Exploring User Needs and Solution Concepts for AI-Assisted Prostate Cancer Diagnosis

Combining Contextual Inquiry and Prototyping in a Design Thinking Approach

Master Thesis

in the study program Human-Computer Interaction

Kieu Tran

[redacted]

August 16th, 2023

at the University of Siegen

First examiner: Univ.-Prof. Dr. Volkmar Pipek

Second examinar: Univ.- Prof. Dr. Volker Wulf

Abstract

With the continuous advancement in artificial intelligence technology, its significance is growing within the medical field, particularly in radiology. However, achieving high algorithmic performance alone is not enough to offer valuable assistance to clinicians. This study adopts a human-centered approach to contribute to the development of impactful AI-driven medical solutions. Employing a design thinking process, this research comprises two consecutive studies aimed at understanding user needs in AI-assisted prostate MRI diagnosis. The first study involved a contextual inquiry to deeply comprehend real work practices, while the second study focused on creating a prototype based on insights from the first study, which was then evaluated with radiologists. The outcomes provide comprehensive descriptions and valuable insights that are crucial for designing effective AI support for prostate cancer diagnosis.

List of Abbreviations

ADC	Apparent Diffusion Coefficient
AFS	Anterior Fibromuscular Stroma
AI	Artificial Intelligence
CAD	Computer-Aided Diagnosis
CI	Contextual Inquiry
CZ	Central Zone
DCE	Dynamic Contrast Enhancement
DICOM	Digital Imaging and Communications in Medicine
DWI	Diffusion Weighted Imaging
EPE	Extraprostatic Extension
HIS	Hospital Information System
HMW	How-Might-We
JTBD	Jobs to Be Done
mpMRI	Multiparametric Magnetic Resonance Imaging
MRI	Magnetic Resonance Imaging
PACS	Picture Archiving and Communication System
PI-QUAL	Prostate Imaging Quality
PI-RADS	Prostate Imaging - Reporting and Data System
PSA	Prostate-Specific Antigen
PZ	Peripheral Zone
RIS	Radiology Information System
ROI	Region of Interest
T2W	T2-Weighted

TZ Transition Zone

XAI Explainable Artificial Intelligence

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Given the ongoing advancements in artificial intelligence (AI) research, incorporating AI-driven solutions into medical workflows is becoming increasingly relevant. The field of radiology, in particular, has been regarded as a promising arena for such integration, primarily due to its generation of substantial image data that can be harnessed for training and processed by machine learning (ML) algorithms. Additionally, the potential for AI to assist radiologists in managing the exponential surge in medical data volume and complexity while simultaneously enhancing diagnostic accuracy renders this technology highly appealing. Clinical trials have already showcased AI algorithms achieving performance levels on par with or surpassing those of medical practitioners in specific tasks (Q. Zheng et al., 2021). Consequently, many hold high expectations for AI's potential to reshape the radiology landscape, improving both clinician working conditions and patient outcomes.

While the efforts put into developing ever more sophisticated, AI models have shown promising results, high algorithmic performance alone is insufficient to provide clinicians with meaningful assistance. In reality, only a small number of AI-driven systems have transitioned from controlled laboratory environments to the realm of real-world clinical workflows (Cabitza, Campagner, & Balsano, 2020). The obstacles impeding the integration of AI into clinical practice are intricate and diverse, encompassing challenges such as limited access to appropriate data for model training (Cabitza et al., 2020), ambiguities concerning legal and policy frameworks (Strohm, Hehakaya, Ranschaert, Boon, & Moors, 2020), and a prevailing sense of skepticism toward automation (Cabitza, 2019). Amidst the array of considerations, one aspect that has often been overlooked is the user experience.

The field of human-computer interaction has long established the importance of a human-centered approach, countering the technocratic mindset often accompanied by cutting-edge technologies. Acknowledging the importance of the human experience within its contextual environment, a broader spectrum of factors relevant to the success of any technical solution is emphasized. Nevertheless, the socio-technical dimensions have been considerably underappreciated, particularly at the intersection of AI and medicine, where the focus lies on technical performance and clinical outcomes. However, not accounting for these factors has been demonstrated to factor into the non-adoption of technology in healthcare (Greenhalgh et al., 2018). In this context, this research aims to contribute to efforts to strengthen the human factor of AI-assisted medical tools.

This thesis was written within the research project PAIRADS (PAIRADS, 2023). The project is dedicated to integrating AI into radiological practice, focusing on creating a demonstrator for prostate cancer diagnosis. Employing a human-centered approach, the project involves radiologists in the development process, aims for explainable AI results, and leverages established radiological guidelines like the Prostate Imaging Reporting and Data System (PI-RADS) standard. In collaboration with Gemedico (Gemedico GmbH), PAIRADS aims to develop an AI solution to support radiologists with the diagnosis of prostate carcinoma with magnetic resonance imaging (MRI) data (Bijl, Blaumer, & Matuschek, 2022). The aim is to improve radiologists' workflow and enhance patient outcomes by seamlessly integrating AI solutions into radiology practices.

With a focus on promoting a human-centered approach, this thesis aimed to delve deeply into the work practices and explore the potential impact of AI-based solutions. Through detailed description and analysis, this study seeks to offer insights that can inform the design of AI solutions to facilitate seamless integration into the workflow and achieve effective radiologist support. Moreover, this work contributes academically to the expanding knowledge of Human-AI Interaction in healthcare.

The research was structured following a design thinking approach to reach this goal and comprised two consecutive studies. The initial study involved conducting a contextual inquiry (CI), accompanying three radiologists as they carried out their work practice. The primary aim of this phase was to obtain an in-depth understanding of the process, contextual factors, and practitioner needs within the domain. Subsequently, the second study involved developing a prototype of a potential AI solution, drawing from the

insights acquired from the previous study, and evaluating it with the radiologists. This phase enabled the collection of additional insights regarding user needs and other implications when introducing AI solutions to the practice.

The thesis is organized as follows: Chapter 2 offers an overview of the relevant medical, technical, and design elements that underpin this research. Chapter 3 explores related studies and pertinent research that delves into the interaction between humans and AI, particularly in the medical field and radiology. Next, Chapter 4 details the contextual inquiry (CI) study. Chapter 5 elaborates on the prototype's design and development, while Chapter 7 is dedicated to the evaluation of the prototype. Finally, Chapter 7 draws the thesis to a close, discussing its contributions and outlining avenues for future research.

Chapter 2

Background

2.1 Prostate Cancer Diagnosis in Germany

Prostate cancer is a prevalent and significant health concern in Germany, accounting for a considerable number of new cases and associated mortality. In 2018 alone, approximately 65.200 new cases of prostate cancer were registered, leading to approximately 15.000 deaths (Erdmann et al., 2021). This malignancy predominantly affects elderly men, with the median age of diagnosed patients at 71 years; and cases rarely occur before the age of 50 (Erdmann et al., 2021). In addition to age, genetic factors also play a role in increasing the risk of the disease, particularly among certain ethnic groups and individuals with a family history of prostate cancer (Rawla, 2019). These epidemiological factors underscore the importance of effective screening, early detection, and appropriate management strategies to address this significant health issue.

Early detection and risk stratification are key elements in effectively managing prostate cancer. By identifying prostate cancer at an early stage, it becomes possible to initiate timely and appropriate interventions that can improve patient outcomes. In contrast, delayed or missed diagnosis of prostate cancer can lead to the progression of the disease and potentially worse prognosis as well as the requirement of more aggressive treatments (Dunn & Kazer, 2011). On the other hand, some prostate tumors may exhibit indolent behavior and remain dormant, never causing harm within a patient's lifetime. Overdiagnosis and overtreatment of such cases would subject patients to unnecessary distress, including the risk of experiencing adverse treatment ef-

fects such as incontinence and erectile dysfunction. The clinical significance of a tumor hinges on various factors, including cancer stage, aggressiveness (how fast it grows and spreads), and its potential to invade and metastasize to other organs. Consequently, meticulous patient assessment and diagnosis are crucial in identifying the optimal strategy for each unique case (Loeb et al., 2014).

In response to the variability of prostate cancer and the need for accurate risk assessment, several diagnostic methods and tools have been developed to support the diagnosis and further management. This section introduces a selection of diagnostic approaches commonly used in prostate cancer evaluation. It is important to note that the inclusion of the specific methods and tools in this section does not represent an exhaustive list of all available diagnostic approaches for prostate cancer. Some various other methods and tools are utilized in clinical practice, and their omission from this section does not diminish their potential importance or clinical relevance. The selection presented here was specifically chosen due to their direct relevance to the research project and its objectives.

2.1.1 Prostate-Specific Antigen Testing

Prostate-specific antigen (PSA) testing is a widely used diagnostic tool for early prostate cancer detection and evaluation. PSA is a protein produced by the prostate gland, which at elevated levels, can indicate the presence of prostate abnormalities, including cancer. PSA testing involves a simple blood test, where a blood sample is taken from the patient and analyzed for PSA levels. If a certain threshold or conspicuous increase is measured, a prostate biopsy is recommended (Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF, 2021). However, the efficacy of PSA testing in improving patient outcomes has stirred controversy due to its inability to discriminate between low-risk tumors and clinically significant cancer, which can result in overdiagnosis and overtreatment (Pezaro, Woo, & Davis, 2014).

2.1.2 Biopsy

Biopsy plays a crucial role in prostate cancer diagnosis and risk assessment. It involves collecting and examining tissue samples from the prostate gland to determine if cancer is present and to characterize its aggressiveness. A

biopsy is typically recommended when there are indications of prostate cancer based on screening tests, such as elevated PSA levels or abnormal findings on imaging studies like mpMRI (Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF, 2021). The procedure is typically performed by a urologist, who extracts 10 to 12 small tissue samples using a thin needle inserted into various prostate gland areas. The collected tissue samples are then sent to a pathology laboratory for examination. The sample is analyzed to identify the presence of cancer cells, determine the Gleason score, which indicates the aggressiveness of the cancer, and assess other important features of the tissue. While biopsies are integral for the histopathological assessment of prostate cancer, the intervention can also lead to complications and false negative results (Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF, 2021).

2.1.3 Multiparametric Magnetic Resonance Imaging

Multiparametric magnetic resonance imaging (mpMRI) has emerged as a valuable diagnostic modality for prostate cancer (Murphy, Haider, Ghai, & Sreeharsha, 2013). Unlike traditional MRI, which relies on anatomical images alone, mpMRI combines multiple imaging sequences to evaluate the prostate gland comprehensively. These sequences include T2-weighted (T2W) imaging, diffusion-weighted imaging (DWI), and dynamic contrast-enhanced imaging (DCE). T2W imaging provides detailed anatomical information, allowing for the visualization of the prostate and surrounding structures. DWI measures the diffusion of water molecules within tissues and can highlight areas of restricted diffusion, which may indicate the presence of cancer cells. The apparent diffusion coefficient (ADC) map, derived from the DWI sequence, quantifies the diffusion characteristics of the prostate tissues, further aiding in the detection and characterization of lesions. Lastly, DCE imaging involves the injection of a contrast agent to assess the prostate's vascularity and blood flow patterns. Areas of increased enhancement may indicate regions of increased vascularity, suggesting the presence of tumors.

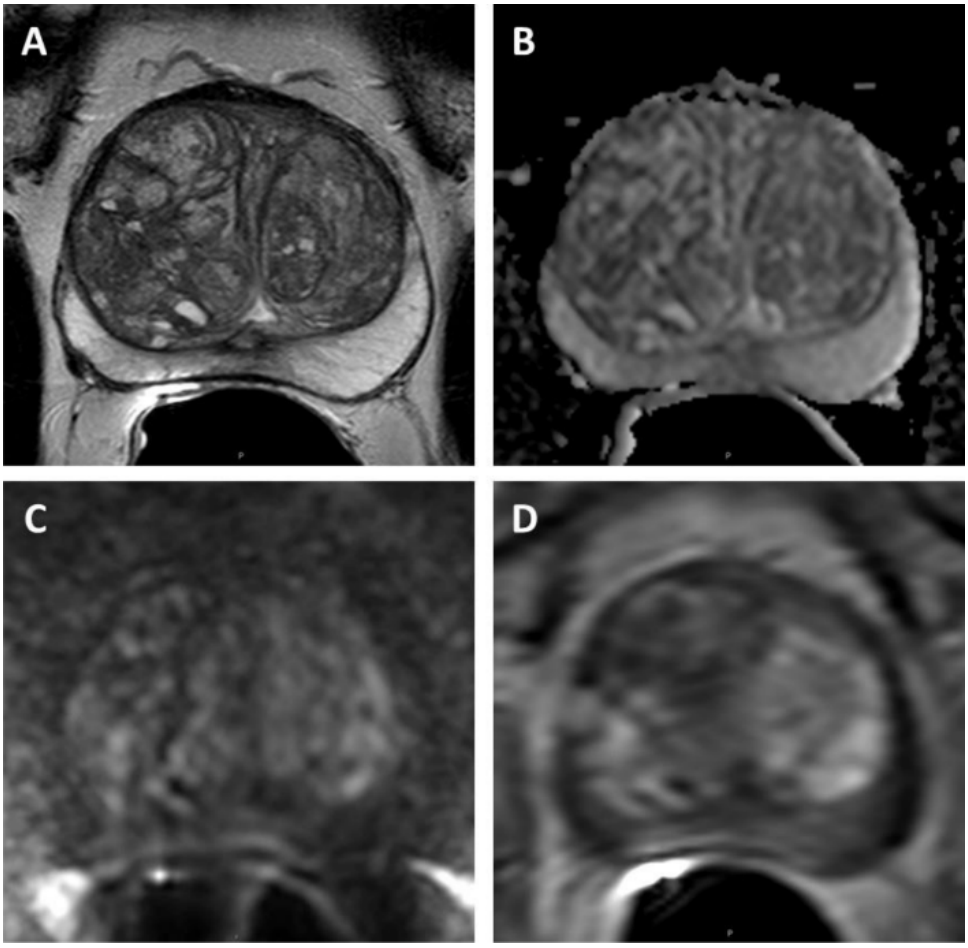


Figure 2.1: Images from a prostate mpMRI exam: A) T2W image, B) ADC map, C) DWI image, and D) DCE image (American College of Radiology, 2019).

Beyond detecting cancer lesions, mpMRI offers a wide range of applications in the context of prostate cancer diagnosis and treatment (Schlemmer, 2018). Among these applications are imaging analysis of suspicious lesions, supporting targeted biopsies, and conducting progress controls as part of active surveillance. The analysis of detected lesions is supported by diverse data available through the various imaging sequences. These sequences enable the determination of various lesion characteristics, including its position within the prostate gland, dimensions, and tumor stage. Acquiring these determinations gives critical insight for guiding subsequent treatment decisions.

Moreover, mpMRI data can be utilized during biopsies to guide the sampling

of additional areas from the prostate where suspicious findings were identified. Several techniques are available for this purpose. In a cognitive-targeted biopsy, the urologists mentally assess the targeted location within the gland based on the radiological report or the original MRI images. On the other hand, fusion biopsy involves the integration of MRI images with real-time ultrasound imagery, creating a precise alignment that directs the urologist to the designated target. Unlike cognitive-targeted biopsy, fusion biopsy necessitates specialized software. By harnessing these techniques, MRI images can support further interventions, improving the accuracy and precision of the diagnostic process.

Furthermore, mpMRI plays a crucial role within active surveillance protocols designed for patients diagnosed with low-risk prostate cancer (Barrett & Haider, 2017). Instead of immediate intervention, individuals under such protocols are subject to periodic monitoring to track the tumor's evolution. This approach enables postponing surgical treatment until a significant risk is identified, thus mitigating the overtreatment for benign findings.

These applications and many more have made mpMRI an invaluable tool in guiding treatment decisions and improving patient outcomes in the context of prostate cancer. Its ability to provide multi-dimensional and multiparametric information data through a non-invasive approach is a distinct advantage.

2.1.4 PI-RADS

Prostate Imaging Reporting and Data System (PI-RADS) is a standardized scoring system developed to assist in interpreting and reporting prostate mpMRI findings (American College of Radiology, 2019). It provides a structured framework for radiologists to assess and categorize the likelihood of clinically significant prostate cancer based on imaging characteristics. Moreover, the framework provides technical specifications for acquiring mpMRI images, a template for creating a structured report, and a lexicon defining relevant terms for a uniform language. Through its guidelines, PI-RADS aims to improve the consistency, accuracy, and communication of mpMRI results, ultimately leading to enhanced diagnosis, treatment planning, and patient care.

Since its introduction in 2012, PI-RADS has evolved through multiple versions. PI-RADS v1 established standardized reporting for prostate mpMRI,

while PI-RADS v2, released in 2015, introduced a simplified scoring system and guidelines for incorporating DCE and the ADC map. The most recent version, PI-RADS v2.1, released in 2019, further refined scoring categories, lesion characterization, and reporting requirements.

In order to consistently determine and communicate the importance of focal findings, PI-RADS employs an algorithm that assigns a PI-RADS score indicating the likelihood of a lesion being of clinical significance. This score is on a scale from one to five, where one signifies a highly improbable presence of clinically significant cancer, and five indicates a high likelihood. This scoring is established based on the lesion's assessment of key mpMRI sequences.

For instance, on the T2W sequence, a lesion's morphological characteristics determine its rating within the one-to-five scale. Signal hypointensity influences a rating within the same scale on the DWI sequence and ADC map. In the DCE sequence, focal enhancement presence or absence is noted as positive or negative, respectively. The PI-RADS score for a lesion is then determined according to these ratings. However, the sequence contributing to the score varies depending on the prostate zone in which the lesion is located. These zones encompass the peripheral zone (PZ), transition zone (TZ), central zone (CZ), and anterior fibromuscular stroma (AFS), although the scoring algorithm does not account for lesions in the CZ or AFS due to their rarity in causing cancer (American College of Radiology, 2019).

For lesions in the PZ, the DWI rating mainly determines the score, which can elevate from three to four if DCE is positive. Conversely, the T2W rating is adopted for TZ-lesions but changed from a three to four if the DWI rating is five. Additionally, any lesion score is increased from a four to a five if its maximal diameter is at least 1.5cm or if an extraprostatic extension (EPE), i.e., spreading to other organs, is identified. A detailed depiction of this algorithm can be observed in Figure 2.2.

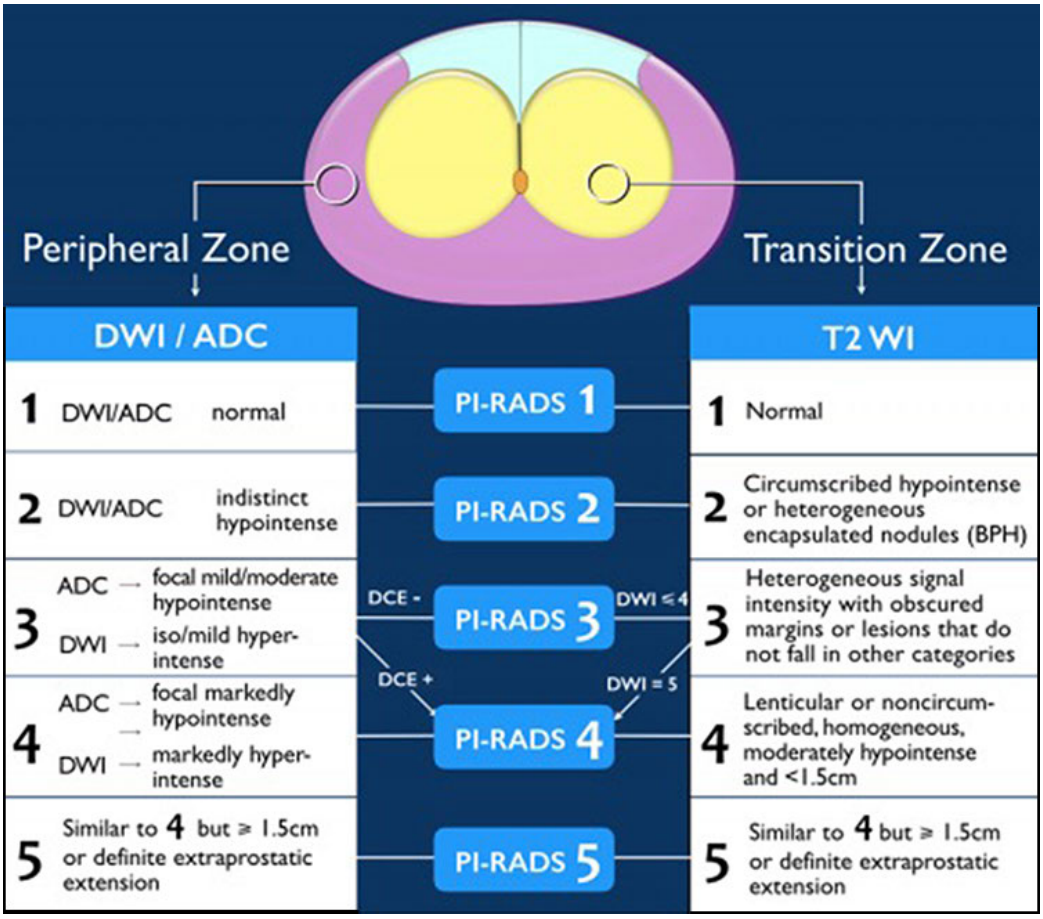


Figure 2.2: PI-RADS v2.1 scoring system (Panzone et al., 2022).

