceived as *useful*; it should align with clinical context-of-use and communicate the right information at the right time.[71]. Specifically, in a health context, HCPs should perceive the explanations and understandings derived from a cDSS as scientifically correct and clinically valid.

Next to supporting perceived usefulness, recent XAI studies additionally foreground the need for ease-of-use,[46, 50, 71]. From the perspective of the HCP, who is most often not an AI expert, the cDSS should be designed with *ease-of-use* for the clinical context. A cDSS that lacks ease-of-use reduces users' uptake [84]. Moreover, for high-stakes applications like healthcare, ease in understanding how the output is established not only raises the uptake of HCPs but is also reinforced by medico-legal and ethical requirements such as the European GDPR [41].

Finally, as a cDSS incorporating (digital) biomarkers, establishing *trust* is crucial [72]. In clinical decision-making, decisions made by HCPs have a significant impact on patients' lives [41]. Therefore, HCPs need to understand the accuracies of algorithms' output and the margins of uncertainty associated with any biomarker. Establishing *appropriate* trust [42] may even be more relevant in the context of novel digital biomarkers, such as those derived from games. Researchers have also shown that trust is not a static but rather a dynamic factor [47, 70, 73]. Users' trust in a system changes throughout their experience with the system. That means some users might be skeptical of the prediction of a cDSS at the beginning, but as they gain an understanding and build a relationship with the system, their trust could evolve [86].

To improve health experts' perceived ease-of-use, perceived usefulness, and trust, previous studies have formulated guidelines regarding the suitability of different types of explanations; in particular, three types of explanations have been foregrounded: feature attribution, counterfactuals, and data-centric explanations [95, 106]. Feature attribution highlights the importance of individual input features in a model's prediction, enabling users to understand which factors contribute most to the outcome [17]. Counterfactual explanations illustrate how changing certain inputs could alter the prediction, thereby offering actionable insights [61]. Data-centric explanations focus on comparing a given instance to similar examples in the training data, helping users contextualize predictions by linking them to known patterns. Moreover, researchers have promoted contextualizing data [17] and combining local (at the level of the individual person) with global information (data at the level of the population, e.g., means, variability, distributions) [14].

2.3 Mild cognitive impairment and screening MCI by game-based digital biomarkers

MCI is defined as *a syndrome of self-reported cognitive complaint with one or more objective cognitive impairments but preserved independence in functional abilities without dementia* [78]. Hence, in the stage of MCI, despite experiencing cognitive impairments, persons still successfully perform activities of daily living (cooking, toileting, walking, visiting friends, etc.). Nevertheless, MCI indicates an elevated risk of future dementia. Therefore, a timely assessment of MCI is beneficial. The current golden standard to support the clinical diagnosis of MCI is a detailed anamnesis followed by a neuropsychological examination, complemented by brain imaging to look for structural changes and blood and cerebrospinal fluid tests [32]. This comprehensive assessment necessitates a team of medical specialists, consuming a significant number of medical resources [15]. Therefore, before a detailed examination is performed, it is advised to administer a quick screening test and to continue with a full examination only if necessary [24]. The Mini-Mental State Examination (MMSE) is currently the most widely used instrument for rapid screening of Alzheimer's and MCI [76]. It is fast to administer, taking only 10 minutes, but it has also shown a lack of sensitivity, misclassifying older adults with MCI as healthy [67]. The Montreal Cognitive Assessment (MoCA) is another popular instrument, evaluating similar cognitive domains, taking up to 20 minutes. Several studies have shown MoCA to be superior to the MMSE in terms of sensitivity for the MCI [80]. Yet, the MoCA has shown lower specificity, misclassifying healthy oldest older adults (with natural cognitive decline) as at risk for MCI. Review studies find AUC values range from 0.71 to 0.99 for MoCA (mean AUC: 0.883) and from 0.43 to 0.94 for MMSE (mean AUC: 0.780) [6]. Therefore, the outcomes of these tests are never used in isolation to come to a diagnosis, and additional tools are welcomed. Recent research showed that utilizing digital technology and commercial out-of-shelf games also has the potential to complement the screening of MCI and assist in diagnosis [63]. Game-based digital biomarkers could be a complementary resource to existing tests [57, 98, 104]. In a series of studies, Gielis et al. [36, 39] explored the potential of the popular card game Klondike Solitaire, played on a tablet, to screen for MCI. First, twenty-three potential digital biomarkers were linked with eleven cognitive functions, involving eleven HCPs [36]. Next, data was collected from 23 healthy participants and 23 participants living with MCI. Machine learning models were trained on this data. Among the 19 models trained and optimized for the assessment of MCI, the best three models (Extra Trees model, a Gradient Boosting model, Nu-Support Vector Model) had F1 scores ranging from 0.811 to 0.824 [39]. These results suggest that game-based digital biomarkers obtained via Solitaire gameplay are comparable in performance (AUC >0.877) to the aforementioned widely adopted cognitive MCI screening tests. However, the authors also called for more research on how to communicate the results of such models to clinical practitioners. In this paper, we address this call. In particular, we focus on the design and evaluation of an XAI-enriched clinical dashboard to communicate game-based digital biomarkers for diagnosing MCI.

3 Materials and Methods

This section first presents the prototype of the XAI-enriched clinical dashboard used in the study. Therefore, we detail the specific user interfaces of the cDSS, along with the design rationales. Second, we detail the procedure used for the qualitative study. Lastly, we describe our analysis method on the data collected from the interview.

3.1 Prototype of the XAI-enriched clinical dashboard | Solitaire DSS

We designed the prototype of our XAI-enriched clinical dashboard for game-based digital biomarkers (Solitaire DSS) in Figma, a platform for designing interactive, high-fidelity prototypes. The prototype visualizes risk prediction for MCI, based on the 23 potential game-based digital biomarkers for assessing the MCI by playing the card game Solitaire based on previous works by Gielis et al. [36, 39].

Based on the prior findings in game-based digital biomarkers for the assessment of MCI, the cDSS is designed as a web-based service that supports a wide range of devices used by HCPs. The digital biomarkers normative data shown in the prototype is simulated, yet based on the data from the previous studies [36, 39].

3.1.1 Design rationales for the XAI techniques and visual components. In the main *Overview* tab (see Figure 1), there are four visual components (VCs). According to the design recommendations of Bhattacharya et al., [14], color-coded representations are found to be more useful than graphical representations or textual explanations for HCPs and patients. Therefore, we consistently used color coding across the different visual components, with red signifying high risk and blue signifying low risk. Additionally, we prioritized data-centric visualizations, inspired by recent studies showing that such visualizations impact trustworthiness [8] and that health professionals prefer data-centric explanations over other visuals [14].

The *Personal Information* section (VC1) encompasses basic demographic data, established neuropsychological screening test scores (e.g., MMSE and MOCA), and recent self-reported assessments of anxiety and depression, which are frequently associated with MCI.

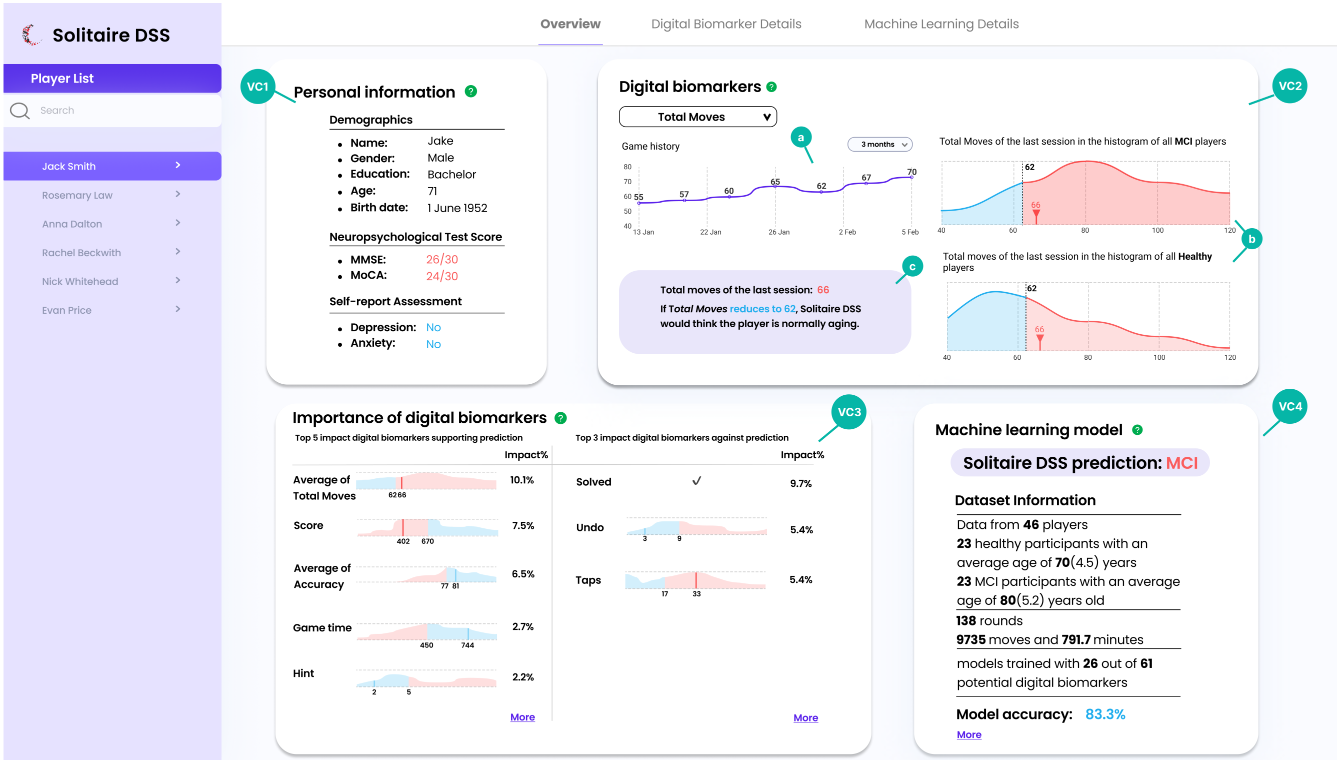

Figure 1: Overview tab

VC1 offers a succinct textual overview of cognitive health information. Additionally, we have implemented a subtle risk indicator, highlighting scores in red if they surpass the recommended threshold, signaling a potential risk for MCI.

The *Digital biomarker* component (VC2) consists of three subcomponents. When a specific digital biomarker is selected, it first displays the evolution of that biomarker (VC2a) for the patient, allowing HCPs to track the evolution over time. This visualization method provides time-oriented information, as observing the evolution of a health parameter over time has been found to be more meaningful than observing a single point in time [4].

VC2 also includes histograms of a digital biomarker, separated for individuals with MCI and cognitively healthy individuals (VC2b). The colors are coded based on the risk prediction. Additionally, we provide the position of the specific patient, contextualized with respect to the scores of MCI patients and healthy individuals [17]. These histograms combine a local explanation (the predicted risk for the individual patient based on the biomarker score) with a global overview (for the entire patient population). Research has shown that such contextualizations and the combination of local and global explanations are crucial for healthcare professionals to understand and trust the decision-making of the model [14] and improve satisfaction [17].