After assigning scores to each lesion identified in the mpMRI images, the overall PI-RADS score is determined, equivalent to the score of the highest-rated lesion. In general, PI-RADS v2.1 suggests a biopsy for scores exceeding three but not for scores below three. In the case of PI-RADS three, no recommendation is given since additional external factors need to be considered to assess the appropriateness of a biopsy.

Besides the scoring system, PI-RADS v2.1 provides guidelines for reporting prostate mpMRI findings. These encompass the reporting of the prostate volume and the PSA density. In the absence of computerized volumetrics, the document recommends utilizing the ellipsoid formula to compute the prostate's volume, which entails inputting the measurements of length, width, and height obtained from MRI images. Subsequently, PSA density can be determined by dividing the PSA level by the prostate's volume. This param-

eter serves as a valuable metric, as it evaluates the PSA value in relation to the prostate's volume.

Another essential guideline for crafting the report involves the mapping of potentially clinically significant lesions. To facilitate this, PI-RADS v2.1 introduces a sector map, which serves as a standardized graphical representation of the prostate gland (refer to Figure 2.3). This map effectively divides the prostate, seminal vesicles, and external urethral sphincter into 41 distinct regions, streamlining the communication of lesion locations. The provided recommendations include assigning up to four lesions with a PI-RADS score of three or higher on the sector map, as well as identifying the index lesion, which holds the highest risk of clinical significance. With the sector map, PI-RADS v2.1 introduces a visual aid that enhances the communication of lesion locations and facilitates more targeted diagnostic and therapeutic interventions.

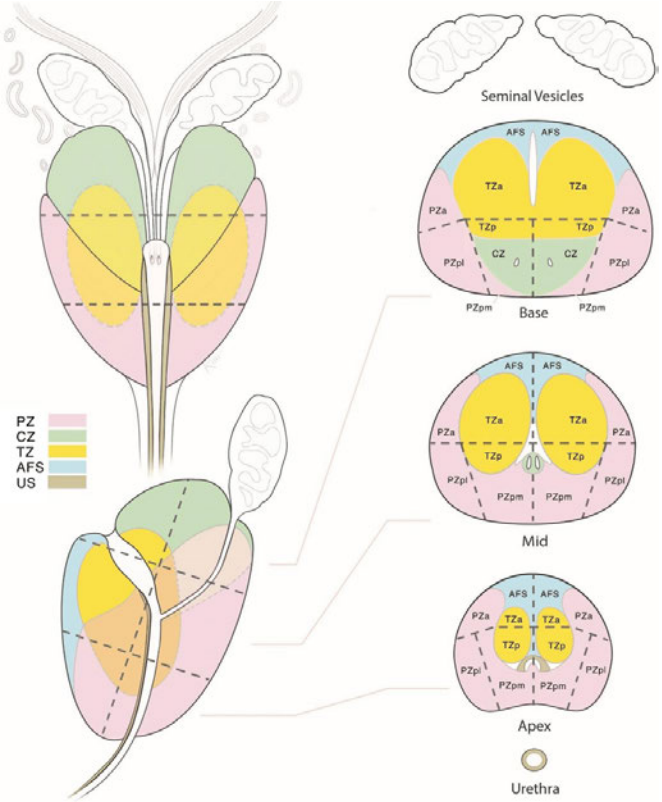


Figure 2.3: PI-RADS v2.1 sector map (American College of Radiology, 2019).

Furthermore, the PI-RADS paper provides a template for the written ra-

diological report. The recommended report structure encompasses multiple sections, including the indication for the MRI exam, the technical specification for the MRI image acquisition, a comparison to previous exams, findings, and the final impression. Through this structured approach, the authors aim to improve communication between healthcare professionals further.

In summary, the comprehensive guidelines and standardized scoring system of PI-RADS v2.1 have positioned it as an invaluable tool for the interpretation and communication of prostate findings on mpMRI images. Undoubtedly, the advantages it offers have driven the widespread adoption of the framework, solidifying its significance as a pivotal instrument within the field of radiology practice.

2.2 The Radiology Work Practice

The practice of radiology holds a vital position within the medical field and is characterized by distinct attributes in contrast to other medical specialties. To begin with, rather than engaging directly with patients, the responsibilities of radiologists predominantly involve interpreting visual data. In the acquisition and analysis of this data, the practice heavily relies on sophisticated scanning and imaging technology, resulting in a certain degree of dependency on these tools and the need to adapt to emerging technological advances. Additionally, radiologists are not directly responsible for patient care; instead, they provide comprehensive reports to referring physicians, furnishing them with invaluable insights into the patients' cases. These characteristics are important to consider for the implementation of new solutions. In the following, the core technologies, as well as general challenges within the work practice, are presented to provide additional context for this work.

2.2.1 Core Technologies

For the effective conduction of their day-to-day tasks, radiologists rely on a hand of fundamental technologies, including the Radiology Information Systems (RIS), Picture Archiving and Communication Systems (PACS), and the Digital Imaging and Communications in Medicine (DICOM) format. Radiology Information Systems (RIS) serve as databases and management systems responsible for managing patient-related information, scheduling examinations, and storing radiology reports. Functionally, RIS have similarities to

Hospital Information Systems (HIS), although RIS are designed to specifically cater to radiologists' requirements, whereas HIS are more broadly applicable and used by diverse hospital staff. However, when both systems are integrated within an institution, they are usually interconnected. Both facilitate the smooth flow of information within a department or institution, ensuring efficient patient management and seamless coordination of tasks.

Conversely, PACS is a technology infrastructure used to electronically store, manage, and distribute medical images and associated data. It includes storage servers, networking infrastructure, and image-viewing workstations. Images acquired via various modalities, e.g., MRI machines, are sent and stored on the PACS and can then be retrieved by the radiologist. It is important to note that PACS, RIS, and HIS are not singular software entities but rather categories of systems that encompass various software solutions from various vendors.

To support the interoperability and compatibility between different medical imaging devices and systems DICOM was introduced, which is a standard protocol for the storage of medical image data. The protocol allows various digital medical images such as X-ray or MRI images and associated metadata to be saved as a standard format, enabling the seamless exchange of medical data across various healthcare environments. The DICOM files can be viewed using so-called DICOM viewer software.

Together, these core technologies form the backbone of modern medical imaging, empowering healthcare professionals with advanced tools to provide accurate and timely diagnoses for better patient outcomes.

2.2.2 Challenges

The technology-dependent and referrer-based nature of radiology presents distinctive challenges for the practice. For example, the constant advancements in scanning technology have resulted in a surge in the volume, size, and complexity of medical imaging exams, leading to a growing demand for radiologists, who may experience heightened levels of stress and fatigue (Reiner & Krupinski, 2012). Additionally, due to the ever-growing complexity of the work environment and the multifarious responsibilities shouldered by radiologists, the field is especially vulnerable to the impacts of workplace disruptions (John-Paul, Kansagra, & Mongan, 2014).

Another significant challenge in radiology stems from the importance of effective communication between radiologists and referrers. The one-way communication flow inherent in radiological reports makes it essential to maintain a high standard of quality for effective information transmission to the referrer. As the primary medium of communication, the radiological report must be carefully crafted, adhering to multiple quality criteria such as completeness, brevity, and consistency, while still being delivered on time (Reiner, Knight, & Siegel, 2007). To address these challenges, standardized report systems such as PI-RADS have implemented a shift from free-text reports to more structured formats, thereby ensuring a certain level of quality and consistency in communication. However, despite these efforts, the presence of system errors and limitations can introduce an additional technical layer of technical challenges, further hindering seamless communication between radiologists and referrers (Larson, Froehle, Johnson, & Towbin, 2014).

2.3 Artificial Intelligence in Radiology

AI has made significant advancements in various fields, including radiology, and its application in medical imaging has shown promising potential (Sorantin et al., 2021). At the heart of AI in radiology lies Machine Learning (ML), a subfield of AI that focuses on building algorithmic models capable of learning patterns from data and generalizing their knowledge to make predictions on new, unseen data. Deep Learning (DL), a subset of ML, utilizes artificial neural networks with multiple layers to learn complex patterns and representations from vast amounts of data. Convolutional Neural Networks (CNNs) are a specific type of DL model widely used in medical image analysis, particularly in radiology, due to their ability to automatically learn hierarchical features from images (Yamashita, Nishio, Do, & Togashi, 2018).

Radiology presents an attractive domain for AI because it generates large volumes of image data that can be harnessed to train ML models effectively. Moreover, AI has the potential to address several challenges in radiology, including the increasing volume and complexity of medical imaging data, by providing efficient, qualitative, accurate, and reproducible automated image analysis and interpretation. In fact, a large range of application areas for AI within the radiological workflow have already been identified. At the same time, there are a large number of challenges that still lie in the way of successfully integrating AI solutions into clinical work practice.

2.3.1 Application Domains

AI has the potential to revolutionize radiology practice by offering valuable support to radiologists in various aspects and steps of their workflow (Choy et al., 2018; Hosny, Parmar, Quackenbush, Schwartz, & Aerts, 2018). With the ability to process large volumes of medical images rapidly and recognize intricate patterns, AI can be applied in several areas to enhance diagnostic accuracy, efficiency, and patient outcomes.

One of the primary applications of AI in radiology is the detection of abnormalities in medical images. While radiologists scan images for findings based on qualitative criteria, relying on their education and experience, ML algorithms can augment this process by leveraging quantitative patterns and data-driven insights. These algorithms have the capacity to analyze vast amounts of medical image data, identify subtle anomalies, and detect patterns that might elude human perception. Studies examining the performance of such AI models have been promising.

Another potential area in which AI could support radiologists is in the characterization of detected findings. Characterization refers to the determination of certain features of findings in the medical image, such as size, extent, and features. One example of the support in characterization is, for example, the segmentation of potentially cancerous tissue by CNN models. This segmentation of the tissue can then be utilized to assess the extent and stage of cancer.

In the realm of medical imaging, monitoring patients' conditions over time is a critical aspect of radiology practice. AI can play a significant role in monitoring disease progression or response to treatment. By analyzing sequential imaging data, AI algorithms can detect subtle changes in pathology, enabling early identification of potential treatment outcomes or disease progression. This continuous monitoring allows radiologists to assess treatment effectiveness more efficiently and make necessary adjustments to optimize patient care.

AI is not only valuable in the analysis of medical images but can also streamline radiology workflows. By automating routine tasks, such as image sorting, image pre-processing, and report generation, AI can free up radiologists' time, allowing them to focus on more complex and challenging cases. Additionally, AI can help prioritize urgent cases, ensuring that critical findings receive immediate attention, ultimately enhancing patient care and reducing wait times.

While the mentioned examples provide a glimpse into the diverse application areas of AI in radiology, numerous other possibilities have already been explored. These encompass optimized patient scheduling and image acquisition, image quality analysis, and automated report generation (Choy et al., 2018). Future innovations and technical advances will also likely lead to the emergence of further opportunities for AI to augment and streamline the radiology work practice.

2.3.2 Challenges

While the potential for AI to significantly enhance the field of radiology is significant, there are also some innate challenges associated with working with AI, especially in high-stakes environments such as radiology.

The most significant challenge when working with AI is its imperfect performance. Although AI models can demonstrate a high and consistent level of accuracy based on their training data, they are still subject to inherent limitations and uncertainties. These models rely on patterns within the data they were trained on and may not always account for unforeseen scenarios or edge cases. Moreover, if the training data is not representative or lacks diversity, the AI model may perform well in some instances but poorly in others. Furthermore, the trade-off between a model's sensitivity and specificity means that an AI may tend more towards false-positive than false-negative or vice-versa, which also needs to be accounted for by the user and perceived as inconsistent performance.

Closely related to the imperfect performance of AI is the challenge that comes from the human tendency towards over- or under-reliance. Striking the right balance between human judgment and AI recommendations is crucial to ensure optimal outcomes. However, radiologists have often shown not to show the appropriate level of trust towards automated results (Jorritsma, Cnossen, & van Ooijen, 2015). Over-reliance on automation may lead to the acceptance of erroneous results that would not have been made without the automated support (B. Zheng et al., 2001). On the other hand, too little trust in automation is likely to result in disuse, not taking advantage of the AI's power, and a sub-optimal performance overall.

Another issue that can amplify inappropriate trust is the black-box problem which refers to the challenge of understanding how AI algorithms arrive at their decisions or predictions. Many AI models, particularly deep learning

networks, can be highly complex and consist of numerous layers and parameters, making it difficult for humans to interpret their inner workings. This lack of transparency can be concerning, especially in critical applications like healthcare, where the consequences of errors can be significant. The inability to fully comprehend AI's decision-making process can lead to mistrust and hinder the broader adoption of AI technologies.

2.4 Design Thinking

Design thinking is a problem-solving and innovation approach that emphasizes empathy, creativity, collaboration, and an iterative process to develop user-centered solutions (Grots & Pratschke, 2009). It revolves around first understanding and empathizing with the needs, wants, and pain points of the end-users, typically through user research. This human-centric approach helps uncover valuable insights and allows framing the problem from the user's perspective. This understanding then informs the creative process of generating ideas for potential solutions. For this, design thinking encourages a multidisciplinary and inclusive environment where diverse perspectives are welcomed, fostering creativity and the potential for groundbreaking innovations. The approach also acknowledges that solutions will not be perfect right away and therefore consist of an iterative process of ideation, prototyping, testing, and refining solutions. The focus here is on generating multiple ideas, thinking "outside the box," and encouraging a culture that embraces experimentation and learning from failure. By prioritizing user needs and iteratively refining solutions based on user feedback, design thinking ensures that the final outcome addresses real-world challenges effectively and resonates with the target audience.

Although the principles of design thinking are widely accepted, there is no universally standardized procedure, and multiple models illustrating the process are available (Grots & Pratschke, 2009; Wolniak, 2017). The following section will introduce the Double Diamond framework, which served as a structural guide for this research endeavor.

2.4.1 Double Diamond

The Double Diamond is a widely used design thinking framework that emphasizes a structured approach to problem-solving and innovation (Design

Council, 2023). Its name derives from the two diamond shapes that symbolize the diverging and converging process while traversing the problem and solution space (see Figure 2.4). The first diamond is situated in the problem space, which encompasses all aspects relevant to a specific challenge. On the other hand, the second diamond resides in the solution space, which entails the range of potential solutions to a specific problem. Each diamond consists of a diverging stage and converging stage, where the space is first broadened to many elements and then narrowed down to one or few ones (Productboard, 2023). The resulting four phases are referred to as discover, define, develop, and deliver.

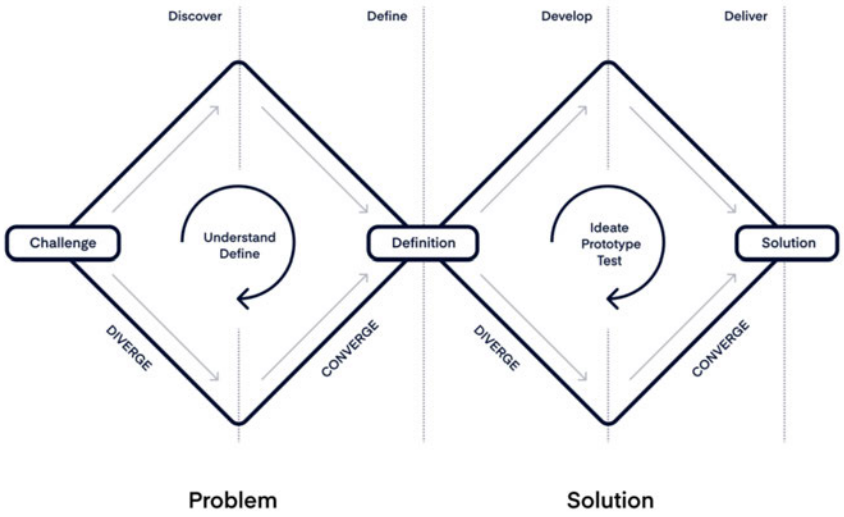


Figure 2.4: Double diamond model (Productboard, 2023).

1. **Discover:** The first phase aims to explore and get an extensive understanding of the problem space. The problems are diverged on by examining its various aspects, such as stakeholders, contextual factors, and pain points experienced by the target audience. Usually, this is achieved through explorative research methods such as interviews and observational studies.

2. **Define:** In this stage, the gathered research insights are synthesized and converged into a concrete definition. By selecting specific aspects within the problem space to concentrate on, a clear direction can be established following steps within the solution space.

3. **Develop:** Once the problem has been defined, the third stage consists

in generating ideas and exploring different solutions. This ideation phase encourages experimentation, collaboration, and the exploration of diverse possibilities to find the most innovative and viable solutions. The creative process also involves developing the ideas into testable prototypes.

4. **Deliver:** In the final phase, the diverse concepts are narrowed to a single solution through evaluation. By examining the performance and reception of the diverse ideas in user testing, an understanding emerges regarding how effectively these solutions address the specified problem, which facilitates subsequent selection and refinement of the concepts.

In summary, the Double Diamond framework provides a systematic and structured approach to design thinking, enabling the navigation through complex challenges with empathy, creativity, and human-centricity.

2.4.2 Application in Health Care

Design thinking has also gained prominence as an effective approach across diverse healthcare domains, offering a flexible framework for generating innovative and user-centered solutions (Ku & Lupton, 2022). Its effective application in the complex medical environment is attributable to its careful consideration of contextual factors, including the needs of users, stakeholders, and available resources, along with the integration of clinical evidence (Oliveira, Zancul, & Fleury, 2021). In fact, the outcomes of design thinking in healthcare have proven to be superior in terms of usability, satisfaction, and overall effectiveness compared to traditional interventions in multiple cases (Altman, Huang, & Breland, 2018). Seeing that usefulness and ease of use have been demonstrated to be critical factors for the adoption of medical technologies (Gagnon, Ngangue, Payne-Gagnon, & Desmartis, 2016), Design Thinking offers an appealing method to increase the likely hood of success.

Chapter 3

Related Work

3.1 Human-AI Collaboration

While there have been speculations on the possibility for AI to eventually completely replace the role of the radiologist, most approaches have focused on ways in which AI can be used as a complementary tool to augment the radiologist (Sorantin et al., 2021). This strategy rests upon the premise that human oversight remains essential and that human and artificial intelligence can synergize effectively. While AI excels in rapidly processing vast datasets and identifying nuances evading the human eye, the radiologist's strength lies in domain expertise and the ability to identify nearly optimal solutions (Sorantin et al., 2021). Moreover, radiologists can factor in additional variables unavailable to AI, such as image data across modalities, medical history, and patient-specific circumstances influencing the appropriateness of individual treatments. This collaborative approach not only holds the potential for improved performance but also elicits greater acceptance from practitioners. Nevertheless, as the tendency to place inappropriate trust in the automated results demonstrates, facilitating and refining human-AI collaboration presents its own challenges, a focal point of ongoing research.

One crucial element that has emerged as highly significant for establishing effective collaboration with AI is the accuracy of the user's mental models of an AI algorithm (Bansal et al., 2019). To harness the synergistic potential of human-AI collaboration, it is crucial to leverage their individual strengths in appropriate contexts. For example, in a scenario where an AI might yield inaccurate results, users might benefit from relying on their own expertise.

However, to correctly asses such situations, users must possess a comprehensive understanding of the AI's strengths and limitations. Here, one major issue that has been pointed out is that both actors in the radiologist-AI interaction tend to operate independently of each other, without taking into account their individual capabilities and biases (Nishikawa & Bae, 2018).

Addressing this challenge involves enhancing users' comprehension of algorithmic functionality. Notably, Bansal et al. (2019) have demonstrated that comprehending the error boundaries of AI models enhances performance, emphasizing the need for models with easily understandable error boundaries. Furthermore, grasping AI's limitations brings the advantage of aligning expectations, ultimately leading to increased satisfaction and acceptance (Kocielnik, Amershi, & Bennett, 2019).

In addition to improving the interpretability of ML models, various strategies aimed at enhancing user understanding prior to the interaction. Cai et al. (2019), for instance, have revealed clinicians' preference for receiving insights into diagnostic algorithms prior to utilization to better asses result accuracy. The physicians in their study evaluated AI within the context of their collaborative mental models, akin to seeking a second opinion. In response to this, the suggestion of clinician training in ML has emerged to counteract susceptibility to AI biases (Rubin, 2019). A deeper technical understanding could also mitigate interpreting AI performance that is more inclined towards sensitivity or specificity as inconsistent (Strohm et al., 2020).

Another approach to conveying the limitations of an AI model is through communication via the output. For example, displaying the result confidence level has been shown to enable users to calibrate their trust in the results proposed by an AI (Jorritsma et al., 2015). Another effective strategy involves attaching explanatory demonstrations to results, alerting users to instances where lower performance might be expected (Cabrera, Perer, & Hong, 2023). Additionally, Cai et al. (2019) demonstrated the benefits of giving clinicians the possibility to refine AI output. Their study involved an AI prototype that supported pathologists in the examination of organic tissue by providing similar images based on the reference image and variables adaptable by the user. The study highlighted how participants utilized the refinement options in order to experiment with the algorithm and thereby incrementally enhanced their understanding of the underlying mechanisms of the model.

While most of this work has focused on the effects of human-AI collaboration on the combined performance, less attention has been put on the user

perspective and what their needs are for an effective collaboration

3.1.1 Human-Centered Explainable AI

In response to the escalating complexity and opacity of modern DL algorithms, the field of explainable AI (XAI) has gained significant relevance in the integration of AI into medicine. XAI is a field of research and development focused on creating AI systems capable of providing human-understandable explanations for their decisions and actions (Gunning et al., 2019). While the demand for transparent and interpretable AI models may not be crucial in all applied areas, it holds great significance in high-stakes environments such as radiology, as it allows users to validate and comprehend the rationale behind AI-generated outcomes. However, most research in XAI has focused on technical implementation and has neglected the role of the user receiving explanation (Adadi & Berrada, 2018). This algorithm-centric approach was criticized because the development and integration of the XAI techniques were mostly based on the needs and intuition of the programmers and therefore run the risk of being of less value to the end-user (Miller, Howe, & Sonenberg, 2017). To address this issue, parts of the research community have called for a shift towards human-centered XAI (Liao & Varshney, 2021).

One important principle behind human-centered XAI is that the quality of an explanation is dependent on how well it serves the person who receives it. To underline this point, the distinction between explainability and causability has been proposed for the field of medicine (Holzinger, Langs, Denk, Zatloukal, & Müller, 2019). While the term explainability is tied to the information within the explanation, causality refers to its ability to achieve effective, efficient, and satisfactory understanding with the medical expert. This underscores the necessity to concentrate not solely on the content of the information to be conveyed but also on the manner in which it should be presented.

With an increasing focus on the recipient of explanations, the necessity to comprehend user needs for improved explanation implementation has come to the forefront. Traditional XAI techniques were primarily designed to aid AI researchers in evaluating models, often overlooking the end-user and the contextual environment in which the explanations would be delivered (Miller, 2019). This presents a concern, given that the end user's requirements for explanations might differ from those of a data scientist. In response, frameworks have been proposed that discern between diverse needs contingent

upon stakeholders (Hong, Hullman, & Bertini, 2020). A more advanced approach, as introduced by Suresh et al. (2021), takes into account stakeholders' requirements based on their knowledge and objectives.

Adopting a more bottom-up perspective on comprehending user explanation needs, some researchers have introduced alternative approaches. Liao et al. (2021), for instance, has developed a research and design process with the objective of matching the right XAI technique to suitable users. Their approach involves defining user needs in terms of questions that must be answered to comprehend the model's output effectively. For example, users might seek to understand how the system's predictions change when specific features of the input are altered. By translating these derived inquiries into existing XAI solutions, designers can implement and assess appropriate design strategies through an iterative process.

The authors also emphasize the breadth of data that could be pertinent to users, extending beyond the model itself to encompass aspects like the training data utilized. Their research revealed a wide spectrum of user questions influenced by multiple external and internal factors (Liao, Gruen, & Miller, 2020). This variability was also observed in a study conducted by Calisto et al. (2022), where radiologists exhibited a greater need for explanations with moderate cases rather than low or high cases in breast screening interpretation.

Another human-centric approach, suggested by Eiband et al. (2018), centers on understanding how users' comprehension deviates from an optimal understanding of the AI model, thus identifying areas that warrant explanations. This research process, as well as the subsequent iterative design phase, involves the active engagement of a variety of stakeholders.

The presented human-centered XAI research demonstrates the efforts that are being put into bringing the human aspect into the development of AI-based software. Furthermore, it underlines the importance of gaining a more profound insight into user requirements within their specific contexts of practice.

3.2 Workflow Integration

The disparity between the promising performance of AI models in medical tasks and the limited adoption of these solutions in actual medical practice

has been highlighted by numerous researchers (Cabitza et al., 2020). Given that the value of AI support in the medical field is only realized through successful integration into practice, researchers have devoted their attention to identifying barriers and potential solutions. Much of these endeavors have been directed toward socio-technical factors, which play a pivotal role in the integration of software solutions. For instance, an ethnographic study conducted by Greenhalgh et al. (2018) delved into the reasons for the non-adoption of technology-supported interventions in healthcare. This study unveiled an array of complexities often overlooked, stemming from diverse sources like the unpredictable nature of diseases, the need for seamless integration within existing IT structures, and the challenges posed by new work protocols.

To tackle these hurdles, an approach rooted in human-centered principles is advocated by the authors. This approach would encompass the motivations, values, and clinical norms of the various stakeholders, aiming to address challenges and promote successful implementation. Moreover, it emphasizes the importance of ongoing learning from usage and refining implementation for continued efficacy. Additionally, the support for the appropriation of IT solutions is underscored by the design case study framework proposed by Wulf et al. (2015), which has been structured to encompass socio-technical dimensions through the active engagement of end-users.

Indeed, relying solely on predominately used clinical trials for evaluating technologies designed to support clinicians in their tasks has been deemed insufficient. Complementary to such trials, qualitative assessments of the solution's impact on everyday practice have been advocated. In an ethnographic study exploring the potential of automatic detection to replace the second reader in breast cancer imaging, Hartswood et al. (2003) underscored the software's limitations to be 'socialized' within established practices. Conversely, qualitative evaluations conducted within the actual context have the potential to uncover unexpected challenges. For instance, in the assessment of AI-supported retina diagnosis in Thai hospitals, Beede et al. (2020) identified a range of issues tied to social and environmental factors. Notably, nurses utilizing the AI tool cautioned certain patients against an AI-assisted examination due to potential hardships associated with a system's recommendation. In scenarios of immediate referrals, a process that conventionally takes multiple days, patients would be required to travel to a different hospital. Unfortunately, this was often unfeasible due to their individual circumstances.

Another important aspect in regard to the adoption of technical solutions is the integration into the existing IT systems and infrastructure. Especially in the field of radiology, which heavily relies on the interplay between multiple different technologies, the way in which new solutions are integrated is essential. In their guidelines for selecting appropriate AI solutions for radiology, Omoumi et al. (2021) mention a lack of interoperability as a significant pitfall. The ineffective support for the transfer of medical data between various systems can substantially impede workflow efficiency. In response, the authors propose a range of potential remedies, including opting for tools intractable within established systems like the PACS, rather than deploying standalone solutions, embracing interoperability standards, and harnessing cloud technologies.

Beyond technical integration, effectively assimilating new solutions into the established clinical workflow without causing undue disruption represents another crucial factor to be taken into account. An extensively explored aspect of research revolves around determining the optimal point at which the diagnostic AI output should be introduced to the radiologist's process (Cabitza, Campagner, & Sconfienza, 2021). With regard to human-AI performance, indications suggest that providing immediate AI prompts to diagnosticians might negatively influence their image analysis, possibly leading to the overlooking of findings that would have been detected without AI assistance (Alberdi, Povyakalo, Strigini, & Ayton, 2004). Conversely, research findings demonstrate that a two-step workflow, involving the provision of AI output only after the radiologist has independently conducted a diagnosis, resulting in dissatisfaction with the AI system (Fogliato et al., 2022). This outcome is attributed to the additional cognitive effort required to halt and re-evaluate an already concluded diagnosis in light of the new information. This challenge underscores the intricate interplay and potential trade-offs between optimal performance and seamless integration within the clinical workflow.

In general, usability has been identified as an important aspect of the integration of medical software into the workflow and overall adoption. While usability is seen as a universally favorable characteristic it is especially critical in a high-stakes environment such as healthcare. This was clearly underscored by Ratwani et al. (2018), whose study demonstrated that usability issues within electronic health records directly contributed to errors in drug dose administration. Furthermore, usability holds the potential to enhance day-to-day workflow. In their recommendations concerning effective clinical decision support systems, Bates et al. (2003) emphasize the multifaceted importance of usability, encompassing aspects like swift information delivery

and maintaining visibility within a single screen. Much like other domains, achieving usability in medical technology can be facilitated through the application of human-centered design methodologies.

Moreover, user satisfaction and acceptance have been proven to be positively influenced by user control. In an assessment involving a breast cancer lesion detection prototype, the incorporation of an option to accept or decline AI recommendations yielded a notable boost in acceptance rates (Calisto et al., 2022). Likewise, the introduction of a comparable feature to radiologists generated similar results (Blezek, Olson-Williams, Missert, & Korfiatis, 2021). Here, the possibility to accept, reject, or modify results was praised, as it relieved radiologists from the task of rectifying or removing inaccurately post-processed images that were directly stored in patient records.

Chapter 4

Study 1: Contextual Inquiry

This chapter presents the contextual inquiry (CI) study conducted as the first step of this research project. In the context of the entire design thinking double-diamond process, this study fills the role of the first diamond, discovering the targeted users and their work practices and defining their goals, circumstances, and needs.

4.1 Methodology

The methodology employed in this study aimed to gain a deep understanding of the current work practices of radiologists involved in prostate cancer diagnosis on MRI screenings. Specifically, the study sought to explore the individual steps within the radiologists' workflows, the contextual factors influencing the practice, and the needs that the practitioners have throughout the procedure. This comprehensive analysis aimed to establish a solid foundation of understanding, serving as the basis to inform the development of a prototype solution for a subsequent evaluative study.

To this end, the research employed contextual inquiry (CI) (Beyer & Holtzblatt, 1999) as the chosen method for the initial data collection. CI can be characterized as a condensed version of an ethnographic field study. It entails both observing participants engaged in their practice and conducting interviews to acquire a comprehensive and in-depth understanding. By observing participants performing in their natural environment, it becomes possible to capture elements that might be so ingrained in their routines that they

are subconscious. Conversely, the impromptu interviews enable researchers to capture the underlying motivations as well as tap into participants' expertise to explain concepts. The time-efficient yet thorough nature of CI has contributed to its widespread adoption for shaping product design. This methodology has also demonstrated its effectiveness in influencing the development of medical devices by considering critical factors within clinical work practices (Privitera, 2015).

Participants were invited to join the contextual inquiry (CI) study through a targeted recruitment process. An invitation was sent to a list of radiologists with whom the PAIRADS research project team had established prior contact and who were deemed suitable for the study. Of the invited radiologists, three agreed to participate and were selected as the study's participants. The participants' demographic and professional information is provided in table 4.1 and was as follows: Participant 1 (P1), aged 50-55, worked in a private practice, possessed 20-25 years of overall radiology experience, and had 10-15 years of specific experience with prostate MRI diagnosis. Participant 2 (P2), also aged 50-55, works in a private practice and had 20-25 years of radiology experience, with 5-10 years focused on prostate MRI diagnosis. Participant 3 (P3) was aged 35-40, working in a public hospital, and had 10-15 years of radiology experience, including 5-10 years of experience with prostate MRI diagnosis. All participants were male and practiced their profession in Germany. The composition of these participants promised a diverse range of experiences and perspectives within the field of radiology, particularly concerning prostate cancer diagnosis, and from the work environment of different health institutes.

ID	**Age**	**Gender**	**Institution**	**Radiology Experience (in years)**	**Prostate MRI Experience (in years)**
P1	50-55	Male	Private practice	20-25	10-15
P2	50-55	Male	Private practice	20-25	5-10
P3	35-40	Male	Public hospital	10-15	5-10

Table 4.1: List of participants

The scheduling of the CI days was strategically planned to optimize the efficiency and value of each visit. Given that prostate MRI exams were not conducted on a daily basis by any of the participants, it was important to maximize the opportunities for data collection during these visits. To

achieve this, multiple prostate MRI exams were scheduled on the days of visit, allowing for concentrated and productive CI sessions. This ensured that the CI days were utilized to their fullest potential and allowed the researcher to gather insights from a larger and diverse range of work practice samples. In the end, at least four prostate MRI examinations were scheduled on each day of visit.

The contextual inquiries for this study were conducted at the workplace of the radiologists, providing a firsthand understanding of their clinical environment and practices. Each CI session extended throughout one entire workday, allowing for a comprehensive exploration of the radiologist's activities, tasks, and interactions. To capture a detailed record of the observations, the screens used by the radiologist during the examination process were video recorded with a strategically placed camera. To better get a better understanding of the "what" and "why" behind the work practices, the participants were asked to provide explanations of their actions and decision-making processes during their initial diagnosis. Additionally, semi-structured interviews were conducted to address a combination of pre-defined questions and spontaneous questions that arose during the observation process throughout the sessions. These interviews served as opportunities for the researcher to engage in direct conversation with the radiologist, seeking further explanations, elaborations, or context on specific actions, decisions, or observations made during the CI. To ensure a comprehensive and accurate preservation of the information shared, the interviews were audio recorded. Detailed notes were also taken during the contextual inquiries, documenting key observations, important interactions, and notable insights.

An iterative approach was taken throughout the contextual inquiry studies to ensure a comprehensive understanding of the radiologists' work practices. After each session, the data collected, including video recordings, audio recordings, and notes, were carefully reviewed. This analysis process allowed the identification of areas that required further clarification or exploration. These identified areas were then addressed in subsequent sessions, either through additional observations or targeted follow-up questions during the spontaneous interviews throughout the CI. Additionally, more thorough follow-up interviews were conducted with two of the participants a few days after the respective CI. These follow-up interviews served the purpose of clarifying any remaining open questions and obtaining further insights into specific aspects of their work practices. However, it should be noted that the third participant was not interviewed separately as they had recently been interviewed by the research team of the ongoing research project. Out of respect for

his time, it was decided not to conduct an additional interview with this participant.

For the translation of the collected data into valuable insight the Jobs-to-be-Done (JTBD) framework (Christensen, Hall, Dillon, & Duncan, 2016) was utilized. JTBD is a customer-centric approach that focuses on the goals that people aim to achieve. These goals are depicted as jobs for which people 'hire' products or solutions in order to get them done. For example, in order to check for irregular heartbeats in a patient (job) a physician (job performer) may use or 'hire' a stethoscope (solution). Moreover, the JTBD framework emphasizes the importance of the circumstances in which the job is conducted and the needs of the performer associated with it. As a tool for analysis, JTBD is particularly useful due to its structured approach. Various aspects of a goal are separated into individual components and are formulated following a consistent structure (Kalbach, 2020). The overall aim of the job performer is defined as the main job, which can be subdivided into a series of smaller jobs that need to be completed in order to achieve the main goal. These jobs are formulated as job statements, e.g., "Detect irregular heartbeats in a patient." The contextual factors influencing the job execution are documented as circumstance statements, e.g., "When doing a home visit." Lastly, the needs of the job performer are captured as outcome statements, which describe the desired result, e.g., "Minimize the patient's discomfort." An advantage of how the statements are formulated is that they are solution agnostic. For instance, none of the statement examples required a solution to be mechanical (e.g., a stethoscope) or electric (e.g., an electrocardiogram). This is particularly valuable for informing the design of innovative solutions making JTBD a fit tool for this study.

To translate the observations, interviews, and notes into valuable JTBD statements, a systematic analysis of the collected data was conducted. After transcribing the interviews and audio data from the video footage, the data underwent a systematic coding process, where themes and patterns were identified. The coded data was then organized using affinity diagramming to formulate job, circumstance, and desired outcome statements.

4.2 Results

This section presents the job statements, circumstance statements, and outcome statements identified as part of the radiologist's prostate MRI diagnosis

process. Each item is thoroughly described, supported with detailed observations and participants' statements, providing contextual examples to enhance understanding.

4.2.1 Process

The data analysis of the collected data resulted in multiple sub-goals or steps, which were transformed into job statements. These job statements are presented in Table 4.2 for reference. The job statements presented are organized based on the sequential order in which the radiologists typically performed them during the prostate MRI diagnosis process. However, it is important to note that while there is a general order, it is not strictly linear due to individual differences among the participants and the flexible nature of the diagnostic workflow. The radiologists may adapt their approach based on various factors, such as the specific case, patient history, or their own preferences. Therefore, the order of the job statements reflects a typical sequence observed, but it is important to recognize the inherent variability in practice.

To provide a comprehensive understanding of the participants' diagnostic approaches and to ensure that the subsequent descriptions are well-contextualized, it is important to mention the two general approaches observed during the diagnosis process. Among the participants, P1 and P3 primarily conducted their diagnoses within the DICOM-viewer, while P2 followed a different path, utilizing a specialized workstation program designed for prostate diagnosis. This program provided a guided structure that the radiologist followed step by step. Although this approach presented an apparent distinction from the other two participants, the individual job or sub-goals remained largely consistent. Naturally, some individual differences were observed between all participants to a certain extent.

MJ: Provide the referrer with diagnostic insight based on prostate MRI data and in regards to the medical question. The overall goal of the radiologist when performing a prostate MRI diagnosis is to provide the referring physician with accurate, relevant, and timely data to support them in the subsequent treatment and management of their patients in regard to their prostate. This may include information on any suspicious findings, the development of past-diagnosed conditions, recommendations on follow-up steps, and data supporting the conduction of interventions, such as biopsies or operations. To do so, the radiologists relate to the medical question pro-

ID	Job Statements
MJ	Provide the referrer with diagnostic insight based on prostate MRI data and in regards to the medical question
J1	Access the medical data
J2	Understand the clinical context
J3	Get a general overview of the images
J4	Assess the image quality
J5	Determine the prostate volume
J6	Determine the PSA density
J7	Determine anatomical characteristics of the prostate
J8	Detect suspicious focal lesions
J9	Characterize detected lesions
J9.1	Determine the boundaries
J9.2	Determine the location
J9.3	Determine the size
J9.4	Determine the ADC value
J9.5	Determine visual characteristics
J9.6	Determine local spreading
J10	Assess the significance of detected lesions
J11	Prioritize detected lesions
J12	Detect and assess metastases
J13	Detect and assess additional findings
J14	Make an overall assessment
J15	Generate the radiological report
J15.1	Generate the structured graphic
J15.2	Generate the written report
J16	Share the report

Table 4.2: List of job statements

vided by the referrer. In this role, the radiologists act as a kind of service provider for other physicians.

In this particular work practice, the radiologist assimilates the relevant information from acquired MRI images, complemented with additional patient data. In using their specialized skills and knowledge, they assimilate the available information in a radiological report, which is then shared with the referring physician. Getting to this point requires multiple steps, which will be described in the following.

J1: Access the medical data. The initial step in the work process is to retrieve all the available data relevant to the assessment and evaluation of the MRI exam. This naturally includes the MRI images but also other information such as the medical questions, clinical notes, PSA levels, biopsy results, medical history, and previous diagnoses and treatments. During the CI, the participants retrieved this data from various means. The MRI images from present and previous exams were accessed on the DICOM viewer via the PACS, which were either acquired during an on-site MRI scan or received from external sources. In P2's case, the current MRI images were also displayed on the specialized workstation software. Much of the other information was retrieved from the RIS. Here, the participants had access to scanned documents, such as the patient's referral slip, which contains the medical question, laboratory results, and questionnaires that the patients had to fill out before the MRI examination. Other data and digital documents were also available in the RIS, e.g., previous reports. Additionally, P1 accessed details on the patients' PSA levels via the online booking application Doctolib (Doctolib SAS), where the referrer can input the PSA test results when booking an appointment. P2 also collected further patient data, such as medical history, during the consultation with the patient. During one assessment, P3 was handed the physical referral slip and other documents by his chief medical doctor.

J2: Understand the clinical context. Once the medical data was available, the radiologists familiarized themselves with the clinical context before assessing the MRI images. This included reviewing the medical question, the reasons for the examination, i.e. the medical indication, PSA levels, reports from preliminary exams, and laboratory results. This preliminary assessment helps guide the subsequent image interpretation and ensures that the radiologist takes into account the specific clinical concerns and objectives. Furthermore, the information gives additional context when analyzing the imaging findings, helping to achieve a final conclusion. *(P3:) "At the end*

of the day, all the parameters, I always do that in the back of my mind also when I then make the findings, these things of course always rattle along and so an overall picture then also arises in my head."

J3: Get a general overview of the images. Before analyzing the MRI images in more detail, the participants tried to get a more general overview of the examination to gain an initial understanding of the overall anatomy and any notable abnormalities or findings. The purpose is to familiarize themselves with the imaging data, identify any gross abnormalities, and develop a mental framework for subsequent detailed analysis. This process involved scrolling through the image series, examining different imaging sequences, and evaluating different planes. *(P1:) "That means I first scroll through the examination without any expectations to mentally prepare for what I have to deal with. And then I see immediately that he has a carcinoma. Well, but now I don't go into more detail on this [...]"*

J4: Assess the image quality. A crucial step in the assessment of the MRI examination is the evaluation of the image quality of each sequence and how it might affect the image interpretation. In doing so, the radiologist evaluates if the images provide sufficient diagnostic data for a reliable diagnosis or if the screening process needs to be repeated. The documentation of any limiting factors, e.g. visual artifacts, in image quality is also important for the communication to the referring clinician as it puts the results of the diagnosis in the appropriate context. *(P1:) "So the diffusion is very, very susceptible to artifacts. The T2-weighted images are much less susceptible to artifacts and of course, you have to call it out and say 'Okay, the diffusion-weighted images are massively limited due to so and so. That means the informative value, especially regarding the peripheral zone, is very limited.' You have to communicate that somehow."* To assess the image quality the participants scrolled through the available images, taking note of several factors such as the signal-to-noise ratio or the presence of visual artifacts. The overall quality was then given a qualitative score, e.g. "very good".