Finally, in VC2c, there are counterfactuals [2] that indicate the biomarker value at which the risk prediction changes. Counterfactual explanations are a type of example-based method [106] that outlines the minimum conditions necessary to yield an alternative decision. In addition to providing an explanation, counterfactuals also aim to offer personalized, actionable insights [40], particularly when they involve actionable health variables [14].

The *Importance of the digital biomarkers* visual component (VC3) offers feature-importance explanations to help HCPs pinpoint the most influential biomarkers for the prediction model. On the left (VC3a), HCPs can find the top five digital biomarkers that support the prediction, along with their impact factors. On the right (VC3b), HCPs can find the top 3 biomarkers that challenge the prediction, along with their impact factors. For each of these biomarkers, a histogram is provided, showing the distribution of all players, with the position of the selected player. The color-coded regions also indicate the risk prediction. In this way, VC3 offers both data-centric information and individualized (local explanation) while providing a global overview for the entire population [14]).

The *Machine learning model* serves as the final visual component (VC4). Positioned at the top of the component, it presents the risk prediction of the player, indicating whether they are at risk for MCI or not, color-coded for clarity. In accordance with the data-centric approach [8], it furnishes details about the dataset and models used for training, along with model accuracy and basic demographic information such as age. This level of transparency is essential for users to evaluate trustworthiness and apply *appropriate* trust [110].

Adhering to the overall mantra in information visualization of "Overview first, details on demand" [89], it was a conscious choice to move extra detailed information on biomarkers and the machine learning models to the additional tabs.

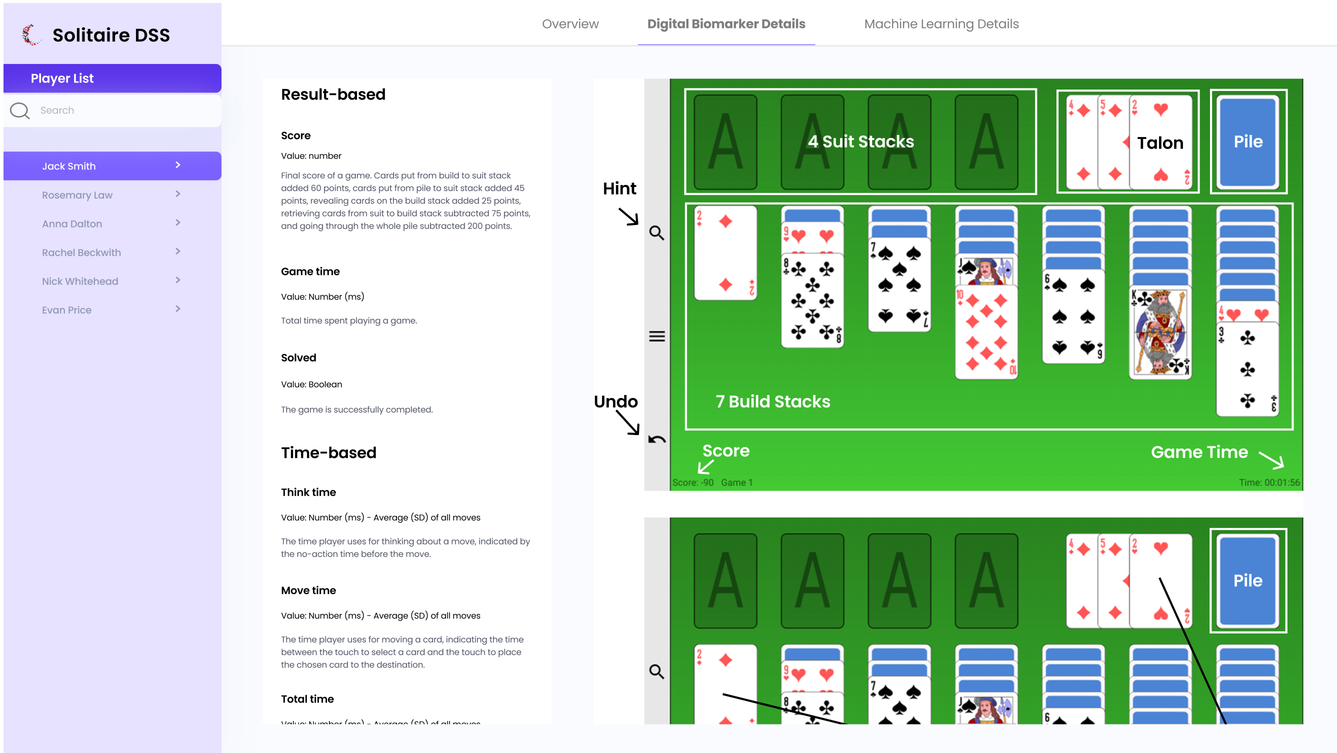

Figure 2: Digital biomarker details tab

The *Digital Biomarker Details* tab (see Figure 2) serves as a dedicated section adjacent to the overview tab. It aims to provide a comprehensive explanation of the game metrics utilized for generating digital biomarkers. The decision to create this separate section, accessible through the second tab, stems from the recognition that HCPs may not be well-versed in the concept of game-based digital biomarkers. Simply encountering the name may not equip them with a thorough understanding. Additionally, they may be unfamiliar with the rules and terminology of the Solitaire card game, which serves as the basis for generating game-based digital biomarkers. Therefore, this section is designed to elucidate game-based digital biomarkers within the context of the Solitaire game and their correlation with cognitive functions, as specified in [36].

The left component of this tab contains textual annotations for game-based digital biomarkers, accompanied by the value type of digital biomarkers used in the models. The right component offers a graphical manual outlining the rules of the Solitaire game. To ensure accessibility for all HCPs, this section also includes screenshots of the game and utilizes visual cues to introduce the rules.

The last tab provides access to the *Machine learning details* (see Figure 3). This section is included for cases where additional information is required. In this third tab of the prototype, demographic information of groups, best-performing models, data preprocessing methods, and confusion matrix are available to provide further insight into how the machine learning models generated the predictions.

3.2 Procedure of the study

In our research, we engaged 12 HCPs with backgrounds in neuropsychology to participate in a user evaluation study. Initially, we recruited participants through the researchers' contacts, followed by snowball sampling through the networks of those participants. Each study session lasted approximately 1.5 hours.

We collected qualitative data using a think-aloud protocol as participants completed five tasks representing typical usage scenarios. Additionally, we observed their interactions with the prototype and assessed their levels of success. To gain deeper insights into their experiences, we conducted semi-structured interviews.

The first participant, a senior geriatrician referred to as P1, served as a pilot tester. Due to her insightful verbalizations during the tasks and feedback in the interview, we included their data in the qualitative (thematic) analysis. The study was approved by the ethical committee of KU Leuven with the number G-2023-6194.

3.2.1 Study design. **A. Introduction** The study commenced with an introduction aimed at acquainting HCPs with the concept of game-based digital biomarkers. The researcher elucidated the essence of game-based digital biomarkers and expounded on the prior scientific studies that form the basis of Solitaire DSS, following a prepared script. Subsequently, the researcher fielded any remaining questions pertaining to background knowledge. Once the HCPs had no further queries, the researcher expounded on the study's purpose in alignment with the research questions.

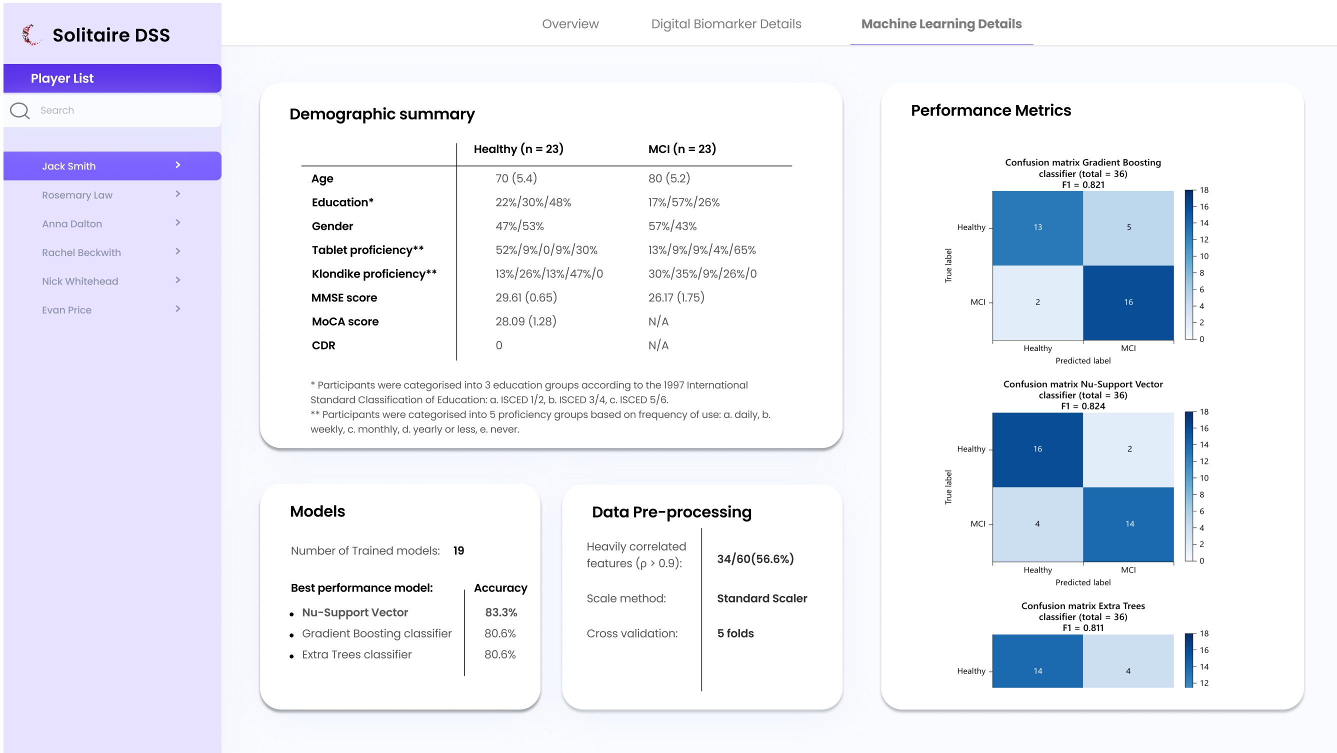

Figure 3: Machine learning details tab

B. Scenarios with tasks The next step involved asking the participants to complete scenarios with tasks by interacting with the Solitaire cDSS. While completing the tasks, they were asked to think aloud. The researcher refrained from providing guidance or answering any questions, following a typical usability protocol. The tasks were designed to simulate real-world decision-making scenarios, ensuring alignment with clinical workflows by mimicking common activities such as identifying the patient, interpreting patient data, assessing predictions, and evaluating model outputs. The participants were asked to complete five tasks:

(1) Could you tell me whose data is shown on the prototype and what's the Solitaire DSS's prediction of this player?
(2) Could you tell me what would be the prediction of Solitaire DSS in case the player's score becomes 700 (and other behaviors remain the same)?
(3) Could you tell me the second most important digital biomarker supporting the current prediction?
(4) Could you tell me the meaning of a "Final beta error"?
(5) Could you tell me which model has the highest accuracy?

After completing all the tasks, the participants were given additional time to explore the Solitaire DSS. [1]

C. Semi-structured interview As the last step, the study ended with a semi-structured interview, during which four probe questions were used to guide the interview.