J5: Determine the prostate volume. The determination of the patient's prostate volume is an important step in the diagnosis for several reasons. Monitoring changes in prostate volume over time can be useful in tracking disease progression or response to treatment. Furthermore, prostate volume plays a key role in calculating important parameters, such as PSA density, which are essential in evaluating the likelihood of prostate cancer presence and its potential aggressiveness. All participants in the study employed the ellipsoid formulation to calculate the prostate volume. The measurements of

the length, width, and height of the gland were taken on the T2W images in transversal and sagittal planes. P1 and P3 conducted these measurements directly within their DICOM viewer software. They then manually applied the ellipsoid formula using a calculator program on their computer to calculate the volume. In contrast, P2 utilized the workstation program that integrated the measurement process as a step within its guided diagnosis procedure. After drawing the three measurements on the images in the program, the prostate volume was automatically calculated and displayed.

J6: Determine the PSA density. The subsequent calculation of the PSA density is standard procedure since the value puts the PSA levels in proportion to the prostate volume and therefore acts as a more reliable biomarker. This step was always conducted right after the determination of the prostate volume. As in the previous step, P1 and P3 used a calculator program to obtain the result. In one instance, P1 utilized a speech assistant on his smartphone to obtain the results, instead. In P3's case, the PSA density was automatically calculated by the workstation software on completion of the prostate volumetry and with the PSA value which had to be inputted in the previous step.

J7: Determine anatomical characteristics of the prostate. One of the first steps in the actual image analysis was the evaluation and description of the prostate's features, properties, and condition based on the clinical data. Typical features include size, shape, texture, and any potential abnormalities. This characterization process gives valuable data not only in regard to potential cancer lesions but also other implications. For example, an enlargement of the prostate may lead to impaired urinary flow. During the CI, the participants characterized the PZ and TZ separately due to the anatomical differences between the zones. The following characterization could be observed for the TZ: The determination between the three types of dominant glandular, dominant fibrous, and mixed. *(P1:) "Grateful in principle are the glandular types, because within these changes, you can actually identify carcinomas very well."* The elevation of the urinary bladder floor (Intravesical prostatic protrusion) was another characteristic that was identified in multiple cases. However, P1 also measured the extent of the elevation using the DICOM viewer's measuring tools. For the PZ, signal patterns from the DWI were often mentioned as an example.

J8: Detect suspicious focal lesions. A prerequisite for the evaluation of any present cancerous lesions is first to detect them. To this end, the participants carefully went through the individual sequences in order to find

any signs that might point to a cancerous lesion. The participants usually performed this step simultaneously with the evaluation of the individual prostate zones as described in the previous job. For the detection task, the DWI sequence and ADC map were preferred, due to the well-recognizable signal diffusion. However, it was also mentioned that the individual searching algorithm also depends on the specific medical question. *(P3:) "As I said, I'll start with DWI. DWI and peripheral zone. [...] If I now have a very general "Find a carcinoma," then, of course, I first go through it very roughly. There is a difference if I now have a patient who has already had a negative punch and who is to be operated on, so I would proceed a little differently in terms of the algorithm."*

J9: Characterize detected lesions. Once a lesion has been detected, the participants measured and evaluated multiple characteristics, which could give indications into the properties of the finding. To execute this step, the participants performed a number of individual evaluations, which are described in the following.

J9.1: Determine lesion boundaries. Once a lesion has been detected, it is important to define the lesion's outline, delineating it from the surrounding tissue. This step, usually referred to as segmentation, is crucial for localizing the lesion and quantifying its extent, which has strong implications for the lesion's malignancy. During the CI, the participants performed the segmentation mentally by examining the images on which the lesion is visible and on multiple series.

J9.2: Determine the location. After the boundaries of a lesion have been defined, the radiologist can localize it within the prostate gland. Determining from which sector a lesion originates has multiple implications, e.g., for its aggressiveness. *(P3:) "And the problem in TC is first of all that in 95% of cases prostate cancer is a slow growing carcinoma."* Moreover, being able to refer to the specific sector in which the lesion is present allows for communicating the location of the lesion on a standardized map. During the process, P2 had the sector map of the workstation program open, which might have helped classify the correct sector.

J9.3: Determine the size. Another crucial step for the evaluation of a lesion is the determination of its size. In order to do so, the participants used the integrated measuring function of their DICOM-viewer or workstation software to measure the maximal diameter of the lesion. In most of the cases, the transversal T2W was used for this purpose, although in one

instance, the DWI was used. However, it was mentioned by one of the other participants that, generally, the DWI should not be used for this purpose, as there can be anatomical distortions in this sequence.

J9.4: Determine the ADC value. One step that does not appear in the recommendations of PI-RADS 2.1 but was nevertheless done by all three participants was the measurement of the lesion's ADC value. This process involved selecting an ROI within the lesion boundaries to be measured. Afterward, multiple variables were displayed by the software, such as the maximal, minimal, and mean ADC values within the selection region.

J9.5: Determine visual characteristics. Besides capturing measurable aspects of the lesion, the participants also identified visual characteristics of the lesion using the various sequences at their disposal. On the T2W images, the participants looked at the morphology of the lesion, including its shape, internal texture, and whether the edges were well-defined, irregular, or ill-defined. On the DWI sequence and ADC map, the radiologist could evaluate the signal intensity indicating the density of the tissue. Finally, the radiologists analyzed the temporal changes in signal intensity within the lesions visible in the DCE images. Here, aspects such as the time taken for the contrast agent to pass through the vasculature are taken into consideration.

J9.6: Evaluate local spreading. Another important step in the evaluation of a prostate lesion is the detection and evaluation of its spreading beyond the prostate gland into the surrounding tissues or organs. Such extraprostatic extension (EPE) indicates a more advanced stage of prostate cancer and therefore is an important factor in determining the extent of the disease and guiding treatment options. During the diagnosis process, the participants determined the presence or absence of local spreading by viewing the images.

J10: Assess the significance of detected lesions. The last step in the evaluation of the individual lesions is the assessment of their clinical significance. This information naturally has strong implications for the subsequent management and treatment of the patient. The participants conducted this step by evaluating the detected lesion on the T2W, and DWI in combination with the ADC map, and the DCE sequence. The participants scored the lesions on each sequence based on the PI-RADS 2.1 criteria. Afterward, based on the scores of the individual sequences, the participants determined the PI-RADS score for the lesion and, thereby, its likelihood to be of clinical significance. While P1 and P3 determined the lesion score in their mind,

P3's workstation software allowed him to input the score for each lesion to get the lesion score automatically.

J11: Prioritize significant lesions. In the case of multiple detected lesions, ordering them based on their medical significance needs to be done, with the aim of identifying the most suspicious or clinically relevant lesions that require further assessment or intervention. Participants did this by ordering them based on their PI-RADS score, defining the first one as the index lesion. However, P3 mentioned that his department does not refer to any index lesion. *(P3:) "Yes, I don't use the term so much now, because, from my point of view, it is always focused only on this lesion. But we know that a prostate carcinoma is multifocal."*

J12: Detect and assess distant metastases. Besides the detection and assessment of local spreading, the radiologist needs to look if cancer has spread to distant sites in the body, which typically happens through the bloodstream or lymphatic system. For this the participants, the participants used a wide-field-of-view sequence to be able to check the entire pelvic region. Special attention was placed on the bones and lymph nodes as these are typical areas for metastases to form. While discussing this stage, P1 expressed scepsis toward the reliability of the visual recognition of metastasis. Another observation was that P2 conducted this step in his DICOM viewer, as the overview sequence after contrast was not included in the workstation software. *(P2:) "but then the weighting is different again and you can still look to see if there is anything else that lights up that is not noticeable here. Such as a double safeguard."*

J13: Detect and assess additional findings. While prostate MRI diagnosis mainly focuses on cancer, it is still vital to scan the images for other findings that might be relevant for patient management. This was done by the participants throughout the process. For example, while analyzing the prostate on the T2W before doing the characterization, the participants also checked for hemorrhages in the gland or the seminal vesicles. Moreover, while looking for metastasis in a wide field of view, the participants also checked for other secondary findings. In one instance, for example, P1 detected fat tissue in the groin channel, which he marked as an indication of a minor bilateral inguinal.

J14: Make an overall assessment. Once all relevant findings have been deemed found, the radiologist needs to assimilate them with the other medical information into an evaluation that refers to the referrer's medical question.

As a part of this step, the participants expressed the evaluation in the form of a final score. For primary diagnoses, the participants used the PI-RADS score, which they determined based on the scores of the found lesions. However, the participants also utilized other scoring systems based on the medical indication. For example, P1 used the PRECISE score (Harder, Heming, & Haider, 2023) to evaluate the progression of a histologically captured cancer lesion during a progress control. *(P1:) "In the case of a histologically confirmed prostate carcinoma, it is pointless to say that there is a carcinoma that is PI-RADS 4 because it is histologically confirmed. This means that we have no option. What is much more interesting now is to say, the histologically confirmed carcinoma, does it remain unchanged? Does it present itself at all? Has it become larger? Is there a new finding?"* Moreover, for preoperative staging, P3 utilized the TNM classification (Brierley, Gospodarowicz, & Wittekind, 2017) to better communicate the extent of cancer in order to support the subsequent operation. In addition to the score, the participants also made recommendations for further actions or follow-up steps. These recommendations included additional imaging, targeted biopsies, or consultations with other specialists, such as urologists or oncologists.

J15: Generate the radiological report. Once all information has been gathered and considered for the final assessment, the results are documented in the radiological report for communication to the referrer. In the CI, each participant had a slightly different approach to creating the final report. However, two elements were consistent throughout all of them: The structured graphic and the written report.

J15.1: Generate the structured graphic. As a supplement to the written report, all participants created a lesion graphic aimed a further supporting the radiologist in a structured and visual way. While the process of generating the graphic, as well as the graphic itself, varied among the participants, there were some main commonalities. Each contained a table listing all detected lesions with a PI-RADS score of three or more. The table also contained detailed information about the lesions, such as the score on the individual sequences and the location of the lesion. Besides the table, the graphic also contained a sector map on which the lesions were drown in their respective region. The biggest difference between the participants was in the graphic creation process. In the case of P1, a sector map was included in the report template in the RIS, in which the lesions were manually drawn using the mouse. The table was also part of the report template and was filled manually for each lesion. On the other hand, P3 had a separate template for the graphic containing the sector map, table, and other infor-

mation, such as the PSA density. After filling out the graphic, P3 captured the graphic by taking a screenshot and uploading the image file to the PACS. To him, this was a workaround since attaching image files to the report was not supported by their RIS software. Finally, P2's graphic was automatically generated by the workstation software, which applied the input data to the table and sector map. Before this, P2 had to mark the center of the lesion on a displayed sector map, which then registered the marked region and drew a lesion around the point. This graphical representation serves as a valuable reference for further analysis, treatment planning, and communication with other healthcare professionals involved in the patient's care.

J15.2: Generate the written report. The written report is the official form of communication in the healthcare sector a form to communicate the diagnosis and therefore needs to be created as part of the overall process. Besides serving as a communication medium to the patient and other physicians, it also acts as a legal document, providing a basis for accountability and evidence of the health service provided. Overall, the process of generating the written report was similar in all CI sessions. All participants used an integrated speech-to-text function to dictate the report directly into the RIS. However, two distinct approaches could be observed in regard to when the process of creating the report. One participant dictated the findings as he was examining the MRI images. The other two participants went through the entire diagnosis examination process first, created the graphic, and then created the written report concurrently. They used the information that they already inputted in the graphic as an aid for dictation.

All participants followed a similar structure to their written report, which was divided into a descriptive part and a final assessment. In the descriptive part, the radiologist provides detailed descriptions of the imaging findings. Typically, it was subdivided into general findings, local findings, and lesion findings. The general findings section provides a brief overview of the imaging protocol used for the MRI examination and any relevant patient information, such as age, previous medical history, and reason for the exam. The local findings section focuses on the specific observations related to the prostate and its surrounding structures. It includes details about its characteristics as well as any notable findings in the surrounding tissues or organs. In the lesion findings section, the radiologist provides a detailed description of any detected lesions or abnormalities within the prostate. While this division was portrayed as the norm, it was also mentioned that the structure may vary depending on the medical question. In contrast to the descriptive part, the final assessment summarizes the radiologist's overall impression and con-

clusion based on the imaging findings, which may include a score as well as recommendations for follow-up actions. *(P1:) "The nomenclature 'high suspicion of prostate carcinoma' therefore does not belong in the findings section, but that is then the assessment, where I then summarize my descriptive findings, so to speak."*

J16: Share the radiological study. Once the radiological report for a prostate MRI is finalized and signed, the radiologist shares it together with the MRI images and other relevant files with the referring physician or other relevant healthcare professionals involved in the patient's care. The participants described different approaches to doing so. The standard approach was to send the radiological report via fax and the digital images on CDs via postal service. While certain media, such as E-mail, were not utilized due to data security reasons, various secure electronic communication systems were also implemented by the intuitions. In P1's practice, for example, the patients and radiologists would receive a QR code with which they would get access to the PACS files on a patient or referrer platform, respectively. P2's practice utilized secured connections with other intuitions in order to share their reports. *(P2:) "We are relatively widely networked with the surrounding practices and clinics so that a VPN tunnel is created accordingly and is always just sent back and forth. We have unlocked images for you. And then you can actively pull them. This is done via DICOM shipping, it is called."* In P3's case, other colleagues within the hospital could access the report directly through the HIS and the image data via the local PACS. While the radiological studies still needed to be sent to external receivers via fax and postal service, he also mentioned plans for the establishment of a cloud solution that would connect hospitals and other institutions in the region.

4.2.2 Contextual Factors

While steps within the diagnostic practice were relatively constant, multiple contextual factors were observed and discussed during the CI, which could influence the approach of the radiologist. All identified factors were turned into circumstances statements (see Table 4.3). In the following, each circumstance is presented in more detail.

C1: When working with various medical questions and indica-

ID	Circumstance Statements
C1	When working with various medical questions
C1.1	When the exam is a primary diagnosis
C1.2	When the exam is a secondary diagnosis
C1.3	When the exam is a process control
C1.4	When the exam is a preoperative staging
C2	When having access to preliminary examinations
C3	When working with incomplete patient data
C4	When working with restricted image data
C5	When examination stems from different scanners
C6	When doing second opinions
C7	When the patient has private insurance vs. public insurance
C8	When a case is clear vs. ambiguous
C9	When a case is not in the guidelines
C10	When the guidelines conflict with one's estimation
C11	When having various levels of experience
C12	When utilizing various scoring systems
C13	When working with various core technologies
C14	When working with various additional IT solutions
C15	When working with various input and output devices
C16	When sharing reports via various platforms, e.g. fax or cloud-based solutions
C17	When working in various types of institutions
C18	When working collaboratively
C19	When working under time pressure
C20	When working in a multidisciplinary environment
C21	When being interrupted during the process, e.g. by a telephone call
C22	When being unfocused
C23	When the referrer is biased towards new technologies

Table 4.3: List of circumstance statements

tions. As evidenced by participants' comments, different medical queries and indications significantly influence the radiologist's objective and approach. Throughout the sessions, multiple cases were identified, which will be detailed subsequently. While not an exhaustive collection, this provides a comprehensive overview of the diverse scenarios radiologists encounter.

C1.1: When the exam is a primary diagnosis. A primary diagnosis involves an MRI screening conducted prior to a biopsy. A typical indication for a primary MRI diagnostic is a suspicious PSA level. In such cases, the main objective of the examination is to detect any potential lesions.

C1.2: When the exam is a secondary diagnosis. Conversely, a secondary diagnosis involves an MRI examination that complements at least one negative biopsy result. While lesion detection remains significant, it becomes essential to interpret potential findings within the context of existing information.

C1.3: When the exam is a process control. Process control involves an examination primarily aimed at assessing the progression of a confirmed tumor. These controls can be incorporated into an active surveillance protocol, where the comparison of medical data with previous examinations assumes a significant role.

C1.4: When the exam is a preoperative staging. Preoperative staging involves assessing the tumor's progression prior to surgically removing the prostate. Determining the precise extent of the cancer provides essential insights to guide the surgical procedure.

C2: When having access to preliminary examinations. Strongly related to C1.3, the availability of preliminary examinations constitutes a significant factor in the diagnostic process. Throughout the sessions, participants inserted MRI sequences from previous exams and linked them with the current ones to simultaneously navigate and compare images. When multiple preliminary exams are available, the amount of accessible information that needs to be Incorporated can become notable. *(P1:) "Maybe you have four preliminary examinations, if it's active surveillance over a longer period of time [...] that gets very time-consuming. You have to hang everything up, you have to synchronize everything. That can cost time then."*

C3: When working with incomplete patient data. While the radiologist usually receives the relevant patient data, such as the PSA level history, from the patient or the referrer, in some cases this information may be missing or contradictory. As this information gives important context for the evaluation of the findings, the absence of them can make the diagnostic process more challenging. *(P3:) "That's why it's also important that we get all this clinical information very adequately. And that is often a problem, especially in the outpatient department. As I said, this is possible in our clinic. We get most of the information very reliably. But in the outpatient clinic it often just says 'PSA elevation, question mark. So MRI, please.'"* Different information sources may also lead to contradictory information. In one example, P1 was faced with two different PSA values, one indicated by the patient in the survey and another provided by the referrer in Doclib. Ultimately, the radiologist decided on choosing the value indicated by the patient.

C4: When working with restricted image data. Because of various reasons, the radiologist might have to work with a limited number of sequences than normally. During the CI, a few such cases were observed. In one instance, the MRI scanning process had to be aborted due to the patient's claustrophobia, which meant that not all sequences were able to be completed. Without a DWI sequence then, the participant said he could not satisfactorily do an evaluation based on the images. *(P1:) "You can already see these changes here in the T2 image, but it's hard to do because you can also see benign changes on the other side that look similar. So you can forget about that without DWI."* Instead, he had to resort to evaluating the exam and giving a recommendation based on the available information. Specifically, a biopsy was recommended based on a high PSA density and ... Other examples of limited image data included the presence of significant motion artifacts in the DCE due to the non-injection of Buscopan during the scanning.

C5: When doing second opinions. One factor that was only relevant to P3 during the CI, was that fact the examination is part of a second opinion diagnosis, which is done when a patient seeks second opinions for various reasons, such as confirming the accuracy of the initial diagnosis, exploring alternative treatment options, or gaining more confidence in their healthcare decisions. In this case, the MRI images as well as a written report is made available by another radiologist.

C6: When examination stems from different scanners. Another influencing factor is the type of MRI device that was used to acquire the images. In terms of magnetic field strength, participants preferred images taken by 3

Tesla over 1.5 Tesla machines as they produce more detailed images. In one practice, for example, prostate MRI screenings were exclusively performed on 3 Tesla devices. Besides image quality, another aspect that is influenced by the device is the cutoff threshold of the ADC value, which can also vary depending on the manufacturer. For this reason, the interpretation of the ADC value needs to consider the specific device on which the images were taken.

C7: When the patient has private insurance vs. public insurance. While the insurance type of the patients has mostly an impact on the billing on the site of the physicians, one instance was captured where this factor influenced the diagnostic process. More precisely, the need for the patient to pay for another MRI in the case of active surveillance was considered for the treatment recommendation for a PI-RADS 3 lesion. *(P1:) "But then it also plays a bit of a role if it's a publicly insured patient because he'll have to pay another €600 in a year's time. Is it then not legitimate to say, "Come on, we will biopsy him now"? Because in 10% of PI-RADS-3 cases, or 10 to 15% of PI-RADS-3 cases, a relevant tumor comes out anyway."*

C8: When a case is clear vs. ambiguous. While there is a large number of variances within each case that might affect the process to a certain degree, one aspect that seemed to have a large impact on the process was whether a case is clear or ambiguous. While cases where the diagnosis is clear, be it positive or negative, can be completed relatively quickly, more ambiguous cases require more time and effort. For example, the radiologist might spend additional time viewing the images or asking for the opinion of a colleague. P1 also mentioned that he might wait until the next morning before sharing an ambiguous finding in order to sleep over it for one night. Causes for ambiguity are manyfold, including unclear outlines.

C9: When a case is not in the guidelines. Another situation that can arise is that a case or aspect of a case is not covered by the guidelines. The specific case which was discussed with the participants was when a lesion is localized in the CZ or AFS. While the PI-RADS paper does not provide any guidelines in these cases, the participants mentioned that these cases can occur. In these situations, the participants said that they have to improvise and often use the algorithm for another zone to derive the score.

C10: When the calculated score does not match one's estimation. A tricky situation that was discussed was when the determined score PI-RADS score of a lesion based on the image data, does not match the radiologist's

estimation. While the PI-RADS scoring system was seen as an invaluable tool for the evaluation of prostate lesions by the participants, it was also mentioned that there exist certain situations, where the derived score might over- or underestimate the clinical significance of a lesion. Reasons for disagreeing with the score might stem from suspicious PSA value which is not considered in the determination of the score, or the experience that lesions are often misjudged in certain zones. *(P3:) "I see this quite often, I have very small 5-millimeter foci in a large prostate with a PI-RADS 4 lesion where in hindsight you say 'well, it probably won't be anything.'"* Two strategies were identified in such a case. The first one consisted of reevaluating parameters which led to the score in order to alter it. *(P1:) "[...] let's say according to PI-RADS it would be a 3, but I think it's carcinoma and I would like him to be biopsied for the reasons because the PSA density is unusually high or because the father has prostate cancer. [...] Then you can make a 4 out of it. That is absolutely legitimate. So you have this freedom in PI-RADS. You don't have to flagellate yourself."* For the other strategy, followed by P3, the discrepancy was communicated to the referring clinician through the written report as well as direct communication, which is easier in a hospital setting.

C11: When having various levels of prostate diagnosis experience. While all three participants had multiple years of experience working in the field of radiology and diagnosing prostate MRI images, some discussions also highlighted how less experience might affect the process. For example, as P3 mentioned that he prepared a document with work instructions and explanations, e.g. for the PI-RADS classification, for new beginner colleagues. *(P1:) "It is really important to define a very clear workflow in order to at least maintain the standard within the department."* This points to the need for a more guided process when starting out with this specialized diagnostic practice and being more dependent on external information in comparison to more senior colleagues who have already internalized this knowledge. Besides the time spent conducting prostate MRI diagnoses and the number of cases completed, other factors may also play an important role in the radiologist's experience level. One factor might be the completion of specialized, non-mandatory training, such as the Q-1 or Q-2 certification. Another factor is the amount of feedback the radiologist receives on their reports, as it allows them to calibrate their approach based on the accuracy of their past evaluations. In general, it is easily conceivable that the radiologist's experience will have a strong impact on the confidence, accuracy, and flexibility during the diagnosis process.

C12: When utilizing various scoring systems. While the PI-RADS

scoring system was used by all participants and is generally viewed as a standard in the diagnosis of prostate MRI, the inclusion of other scoring systems was also observed to impact the work practice. As described in J18, the PRECISE score as well as the TNM classification were utilized by P1 and P3 respectively to score the examination more appropriately in regards to the medical question.

C13: When working with core technologies from various vendors. As mentioned in Chapter X, radiologists work with a multitude of software systems during their work practice, e.g. RIS, PACS, and DICOM-viewer. While each serves a specific role, their functionalities can vary widely depending on the specific product and vendor. In the case of the three participants, none of them shared the same product. In the case of one radiologist, the tools were different depending on whether he was working at the hospital or doing teleradiology from home. The different features offered by the products had a strong impact on the work process of the practitioners. The radiologist working with the workstation, for example, was strongly guided by its step-wise approach. Moreover, some of the steps that the other participants did manually were automated by the software, e.g. the calculation of the PSA density. Another example was the capabilities of the RIS. While one participant inserted the graphic into the RIS report, another only had the possibility to enter a free text, without the option to format, which was seen as a strong inconvenience.

C14: When working with additional IT solutions. Besides various core technology products, the participants also worked with various additional applications to support them in their work practices and thereby affect their process. The usage of Doclib for example, gave P1 an additional source for receiving patient data, besides facilitating the scheduling of patient visits.

C15: When working with various input and output devices In addition to utilizing various software programs, the participants employed distinct sets of hardware, including screens and dictation devices. All of them used high-resolution diagnostic monitors to visualize medical images, accompanied by a smaller screen for non-image-related tasks. However, the number and size of screens varied among the participants. For instance, P2 and P3 worked with two adjacent diagnostic monitors for the DICOM viewer, while P1 used only one. Another hardware distinction was evident in the choice of dictation devices. P1 and P2 opted for handheld devices, while P3 utilized a headset for dictation. This discrepancy had an impact on the participants' ability to scroll through MRI images while dictating, as handheld devices

required one hand for operation, leaving only one hand free to manage other programs.

C16: When sharing reports via various platforms, e.g., fax or cloud-based solutions. Another factor that has implications for the work practice is the means by which one shares the finished radiological study with others. As described for J16, various technologies exist for sharing and depending on the access to it and the access of the receiving person. One effect this has, which was pointed out, was that participants receiving a report by fax might only see it in black and white. Thereby, elements that were highlighted with color in the graphic or report might not be perceived by all referrers.

C17: When working in various types of institutions. A significant factor influencing the work practice that was observed and discussed with the participants was whether one was working in a private practice or in a public hospital. While the specific contextual differences will be isolated as individual circumstances in order to provide a more granular understanding of the factors and are presented in the following, this factor itself was seen as important to be mentioned by itself.

C18: When working collaboratively. A distinctive difference in the process of P3, which he attributed to the public hospital environment, was that the diagnosis was done in collaboration with another radiologist. The collaboration during the session, which was done with the chief radiologist, consisted in evaluating the cases individually and discussing them in person. Multiple times during the CI session, the chief radiologist joined the room for this purpose. This non-mandatory process was introduced to increase the accuracy of the final diagnosis.

C19: When working under time pressure. When comparing the radiologists working in private practice versus the one at the public hospital, one striking difference that could be observed was how much time was taken for each prostate MRI assessment. While P1 and P2 conducted the diagnosis in one continuous process, P3 spent more time analyzing the images and collaborating with colleagues on individual cases. This demonstrates the greater emphasis on productivity and throughput in the private sector. *(P3:) "I have the advantage that we are not in a private practice, so we don't have to pay quite so much attention to speed."*

C20: When working in a multidisciplinary environment. Another important difference that was observed in the public hospital setting in con-

trast to the private practices, was the closer exchange between physicians from different fields. One way this was clearly demonstrated was in the attendance of a tumor board in which cases were discussed with various clinicians. While this multidisciplinary environment can have significant implications for the diagnosis, especially for the communication with the referrer, since the communication goes beyond the report. While this emphasis on interdisciplinary work was very visible in the hospital context, similar exchanges were also mentioned by participants in the private practices. For example, P1 mentioned him joining the urologist's office in order to prepare and accompany the fusion biopsy. While communication via telephone is always possible, closer collaboration with the physicians also involved with the patient's case leads to more touch points and opportunities for information exchange.

C21: When being interrupted during the process, e.g. by a telephone call. Radiologists often face numerous interruptions during their work practice, which can significantly impact their workflow and concentration. During the sessions, multiple sources of interruptions could be observed. The most common interruption came from telephone calls, which the participants always answered. These calls came from colleagues, referring physicians, and other people with inquiries. These interruptions require immediate attention and can disrupt the radiologist's ongoing tasks. In one instance, P2 was called while dictating a report. He proceeded to finish the started sentence and then received the call.

C22: When being unfocused. Another influencing factor that was brought up during the discussions was the condition of being less concentrated than usual. This state might be caused by different factors inside or outside of the radiologist's control. *(P1:) "[...] there are so many influencing factors: bad sleep, phone ringing all the time, queries. You may not be aware of it at the time, but it really affects you."* Naturally, feeling unfocused or being distracted by environmental aspects can result in decreased accuracy and efficiency in analyzing medical images and making clinical decisions.

C23: When the referrer is biased towards new technologies. The availability of technology might also be affected by their openness to new technologies, which may have implications when sharing the radiological study. For example, when describing the portal where patients and referrers can access the report online, P1 mentions the dismissal that came from certain referrers. *(P1:) "Many referring physicians have a problem with this. But unfortunately, that is human nature and also above all the human who*

is not particularly devoted to technology, that he rather closes himself to it and simply does not want to go with the times, for whatever reason."

4.2.3 Needs

Finally, the radiologists' needs associated with the prostate MRI diagnosis practice were identified and formulated into outcome statements. Table 4.4 lists all outcomes statements that were identified based on pain points, wishes, and preferences.

O1: Maximize the accuracy of the diagnosis. The most obvious desired outcome is to provide the referrer with an accurate diagnosis of the patients in order to adequately inform the following steps. Technically, this means maximizing the sensitivity and specificity of the process, i.e. minimizing the number of false negatives and false positives, respectively. However, many factors can influence the accuracy of the final diagnosis.

O1.1: Maximize the accuracy of detection. As a first step, the radiologist needs to detect true findings in order to take them into consideration for the final assessment. This includes benign findings, lesions, and metastases.

O1.2: Maximize the accuracy of localization. Identifying the location of a finding can be crucial for its assessment, for example when applying the PI-RADS score to a lesion. The participants mentioned that the localization of lesions in the prostate can be challenging in certain situations. *(P1:) "What's also difficult are findings that are right on the border between the transition zone and the peripheral zone. Where you don't know, 'Is this a prolapsed node from the transition zone or is this a finding that comes from the peripheral zone?' There are a couple of classic localizations like that where it can be difficult."*

O1.3: Maximize the accuracy of quantification. In multiple instances, during the diagnostic process, the radiologist measures values relevant to the diagnosis, e.g. the maximum lesion diameter or mean ADC value. However, some of the quantifications can be susceptible to inaccuracy. For example, when calculating the prostate volume using the ellipsoid formula, the result strongly varies depending on where the radiologist has measured the axis of the prostate. (P1:) "Where do I draw which line? It's so individual and

ID	Desired Outcome Statements
O1	Maximize the accuracy of the diagnosis
O1.1	Maximize the accuracy of detection
O1.2	Maximize the accuracy of localization
O1.3	Maximize the accuracy of quantification
O1.4	Maximize the accuracy of visual characterization
O1.5	Maximize the accuracy of staging
O1.6	Maximize the accuracy of assessment
O2	Maximize the quality of the radiological report
O2.1	Maximize the clarity of the report
O2.2	Maximize the conciseness of the report
O2.3	Maximize the completeness of the report
O2.4	Maximize the added value of the report
O2.5	Maximize the traceability of the report
O3	Maximize the efficiency of the process
O3.1	Minimize the duration of the process
O3.2	Minimize the cognitive effort required for the process
O3.3	Minimize the time/effort spent on operational tasks
O3.4	Minimize the consumption of resources
O4	Minimize the fragmentation of the process
O5	Maximize the robustness of the process
O6	Minimize the occurrence of execution errors
O7	Maximize the ease of data synthesis
O8	Maximize adherence to protocols
O9	Maximize the potential for conflicts
O10	Maximize the probability of receiving valuable feedback

Table 4.4: List of desired outcome statements

depends on the person making the diagnosis." Seeing that these variances can affect if the value has reached a suspicious threshold or not, maximizing the precision of such quantification may improve overall diagnostic accuracy.

O1.4: Maximize the accuracy of visual characterization. Besides quantifiable aspects, improving the precision of qualitative features can also positively affect the outcome. The importance of this was highlighted by cases the categorization of these characteristics was seen as challenging. *(P1:) "So let's put it this way, what is really problematic are carcinomas in the transaction zone. They are usually problematic. Now really committing there on the T2 image, 'Isn't that a little fuzzy bounded there at that point? Does that really have a completely undescribed capsule?' And then you're really brooding and looking at that thing again for 5 minutes from all planes in the T2 image."*

O1.5: Maximize the accuracy of staging. Correctly determining the stage and extent of a tumor naturally is of vital importance for the selection an appropriate and effective treatment.

O1.6: Maximize the accuracy of assessments. Finally, an accurate diagnosis requires the correct assessment of the collected information.

O2: Maximize the quality of the radiological report. A common observation during the CI sessions was that the participants put a lot of emphasis on creating a radiological report that would be as valuable to the radiologist as possible. In order to achieve this, by following a couple of aims:

O2.1: Maximize the clarity of the report. Firstly, As the radiological report serves as the main medium of communication to the referrer, it is of vital importance the reader fully understands the information that is being conveyed. Any room for misinterpretation or misunderstanding may require the referring physician to inquire about the radiologist, or worse lead to the mistreatment of the patient. To avoid this it is also important that the report is tailored to its audience as they are likely to not share the same specialized knowledge as the radiologist. *(P1:) "You must always try to put yourself in the position of the referring physician. He or she is not a radiologist. This means that the findings must be clear, unambiguous, and comprehensible. And people must be able to make sense of it."*

O2.2: Minimize the time required to understand the report, Next to being unambiguous, it should also be easy and fast for the referrer to capture

the information relevant to them. The participants displayed multiple efforts in consideration of this fact. For one, the participants tried to keep the final assessment as concise as possible. *(P1:) "It is a fact, they all have things to do. And if they then need a quarter of an hour to read through a report... No one does that. They don't want to do that. They'll get a fit."* While the completeness of the descriptive part was seen as more important than its briefness, it was understood that the referrer would primarily read the final assessment and only look at the detailed findings if needed. Besides this, another aspect that was utilized for improving the readability of the report was the use of formatting, for example highlighting important information boldly. Therefore, the absence of formatting options in the hospital RIS was seen as a major inconvenience to P3. Yet another desired feature that would help in this regard is the establishment of a standardized structure for the report and graphics. *(P3:) "And every teleradiologist writes his own findings and then they have to adjust, again and again, the recipients, where does it say what now?"*

O2.3: Minimize the chance of missing information in the report, Another important characteristic of the report is its completeness. Forgetting to look for or document certain aspects of the image may result in the absence of important information for the referrer. For this reason, the structured approach that was prompted by filling out the graphic was highly appreciated by P3. "Then you have a focus where you always look at and then you have something around it that you might overlook. That's why such a standardized procedure, with such a standardized report, where I can say "I can check everything again" and then look at it in conclusion."

O2.4: Maximize the valuable data The primary objective of the diagnostic process is to furnish the referring physician with pertinent patient data, making data value maximization crucial. While participants shared similar approaches resulting in comparable data, certain variations were noticed, enhancing the value of the reports. For instance, instead of including the sector map provided by the PI-RADS v2.1 paper in his graphics, P3 utilized a custom sector map that included additional surrounding organs like the rectum and bladder. This visual representation proved valuable for drawing and visualizing the invasion of large cancer lesions into these areas, supplying essential information for potential surgical interventions. *(P3:) "Either the surgeon says 'Okay, this is such a large finding. I can see that at a glance, that's not really anything to operate on anymore and I have to think about something else, or I have to know, when I operate, that just from the graphic, okay, where do I have to be especially careful with my surgery?'*

That's why something like this is important."

O2.5: Maximize the report's traceability. Besides needing to be quickly understandable to the referrer, the report also needs to be quickly understood by the radiologist or colleagues in combination with the MRI images. This is important because the MRI study might still be relevant to the radiologist after the report has been shared with the referrer. For example, the referrer might contact the radiologist because of a new finding, requiring him to go back to the study and quickly make sense of it. Another situation where past studies become relevant is in the context of active surveillance, where radiologists need to compare the new evaluation with the previous ones. One way in which the participants increased the traceability of the report was by adding references to the specific image in which the lesion findings are visible. These references allow not only the referrer but future radiologists to quickly locate the documented lesion findings on the MRI images. *(P1:) "It could be that the patient comes back in 2 years for another MRI of the prostate and my colleague reads the old findings. And then of course the question arises, where did he see it? Just to simplify things a bit."*

O3: Maximize the efficiency of the process. Naturally, increasing the efficiency of the entire process is a desirable goal. However, there are multiple factors that play into what contributes to this aspect.

O3.1: Minimize process duration. One obvious factor that affects the productivity of the process is the total duration. While speeding up the process seems especially valuable for radiologists working in private practice, it presents a general benefit if it does not compromise other factors such as the precision of the diagnosis. Benefits associated with a short process duration include higher productivity, less time pressure, and experienced stress.

O3.2: Minimize cognitive effort. Another important factor is the cognitive effort required throughout the diagnostic process. With reduced cognitive burden, radiologists can focus more on critical thinking and decision-making, leading to improved diagnostic accuracy. During the CI, multiple strategies were observed that reduced the cognitive burden. One such strategy was the establishment of repeatable actions, such as reusing sentences for the creation of the written report. The importance of having the same arrangement of viewports in the diagnostic software was also mentioned to ensure less cognitive friction. Another factor is to not have to think about how to operate the software. *(P3:) "So for routine work, I think it's extremely important that such processes run subcortically, so that you can really think*

about the referrer question and don't have to constantly think about 'How do I operate what?'" Heighten cognitive effort might also come from external sources such as a loud environment or interruptions, which break the natural flow of the practice and require refocus on the task. *(P1:) "And if I know that I have my peace, then I can dictate 80 MRs a day without getting tired. But what tires me out are the things that come along with it."*

O3.3: Maximize the focus on the assessment The essential task for which the radiologist's skill and knowledge are required is the detection and assessment of radiological findings. However, the process requires additional actions which are not directly related to the medical question, e.g. getting access to the relevant information or generating the radiological report. The participants often mentioned that doing the actual diagnosis is an interesting and enjoyable part of the work. In contrast, many of the low-cognitive tasks are seen as tasks that could also be automated. *(P3:) "If you have now seen everything I still have to do manually, if everything is automated, of course, if I only have to select the lesion or even the AI already does that, then I really only have to worry about the overall context."*

O3.4: Minimize the costs. While the cost generated throughout the process was generally not thematized, it was mentioned when discussing the handing out of CDs to the patients. *(P1:) "In the last 10 or 20 years, it was basically the case that mainly CD-ROMs were burned. We still do that, but we want to move away from that because it is very, very cost-intensive. It is very, very expensive to operate such a burning robot, which also has a corresponding performance and it is also very, very susceptible and maintenance-intensive as a rule."* While the CI did not uncover many points where costs play a factor, reducing these where possible can be seen as a positive outcome.

O4: Minimize the fragmentation of the process. As already mentioned, radiologists can have an array of software products integrated into their workflow. As they can stem from different providers they may not have been designed to Therefore compatibility between the different IT solutions would improve the workflow overall. *(P1:) "In general, this is a problem of all computer-assisted programs. The transfer of the results from the program into the respective radiological findings. This is a huge problem because it is not standard."*

O5: Maximize the robustness of the process. As discussed under C21, radiologists are prone the be interrupted during their work practice. Getting distracted from the current workflow and having to return to it afterward re-

quires cognitive resources and can be draining according to one participant. Moreover, studies have shown that distractions can lead to reduced reading accuracy during abnormal cases (Balint et al., 2014). Therefore, minimizing the negative impact of interruptions was defined as another outcome.

O6: Minimize the occurrence of execution errors. Throughout the diagnostic process, multiple factors were mentioned, which allowed the occurrence of execution errors. These were mostly associated with the transfer of data from one program to another. *(P3:) "The other day I entered a date of birth with 1665, or something like that. That is of course also embarrassing. But that's what happens when you have to do it manually. Such things should all be transferred automatically."* In another instance, P2 changed the PSA value he input in the workstation software after quickly realizing that he used the value of the wrong patient.

O7: Maximize the ease of data synthesis. One of the major skills required for the diagnostic practice is to synthesize information from various sources to inform one's assessment. This may entail correlating a suspicious area on the DWI sequence with the ADC map in order to assess the potential clinical significance of the lesion. Another example is when comparing the images from two different exams to evaluate the progression of a finding. Facilitating the comparison and synthesis of various data is, therefore, a relevant and desirable outcome. During the CI, participants enhanced this process by synchronizing various MRI series, enabling seamless navigation through them simultaneously.

O8: Maximize adherence to protocols. As described in CX, the participants did not go outside of the guidelines even when disagreeing with the result, rather opting for strategies that let them stay within the guidelines but still address their personal opinion. One reason that was given for this was that going against the guidelines can result in legal issues. *(P2:) "Let's say this is a PI-RADS 2 and I say this needs to be punctured, but the urologist says 'Nah, that's nonsense, doesn't need to be punctured. Don't puncture it.' The patient goes to the urologist three years later. There is a carcinoma at the site, by accident, so to speak. This creates a rather tricky legal situation."*

O9: Maximize the potential for conflicts. The participants followed multiple strategies to minimize potential conflicts besides the adherence to guidelines. For example, P1 mentioned only engaging in consultations upon patient request and attempted to keep them brief to avoid overwhelming the patient with information as this might prime the patient and make it more

difficult to work with for the referring physician. *P1:) "I just want to avoid them coming in there and somehow getting some information from us that they don't understand and then it's incredibly difficult for the urologist to recapture them or get them back on track."*

O10: Maximize the feedback on the report. Since the radiological evaluation can differ from the histological results, radiologist benefit from receiving feedback in terms of the accuracy of their findings and assessment. While P3 mentioned regular exchanges with the urologists and pathologists working in the hospital, P1 and P2 were more dependent on getting the histological sent from the referring physician. To increase the likelihood of this, they would add a request for the results in their written report. However, the response rate was rated low. In response to this, P1 tied connections with certain referrers. *(P1:) "I also always write "please inform me of the histology" in the report; this is usually never done. I have to be honest with you. And in the meantime, I have my fixed referral clientele of urologists and fortunately, I also get, let's say, at least feedback from them. And that helps in any case."*

4.3 Discussion and Conclusion

In general, the CI study demonstrated consistent trends across the sessions, while also showcasing variations among individual participants. A factor that contributed to greater uniformity in the process was adhering to the PI-RADS 2.1 guidelines (American College of Radiology, 2019). In fact, numerous parallels emerged between the identified job statements and the proposed report structure template, as several individual job steps correlated with specific sections of the report template. However, the methods participants employed to achieve these objectives could diverge significantly. The execution of individual steps was notably influenced by the software solutions employed by participants and the spectrum of possibilities that they offered. On one hand, these solutions could enable improved ways to complete tasks, such as the automatic calculation of the prostate volume based on the measurements on the images. On the other hand, these tools could also impose limitations that required circumvention, like the need to send graphical elements separately due to the RIS's image data constraints. This underscores the extent to which tools can shape the overall process execution

and the need for a thoughtful approach to their development.