- To what extent did you find the system easy to use? Why or why not?
- To what extent did you find the system useful? Why or why not?
- To what extent could you understand the game-based biomarkers presented in the DSS? Why or why not?
- To what extent would you trust the recommendations from this tool? Why or why not?

3.3 Data analysis

To thoroughly analyze the feedback provided by the participants, we captured the participants' utterances during their think-aloud tasks and the interviews. The qualitative data underwent thematic analysis using Braun and Clarke's 6-phase method [23]. Initially, we developed a codebook based on two related studies on DSS in different topics [14, 74], i.e., deductive coding. Our researchers then iteratively refined these codes, removing irrelevant ones, adjusting definitions, and incorporating new codes identified in the transcripts.

Given that participants used two languages, we thoroughly examined the original transcripts in both languages to uphold the authenticity of their feedback. In the results section, participant quotes were translated into English by a researcher fluent in the other language. All quotes were also reviewed to rectify any grammar errors and eliminate irrelevant mood words.

[1] We also conducted a survey with questionnaires before and after the scenario with tasks. This survey is outside the scope of this paper and will not be discussed in the results.

4 Results

The following section presents the findings from our study, which involved 12 participants. The participants represented a wide range of career stages in the field of neuropsychology, including psychology college students and senior geriatricians from both Western Europe and East Asia. A summary of the participants' demographic information can be found in Table 1. The participants reported being novices in machine learning, and the majority stated that they either seldom played Solitaire or never played it at all.

4.1 Perceived ease-of-use

On the basis of the number of successfully completed tasks, the results paint a nuanced picture. As indicated in Table 2, Tasks 2, 3, and 5 were not challenging for the participants; only one or two participants were unable to complete the task. Tasks 1 and 4, however, were challenging for participants. Three participants in task 1 and four participants in task 4 were unable to complete it, indicating that there is room for improvement in visualizing model predictions and providing textual annotations for game-based digital biomarkers.

In the thematic analysis, three themes related to perceived ease-of-use were raised: *"Immediate access to key information for supporting case-finding"*, *"Challenges in understanding and interpreting game-based digital biomarkers"*, and *"Counterfactuals are found too complex"*.

4.1.1 Immediate Access to Key Information for Supporting Case-Finding. The utilization of colors facilitated participants in acquiring information more effectively. For example, they could readily discern that the red area signified a disease, or that the red line within an area indicated that a player has a disease. Consequently, the current color scheme aided participants in efficiently obtaining information.

> "*Colors in the prototype look good for me, very nice.*" -P2

> "*The colors are very clear, and because the colors keep coming back, I just assumed some sort of pattern.*" -P12

However, despite the clear color coding, a concern that six participants mentioned is that they found the overview tab overwhelming due to excessive information. In the design, researchers intentionally chose a tab layout to prevent overwhelming HCPs with excessive information on a single screen, as detailed in section 3.1. However, even with the information divided over three tabs, the main screen still presented an overwhelming amount of data, causing participants to overlook crucial details.

> "*The interface is too complex for me. I was distracted by details [in components].*" -P2

> "*I think there is a little bit too much information here [overview] and like I don't know where to focus on.*" -P5

The overload of information caused participants to overlook seemingly insignificant but useful details. This also impacted their performance on the tasks. P10 did not discover *this (tab menu) is clickable* until *one of the tasks' answers was not on the screen.* P12 also ignored the subtitles in VC3 of *"... against prediction"*. P11 had the same experience and said

> "*I just ignored it. I don't know why, but I ignored it completely.*" -P11

Participants overlooked not only minor but sometimes also major designs, such as tabs.

> "*I only realized halfway through that there could be some information hidden somewhere because I also didn't really notice that there were three tabs.*" -P12

In conclusion, participants preferred to have key information readily available to help with their practice in diagnosis. Providing too much information was found to be distracting, but color coding was deemed helpful.

4.1.2 Challenges in understanding and interpreting game-based digital biomarkers. Game-based digital biomarkers, as a new method of MCI screening, differ from traditional pen-and-paper tests. HCPs remarked that they were not familiar with how the prediction is generated and, therefore, were confused about the meaning of certain digital biomarkers.

> "*I need to understand what is assessed in the game. If you give me data from a test I haven't used, I won't understand what happened.*" -P11

> "*The software system must be learned beforehand. I need some time to get familiar with it.*" -P2

During the interviews, participants provided feedback indicating that they needed more information on what specific metrics implied.

> "*I think clinicians could use more explanations about what each move means and why you would choose that to be a marker.*" -P12

> "*It's good that healthcare professionals play the game once. It's the most direct way to understand what a 'score' or 'time' means, what constitutes a normal value of these parameters.*" -P4

Even though interactive designs were implemented to allow users to request extra information on demand, such as tooltips to explain various player metrics, none of the users actually used the tooltips while using the cDSS. Interestingly, rather than presenting all explanatory information of the game-based digital biomarkers along with data, HCPs indicated they would prefer to receive this information separately, and even beforehand. As an illustration, three participants requested a user manual or a tutorial for the Solitaire DSS, even though such explanations could be found on the second tab, likely as this is the format HCPs are familiar with.

4.1.3 Counterfactuals are found too complex. Counterfactuals, a model-agnostic method in XAI, have the potential to offer valuable insights for decision-making [14]. In this cDSS, a counterfactual example (VC2c) was implemented in the *overview* tab, highlighting what scoring of the specific player metric would result in a different prediction. However, during the scenarios with tasks, it was observed that many HCPs used this information infrequently or avoided it altogether. While participants were expected to utilize the counterfactuals in Task 2, only three of them actually used VC2c

Table 1: Demographic information of participants

ID	Nationality	Occupation	Solitaire Prof.	ML Knowledge
P1	Belgian	Senior geriatrician	Never	Novice
P2	Chinese	College student in psychology	Yearly or less	Novice
P3	Chinese	Master's student in neurology	Yearly or less	Novice
P4	Chinese	Master's student in neurology	Yearly or less	Novice
P5	Belgian	Neuropsychology researcher	Yearly or less	Novice
P6	Chinese	Neurologist	Yearly or less	Novice
P7	Chinese	Neurologist	Yearly or less	Novice
P8	Belgian	Senior geriatrician	Daily	Novice
P9	Chinese	Neurologist in training	Yearly or less	Novice
P10	Chinese	Neurologist	Yearly or less	Novice
P11	Chinese	Neurology researcher	Yearly or less	Novice
P12	Belgian	Neuropsychology researcher	Yearly or less	Novice

Table 2: Task Performance Summary Across Participants. VC: visual component in overview tab, DB: digital biomarker details tab, ML: machine learning details tab

Participant	Task 1	Task 2	Task 3	Task 4	Task 5
P2	✓ (VC1, VC3, VC4)	✓ (VC2c)	✗	✓ (VC2, DB)	✓ (VC4, ML)
P3	✗	✓ (VC2c)	✓ (VC3)	✗	✓ (VC4, ML)
P4	✗	✓ (VC3)	✓ (VC3)	✗	✓ (ML)
P5	✓ (VC1, VC4)	✓ (VC2b)	✓ (VC3)	✓ (DB)	✓ (ML)
P6	✓ (VC1, VC4)	✓ (VC3)	✗	✓ (DB)	✗
P7	✓ (VC1, VC2, VC3, VC4)	✓ (VC3)	✓ (VC3)	✓ (DB)	✓ (ML)
P8	✓ (VC1, VC4)	✓ (VC2c)	✓ (VC3)	✓ (DB)	✓ (ML)
P9	✗	✓ (VC3)	✓ (VC3)	✗	✓ (ML)
P10	✓ (VC1, VC4)	✓ (VC2b)	✓ (VC3)	✓ (DB)	✓ (ML)
P11	✓ (VC1, VC4)	✓ (VC2b)	✓ (VC3)	✗	✓ (ML)
P12	✓ (VC1, VC2c, VC4)	✗	✓ (VC3)	✓ (DB)	✓ (ML)

to complete Task 2. It was noted that many HCPs required some time to grasp the meaning of the counterfactuals and mentioned that the counterfactual example was confusing.

> "*I'm not sure what is the what-if [counterfactual example]. I think it might be a little bit confusing. [...] But I think that's too indirect, basically because you don't know the probability that this person would score 62, so it doesn't really reflect the uncertainty. And I don't think clinicians would. I mean, I would not put this in a report because you don't really know what it says about the data.*" -P12

4.2 Perceived Usefulness

Two themes were identified under the category of perceived usefulness in the thematic analysis: *"Investigating MCI risk via comparison"*, and *"Details of ML algorithms are not that interesting, reports of neuropsychological results are"*.

4.2.1 Investigating MCI risk via comparison. The nature of video game play makes it much easier to collect multiple sessions of player data to generate game-based digital biomarkers for prediction, compared to pen-and-paper screening tools. As a result, the digital biomarker component presents a graph showing the evolution of the player's digital biomarker over time (VC2a), which captured the interest of HCPs. Participants found the intra-person evolution data valuable as it provided additional information useful for diagnosis and individual case-finding.

> "*It may be more reasonable as a relevant marker of disease progression. For example, I have the measurements in a year. And then look at the rate at which his disease progresses.*" -P6

> "*I think the long term [evaluation] is needed*" -P9

Participants also concurred that the normative data in the DSS was valuable, in addition to the intra-personal evolution. By using normative data visualization, participants can readily compare a player's performance with that of all other players.

> "*What you want to see is the position of the person relative to the MCI and the healthy, I guess. That's what will come here.*" -P5

> "*That's the normative data with which it compares. So I would assume that that's why that prediction is made.*" -P12

The preceding quotes underscore the importance of evolutionary and normative data for HCPs. These two dimensions allow them to compare data, aiding in the assessment of patients' performance both internally and in comparison to others. During the tasks, participants it was also observed that HCPs utilized this method to comprehend predictions in the cDSS. It was mentioned that evaluating patients' metrics by comparing them to normative data supports the standard practice for HCPs.

HCPs also gained an understanding of game-based digital biomarkers by comparing them with other screening tools (e.g., MMSE and MoCA). They utilized the results from these screening tools to compare with players' performance in game-based digital biomarkers, allowing them to gauge the players' performance.

> "*For the same patient, I can compare his MoCA score with them [digital biomarkers]*" -P11

> "*There are also some other games used for the evaluation of MCI, but not on digital devices, like playing with legos*" -P3

4.2.2 Details of ML algorithms are not that interesting, reports of neuropsychological results are. It was observed that participants spent less time exploring the *machine learning details* tab compared to the other two tabs. Additionally, they verbally expressed less interest in machine learning outcomes (such as F1 scores). Some HCPs explicitly stated that they considered the machine learning model to be a "black box."

> "*I don't really know if people would have issues with the fact that the model itself is black box because I think people do understand that it's a very complex mathematical thing.*" -P12

> "*I just input the clinical data [into the model]. It gave me a result based on all the information.*" -P10

Furthermore, one participant mentioned that it was possible that some HCPs would not be interested in understanding machine learning as it falls outside their area of expertise and their time is limited.

> "*[...] with the really limited time they would drop everything they're doing to learn this new thing and pay a lot of money to do it. And I think that's the problem that we have in general now with the gap between practice and research.*" -P12

Compared to solely understanding machine learning models, HCPs expressed a preference for learning more about the connection between *neuropsychological* information and the new screening tool, i.e., how it relates to cognitive functions.