To comprehensively capture the essential facets of the practice, the JTBD framework and its individual components exhibited considerable practicality. Regarding the job statements, their emphasis on goals rather than execution permitted the portrayal of a process marked by high variability in a more universally applicable manner. When it comes to inspiring concept development, the solution-neutral nature of the approach also fosters space for innovation. Additionally, the collection of job statements serves as an informative summary for identifying which steps of the process one wishes to support through a solution. For instance, it is possible to opt for specialization in specific tasks, such as focusing solely on lesion detection, or opt for a broader coverage of the process.

The compilation of circumstance statements also provides valuable perspectives when considering the most effective ways to assist radiologists in their work practice. These diverse factors originating from various sources, including the medical case, work environment, institution, and IT setup, highlight the multitude of influences impacting the work practice. This approach encourages viewing the process not in isolation, but rather as a dynamic interplay shaped by various contextual elements, facilitating thoughtful consideration and planning for their integration. For instance, recognizing that radiologists often work with limited imaging data should be a consideration when designing a solution to process multiple mpMRI sequences as ML inputs, ensuring alignment with practical constraints.

Lastly, the identified user needs play a critical role in exploring and defining the potential value a solution could offer to the practice. While some desired outcomes like maximizing diagnostic accuracy and process efficiency might be self-evident, breaking down these goals into specific needs provides a more precise and comprehensive understanding. Examining each desired outcome statement individually could lead to the discovery of new and additional methods to fulfill these needs. To this end, it is important to acknowledge that some of the listed needs might be interconnected or even conflicting. For instance, focusing on minimizing the duration of the process could potentially impact diagnostic accuracy. This is where prioritization could be valuable. In fact, the JTBD framework often includes a quantitative prioritization process for customer needs (Ulwick, 2017). While this step was considered for this research, it was ultimately discarded due to the complexity of involving multiple radiologists and exceeding the scope of this study. Additionally, the significance of individual needs may depend on various contextual factors for

radiologists. For instance, the importance of minimizing process duration might vary between a public hospital and a private practice, where time pressure could differ.

While the study provided an in-depth understanding of the practice, it also had some limitations. The foremost concern is that, despite the utilization of data-driven analysis, the outcomes remain reliant on the researcher's interpretation. This inherently subjective aspect introduces a potential source of bias that could impact the accuracy and objectivity of the findings. A potential remedy to this constraint could involve a validation process conducted in partnership with the practitioners. Such an approach might uncover misinterpretations as well as any gaps in the gathered insights. While the inclusion of such a process was contemplated, the notion was rejected due to the additional expenditure it would have imposed on the participants.

Another constraint that should be acknowledged considering the diversity among the radiologists is the study's relatively small sample size. As was observed, participants displayed substantial variations in specific aspects based on factors like their institutional work environment. It is conceivable that involving a larger number of radiologists could unveil further dimensions of their work practices. Furthermore, it's worth noting that all participants in the study possessed considerable years of experience in the field. Consequently, the requirements of less experienced radiologists might not have been adequately captured.

On the whole, the CI study has laid a solid groundwork of insights into the work practice of prostate MRI diagnosis, exploring the problem space and defining its most relevant aspects. This provided a crucial basis for the subsequent creation and deployment of a solution prototype, facilitating a targeted exploration of how AI technology can be effectively integrated into this practice.

Chapter 5

Study 2a: Prototype Creation

The second phase of this research project entailed the creation and qualitative evaluation of a prototype designed to explore radiologists' interactions with and reactions to a prospective AI-assisted software solution for prostate cancer diagnosis. Building upon the insights from the previous CI study, which illuminated the existing work practices in prostate MRI diagnosis, this subsequent study sought to investigate the potential impact of introducing AI-based solutions. By engaging the target users, namely radiologists involved in the diagnostic process, with these solutions, the aim was to uncover the requirements and further aspects related to the introduction of AI technology into the work practice. Through this exploration, a deeper understanding of how AI can enhance or alter the diagnostic process could be gained.

To enhance clarity and organization, the study's presentation was divided into two chapters. The current chapter delves deeper into the process by which the insights gathered during the CI were translated into an interactive and testable prototype. First, the creative ideation and design phase is presented, where solution concepts were generated and visualized in an iterative approach. Then, the technical implementation is presented, including the technologies and real-world data that were utilized. Finally, the resulting prototype as well as elements that did not make it in the final implementation, are presented and briefly discussed. Subsequently, the following chapter offers a comprehensive account of the prototype evaluation with radiologists, providing detailed insights into the outcomes and findings of the assessment. By dedicating a separate chapter to the creative process behind the prototype, this work underscores the significance of adopting a system-

atic, methodological, and insights-based approach to prototype design.

5.1 Methodology

Prototypes have long played a vital role in the design process, offering a rapid and resource-efficient way to test specific aspects of an interactive product before its final implementation. Their primary objective is to gain valuable insights and answer specific questions during the evaluation process (Houde & Hill, 1997). Moreover, by involving the intended audience, prototypes serve as an effective tool for a user-centered approach. Recognizing these advantages, the development of a prototype was deemed essential for delving deeper into the understanding of the human-AI interaction in the area of prostate MRI diagnosis.

To ensure that the prototype effectively fulfills its purpose, a thoughtful design and implementation approach in accordance with its objectives is necessary (Houde & Hill, 1997). For this study, the focus was to explore how a potential solution could best meet user needs and uncover crucial factors that need to be considered when designing for the actual work practice. To achieve this, an evidence-based ideation and design process was utilized in order to generate relevant solutions to the users' needs. Additionally, emphasis was placed on implementing the design concept into the prototype in a way that closely simulates the actual user experience. This authentic representation aimed to capture important nuances that could prove crucial in the final implementation of the solution. By following this purposeful approach, the prototype was poised to yield valuable observations and facilitate meaningful discussions in the evaluation process.

5.1.1 Ideation and Design

The ideation and design process of the prototype involved a creative and iterative approach. To meet the user's needs effectively, brainstorming sessions were conducted to generate diverse ideas. For this, the insights from the former contextual inquiry (CI) study were utilized as valuable input, informing the design of relevant and practical solutions. Additionally, guidelines for human-AI interaction were considered to address known design implications of AI-based programs. An iterative approach was adopted, where ideas were continuously generated and refined by incorporating insights from the

CI study and adhering to the human-AI interaction guidelines. Moreover, to visualize and concretize the evolving ideas throughout the iterative process, sketching was employed, providing clear and tangible representations of the software's potential functionalities and user interactions. This iterative and user-focused approach allowed for a comprehensive exploration of ideas, ultimately resulting in a prototype that aligns closely with the specific requirements and preferences of real-world radiologists performing prostate MRI diagnosis.

To facilitate and frame the ideation process, the How-Might-We (HMW) technique was employed. This widely-used design thinking tool involves the formulation of questions that start with "How might we," which guide teams to focus on the right problems while generating creative solutions (Rosala, 2023). To formulate precise HMW questions, the outcome statements derived from the CI study were used as a basis, with the simple addition of "HMW" at the beginning of each statement. For example, 'Minimize the process duration' was turned into the HMW question 'How might we minimize the process duration?' Framing the user needs as questions allowed to direct focus in the right direction effectively. The resulting HMW questions served as a valuable starting point for generating new ideas and kept the focus directed toward addressing the radiologists' needs throughout the process.

The same procedure for formulating outcome statements into HMW questions was also applied to the guidelines for human-AI interaction proposed by Amershi et al. (2019) (see Table 5.1). For instance, the guideline "Support efficient dismissal" (G8) was reformulated into the question "How might we support efficient dismissal?" The inclusion of these guidelines into the ideation process served to address known design implications of AI-based programs. The adherence to these guidelines has already seen success in the development of an AI-based breast screening assistant prototype (Calisto et al., 2022).

In addition to the HMW questions, the ideation process also incorporated the job statements and circumstance statements to guide the generation and refinement of ideas. During the iterative process, each of these statements played a significant role. The job statements were instrumental in exploring how the user needs could be addressed at different stages of the work practice. Conversely, the circumstance statements allowed for careful consideration of how these needs might be influenced by contextual factors and how solutions could cater to these diverse requirements. By taking into account both job and circumstance statements, the ideation process gained valuable insights,

ID	AI Design Guidelines
G1	Make clear what the system can do.
G2	Make clear how well the system can do what it can do.
G3	Time services based on context.
G4	Show contextually relevant information.
G5	Match relevant social norms.
G6	Mitigate social biases.
G7	Support efficient invocation.
G8	Support efficient dismissal.
G9	Support efficient correction.
G10	Scope services when in doubt.
G11	Make clear why the system did what it did.
G12	Remember recent interactions.
G13	Learn from user behavior.
G14	Update and adapt cautiously.
G15	Encourage granular feedback.
G16	Convey the consequences of user actions.
G17	Provide global controls.
G18	Notify users about changes.

Table 5.1: List of human-AI interaction design guidelines (Amershi et al., 2019).

leading to the development of well-rounded and contextually relevant ideas for the AI-based software prototype.

Throughout the iterative process, another supplementary technique employed was the use of sketching to visualize ideas. Sketching involves creating quick visual representations of ideas and has long been utilized in early design phases to support the creative process by providing visual imagery (Verstijnen, van Leeuwen, Goldschmidt, Hamel, & Hennessey, 1998). In this ideation process, ideas were sketched using the design tool Figma (Figma, Inc.). Although sketching is traditionally associated with drawing on paper, the digital platform offers the same benefits, allowing for the quick externalization of ideas. Moreover, the digital nature of the tool provided additional advantages, such as the ability to include images like the sector map from the PI-RADS v2.1 paper or MRI images, as depicted in the sketch in Figure 5.1. Overall, the utilization of sketching played a pivotal role in driving the ideation process forward and creating a well-informed and user-focused design.

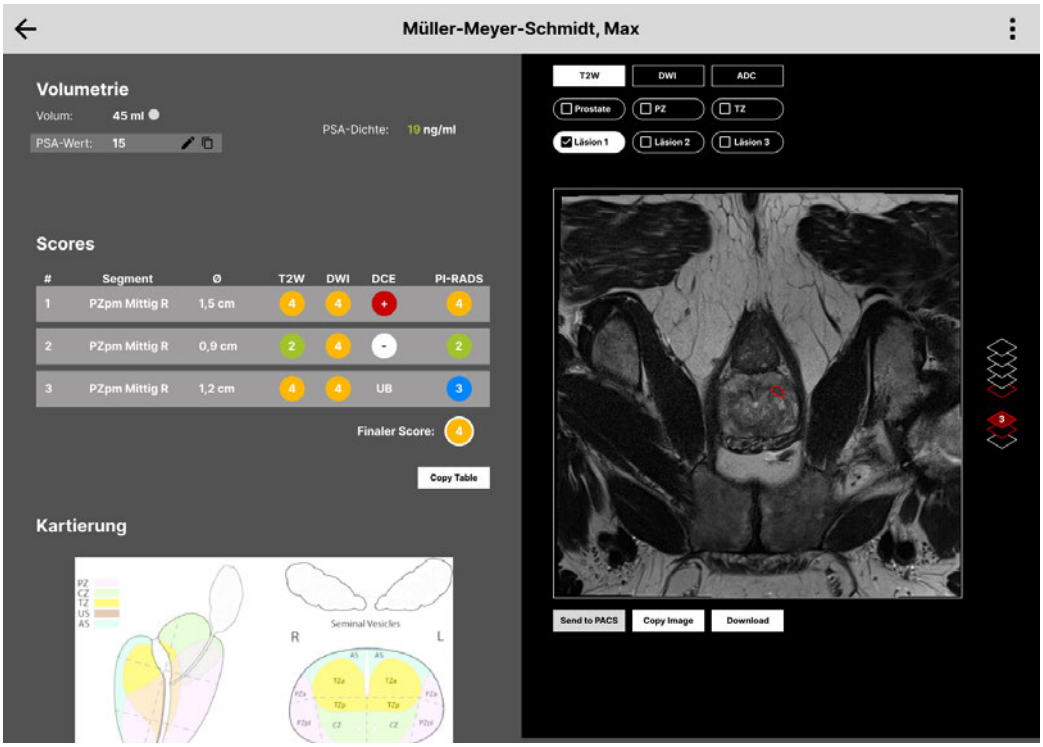


Figure 5.1: Prototype sketch including the PI-RADS 2.1 sector map (American College of Radiology, 2019) and a T2W image.

Following several rounds of iterative ideation and design, a diverse array of ideas was carefully curated to create an AI-based solution. Beyond evaluating individual ideas independently, great emphasis was placed on how these ideas could synergistically merge to form cohesive and comprehensive concepts. Once a stage was achieved where the overall concept was deemed adequate for the scope and purpose of the study, the transition to the implementation phase was made.

5.1.2 Implementation

The primary goal in translating the design concept into a prototype was to create a user experience that closely resembled interacting with a finished product within the limitations of this project. (This was seen as important, since) To achieve this, the prototype was implemented as a front-end web application using AngularJS. This decision allowed the radiologists to access a fully interactive application through their web browsers on their office computers, enabling them to evaluate the prototype in their natural work environment. By avoiding the need for installation of external software, which can be restricted for security reasons, the web-based approach offered a realistic and accessible means of interacting with the solution, aligning perfectly with the design concept of being a web application.

To enhance the realism of the prototype interaction, real MRI examination data and actual AI-generated output were utilized. A selection of anonymized prostate MRI studies, including mpMRI images and their corresponding written reports, was provided by a fourth radiologist. This real data enabled the creation of realistic and coherent cases within the prototype. Additionally, real AI output based on the MRI images was integrated into the system through collaboration with Gemedico. The selected MRI images were sent to Gemedico, where they were processed by their current AI algorithm. The resulting output included image series with ROI masks of the segmented prostate, peripheral zone (PZ), and any detected lesions. Furthermore, various determined parameters, such as the volume of the segmented areas, were made available. Through image editing using the software Photoshop (Adobe Inc.), the medical and AI-based data could be incorporated into the prototype in a streamlined way. By displaying real medical data in combination with real AI output in the prototype, a more genuine user experience was achieved, enhancing the potential for relevant insights during the evaluation phase.

5.2 Results

5.2.1 Design Concept

The overarching objective of the envisioned design concept is to aid radiologists in the prostate MRI diagnosis process (MJ) through automated image analysis, data extraction, and report generation. Once the AI algorithms have processed the MRI images, the results are made available via a web-based front-end application, allowing users to examine each of them for each exam in greater detail. The AI primarily focuses on detecting clinically significant lesions (J8) and accurately delineating both the whole prostate and the lesions (J9.1), which are presented to the user in a simplified DICOM viewer. Additionally, the solution offers an input form to store relevant diagnostic data, such as the PSA levels. Some data, like prostate volume derived from the delineation, is automatically filled (O3), while the user can manually enter other information. Using the stored data, the solution generates a structured and informative graphic (J15.1) to be included in the radiological report. It is essential to note that the solution is not intended to replace the existing DICOM-viewer but rather complement it, operating alongside it to enhance the radiologist's support. While this high-level description only provides the essence of the design concept, the following section gives a more detailed account of the individual features and visual implementation.

5.2.2 Implementation

In order to give the user access to the results from different exams (J1), the prototype starts on an overview page (see Figure 5.2), which lists all the exams processed and being processed by the AI algorithm. To facilitate the identification of the correct exam (O3.2, O6), details on the date of the examination and the patient details are given. Furthermore, details the list is sorted from most recent to oldest. The status of the AI's progress on the exam is also displayed. By clicking on any of the listed exams, the user enters the details page of that particular examination.

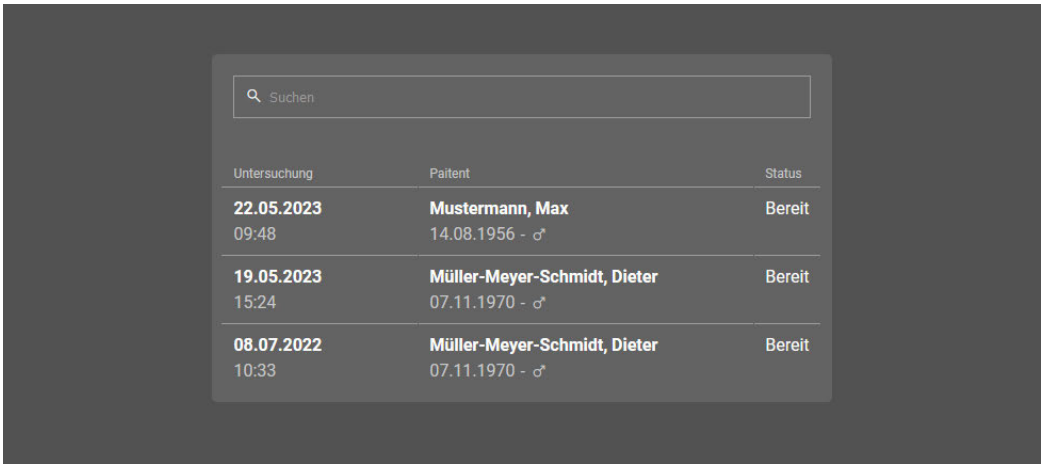


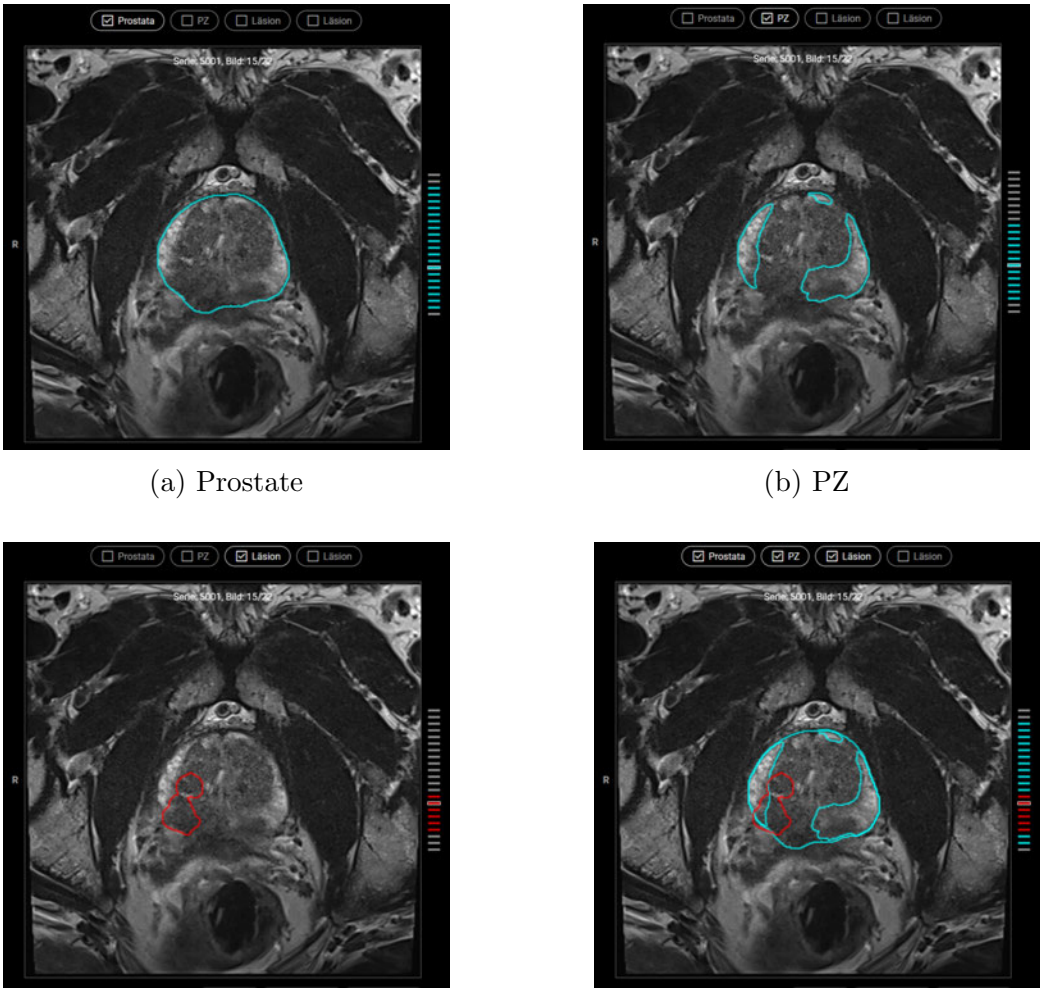
Figure 5.2: List of exams

The details page, as shown in Table 5.3, provides the user with all the information related to a particular exam and is where all other functions reside. In the header, on the top of the page, the patient details are displayed to minimize the probability of working on the wrong patients (O6). The main body of the page is divided into a left and right side. The left side is where the input form is situated, and the non-image data is displayed. In contrast, the right side acts as a viewer for the graphical data, such as the MRI images and lesion graphics. Displaying the visual and graphical data simultaneously allows the user to easily compare, make connections and mentally switch between the two (O7). By avoiding the need to switch between the views manually, minimizing the required cognitive effort (O3.2).