> "*I think it would also help if the more traditional neuropsychological information is also present because now it's just of the prediction of the machine learning model and the model accuracy.*" -P5

Some HCPs also expressed interest in receiving players' results in a format resembling a neuropsychological test report. For example, they suggested presenting digital biomarkers grouped by cognitive functions, patients' percentiles for each biomarker, performance comparisons with normative data, and a concise summary of findings accompanied by clear statements on clinical implications and recommended next steps. Notably, they perceived the visualizations in the prototype as a representation of such a report.

> "*I think it would be nicer if we would read it as a report with a little bit more text to guide what was done in the context of data and interpretation.*" -P4

> "*You can report the result like a clinical report.*" -P10

> "*This [the overview tab] looks like the general report after you finished the test.*" -P9

4.3 Trust

Four themes were identified in trust: *"Trusting the unfamiliar machine learning algorithms"*, *"Novelty of game-based biomarkers may cause ambiguity"*,*"Being critical of dataset quality and outcome"*, and *"Trust building starts before using the tool"*.

4.3.1 Trusting the unfamiliar machine learning algorithms. From the prior theme ("Details of ML algorithms are not that interesting"), it was already evident that HCPs were not keen on delving into understanding the intricacies of machine learning algorithms. At the same time, HCPs demonstrated a high level of trust in the algorithms, and perhaps an overly positive attitude toward machine learning. Additionally, participants displayed respect for knowledge beyond their expertise. For example, P12, said that

> "*I do trust machine learning [models]. I don't understand it enough to be critical of it.*" -P12

P3, who is pursuing a master's degree in clinical neurology, remarked

> "*The computer [algorithms] must be more accurate. [...] It's much more accurate in addition to the [descriptive] analysis*" -P3

Only one participant voiced her distrust in machine learning algorithms due to her uncertainty about how to evaluate a machine learning model.

> "*To me, machine learning is only assistant tools. I can't trust it. [...] Honestly. If you just present me with this and this machine learning idea. I don't know what it is, so I have no idea. [...] just to wish that is good maybe, you have a threshold of ninety-five, [which] is good, right?*" -P11

4.3.2 Novelty of game-based biomarkers may cause ambiguity. While HCPs voiced trust in the unfamiliar machine learning models, at the same time, they were critical of the novelty of the biomarkers themselves. Playing on tablets may be a new experience for older adults, moreover they may be unfamiliar with card games such as Solitaire. In our study, HCPs expressed concerns that this new experience may lead to biased and inaccurate predictions.

> "*It's possible that for older adults, playing [a game on a tablet] is a bit hard. [...] Also, I feel that Solitare is too complex.*" -P3

HCPs were also concerned that data in the learning stage may not accurately reflect players' actual cognitive functions. The game's learning effect might be significant, prompting HCPs to recommend a tutorial for all players before data collection.

> "*The situation for an older adult to play the game first time, ten times, and twenty times must be different. Their proficiency in the game gains gradually. [...] Does this affect [the prediction]?*" -P4

> "*We do research with VR and also with older adults. [...] but what we definitely do is we have a session where we show it to them and kind of explain how it works. [...] I think it would be really worthwhile for older adults to be able to practice a little bit before they get thrown into this thing that's gonna diagnose them.*" -P12

Despite the cDSS providing explanations of Solitaire rules through screenshots and textual annotations of the digital biomarkers, the actual gameplay also remained unclear to some HCPs. As a result, the player metrics captured were also perceived as ambiguous. Additionally, HCPs expressed a desire for more clarity on how digital biomarkers are connected to cognitive functions.

> "*Here it's just the summary measure of what the person did, but there's no interpretation of what that might mean about the cognitive process.*" -P5

> "*Like final beta error, as words together, as a clinician doesn't really mean very much to me. And so I think that kind of loses its worth, especially if that's something that really impacts the model, and if it was described in more, I guess, everyday words, maybe it would be a bit easier to interpret then.*" -P5

4.3.3 *Being critical of dataset quality and outcome.*

HCPs also expressed a critical attitude towards the data set quality and underscored the need for (more) information on the uncertainty of the prediction, which is vital for their diagnosis.

> "*I think what you really want to know as a clinician is the probability that somebody has mild cognitive impairment given the test scores. And the uncertainty of that probability, and not just a label, MCI or not?*" -P5

> "*I'm not sure if you have more information than this, but the model accuracy is 84%. It's unclear whether that means the probability that this patient has mild cognitive impairments given the test data, or whether it means that there's a 20% chance it's a false positive diagnosis.*" -P5

In addition, participants expressed a desire for more comprehensive information regarding the dataset's quality, which was provided by the cDSS on the *"machine learning details"* tab. They sought details on the sample size, educational level, results from other screening tools for the same player, and even clinical imaging evidence.

> "*I think it's also important to know what the education level was of the Healthy control group. So, a bit more information about the control group, I think, is valuable as well.*" -P5

> "*I think the clinical imaging evidence is required.*" -P9

Some participants also expressed concern about the sample size used for generating predictions, which involved 46 players and 138 game sessions in the cDSS. The sample size was relatively small compared to other clinical studies, leading some participants to feel that their trust in the system was greatly diminished.

> "*Based on such a small sample size, how did you ensure the prediction is correct?*" -P9

> "*I think the more data you can show that it's really worthwhile. [...] Then I think that there are people that are looking for ways to kind of improve their practice. [to use a new screening tool]*" -P11

> "*I think now it lacks the data in practice, that data generated in practice. It's better to use it [collecting more data].*" -P3

4.3.4 *Trust building starts before using the tool.*

HCPs emphasized that data from older adults playing the Solitaire game could be beneficial, e.g., P3 mentioned, *"It's great that we can play the game on a real device."*

Yet they also stressed that HCPs needed to better understand game-based digital biomarkers and the specific game at hand. Engaging in a session of Solitaire and observing the results helps to form a direct impression of each digital biomarker.

Moreover, several HCPs expressed that the most important trust-building should happen *before* working with the Solitaire DSS. HCPs expressed a desire to review a manual of game-based digital biomarkers and its cDSS before the study, but the half-hour background introduction served as the substitute. However, it was not comprehensive enough, and participants may not have had sufficient time to digest the information and establish trust within such a brief period. For HCPs, having a manual beforehand is preferable to presenting textual annotations to explain the concepts and functions among the digital biomarkers data in cDSS. Reviewing the manuals before using the screening tool also aligns with HCPs' habits for learning new assessment tools.

> "*At the very beginning, if you give me a manual and I get familiar with everything, then everything is fine. [...] You stop the trust in this before you use that right. Why I do this [using the Solitaire DSS] because I think it's reliable. That's why I use that. If trust is built beforehand, you won't like it*" -P11

5 Discussion

In this study, we first presented the design rationale for a cDSS, incorporating various XAI methods (such as color coding, counterfactuals, local versus global data, etc.) to aid in the detection and diagnosis of MCI. Subsequently, we assessed this cDSS with 12 HCPs. As an early-stage investigation, the focus is on understanding how different explanation methods impact adoption readiness rather than establishing clinical efficacy. In this section, we revisit the research questions through the themes identified in the thematic analysis, addressing *perceived ease-of-use*, *perceived usefulness*, and *trust*. We conclude with design considerations for the cDSS, informed by the study's findings and aligned with insights from prior research.

5.1 Will health experts adopt a clinical decision support system for game-based digital biomarkers?

We investigated the impact of different visualizations and interaction functionalities on the willingness of HCPs to adopt an XAI-enriched cDSS for the assessment of MCI.

Our detailed qualitative analysis provides a nuanced understanding. In terms of *perceived ease-of-use*, we found that HCPs welcomed receiving key diagnostic information in an easily digestible format within the cDSS using color coding. However, participants also expressed feeling overwhelmed when examining the various visual components, despite our efforts to simplify the user interface by moving certain (more elaborate) details to separate tabs. Specifically, for more information explaining game-based digital biomarkers, HCPs preferred to review it separately from the actual patient data, and preferably beforehand. Furthermore, in the context of this cDSS, the implementation of the counterfactuals did not enhance understanding but rather complicated it. Participants avoided making use of this visualization method altogether.

When considering *perceived usefulness*, participants found visualizations of both intra-person evolution data and normative data valuable for comparing a patient's performance over time and against others. However, HCPs showed less interest in the machine learning aspects of the tool and expressed a greater preference for understanding neuropsychological information and practical clinical applications. Additionally, they recommended presenting the information in a report format that aligns with clinical practice.

In relation to *trust*, HCPs expressed trust in machine learning algorithms despite a limited understanding of them, with most believing the algorithms to be more precise than manual methods. However, they expressed concerns about the novelty of game-based biomarkers, fearing that unfamiliarity with tablet games could potentially bias results, particularly for older adults. HCPs also raised questions about the quality of the dataset, seeking more comprehensive information on the sample size, education levels, and the uncertainty of predictions. Lastly, HCPs underscored the importance of establishing trust before using the tool, suggesting that a detailed manual or tutorial would be beneficial for better comprehension of game-based digital biomarkers.

5.2 Considerations for Design

On the basis of the findings of our study, we reflect on the different explanation methods used in the cDSS, contrast them with prior research findings, and formulate considerations for the design of the XAI-enriched cDSS to communicate digital biomarkers.

5.2.1 Limit the use of textual annotations. When it comes to ease-of-use, HCPs emphasized the importance of having immediate access to data that supports their case-finding in the cDSS. They require a comprehensive understanding of game-based digital biomarkers to effectively interpret this data. In our design, color coding was considered useful for data representations, which is in line with the recommendation of *identifying coherent factors* by Wang et al [106]. Our research also revealed that lengthy textual annotations, combined with game-based digital biomarker data, tended to divert the attention of HCPs. They did not fully utilize these functions and often chose to ignore textual annotations, expressing an overload of information on the screen.

Here we like to point to findings by Szymanski et al. [93], users expressed a preference for a hybrid approach incorporating both *textual* and *visual* explanations. However, in this case, the textual annotations were brief and presented alongside visual descriptions. In our study, the textual annotations were not concise, for example, the explanation of the game rules and the description of game-based digital biomarkers spanned several sentences. This finding is consistent with another study on DSS [14]. Striking a balance between visual and textual annotations is crucial for improving the perceived ease-of-use of the system. Textual annotations/explanations that are not succinct when paired with visualization should be reconsidered, even if they are important for providing explanations. In such instances, providing a manual or tutorial in advance may be more effective in saving time for HCPs during consultations, when using the cDSS.

5.2.2 Consider data-centric methods to support the interpretation of data through comparison. This study revealed that HCPs favored the data-centric method, which significantly influenced their perceived usefulness. HCPs expressed interest in comparing individual patient data to normative data and investigating the evolution of game-based digital biomarkers in older adults. The latter was considered novel, and participants believed it would be a valuable addition for decision-making. The finding that the usefulness of the data-centric method in XAI is consistent with many other studies across various domains [8, 60, 109, 112]. This finding is also consistent with the theory that making comparisons can contribute to the development of knowledge [16]. In our specific case, HCPs analyzed the data both globally and locally. First, similar to many other neuropsychological tests, normative data helps in understanding a player's performance compared to others globally and provides an idea of how well the player is doing in relation to a specific metric. Additionally, comparing the distribution of scores between MCI players and healthy players also provides HCPs with the quality of the metrics, whether it is statistically distinct to distinguish the two groups of players. This is in accordance with utilizing *contrastive causal explanation* proposed in the framework by Wang et al. [106]. Second, HCPs' feedback indicates that they were also interested in local intra-person data comparisons. This presents an opportunity for game-based biomarkers [63]: due to the entertaining nature of video games, it is much easier to collect data over a period of time. This supports comparing the performance of a player longitudinally, which is often improbable in other MCI screening tests.

5.2.3 Transform feature importance explanations into interactive what-if analysis. In our study, task 2 evaluated how HCPs experience feature importance explanations in the cDSS. Most HCPs were able to access information about which game-based digital biomarkers are important for the models' predictions. However, some HCPs found the concept of *'biomarkers against prediction'* confusing in the conventional representation of feature importance explanations, given the static nature in our CDSS. Understanding how certain biomarkers negatively contribute to the prediction was challenging for them. Bhattacharya et al. [14] also made similar observations. They addressed this challenge by enhancing *actionability* through interactive what-if analyses. The interactive design elements enable HCPs to modify the feature value and observe the resulting change in the overall prediction. This approach provides more actionable explanations for HCPs through *hypothesis generation*, as proposed by Wang et al. [106].

5.2.4 Be careful with counterfactuals that are not actionable. In our cDSS, we tried to explain MCI risk prediction with counterfactuals, which have been found useful model-agnostic post-hoc explanations. However, this explanation method caused confusion in our study, different from the result of prior studies that claimed that counterfactual examples might help users' understanding [18, 61, 96]. However, as emphasized by [53], likely, this is because the counterfactuals were not formulated as 'directives'; thereby, they were not actionable.Additionally, we attempted to explain concepts using counterfactual examples. However, this approach caused confusion for some HCPs, which contrasts with other studies that suggest counterfactuals help users' understanding [18, 61, 96]. Based on participant feedback, two potential reasons emerged for why counterfactuals were not well received. First, the way we visualized the counterfactuals may not have been optimal. HCPs preferred more direct communication, and our method of using textual descriptions with highlighted key numbers may not have been sufficiently clear. Second, HCPs may not be familiar with counterfactuals, as they are not commonly used in clinical screening reports, making them more difficult to interpret.

5.2.5 Build trust before using the system. Appropriate trust is essential for HCPs when adopting cDSS in real-world practice [110]. As Han et al. [43] emphasized, avoiding both mistrust and misdistrust is crucial. Our qualitative results indicated that trust is multi-faceted, consistent with previous studies [45, 70, 73]. On one hand, most HCPs expressed trust in the algorithms, respecting the mathematical and computational knowledge behind them, even though they were not familiar with these areas. On the other hand, HCPs remained cautious about the model outcomes, highlighting the need for transparency regarding uncertainty and the robustness of the datasets used to train the models. These factors, which significantly influence trust, have also been similarly noted in other research [13, 21].

Moreover, HCPs emphasized that trust in the Solitaire DSS should be established *prior* to its use. HCPs tended to understand game-based biomarkers by linking them to cognitive functions, an area with which they are familiar. Therefore, simply explaining how player metrics generate these biomarkers may not be the most effective approach, as the metrics themselves are new to HCPs, most of whom do not frequently play Solitaire (as was the case for most of our participants). This reflects the complexity of labeling participants as "non-experts" or domain experts [17, 94]. Despite not being machine learning experts, HCPs have a solid grasp of statistics, enabling them to critically evaluate models based on data and outcomes. While HCPs are well-versed in cognitive functions and existing screening tools, they were not familiar with player metrics derived from Solitaire. This distinguishes them from participants in other studies with less knowledge of statistics.

6 Limitations and future work

Firstly, we acknowledge the limitation of a small sample size, comprising only 12 HCPs. This highlights the inherent challenge of recruiting experts specializing in the assessment and diagnosis of neurodegenerative diseases among older adults. The findings could be strengthened by including participants from diverse geographical locations and with varying levels of experience using Solitaire. Future work should aim to involve a broader and more diverse population of HCPs to capture a wider range of perspectives, ensure generalizability, and better understand the tool's applicability across different clinical contexts.

Secondly, the introduction of game-based biomarkers, particularly those derived from Solitaire, presented unfamiliar concepts for many HCPs. They required ample time to comprehend the game and its rules. However, participants were only given limited time to assimilate these new concepts, potentially hindering their full understanding. In the future, researchers may consider dividing such studies into two sessions. The first session could be used to introduce the game, player metrics, and potential biomarkers, followed by a second session a few days later to conduct the actual measurements, scenarios, tasks, and interviews. Lastly, the data presented in the cDSS prototype was partially synthesized, and only one player's data was shown to the participants. We are currently developing a real web application based on the feedback from this study. We will be using real player data, and we anticipate that this will lead to more insights being discovered.

7 Conclusion

This study explored the adoption of an XAI-enriched cDSS utilizing game-based digital biomarkers for diagnosing MCI. We conducted a qualitative evaluation with 12 HCPs to assess the system's ease-of-use, usefulness, and trust. Our findings reveal that HCPs prioritize seeing the most critical information upfront, emphasizing the value of providing a manual or tutorial prior to using the cDSS. We found that both the evolution data of individual players and normative data explanations helped HCPs better understand the system by enabling comparisons. Additionally, linking cognitive functions with digital biomarkers significantly improved their comprehension. In terms of trust, HCPs, as domain experts, were not particularly interested in understanding the system's inner workings or "black box" algorithms. Instead, they built trust by assessing the quality of input data and the reliability of the predictions. Presenting the uncertainty of predictions played a crucial role in building trust. Additionally, we discovered that trust-building should commence well before the system is presented to HCPs. These insights contribute to the broader application of digital biomarkers and XAI in healthcare. Future work should focus on refining the explanations of digital biomarkers and ensuring these tools are seamlessly integrated into clinical workflows to maximize their potential impact.

References

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–18. https://doi.org/10.1145/3173574.3174156

[2] Amina Adadi and Mohammed Berrada. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6 (2018), 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

[3] Muhammad Afzaal, Jalal Nouri, Aayesha Zia, Panagiotis Papapetrou, Uno Fors, Yongchao Wu, Xiu Li, and Rebecka Weegar. 2021. Explainable AI for Data-Driven Feedback and Intelligent Action Recommendations to Support Students Self-Regulation. *Frontiers in Artificial Intelligence* 4 (Nov. 2021), 723447. https://doi.org/10.3389/frai.2021.723447

[4] Wolfgang Aigner, Paolo Federico, Theresia Gschwandtner, Silvia Miksch, and Alexander Rind. 2012. Challenges of time-oriented data in visual analytics for healthcare. In *IEEE VisWeek workshop on visual analytics in healthcare*, Vol. 4. Citeseer, Seattle, WA, USA, 4 pages. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=5b609326557c715d9a195c9a630ef9ad80acff1c

[5] Mahmood Aljumaili, Robert McLeod, and Marcia Friesen. 2019. Serious Games and ML for Detecting MCI. In *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, Ottawa, ON, Canada, 1–5. https://doi.org/10.1109/GlobalSIP45357.2019.8969123

[6] Nayyereh Aminisani, Rasoul alimi, Ali Javadpour, Mohhamad Asghari-Jafarabadi, Mozhgan Jourian, Chris Stephens, and Morteza Shamshirgaran. 2021. *Comparison between the accuracy of Montreal Cognitive Assessment and Mini-Mental State Examination in the detection of mild cognitive impairment.* preprint. In Review. https://doi.org/10.21203/rs.3.rs-136185/v1

[7] J. A. Anguera, J. Boccanfuso, J. L. Rintoul, O. Al-Hashimi, F. Faraji, J. Janowich, E. Kong, Y. Larraburo, C. Rolle, E. Johnston, and A. Gazzaley. 2013. Video game training enhances cognitive control in older adults. *Nature* 501, 7465 (Sept. 2013), 97–101. https://doi.org/10.1038/nature12486 Number: 7465 Publisher: Nature Publishing Group.

[8] Ariful Islam Anik and Andrea Bunt. 2021. Data-Centric Explanations: Explaining Training Data of Machine Learning Systems to Promote Transparency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3411764.3445736

[9] Jeffrey K. Aronson and Robin E. Ferner. 2017. Biomarkers—A General Review. *Current Protocols in Pharmacology* 76, 1 (March 2017), 17 pages. https://doi.org/10.1002/cpph.19

[10] Sander C.J. Bakkes, Pieter H.M. Spronck, and Giel Van Lankveld. 2012. Player behavioural modelling for video games. *Entertainment Computing* 3, 3 (Aug. 2012), 71–79. https://doi.org/10.1016/j.entcom.2011.12.001

[11] H. Baumeister, L. Nowoczin, J. Lin, H. Seifferth, J. Seufert, K. Laubner, and D. D. Ebert. 2014. Impact of an acceptance facilitating intervention on diabetes patients' acceptance of Internet-based interventions for depression: A randomized controlled trial. *Diabetes Research and Clinical Practice* 105, 1 (July 2014), 30–39. https://doi.org/10.1016/j.diabres.2014.04.031

[12] Yaëlle Bellahsen-Harrar, Mélanie Lubrano, Charles Lépine, Aurélie Beaufrère, Claire Bocciarelli, Anaïs Brunet, Elise Decroix, Franck Neil El-Sissy, Bettina Fabiani, Aurélien Morini, Cyprien Tilmant, Thomas Walter, and Cécile Badoual. 2023. *Evaluating AI-Assistance for Pathologists in Diagnosing and Grading Laryngeal Lesions.* preprint. Pathology. https://doi.org/10.1101/2023.07.23.23292962

[13] Umang Bhatt, Javier Antorán, Yunfeng Zhang, Q. Vera Liao, Prasanna Sattigeri, Riccardo Fogliato, Gabrielle Melançon, Ranganath Krishnan, Jason Stanley, Omesh Tickoo, Lama Nachman, Rumi Chunara, Madhulika Srikumar, Adrian Weller, and Alice Xiang. 2021. Uncertainty as a Form of Transparency: Measuring, Communicating, and Using Uncertainty. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21)*. Association for Computing Machinery, New York, NY, USA, 401–413. https://doi.org/10.1145/3461702.3462571

[14] Aditya Bhattacharya, Jeroen Ooge, Gregor Stiglic, and Katrien Verbert. 2023. Directive Explanations for Monitoring the Risk of Diabetes Onset: Introducing Directive Data-Centric Explanations and Combinations to Support What-If Explorations. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. ACM, Sydney NSW Australia, 204–219. https://doi.org/10.1145/3581641.3584075

[15] Marina Boccardi, Andreas U. Monsch, Clarissa Ferrari, Daniele Altomare, Manfred Berres, Isabelle Bos, Andreas Buchmann, Chiara Cerami, Mira Didic, Cristina Festari, Valentina Nicolosi, Leonardo Sacco, Liesbeth Aerts, Emiliano Albanese, Jean-Marie Annoni, Nicola Ballhausen, Christian Chicherio, Jean-François Démonet, Virginie Descloux, Suzie Diener, Daniel Ferreira, Jean Georges, Anton Gietl, Nicola Girtler, Ingo Kilimann, Stefan Klöppel, Nicole Kustyniuk, Patrizia Mecocci, Nathalie Mella, Martina Pigliautile, Katrin Seeher, Steven D. Shirk, Alessio Toraldo, Andrea Brioschi-Guevara, Kwun C. G. Chan, Paul K. Crane, Alessandra Dodich, Alice Grazia, Nicole A Kochan, Fabricio Ferreira Oliveira, Flavio Nobili, Walter Kukull, Oliver Peters, Inez Ramakers, Perminder S. Sachdev, Stefan Teipel, Pieter Jelle Visser, Michael Wagner, Sandra Weintraub, Eric Westman, Lutz Froelich, Henry Brodaty, Bruno Dubois, Stefano F. Cappa, David Salmon, Bengt Winblad, Giovanni B. Frisoni, Matthias Kliegel, and the Consortium for the Harmonization of Neuropsychological Assessment for Neurocognitive Disorders (https://nextcloud.dzne.de/index.php/s/EwXjLab9caQTbQe). 2022. Harmonizing neuropsychological assessment for mild neurocognitive disorders in Europe. *Alzheimer's & Dementia* 18, 1 (Jan. 2022), 29–42. https://doi.org/10.1002/alz.12365