Figure 5.3: Exam details page

One of the graphical elements included in the viewer on the right side of the screen was the MRI series of the exam relevant to the PI-RADS scoring. Users could switch between the different series using the buttons on the top and scroll through the images using the mouse wheel, as is possible within DICOM viewers. Additionally, a scroll bar was placed to the right of the images for more interaction and to visualize which image in the series is selected (O3.2).

Another function the viewer has is to display the AI-generated delineations inside the images. By selecting and deselecting the mask options above the image, the user can activate the masks for the segmentation of the prostate, PZ, and any lesion detected and classified as significant by the AI algorithm (see Figure 5.4). This function would allow the user to activate any combination of masks and view them through the entire stack of images by scrolling through them. To facilitate the overview, the image representations in the scroll bar were colored based on the activated masks to visualize which images masks were visible (O3.2). To further improve the clarity, the organ and zone delineations were colored differently than lesion delineations (O3.2). It should be noted that the masks were only available on the axial T2W images.



(a) Prostate

(b) PZ

(c) Lesion

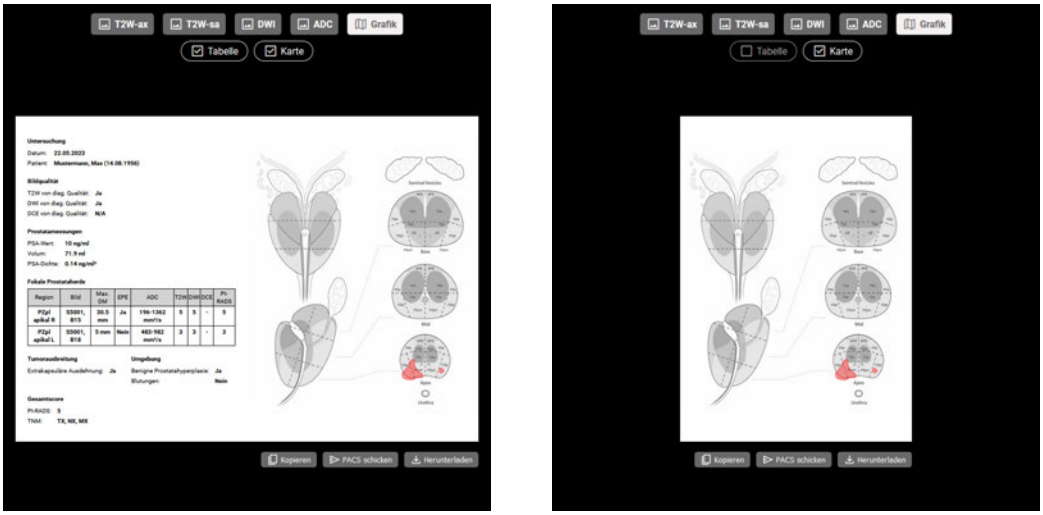
(d) Prostate, PZ, and lesion

Figure 5.4: Segmentation masks

While the viewer offers some functionality similar to a DICOM viewer, the intent was not to replace it as the primary tool for evaluating the MRI images. The reason for this was that the DICOM viewer used by the participants offers an array of specialized functions and would be difficult to achieve at the same level (C13). Moreover, it prevents radiologists from giving up on a familiar part of their established process (C3.2). Instead, the concept intended the solution to run alongside the DICOM-viewer. Considering the various hardware setups, the idea was to run the software on one of the smaller office monitors while the diagnostic monitors were still free for the DICOM viewer.

The design concept entailed multiple ways to transfer the visual data to support the interplay between different software systems. Specifically, the option to download or send the set of images to the PACS, including the activated masks, was defined as a feature, as was the possibility to copy the currently selected images to the clipboard. While these functionalities were not fully implemented, they were hinted at by the buttons under the images. Allowing the user to transfer the data would enable them to view the edited series in their DICOM viewer or add them to the report (O4).

Besides displaying the various MRI sequences, the viewer lets the user preview the automatically generated structured graphic (J5.1). The graphic itself can be seen as the end-product of the software solution, which, once finished, can be transferred to the final radiological reports through the same means as the edited MRI images (O4). This feature automates the creation of the graphic, preventing execution errors such as inserting a wrong value (O6), making the process more efficient (O3), and ensuring a consistent output (O2.2). Moreover, to address the various ways radiologists may share the graphic, the prototype gave the option to export the non-image values and the sector map in one horizontally formatted graphic or separately in a vertical layout (O4). For example, while P3 might export the entire graphic to the PACS, P1 might prefer the insert the vertical layouts into his written report as they would fit better on the pages (C13).



(a) Full graphic (b) Sector map only

Figure 5.5: Structured Graphic

On the left side of the screen, the input form resides, giving the user access to all the values used to generate the lesion graphic. The input form is structured into multiple sections (see Figure 5.6). The sections are ordered based on the general steps of the diagnostic process that could be observed during the CI. This allows the radiologist to keep a familiar procedure, taking less cognitive effort to adapt (O3.2). Moreover, checkboxes were added to each section, allowing the user to document their process. This feature should account for the many disruptions that can occur during the work process (C21), helping the radiologist to more easily return to where they left off (O3.2).

Each section is structured like a table, where the row represents the current exam and each column a specific parameter, e.g., PSA level. The cells are input fields that contain the values. If there are prior exams registered for the same patients (C2), each gets a row with its values displayed in them. Through this, users can easily compare the values between the exams as seen in Figure 5.7a. Additionally, the prototype lets users view how much the values have changed since the last exam (see Figure 5.7b). Together, these features facilitate data comparison (O7) and eliminate the risk of miscalculations (O6).

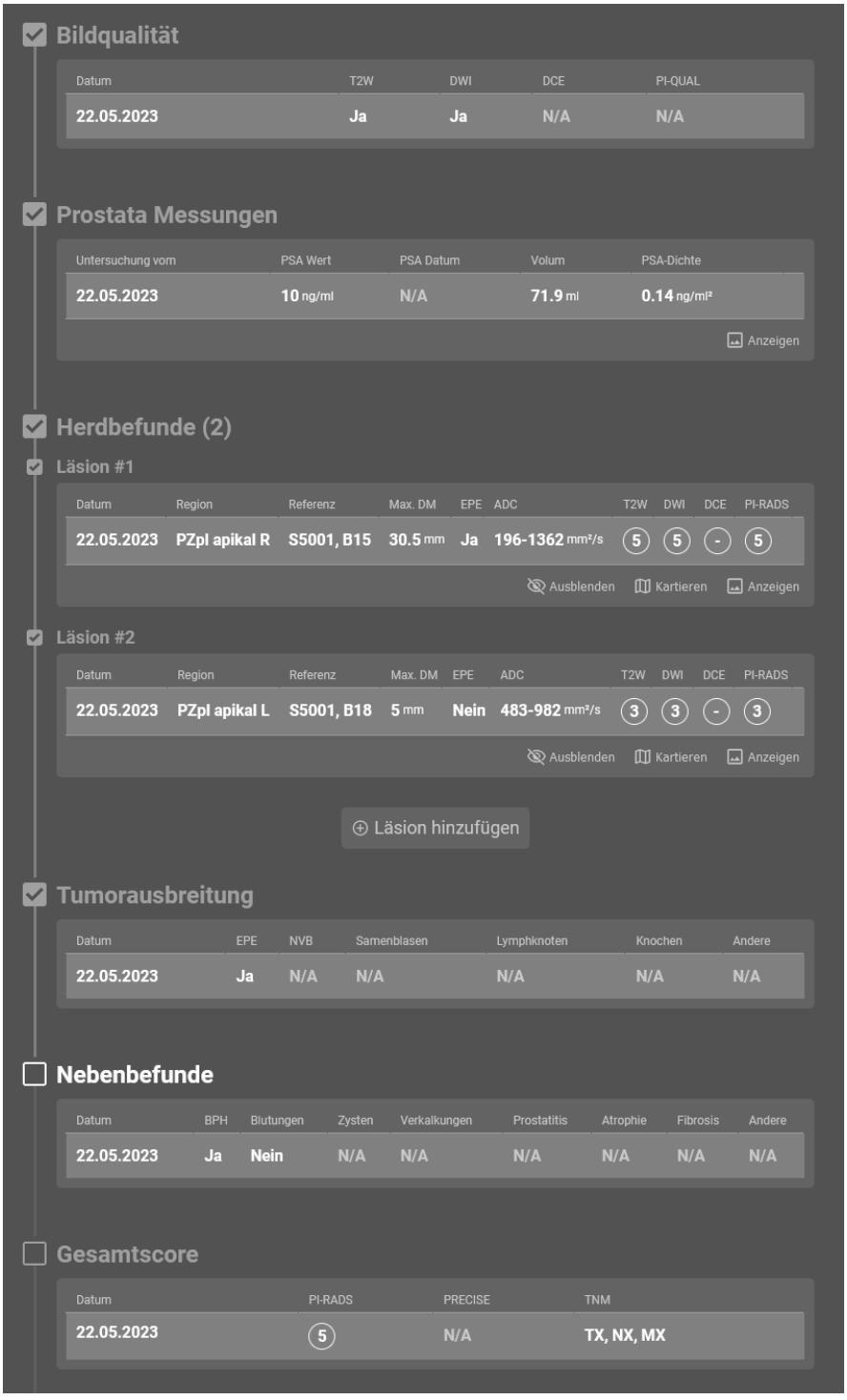
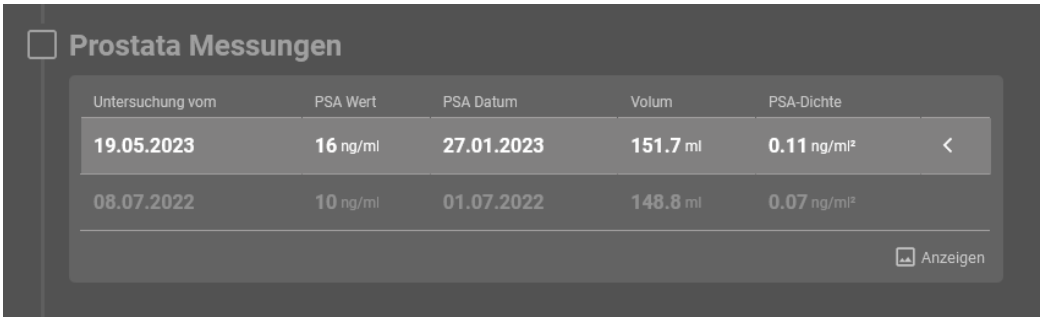
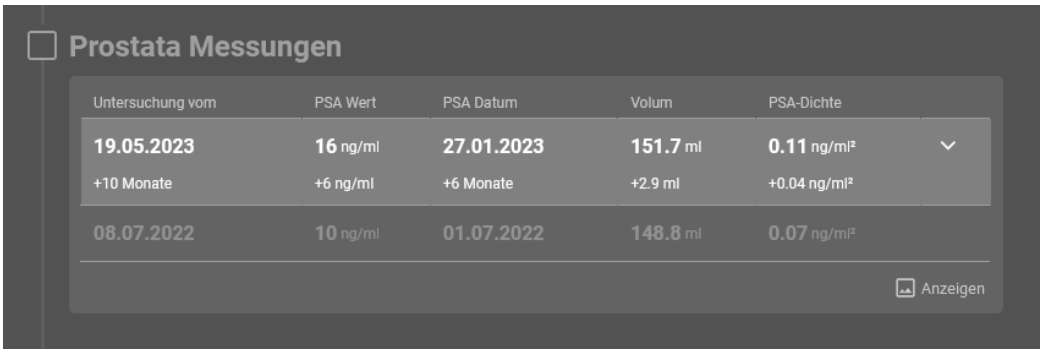


Figure 5.6: Full input form



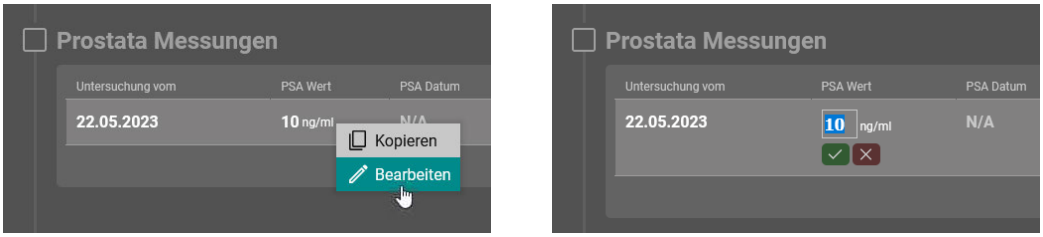
(a) Collapsed



(b) Expanded

Figure 5.7: Comparison between two exams

The user has multiple options to interact with the value fields in the input form. With a few exceptions, the user can edit all parameters, which can be done through the context menu accessed by right-clicking on a value. Alternatively, users can also double-click on an input field to start editing. This option was added to reduce the number of needed actions and make the interaction more efficient (C3). While some of the parameters are pre-filled by the AI, these values can be overwritten by simply editing them, allowing the user to easily dismiss (G8) and correct faulty values (G9). Besides editing the values, the context menu also allows the user to copy any value to the clipboard, making the transfer to various other tools the radiologist might work with easier and less prone to error (O4).



(a) Right-click context menu. (b) Edit mode.

Figure 5.8: Input field editing

The first section in the input form is the image quality (J4). Here, the information about the diagnostic quality of the T2W, DWI, and DCE sequence is documented either as of diagnostic quality or not. Based on the concept, this information was pre-filled from the AI algorithm (O3). However, since this information was not part of the received AI output, it was filled based on the information in the report associated with the particular MRI images. The section also contained a total image quality score in the form of the PI-QUAL score (Giganti et al., 2021). The PI-QUAL score was chosen as it would give a structured (O2.1), objective (1.3), and quantitative way (O1.3) of expressing the image quality.

The prostate measurements section contains information relevant to the determination of the prostate volume (J5) and the PSA density (J6). Specifically, it contains the PSA level, the date on which the PSA test was done, the prostate volume, and the PSA density as parameters. The prostate volume is calculated based on the prostate segmentation and automatically filled. Moreover, once a PSA level is inserted, the PSA density is also automatically displayed. This automation makes the process more efficient (O3) and less prone to calculation or transmission errors (O6). Another thing that was added to this section was a button that, when clicked, displays the axial T2W in the viewer on the right side, with the prostate segmentation mask active and on the first image where the segmentation begins. This allows the user to quickly scroll through the segmentation and assess its accuracy (G11) in order to validate the derived prostate volume. This further support the explanation of automatically derived values.

The lesion-finding section contains the relevant data associated with detected lesions (J8, J9). Under this section, each registered lesion is listed, each one containing a multitude of input fields. The region parameter allows the user to specify one of the prostate sectors for documenting the prostate location (J9.2). The reference field saves the series and image number of the

image where the lesion was located (O2.5). As an additional feature, when editing this parameter, a button lets the user directly transfer the reference of the image that is currently visible inside the viewer (O3). In the next field, the maximal diameter is automatically inputted based on the lesion delineation (J9.3, O3). In the next field, the presence or absence of EPE is automatically documented (J9.6), which was derived by comparing the lesion and prostate delineation. A range based on the minimal and maximal ADC values measured within the lesion is automatically displayed in the ADC field (J9.4). In the last four fields, the scores for PI-RADS scoring are saved, i.e., the T2W, DWI, DCE (J9.5), and the lesions' total PI-RADS score (J10). While the automatic assessment of the scores was implemented as part of the prototype, the PI-RADS score could be automatically derived based on the other registered values. However, it was also made possible to input one's own score in order to address situations in which there are no clear guidelines (C9), such as lesions in the AFS.

Besides the storing of values, the lesion-finding section also included other functions. Firstly, similar to the image quality section, a button was added to jump to the corresponding lesion delineation on the viewer quickly (O3). Furthermore, the prototype allowed users to add new lesions with the click of one button. This would create an empty lesion entry, which could then be filled out by the user. Moreover, registered lesions could be hidden in order to exclude them from the graphic in case the user wants to dismiss a lesion detected by the AI. Finally, for each lesion, there was a feature to edit the marking on the sector map in the graphic. However, this feature was not functionally implemented but only hinted at by a button. All these functionalities were primarily included to support the efficient dismissal and correction of the AI's output (G9).

The next two sections contained the documentation of cancer local and external cancer spreading (J9.6, J12), and the presence of additional findings (J13). Each of these sections included a number of fields in which the presence or absence of a particular case could be documented (O2.3). Specifically, these were EPE, NVB, seminal vesicles, lymph nodes, bones, and others for the tumor spreading section, and BPH, hemorrhages, cysts, calcification, prostatitis, atrophy, fibrosis, and others for the additional-findings section. The ladder was taken from the list of benign findings from the PI-RADS v2.1 paper (American College of Radiology, 2019).

In the last section, the final score for the final assessment can be registered. Specifically, the prototype lets the user define a PI-RADS, PRECISE, or

TNM score, which can also be added in combination. Through this, the design addresses the different medical questions and the specific answers they require.

5.3 Discussion and Conclusion

This chapter has presented the methodical development of the prototype utilized for the evaluation study. Throughout the ideation process, the advantages of the insights-based and systematic approach became apparent. The HMW questions acted as useful starting points to generate ideas in a targeted manner. Furthermore, utilizing the job statements and circumstance statements allowed further refinement of the concepts. In addition, the human-AI collaboration guidelines gave additional points for guiding the design of the AI functionalities. However, due to the specific domain of the nature, and the general applicability of the guidelines, some of the guidelines were not applicable to the solution.

In conclusion, the resulting prototype holds great promise as a robust tool for the evaluation phase. By addressing diverse objectives, requirements, and contextual factors through AI-driven concepts, the prototype offers a robust groundwork for exploring relevant aspects in collaboration with the end users.

Chapter 6

Study 2b: Prototype Evaluation

The second part of the prototyping study represents the final "test" stage within the design thinking process and consisted of the qualitative evaluation of the developed prototype with representatives of the target audience. This chapter presents the methodological considerations taken, as well as a detailed description and discussion of the results.

6.1 Methodology

The general objective of the prototype evaluation was to gain profound insights into the perceived effectiveness of AI-based solutions in supporting radiologists during prostate MRI diagnosis. Specifically, the study aimed to assess how well the proposed features met the user needs and to identify any potential shortcomings. However, beyond simply determining whether a feature was beneficial or not, the focus was on understanding the underlying reasons and factors contributing to its usefulness. This deeper exploration sought to yield valuable, transferable insights that could be applied beyond the scope of this particular prototype design. Additionally, the examination provided an opportunity to enhance comprehension of the work practice itself.

To this end, a qualitative approach has been chosen for the evaluation of the prototype. Although quantitative prototype evaluation is often utilized

to assess the overall effectiveness and performance of a design accurately, a qualitative evaluation offers a more in-depth exploration. By providing participants with the opportunity to share their perspectives and elaborate on their thoughts, context-richer, and potentially unconsidered data can be gathered. While the inclusion of quantitative assessments was contemplated, it was ultimately disregarded due to the limited number of participants, which aligns with the focus on conducting a thorough qualitative analysis and avoids the potential limitations of low statistical power.

The prototype evaluation featured the same three radiologists who had actively participated in the previous study (see section 4.1). Their involvement in the second phase of the research was driven by two primary factors. Firstly, their voluntary participation provided convenient access to the targeted audience necessary for the study. Secondly, having these radiologists on board allowed for seamless references to situations and insights gathered from the previous CI, enhancing the continuity and depth of the evaluation process. All participants signed a consent form prior to the sessions.

The study was thoughtfully designed to emulate the usage of a real product in an authentic work setting. To achieve this, the study was conducted via the video communication platform Zoom (Zoom Inc.), enabling the radiologists to participate remotely from their work computers within their natural work environment. This was made possible by providing them access to the prototype, hosted on a university server, through a standard internet browser. To capture comprehensive data, Zoom's screen-sharing and recording functions were used during the sessions. Additionally, the participants were given access to the anonymized MRI images corresponding to the prototype cases. This allowed them to import the MRI images into their DICOM viewer, allowing them to use the conventional tools during the evaluation.

The study followed a structured procedure that encompassed five main stages. Firstly, an introduction was provided to the participants, where the goals and overall procedure of the study were explained in detail. The limitations of the prototype were also transparently communicated. This was an important step to frame the participants' expectations and direct their focus on the right questions (Houde & Hill, 1997).

In the second stage, the prototype was presented to the participants, with each feature being individually showcased and thoroughly explained in terms of its functionality and purpose. Participants were actively encouraged to share their thoughts, raise questions, give suggestions, and provide feedback during this presentation phase. Initially, the prototype showcasing was con-

ducted using the researcher's computer and the screen-sharing function to demonstrate the features to the participants. However, after the first session, the protocol was revised. The revised approach involved letting the participants interact with the prototype directly and sharing their own screen while the features were being explained. The rationale behind the revision was to allow the participants to explore the features firsthand and develop a deeper understanding of how the prototype could support their work practice.