[16] Lera Boroditsky. 2007. Comparison and the development of knowledge. *Cognition* 102, 1 (Jan. 2007), 118–128. https://doi.org/10.1016/j.cognition.2002.08.001

[17] Clara Bove, Jonathan Aigrain, Marie-Jeanne Lesot, Charles Tijus, and Marcin Detyniecki. 2022. Contextualization and Exploration of Local Feature Importance Explanations to Improve Understanding and Satisfaction of Non-Expert Users. In *27th International Conference on Intelligent User Interfaces (IUI '22)*. Association for Computing Machinery, New York, NY, USA, 807–819. https://doi.org/10.1145/3490099.3511139

[18] Clara Bove, Marie-Jeanne Lesot, Charles Albert Tijus, and Marcin Detyniecki. 2023. Investigating the Intelligibility of Plural Counterfactual Examples for Non-Expert Users: an Explanation User Interface Proposition and User Study. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. ACM, Sydney NSW Australia, 188–203. https://doi.org/10.1145/3581641.3584082

[19] David Buenestado, Javier Elorz, Eduardo G. Pérez-Yarza, Ander Iruetaguena, Unai Segundo, Raúl Barrena, and Juan M. Pikatza. 2013. Evaluating Acceptance and User Experience of a Guideline-based Clinical Decision Support System Execution Platform. *Journal of Medical Systems* 37, 2 (April 2013), 9910. https://doi.org/10.1007/s10916-012-9910-7

[20] Federico Cabitza, Andrea Campagner, Luca Ronzio, Matteo Cameli, Giulia Elena Mandoli, Maria Concetta Pastore, Luca Maria Sconfienza, Duarte Folgado, Marília Barandas, and Hugo Gamboa. 2023. Rams, hounds and white boxes: Investigating human–AI collaboration protocols in medical diagnosis. *Artificial Intelligence in Medicine* 138 (April 2023), 102506. https://doi.org/10.1016/j.artmed.2023.102506

[21] Federico Maria Cau, Hanna Hauptmann, Lucio Davide Spano, and Nava Tintarev. 2023. Supporting High-Uncertainty Decisions through AI and Logic-Style Explanations. In *Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23)*. Association for Computing Machinery, New York, NY, USA, 251–263. https://doi.org/10.1145/3581641.3584080

[22] Silvia Cavedoni, Alice Chirico, Elisa Pedroli, Pietro Cipresso, and Giuseppe Riva. 2020. Digital Biomarkers for the Early Detection of Mild Cognitive Impairment: Artificial Intelligence Meets Virtual Reality. *Frontiers in Human Neuroscience* 14 (July 2020), 245. https://doi.org/10.3389/fnhum.2020.00245

[23] Harris Cooper, Paul M. Camic, Debra L. Long, A. T. Panter, David Rindskopf, and Kenneth J. Sher (Eds.). 2012. *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological.* American Psychological Association, Washington. https://doi.org/10.1037/13620-000

[24] Ellen Elisa De Roeck, Peter Paul De Deyn, Eva Dierckx, and Sebastiaan Engelborghs. 2019. Brief cognitive screening instruments for early detection of Alzheimer's disease: a systematic review. *Alzheimer's Research & Therapy* 11, 1 (Dec. 2019), 21. https://doi.org/10.1186/s13195-019-0474-3

[25] Ine D'haeseleer, Dries Oeyen, Bart Vanrumste, Dominique Schreurs, and Vero Vanden Abeele. 2020. The Unacceptance of a Self-Management Health System by Healthy Older Adults.. In *ICT4AWE*. Springer, Prague, Czech Republic, 269–280. https://pdfs.semanticscholar.org/b82f/305445e7a01039b5db78bc4b65891a92f567.pdf

[26] Anja Dillenseger, Marie Luise Weidemann, Katrin Trentzsch, Hernan Inojosa, Rocco Haase, Dirk Schriefer, Isabel Voigt, Maria Scholz, Katja Akgün, and Tjalf Ziemssen. 2021. Digital Biomarkers in Multiple Sclerosis. *Brain Sciences* 11, 11 (Nov. 2021), 1519. https://doi.org/10.3390/brainsci11111519

[27] James D. Doecke. 2012. Blood-Based Protein Biomarkers for Diagnosis of Alzheimer Disease. *Archives of Neurology* 69, 10 (Oct. 2012), 1318. https://doi.org/10.1001/archneurol.2012.1282

[28] E. Ray Dorsey, Spyros Papapetropoulos, Mulin Xiong, and Karl Kieburtz. 2017. The First Frontier: Digital Biomarkers for Neurodegenerative Disorders. *Digital Biomarkers* 1, 1 (July 2017), 6–13. https://doi.org/10.1159/000477383

[29] Murna Downs, Stephen Turner, Michelle Bryans, Jane Wilcock, John Keady, Enid Levin, Ronan O'Carroll, Kate Howie, and Steve Iliffe. 2006. Effectiveness of educational interventions in improving detection and management of dementia in primary care: cluster randomised controlled study. *BMJ* 332, 7543 (March 2006), 692–696. https://doi.org/10.1136/bmj.332.7543.692

[30] Anders Drachen, Magy Seif El-Nasr, and Alessandro Canossa. 2013. Game Analytics – The Basics. In *Game Analytics*, Magy Seif El-Nasr, Anders Drachen, and Alessandro Canossa (Eds.). Springer London, London, 13–40. https://doi.org/10.1007/978-1-4471-4769-5_2

[31] D. D. Ebert, M. Berking, P. Cuijpers, D. Lehr, M. Pörtner, and H. Baumeister. 2015. Increasing the acceptance of internet-based mental health interventions in primary care patients with depressive symptoms. A randomized controlled trial. *Journal of Affective Disorders* 176 (May 2015), 9–17. https://doi.org/10.1016/j.jad.2015.01.056

[32] Emily C. Edmonds, Carrie R. McDonald, Anisa Marshall, Kelsey R. Thomas, Joel Eppig, Alexandra J. Weigand, Lisa Delano-Wood, Douglas R. Galasko, David P. Salmon, Mark W. Bondi, and Alzheimer's Disease Neuroimaging Initiative. 2019. Early versus late MCI: Improved MCI staging using a neuropsychological approach. *Alzheimer's & Dementia* 15, 5 (May 2019), 699–708. https://doi.org/10.1016/j.jalz.2018.12.009

[33] Upol Ehsan, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz. 2021. Expanding Explainability: Towards Social Transparency in AI systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–19. https://doi.org/10.1145/3411764.3445188

[34] Juliana J. Ferreira and Mateus S. Monteiro. 2020. What Are People Doing About XAI User Experience? A Survey on AI Explainability Research and Practice. In *Design, User Experience, and Usability. Design for Contemporary Interactive Environments*, Aaron Marcus and Elizabeth Rosenzweig (Eds.). Vol. 12201. Springer International Publishing, Cham, 56–73. https://doi.org/10.1007/978-3-030-49760-6_4 Series Title: Lecture Notes in Computer Science.

[35] Luis Garcia-Gancedo and Andrew Bate. 2022. Digital biomarkers for post-licensure safety monitoring. *Drug Discovery Today* 27, 11 (Nov. 2022), 103354. https://doi.org/10.1016/j.drudis.2022.103354

[36] K. Gielis, M.-E. Vanden Abeele, R. de Croon, P. Dierick, F. Ferreira-Brito, L. van Assche, K. Verbert, J. Tournoy, and V.V. Abeele. 2021. Dissecting digital card games to yield digital biomarkers for the assessment of mild cognitive impairment: Methodological approach and exploratory study. *JMIR Serious Games* 9, 4 (2021), 25 pages. https://doi.org/10.2196/18359

[37] Karsten Gielis, Marie-Elena Vanden Abeele, Robin De Croon, Paul Dierick, Filipa Ferreira-Brito, Lies Van Assche, Katrien Verbert, Jos Tournoy, and Vero Vanden Abeele. 2021. Dissecting Digital Card Games to Yield Digital Biomarkers for the Assessment of Mild Cognitive Impairment: Methodological Approach and Exploratory Study. *JMIR Serious Games* 9, 4 (Nov. 2021), e18359. https://doi.org/10.2196/18359

[38] K. Gielis, M.-E. Vanden Abeele, K. Verbert, J. Tournoy, M. De Vos, and V. Vanden Abeele. 2021. Detecting Mild Cognitive Impairment via Digital Biomarkers of Cognitive Performance Found in Klondike Solitaire: A Machine-Learning Study. *Digital Biomarkers* 5, 1 (2021), 44–52. https://doi.org/10.1159/000514105

[39] Karsten Gielis, Marie-Elena Vanden Abeele, Katrien Verbert, Jos Tournoy, Maarten De Vos, and Vero Vanden Abeele. 2021. Detecting Mild Cognitive Impairment via Digital Biomarkers of Cognitive Performance Found in Klondike Solitaire: A Machine-Learning Study. *Digital Biomarkers* 5, 1 (Feb. 2021), 44–52. https://doi.org/10.1159/000514105

[40] Oscar Gomez, Steffen Holter, Jun Yuan, and Enrico Bertini. 2020. ViCE: Visual Counterfactual Explanations for Machine Learning Models. https://doi.org/10.48550/arXiv.2003.02428 arXiv:2003.02428 [cs].

[41] Bryce Goodman and Seth Flaxman. 2017. European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation". *AI Magazine* 38, 3 (Oct. 2017), 50–57. https://doi.org/10.1609/aimag.v38i3.2741 Number: 3.

[42] David Gunning and David W. Aha. 2019. DARPA's Explainable Artificial Intelligence Program. *AI Magazine* 40, 2 (June 2019), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

[43] Wenkai Han and Hans-Jörg Schulz. 2020. Beyond Trust Building — Calibrating Trust in Visual Analytics. In *2020 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX)*. IEEE, Salt Lake City, UT, USA, 9–15. https://doi.org/10.1109/TREX51495.2020.00006

[44] N. Lynn Henry and Daniel F. Hayes. 2012. Cancer biomarkers. *Molecular Oncology* 6, 2 (April 2012), 140–146. https://doi.org/10.1016/j.molonc.2012.01.010

[45] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 57, 3 (May 2015), 407–434. https://doi.org/10.1177/0018720814547570

[46] Robert R. Hoffman, Shane T. Mueller, Gary Klein, and Jordan Litman. 2019. Metrics for Explainable AI: Challenges and Prospects. https://doi.org/10.48550/arXiv.1812.04608 arXiv:1812.04608 [cs].

[47] Daniel Holliday, Stephanie Wilson, and Simone Stumpf. 2016. User Trust in Intelligent Systems: A Journey Over Time. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI '16)*. Association for Computing Machinery, New York, NY, USA, 164–168. https://doi.org/10.1145/2856767.2856811

[48] Baha Ihnaini, M. A. Khan, Tahir Abbas Khan, Sagheer Abbas, Mohammad Sh Daoud, Munir Ahmad, and Muhammad Adnan Khan. 2021. A Smart Healthcare Recommendation System for Multidisciplinary Diabetes Patients with Data Fusion Based on Deep Ensemble Learning. *Computational Intelligence and Neuroscience* 2021 (Sept. 2021), e4243700. https://doi.org/10.1155/2021/4243700 Publisher: Hindawi.

[49] Nicholas C. Jacobson, Hilary Weingarden, and Sabine Wilhelm. 2019. Digital biomarkers of mood disorders and symptom change. *npj Digital Medicine* 2, 1 (Feb. 2019), 3. https://doi.org/10.1038/s41746-019-0078-0

[50] Alon Jacovi, Ana Marasović, Tim Miller, and Yoav Goldberg. 2021. Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 624–635. https://doi.org/10.1145/3442188.3445923

[51] Meredith A. Jones, Warid Islam, Rozwat Faiz, Xuxin Chen, and Bin Zheng. 2022. Applying artificial intelligence technology to assist with breast cancer diagnosis and prognosis prediction. *Frontiers in Oncology* 12 (Aug. 2022), 980793. https://doi.org/10.3389/fonc.2022.980793

[52] Nigel S. Kanagasundaram, Mark T. Bevan, Andrew J. Sims, Andrew Heed, David A. Price, and Neil S. Sheerin. 2016. Computerized clinical decision support for the early recognition and management of acute kidney injury: a qualitative evaluation of end-user experience. *Clinical Kidney Journal* 9, 1 (Feb. 2016), 57–62. https://doi.org/10.1093/ckj/sfv130

[53] Amir-Hossein Karimi, Bernhard Schölkopf, and Isabel Valera. 2021. Algorithmic Recourse: From Counterfactual Explanations to Interventions. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 353–362. https://doi.org/10.1145/3442188.3445899

[54] Michael Khalil, Charlotte E. Teunissen, Markus Otto, Fredrik Piehl, Maria Pia Sormani, Thomas Gattringer, Christian Barro, Ludwig Kappos, Manuel Comabella, Franz Fazekas, Axel Petzold, Kaj Blennow, Henrik Zetterberg, and Jens Kuhle. 2018. Neurofilaments as biomarkers in neurological disorders. *Nature Reviews Neurology* 14, 10 (Oct. 2018), 577–589. https://doi.org/10.1038/s41582-018-0058-z

[55] N. Komal Kumar and D. Vigneswari. 2021. A Drug Recommendation System for Multi-disease in Health Care Using Machine Learning. In *Advances in Communication and Computational Technology*, Gurdeep Singh Hura, Ashutosh Kumar Singh, and Lau Siong Hoe (Eds.). Vol. 668. Springer Nature Singapore, Singapore, 1–12. https://doi.org/10.1007/978-981-15-5341-7_1 Series Title: Lecture Notes in Electrical Engineering.

[56] E.I. Konstantinidis, P.D. Bamidis, A. Billis, P. Kartsidis, D. Petsani, and S.G. Papageorgiou. 2021. Physical training in-game metrics for cognitive assessment: Evidence from extended trials with the fitforall exergaming platform. *Sensors* 21, 17 (2021), 13 pages. https://doi.org/10.3390/s21175756

[57] Evdokimos I. Konstantinidis, Panagiotis D. Bamidis, Antonis Billis, Panagiotis Kartsidis, Despoina Petsani, and Sokratis G. Papageorgiou. 2021. Physical Training In-Game Metrics for Cognitive Assessment: Evidence from Extended Trials with the Fitforall Exergaming Platform. *Sensors* 21, 17 (Aug. 2021), 5756. https://doi.org/10.3390/s21175756

[58] Lampros C. Kourtis, Oliver B. Regele, Justin M. Wright, and Graham B. Jones. 2019. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. *npj Digital Medicine* 2, 1 (Feb. 2019), 1–9. https://doi.org/10.1038/s41746-019-0084-2 Number: 1 Publisher: Nature Publishing Group.

[59] Bum Chul Kwon, Min-Je Choi, Joanne Taery Kim, Edward Choi, Young Bin Kim, Soonwook Kwon, Jimeng Sun, and Jaegul Choo. 2019. RetainVis: Visual Analytics with Interpretable and Interactive Recurrent Neural Networks on Electronic Medical Records. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (Jan. 2019), 299–309. https://doi.org/10.1109/TVCG.2018.2865027

[60] Pierre Le Bras, David A. Robb, Thomas S. Methven, Stefano Padilla, and Mike J. Chantler. 2018. Improving User Confidence in Concept Maps: Exploring Data Driven Explanations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173978

[61] Min Hun Lee and Chong Jun Chew. 2023. Understanding the Effect of Counterfactual Explanations on Trust and Reliance on AI for Human-AI Collaborative Clinical Decision Making. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (Oct. 2023), 369:1–369:22. https://doi.org/10.1145/3610218

[62] Yi-Shan Lin, Wen-Chuan Lee, and Z. Berkay Celik. 2021. What Do You See?: Evaluation of Explainable Artificial Intelligence (XAI) Interpretability through Neural Backdoors. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. ACM, Virtual Event Singapore, 1027–1035. https://doi.org/10.1145/3447548.3467213

[63] Regan Lee Mandryk and Max Valentin Birk. 2019. The Potential of Game-Based Digital Biomarkers for Modeling Mental Health. *JMIR Mental Health* 6, 4 (April 2019), e13485. https://doi.org/10.2196/13485

[64] Regan L. Mandryk, Max V. Birk, Sarah Vedress, Katelyn Wiley, Elizabeth Reid, Phaedra Berger, and Julian Frommel. 2021. Remote Assessment of Depression Using Digital Biomarkers From Cognitive Tasks. *Frontiers in Psychology* 12 (Dec. 2021), 767507. https://doi.org/10.3389/fpsyg.2021.767507

[65] Héder Mihály. 2023. Explainable AI: A Brief History of the Concept. *ERCIM NEWS* 134 (2023), 9–10.

[66] Matthew K. Miller and Regan L. Mandryk. 2016. Differentiating in-Game Frustration from at-Game Frustration using Touch Pressure. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*. ACM, Niagara Falls Ontario Canada, 225–234. https://doi.org/10.1145/2992154.2992185

[67] Alex J. Mitchell. 2009. A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. *Journal of Psychiatric Research* 43, 4 (Jan. 2009), 411–431. https://doi.org/10.1016/j.jpsychires.2008.04.014

[68] Christian Montag, Jon D. Elhai, and Paul Dagum. 2021. On Blurry Boundaries When Defining Digital Biomarkers: How Much Biology Needs to Be in a Digital Biomarker? *Frontiers in Psychiatry* 12 (Sept. 2021), 740292. https://doi.org/10.3389/fpsyt.2021.740292

[69] Daniel R. Murphy, Louis Wu, Eric J. Thomas, Samuel N. Forjuoh, Ashley N.D. Meyer, and Hardeep Singh. 2015. Electronic Trigger-Based Intervention to Reduce Delays in Diagnostic Evaluation for Cancer: A Cluster Randomized Controlled Trial. *Journal of Clinical Oncology* 33, 31 (Nov. 2015), 3560–3567. https://doi.org/10.1200/JCO.2015.61.1301

[70] Mahsan Nourani, Joanie King, and Eric Ragan. 2020. The Role of Domain Expertise in User Trust and the Impact of First Impressions with Intelligent Systems. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 8 (Oct. 2020), 112–121. https://doi.org/10.1609/hcomp.v8i1.7469

[71] Jeroen Ooge, Shotallo Kato, and Katrien Verbert. 2022. Explaining Recommendations in E-Learning: Effects on Adolescents' Trust. In *27th International Conference on Intelligent User Interfaces*. ACM, Helsinki Finland, 93–105. https://doi.org/10.1145/3490099.3511140

[72] Jeroen Ooge, Gregor Stiglic, and Katrien Verbert. 2022. Explaining artificial intelligence with visual analytics in healthcare. *WIREs Data Mining and Knowledge Discovery* 12, 1 (2022), e1427. https://doi.org/10.1002/widm.1427 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1427.

[73] Jeroen Ooge and Katrien Verbert. 2021. Trust in Prediction Models: a Mixed-Methods Pilot Study on the Impact of Domain Expertise. In *2021 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX)*. IEEE, New Orleans, LA, USA, 8–13. https://doi.org/10.1109/TREX53765.2021.00007

[74] Jeroen Ooge and Katrien Verbert. 2022. Visually Explaining Uncertain Price Predictions in Agrifood: A User-Centred Case-Study. *Agriculture* 12, 7 (July 2022), 1024. https://doi.org/10.3390/agriculture12071024

[75] Jerome. A Osheroff, Jonathan Teich, Donald Levick, Luis Saldana, Ferdinand Velasco, Dean Sittig, Kendall Rogers, and Robert Jenders. 2021. *Improving Outcomes with Clinical Decision Support* (0 ed.). HIMSS Publishing, New York. https://doi.org/10.4324/9780367806125

[76] Verna C. Pangman, Jeff Sloan, and Lorna Guse. 2000. An examination of psychometric properties of the Mini-Mental State Examination and the Standardized Mini-Mental State Examination: Implications for clinical practice. *Applied Nursing Research* 13, 4 (Nov. 2000), 209–213. https://doi.org/10.1053/apnr.2000.9231

[77] Sui Peng, Yihao Liu, Weiming Lv, Longzhong Liu, Qian Zhou, Hong Yang, Jie Ren, Guangjian Liu, Xiaodong Wang, Xuehua Zhang, Qiang Du, Fangxing Nie, Gao Huang, Yuchen Guo, Jie Li, Jinyu Liang, Hangtong Hu, Han Xiao, Zelong Liu, Fenghua Lai, Qiuyi Zheng, Haibo Wang, Yanbing Li, Erik K Alexander, Wei Wang, and Haipeng Xiao. 2021. Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study. *The Lancet Digital Health* 3, 4 (April 2021), e250–e259. https://doi.org/10.1016/S2589-7500(21)00041-8

[78] Ronald C. Petersen. 2016. Mild Cognitive Impairment:. *CONTINUUM: Lifelong Learning in Neurology* 22, 2, Dementia (April 2016), 404–418. https://doi.org/10.1212/CON.0000000000000313

[79] Antoine Piau, Katherine Wild, Nora Mattek, and Jeffrey Kaye. 2019. Current State of Digital Biomarker Technologies for Real-Life, Home-Based Monitoring of Cognitive Function for Mild Cognitive Impairment to Mild Alzheimer Disease and Implications for Clinical Care: Systematic Review. *Journal of Medical Internet Research* 21, 8 (Aug. 2019), e12785. https://doi.org/10.2196/12785

[80] Tiago C. C. Pinto, Leonardo Machado, Tatiana M. Bulgacov, Antônio L. Rodrigues-Júnior, Maria L. G. Costa, Rosana C. C. Ximenes, and Everton B. Sougey. 2019. Is the Montreal Cognitive Assessment (MoCA) screening superior to the Mini-Mental State Examination (MMSE) in the detection of mild cognitive impairment (MCI) and Alzheimer's Disease (AD) in the elderly? *International Psychogeriatrics* 31, 4 (April 2019), 491–504. https://doi.org/10.1017/S1041610218001370

[81] Sarah Price, Anne Spencer, Antonieta Medina-Lara, and Willie Hamilton. 2019. Availability and use of cancer decision-support tools: a cross-sectional survey of UK primary care. *British Journal of General Practice* 69, 684 (July 2019), e437–e443. https://doi.org/10.3399/bjgp19X703745

[82] Inter-Organization programme for the sound management of chemicals (Ed.). 2001. *Biomarkers in risk assessment: validity and validation*. Number 222 in Environmental health criteria. World health organization, Geneva.

[83] Carlo Reverberi, Tommaso Rigon, Aldo Solari, Cesare Hassan, Paolo Cherubini, GI Genius CADx Study Group, Giulio Antonelli, Halim Awadie, Sebastian Bernhofer, Sabela Carballal, Mário Dinis-Ribeiro, Agnès Fernández-Clotett, Glòria Fernández Esparrach, Ian Gralnek, Yuta Higasa, Taku Hirabayashi, Tatsuki Hirai, Mineo Iwatate, Miki Kawano, Markus Mader, Andreas Maieron, Sebastian Mattes, Tastuya Nakai, Ingrid Ordas, Raquel Ortigão, Oswaldo Ortiz Zúñiga, Maria Pellisé, Cláudia Pinto, Florian Riedl, Ariadna Sánchez, Emanuel Steiner, Yukari Tanaka, and Andrea Cherubini. 2022. Experimental evidence of effective human–AI collaboration in medical decision-making. *Scientific Reports* 12, 1 (Sept. 2022), 14952. https://doi.org/10.1038/s41598-022-18751-2

[84] David C. Rose, William J. Sutherland, Caroline Parker, Matt Lobley, Michael Winter, Carol Morris, Susan Twining, Charles Ffoulkes, Tatsuya Amano, and Lynn V. Dicks. 2016. Decision support tools for agriculture: Towards effective design and delivery. *Agricultural Systems* 149 (Nov. 2016), 165–174. https://doi.org/10.1016/j.agsy.2016.09.009

[85] Yuri Rykov, Thuan-Quoc Thach, Iva Bojic, George Christopoulos, and Josip Car. 2021. Digital Biomarkers for Depression Screening With Wearable Devices: Cross-sectional Study With Machine Learning Modeling. *JMIR mHealth and uHealth* 9, 10 (Oct. 2021), e24872. https://doi.org/10.2196/24872

[86] Dominik Sacha, Hansi Senaratne, Bum Chul Kwon, Geoffrey Ellis, and Daniel A. Keim. 2016. The Role of Uncertainty, Awareness, and Trust in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (Jan. 2016), 240–249. https://doi.org/10.1109/TVCG.2015.2467591 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[87] Peter Savadjiev, Jaron Chong, Anthony Dohan, Maria Vakalopoulou, Caroline Reinhold, Nikos Paragios, and Benoit Gallix. 2019. Demystification of AI-driven medical image interpretation: past, present and future. *European Radiology* 29, 3 (March 2019), 1616–1624. https://doi.org/10.1007/s00330-018-5674-x

[88] Vrutangkumar V. Shah, James McNames, Martina Mancini, Patricia Carlson-Kuhta, John G. Nutt, Mahmoud El-Gohary, Jodi A. Lapidus, Fay B. Horak, and Carolin Curtze. 2020. Digital Biomarkers of Mobility in Parkinson's Disease During Daily Living. *Journal of Parkinson's Disease* 10, 3 (July 2020), 1099–1111. https://doi.org/10.3233/JPD-201914

[89] Ben Shneiderman. 1997. Designing Information-Abundant Web Sites: Issues and Recommendations. *International Journal of Human-Computer Studies* 47, 1 (July 1997), 5–29. https://doi.org/10.1006/ijhc.1997.0127

[90] Ilina Singh and Nikolas Rose. 2009. Biomarkers in psychiatry. *Nature* 460, 7252 (July 2009), 202–207. https://doi.org/10.1038/460202a

[91] E. Sirály, Á. Szabó, B. Szita, V. Kovács, Z. Fodor, C. Marosi, P. Salacz, Z. Hidasi, V. Maros, P. Hanák, É. Csibri, and G. Csukly. 2015. Monitoring the early signs of cognitive decline in elderly by computer games: An MRI study. *PLoS ONE* 10, 2 (2015), 14 pages. https://doi.org/10.1371/journal.pone.0117918

[92] Reed T. Sutton, David Pincock, Daniel C. Baumgart, Daniel C. Sadowski, Richard N. Fedorak, and Karen I. Kroeker. 2020. An overview of clinical decision support systems: benefits, risks, and strategies for success. *npj Digital Medicine* 3, 1 (Feb. 2020), 17. https://doi.org/10.1038/s41746-020-0221-y

[93] Maxwell Szymanski, Cristina Conati, Vero Vanden Abeele, and Katrien Verbert. 2023. Designing and Personalising Hybrid Multi-Modal Health Explanations for Lay Users. In *IntRS@ RecSys*. Association for Computing Machinery, Singapore, 35–52.

[94] Maxwell Szymanski, Martijn Millecamp, and Katrien Verbert. 2021. Visual, textual or hybrid: the effect of user expertise on different explanations. In *26th International Conference on Intelligent User Interfaces*. ACM, College Station TX USA, 109–119. https://doi.org/10.1145/3397481.3450662

[95] Maxwell Szymanski, Vero Vanden Abeele, and Katrien Verbert. 2024. Designing and Evaluating Explanations for a Predictive Health Dashboard: A User-Centred Case Study. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3613905.3637140

[96] Juntao Tan, Shuyuan Xu, Yingqiang Ge, Yunqi Li, Xu Chen, and Yongfeng Zhang. 2021. Counterfactual Explainable Recommendation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management (CIKM '21)*. Association for Computing Machinery, New York, NY, USA, 1784–1793. https://doi.org/10.1145/3459637.3482420

[97] Ioannis Tarnanas, Nikos Laskaris, Magda Tsolaki, René Muri, Tobias Nef, and Urs P. Mosimann. 2015. On the Comparison of a Novel Serious Game and Electroencephalography Biomarkers for Early Dementia Screening. In *GeNeDis 2014*, Panayiotis Vlamos and Athanasios Alexiou (Eds.). Vol. 821. Springer International Publishing, Cham, 63–77. https://doi.org/10.1007/978-3-319-08939-3_11 Series Title: Advances in Experimental Medicine and Biology.

[98] Ioannis Tarnanas, Nikos Laskaris, Magda Tsolaki, René Muri, Tobias Nef, and Urs P. Mosimann. 2015. On the comparison of a novel serious game and electroencephalography biomarkers for early dementia screening. *Advances in Experimental Medicine and Biology* 821 (2015), 63–77. https://doi.org/10.1007/978-3-319-08939-3_11

[99] Philipp Tschandl, Christoph Rinner, Zoe Apalla, Giuseppe Argenziano, Noel Codella, Allan Halpern, Monika Janda, Aimilios Lallas, Caterina Longo, Josep Malvehy, John Paoli, Susana Puig, Cliff Rosendahl, H. Peter Soyer, Iris Zalaudek, and Harald Kittler. 2020. Human–computer collaboration for skin cancer recognition. *Nature Medicine* 26, 8 (Aug. 2020), 1229–1234. https://doi.org/10.1038/s41591-020-0942-0

[100] Nichole S. Tyler and Peter G. Jacobs. 2020. Artificial Intelligence in Decision Support Systems for Type 1 Diabetes. *Sensors* 20, 11 (June 2020), 3214. https://doi.org/10.3390/s20113214

[101] Jacobus T. Van Wyk, Marc A.M. Van Wijk, Miriam C.J.M. Sturkenboom, Mees Mosseveld, Peter W. Moorman, and Johan Van Der Lei. 2008. Electronic Alerts Versus On-Demand Decision Support to Improve Dyslipidemia Treatment: A Cluster Randomized Controlled Trial. *Circulation* 117, 3 (Jan. 2008), 371–378. https://doi.org/10.1161/CIRCULATIONAHA.107.697201

[102] Srikanth Vasudevan, Anindita Saha, Michelle E. Tarver, and Bakul Patel. 2022. Digital biomarkers: Convergence of digital health technologies and biomarkers. *npj Digital Medicine* 5, 1 (March 2022), 36. https://doi.org/10.1038/s41746-022-00583-z

[103] Venkatesh, Morris, Davis, and Davis. 2003. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly* 27, 3 (2003), 425. https://doi.org/10.2307/30036540

[104] Athanasios Vourvopoulos, Ana Lúcia Faria, Kushal Ponnam, and Sergi Bermudez i Badia. 2014. RehabCity: design and validation of a cognitive assessment and rehabilitation tool through gamified simulations of activities of daily living. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology*. ACM, Funchal Portugal, 1–8. https://doi.org/10.1145/2663806.2663852

[105] Neff Walker, David E. Meyer, and JOhn B. Smelcer. 1993. Spatial and Temporal Characteristics of Rapid Cursor-Positioning Movements with Electromechanical Mice in Human-Computer Interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 35, 3 (Sept. 1993), 431–458. https://doi.org/10.1177/001872089303500304

[106] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing Theory-Driven User-Centric Explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–15. https://doi.org/10.1145/3290605.3300831

[107] Greta Warren, Ruth M. J. Byrne, and Mark T. Keane. 2023. Categorical and Continuous Features in Counterfactual Explanations of AI Systems. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. ACM, Sydney NSW Australia, 171–187. https://doi.org/10.1145/3581641.3584090

[108] Katharina Weitz, Dominik Schiller, Ruben Schlagowski, Tobias Huber, and Elisabeth André. 2019. "Do you trust me?": Increasing User-Trust by Integrating Virtual Agents in Explainable AI Interaction Design. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. ACM, Paris France, 7–9. https://doi.org/10.1145/3308532.3329441

[109] Yu Xiong, Runze Wu, Shiwei Zhao, Jianrong Tao, Xudong Shen, Tangjie Lyu, Changjie Fan, and Peng Cui. 2023. A Data-Driven Decision Support Framework for Player Churn Analysis in Online Games. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23)*. Association for Computing Machinery, New York, NY, USA, 5303–5314. https://doi.org/10.1145/3580305.3599759

[110] Fumeng Yang, Zhuanyi Huang, Jean Scholtz, and Dustin L. Arendt. 2020. How do visual explanations foster end users' appropriate trust in machine learning?. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. ACM, Cagliari Italy, 189–201. https://doi.org/10.1145/3377325.3377480

[111] Hanrui Zhang, Kaiwen Deng, Hongyang Li, Roger L. Albin, and Yuanfang Guan. 2020. Deep Learning Identifies Digital Biomarkers for Self-Reported Parkinson's Disease. *Patterns* 1, 3 (June 2020), 100042. https://doi.org/10.1016/j.patter.2020.100042

[112] Guojing Zhou, Xi Yang, Hamoon Azizsoltani, Tiffany Barnes, and Min Chi. 2020. Improving Student-System Interaction Through Data-driven Explanations of Hierarchical Reinforcement Learning Induced Pedagogical Policies. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '20)*. Association for Computing Machinery, New York, NY, USA, 284–292. https://doi.org/10.1145/3340631.3394848