In the third stage of the study, the participants were asked to simulate a prostate MRI diagnosis using the provided MRI images and the prototype as a supportive tool. They were given a specific scenario that required them to diagnose one of the provided MRI studies, with the additional information that the images had already been processed by the assistive software, in which case they could find the results. Prior to starting the task, the participants were encouraged to share their screen, display the prototype, and were prompted to vocalize their thoughts throughout the process using the think-aloud protocol. This technique is widely used to gain insights into participants' actions and user experiences, providing a deeper understanding of their decision-making and opinions (Jääskeläinen, 2010).

During the fourth stage of the session, the list of potential features that were not implemented in the prototype was discussed. Each feature was described in terms of its function and how it could have been integrated into the prototype. Participants were asked to provide their insights on the potential benefits and downsides of each feature. Despite the lack of direct interaction, having been given the opportunity to familiarize themselves with the prototype is likely to have facilitated envisioning the planned features as potential additions to the software solution.

Finally, in the last stage, the entire process was discussed with the participants, inquiring about the total experience, how well the user needs were met, and what they would change or suggests. Moreover, any remaining questions that came up during the process were also discussed then. During these interviews and throughout the whole session, the why method was employed. This technique consists in continuously asking the participant for the reason behind their statements in order to gain a more profound understanding of the underlying beliefs and motivations. In total, each session was scheduled to last around 90 minutes.

For the data analysis, the sessions were transcribed and analyzed following thematic analysis. Multiple codes were defined in advance to ensure their focus. First, possible gaps in the job map were looked for. The results are

presented in the following section.

6.2 Results

During task completion, participants adhered to the intended setup by running the prototype and their DICOM viewer onto separate screens. Generally, the concept of assessing images within the DICOM viewer while simultaneously utilizing the AI solution on a distinct screen was deemed logical. *(P3:) "I do the diagnosis in the DICOM viewer anyways, and then I would cross-check on the other side."* This configuration also compensated for the prototype's viewer limitations, like the absence of features such as synchronized scrolling through multiple sequences. *(P1:) "Under the premise that you basically always have a DICOM viewer on at the same time where you look at the images, the basic idea of keeping it minimalistic is very good."* Conversely, P2 highlighted the potential advantages of direct integration into the existing PACS system, which could reduce cognitive load. *(P2:) "This is from the workflow more sensible because you stay on the same series with your eye. This way, you switch for the co-registration between two monitors."* However, the option to easily transfer edited images to the PACS was considered a plausible solution.

The presentation of prostate delineations and identified lesions on the T2W images was considered beneficial, enabling users to examine the AI algorithm's outcomes easily. The capacity to navigate through images and observe the complete output facilitated participants in comprehending the AI's conclusions comprehensively. Furthermore, visually indicating the images with active masks aided in the overview and identifying AI findings. *(P2:) "This facilitates and fastens of course a quick overview. If I want to see where did the algorithm find something, I don't need to scroll through the whole image set but can approach it directly in a targeted manner."*

The primary purpose of visual prostate delineation was to validate the automated volumetry. During this process, participants navigated through the T2W series with the segmentation mask enabled, comparing each image to assess if the AI-generated delineation aligned with their expectations. Throughout the evaluation, participants highlighted instances where the segmentation either omitted portions of the prostate or included external areas. In such cases, participants expressed a desire to edit the delineation directly on the image to rectify inaccuracies. They described this potential feature

as similar to the manual delineation process used in preparation for fusion biopsies. P1 and P3 estimated that correcting the segmentation this way would demand minimal time and effort, drawing from their experience with the manual delineation process. Consequently, this approach was deemed preferable over manual recalculation of prostate volume and overwriting the value in the input form. Additionally, the corrected segmentation would offer more precise measurements and ensure data accuracy. In contrast, the segmentation of the peripheral zone (PZ) received limited attention and was not extensively used during the diagnostic task.

The presentation of lesions on the T2W series received positive feedback from all participants due to its various advantages. The identification of suspicious regions within the image stack facilitated an easy assessment of the number, size, and position of detected lesions. Additionally, it enabled participants to validate the accuracy of delineations. Similar to the prostate segmentation, participants pointed out instances where they desired manual adjustments to the delineations. While discussing the necessity for editable AI delineations, the significance of preserving the original AI delineation was emphasized to enable comparison with the user's corrected version. To facilitate this, P3 suggested the possibility of overlaying both delineations for simultaneous viewing.

Regarding the individual mask selection, it was observed that participants would deactivate the prostate and PZ segmentation masks before examining detected lesions. This behavior highlights the practicality of this feature. However, activating lesion masks did present certain challenges. Participants did not always realize that multiple lesion masks could be selected, leading to situations where they believed a lesion was missing.

An aspect that participants felt was missing was the incorporation of visual cues within other sequences besides the T2W. Particularly in the DWI sequence, participants expressed a desire to visualize areas where the AI algorithm detected regions suggestive of clinically significant lesions. One participant specified that the DWI sequence was their primary choice for initiating lesion searches. Thus, having indications there would complement their diagnostic approach. Furthermore, it would enhance the mental process of correlating between various sequences, particularly when simultaneous viewing and scrolling through them are possible. However, transferring the T2W masks onto the DWI images was deemed unfeasible due to potential offsets between the two sequences.

All participants recognized the advantages of automatically generating a le-

sion graphic, citing multiple associated benefits. Primarily, the automation of graphic creation was noted for its time-saving and efficiency-enhancing attributes. Furthermore, it eradicates the potential for transfer errors and guarantees a consistent graphic outcome. Despite these benefits, participants also raised suggestions for enhancing the graphic.

One recommendation involved incorporating details surrounding the examination's indication. This encompasses factors like the exam type (e.g. process control), dates of preliminary exams, and available biopsy results. Incorporating such information would supply crucial context to the reader and streamline the comprehension of the remaining data.

An additional recommendation for enhancing the graphic pertains to including scientific references. The incorporation of these references would serve the purpose of enabling readers to explore potentially unfamiliar components within the graphic, such as the most recent PI-RADS version or the PI-QUAL scoring system. Furthermore, these references lend support to the scientific credibility of the methods employed and, consequently, the subsequent findings. *(P1:) "There are indeed people who are interested in that and look it up. And that underlines this objectionable and scientific aspect."* The significance of incorporating references was underscored by highlighting the potential diversity in the readers' backgrounds. *(P1:) "Such reports are not just seen by urologists, but also family physicians or other referrers."* While urologists are familiar with various prostate-related systems like PI-RADS, a general physician might need to look them up. In order to aid readers who are not acquainted with PI-RADS, it was also recommended to provide an explanation for the PI-RADS scores on the graphic.

The prostate sector map and the marked lesions also received positive feedback as well as suggestions for improvement. An aspect that was well-received involved marking lesions based on their delineation, resulting in a more precise representation. Nonetheless, there were reservations about the accuracy of the markings. For instance, large lesions were expected to be marked on all planes (base, middle, and apex) in which they were present. Additionally, one participant noticed that a delineation from the base was employed as a marking in the apex, leading to a misrepresentation of the actual scenario. Participants stressed the importance of an authentic depiction of lesions. *(P3:) "If I look at it like this, it's wrong or rather incomplete, and I would not share it like this."* Criticism also emerged regarding the lack of markings in the sagittal and coronal planes of the map. However, the significance of including this information varied among participants.

Moreover, P3 expressed a desire for an additional map feature, which involved the option to substitute the official PI-RADS v2.1 sector map with the custom map utilized in their department. Enabling the incorporation of this map would permit the retention of certain elements from the existing workflow while preserving the advantages that were perceived as absent in the PI-RADS v2.1 sector map. *(P3:) "This is a point that I criticized with this graphic is that there is no rectum drawn. Because that is important information for the urologists in my option."*

During the conversation about the advantages of employing colors to emphasize points in the lesion graphic, P1 proposed a feature to enable viewing the graphic in black and white. This suggestion was driven by the recognition that numerous referrers continue to rely on fax as a communication method for receiving reports, with many fax machines not supporting color. *(P1:) "But from experience, 95% of the referrers look at the printed fax reports, and then you wouldn't see this."* By offering a switch to a black-and-white mode, radiologists could verify the comprehensibility of a color-neutral version. Such functionality would be particularly crucial to ensure clear recognition of markings on the map in a black-and-white format.

Participants found the various export options provided for the finalized graphic to be suitable. Favorable feedback was received concerning the capability to present the table and map separately in an upright layout. P1 noted that exporting the graphic in this manner would align better with their written report's upright DINA 4 format. Additionally, providing the option to transmit the graphic as well as the segmentations to the PACS was also seen as practical as this allows to legally retain them for up to 10 years. This functionality was perceived as particularly valuable when distributing the report and medical images through the PACS. *(P1:) "And the uncomplicated sending to the PACS or copying or downloading for the report I like as well. That you can choose yourself which images you want to attach to the report."*

In the first session, the concept of integrating an automatically generated written report alongside the graphic was suggested by P3. Drawing from the input form values, sentences, or even cohesive text segments would be generated to partially or entirely automate the process of crafting the written report. P3 cited Smart Reporting (Smart Reporting GmbH), a documentation software he employed, as an exemplar for his proposal. The other participants also recognized the prospective merit of this feature, acknowledging its ability to save time and effort, enhance report consistency, and ensure a certain level of clarity. *(P1:) "Many, and I am one of them, do dic-*

tate less rather than more, and then there are many who write an incredible amount of prose, which no one reads through anyway. Therefore, automatic text editing or generation via ChatGBT would be cool." P2 expanded on this point, advocating the structured information within the graphic to largely replace continuous text, which should be confined to an executive summary. Through this summarized overview, readers can swiftly grasp the key points and determine whether delving deeper into the report is warranted. *(P2:) "When it says 'PI-RADS 3, PSA screening recommended,' then he reads that and checks it off. That is faster for him than reviewing the entire graphic. But if he reads five, then he'll need to check the entire sheet."*

Regarding the input form positioned on the left side, a unanimous positive response from all participants highlighted the structure's alignment with their current workflow. This arrangement allowed the participants to methodically progress through individual steps and seamlessly proceed once they are completed. *(P3:) "This already represents my approach very nicely, how I would work through the diagnosis. That's what I need. That I can go through this and mentally tick my things. And then I have my finished graphic and report."* Participants emphasized the advantages of time-saving and enhanced consistency offered by this approach. The introduction of checkboxes to mark off completed steps was also deemed beneficial, particularly in scenarios where the radiologist needs to resume the diagnosis process after a disruption. Such situations would be made more manageable by the ability to continue from where they left off.

An additional aspect that garnered unanimous positive feedback was the presentation of values from prior examinations, along with the corresponding differences compared to the current ones. This feature was regarded as a comprehensive overview that facilitates the simultaneous assessment of the progression of various variables. *(P2:) "This is very helpful for estimating the development of findings. Because there are also patients with inflammatory changes where there are always fluctuations."* Moreover, making the data available automatically elevates the user from having to retrieve the data first, e.g., from the RIS. *(P3:) "Especially if I already registered the data, then I don't need to search for it first. That's good."* However, some concern was raised regarding the software's ability to access the data automatically. Moreover, P2 highlighted that in their workflow, they review data from earlier examinations before analyzing the current images. With regard to showcasing the differences, participants recognized the benefit of improved visibility and reduced potential for miscalculation. While this information was viewed as relevant for referring urologists, integrating it into the graphic's table was

met with reluctance due to concerns about compromising its clarity.

Regarding the interaction with the input fields, several observations could be noted. Participants utilized both the right-click and double-click methods to modify values. However, during the task, participants occasionally encountered challenges with the interaction. Following value adjustments, they would at times initiate editing the next value before confirming their selection. Conversely, the ability to copy a value was neither remarked upon nor utilized during the task.

A topic that sparked extensive discussion revolved around the capability to replace the automatically generated values from the AI algorithm, such as the prostate volume. This capacity was deemed crucial for ensuring the software's applicability in diagnostic practice, enabling radiologists to proceed with generating a report featuring values they deem accurate. Furthermore, multiple requirements were highlighted in the event of AI results being overwritten by the user. Firstly, the revised value should be clearly identified as such to prevent users from mistaking the new value for AI-generated. Moreover, this differentiation would prove advantageous when crafting the eventual report, as discrepancies from the AI output could be worth communicating to the referrer.

Secondly, the original AI value should be retained and remain accessible even after being replaced, for instance, in a log file as suggested by P1. Participants conveyed their desire to track the AI's outcomes in relation to their own assessments over time, allowing them to evaluate the AI's efficacy and their own learning trajectory. *(P3:) "And when I see it and that I need to change, then I would like to see retrospectively what did I alter, for the communication with my referrer, my learning curve, and also for you as a feedback loop."* The ability to readily identify altered variables and access the original values also contributes to the retroactive traceability of the report. Ensuring that the report remains comprehensible in the future was deemed significant, especially in cases where the results need to be reevaluated. However, P2 expressed concern about the storing of overwritten AI values, as they might be used against the radiologist in a legal case. *(P2:) "[...] and the first thing the lawyer does is to get the old findings and says, 'Look, the AI already said back then that there was something there.'"*

Regarding the specific components within the input form, the image quality section was perceived as an essential element of the workflow and a critical detail to convey to the referring physician. The binary classification of image quality into diagnostic and non-diagnostic categories was not immediately

evident, but upon clarification, was endorsed by all participants. The structured manner in which quality assessment was approached was regarded as beneficial, as it not only guarantees a certain standard of quality but also offers documentation of the process. *(P2:) "That's good. This is also a procedure that is known for cardiac diagnostics. [...] There there are also tabular prefabricated diagnostic tools available."* The inclusion of the PI-QUAL score garnered positive feedback, despite some participants being either unfamiliar with it or not currently incorporating it into their workflow. This addition was valued for offering a quantitative and more objective measure in contrast to a qualitative score. *(P1:) "Good, bad, medium, that is always subjective."* Furthermore, the automated evaluation of image quality was also regarded as advantageous, as it serves as an indicator of whether the image quality meets the criteria for the AI to produce dependable outcomes.

The participants recognized the prostate measurement section as important to the overall process, and each of the displayed values as being relevant to the graphic. Especially the automated calculation of the prostate volume and PSA density was seen as an effective form of support. *(P1:) "This is in principle what facilitates one's work, that you don't have to type into your calculator or ask Siri."* The possibility of jumping to the prostate segmentation from this section was also appreciated for its practical convenience. An additional feature independently suggested by P2 was the utilization of color to emphasize abnormal values, for example, when the PSA density has reached a certain threshold. However, P1 did not share the same preference for color signaling and instead recommended the option to adjust this setting within a dedicated settings menu.

The subsequent lesion-finding segment elicited substantial feedback from participants. Generally, the integration of variables alongside automatic registration utilizing detected lesions was regarded as advantageous assistance for the diagnostic process. Specifically, the individual input fields garnered a diverse range of comments.

Registering the prostate sector from which a lesion originates was deemed essential information to convey. Yet, a unanimous criticism voiced by all participants was the limitation of selecting only one section per lesion. This limitation raised significant concern due to the potential for cancer lesions to span multiple sectors, which necessitates accurate communication to the referrer. Restricting the selection to just one sector in such cases would inaccurately represent the situation, presenting a substantial concern. *(P3:) "That's important, otherwise you come to this situation and then I don't*

know how to handle this. Then reality and report contradict." However, it was acknowledged that including a list of all impacted sectors of a large lesion in the graphic table might compromise its clarity, presenting an ongoing challenge that remains unresolved in the participants' present workflow.

Adding an image reference to pinpoint the location of detected lesions garnered positive feedback from all participants, particularly those who had not previously included this information in their reports (P1 and P2). *(P1:) "The more I think about it, I think it's a good idea because you give the urologist information: Where do I have to look? Where is the finding? In which image?"* Facilitating the process of finding the documented lesion within the MRI data was recognized as a distinct benefit. To enhance this further, P1 recommended incorporating the relevant sequence name, such as "T2W", into the reference. This modification would enhance the recognition of the reference, aiding in planning procedures like fusion biopsies. Furthermore, P3 also conveyed the desire to choose a range of images from a sequence to document all occurrences of a lesion being visible. However, indicating multiple sequences was considered unnecessary, as it is feasible to navigate to the equivalent location of other sequences in the DICOM viewer by linking the series.

The presentation and automated measurement of the maximum lesion diameter were perceived as beneficial functionalities. However, a requested addition was an indication of the specific location on the image where the measurement was taken. This information was deemed valuable as it would enable users to verify whether the AI accurately identified the correct location for measurement.

In addition to the maximum lesion diameter, P1 and P3 agreed on the advantage of automatically calculating and displaying the lesion volume within the input form. This data was deemed more accurate than the maximum diameter, making it more valuable for monitoring lesion growth. *(P3:) "I could imagine that for when doing progress control after focal therapy."* However, it was suggested that this information should not be included in the graphic. Another mentioned reason for its inclusion was its potential relevance in future guidelines. On the contrary, P2 did not see the practicality of working with lesion volume since it is not yet part of the guidelines and was considered less reliable than the maximum diameter. *(P2:) "And that's what is still the standard in oncology. The maximal diameter is used as a marker if a lesion gets larger or smaller."*

Regarding the EPE variable, P1 highlighted the necessity for an extra "prob-

able" option since the confirmation of EPE's presence is not always definitive on MRI images. *(P1:) "There are studies that say that even if you don't see any extraprostatic growth on the MRI, but the tumor has 5cm contact to the capsule or more, you should assume there is a capsule penetration."*

The automatic determination of the lesions' ADC range and its displaying in the graphic received an array of feedback points. Altogether, the participants agreed that the ADC range alone has only limited usefulness, both for the radiologist and the referring physician. One contributing factor that was mentioned is the absence of guidelines that define which ADC value is significant for characterizing a lesion. *(P1:) "I'm not aware of what ADC value is referred to in order to characterize the extent of the diffusion disorder. Is it the lowest value that I measure in the lesion? Is it the average or highest value? To my knowledge, that is not unambiguously clarified yet."* Furthermore, the size of the range was not seen as specific enough. *(P2:) "Because an ADC value of 1634 is totally normal and an ADC value of 408 is highly pathological. The statement is basically null."* For these reasons, the mean ADC value was proposed as a valuable addition and a superior metric to communicate to the referrer. Generally, however, the ADC values were considered more meaningful to radiologists than to referring physicians.

Another point of discussion concerning ADC measurements was the selection of the region for determining the values. Participants found it unclear which area was utilized for the measurement and desired a clear visualization. This was seen as crucial since ADC values can vary significantly within a lesion, making the decision on where to measure an important factor. *(P2:) "The art is to find where to measure where is the probability highest that you find a tumor. Tendency to take small ROIs and measure the low-intensity areas, since ADC drop are an important criterion for malignancy."* The selected area for measurement also influences the tracking of ADC value changes across multiple exams. To ensure a valid comparison, the measured areas for the two assessments should be the same. Visualizing the selected ROI would enable verification of this consistency. Therefore, the capability to modify the chosen ROI was deemed an important addition. Furthermore, the suggestion to include the standard deviation was put forth to provide additional context to the ADC values.

Incorporating the scores for each sequence, as well as the overall final score, was considered a crucial aspect of the assisted workflow. The automated calculation of the final lesion score was regarded as a practical feature. However, determining the version of PI-RADS used for the final score was not

immediately evident and required clarification in one session. The option to manually input the final score was also valued, particularly for unique cases like those in the AFS or CZ regions, enabling personalized scoring based on the specific situation. The potential inclusion of visual alerts in cases where the manually assigned score contradicts the PI-RADS algorithm was regarded as a beneficial safeguard. *(P3:) "Especially for beginners, that they can cross-check, do I have a thinking error. But also not bad for me. Sometimes you have a knot in your head."* A feature that was missed by the participants was the AI-generated classifications for each sequence, as it would provide a second quantitative evaluation of the image data. *(P2:) "This is the part that is still very subjective, the classification according to PI-RADS. Is this a lesion that is clearly outlined in the T2 image, or is there an area that is blurry?"*

The actions available for the lesion-finding section were favorably received by all participants. The "jump" function, enabling users to navigate to relevant images with a single click, was utilized during the task and regarded as convenient. However, participants faced challenges in associating lesions listed in the input form with their delineations on the T2W image when multiple lesions were recorded. Similar to other AI-generated data, the ability to modify map markings was deemed necessary to integrate the software into the clinical workflow. This also extended to the capability of removing registered lesions as well as adding new ones. However, P1 and P3 proposed an alternative approach to manually registering lesions, which entailed drawing their outline directly on the MRI image. This method would allow the program to auto-populate values based on the image data within the delineated region. The process of drawing the outline was compared to editing existing AI-generated delineations. The anticipated advantages of this method included improved efficiency due to fewer required interactions. *(P1:) "This is defensively easier than doing a thousand clicks."* However, P2 expressed skepticism about the anticipated reduction in effort in practical use, arguing that lesions missed by the AI will likely be more challenging to delineate manually.

The values presented in the staging section were deemed relevant to the graphic. However, for the NVB and seminal vessels, mere presence or absence was seen as insufficient. Distinguishing between left, right, or both sides was deemed crucial. Furthermore, for the "others" category, participants desired the ability to specify the exact organ involved, possibly through a free text field. Overall, the "others" option was valued due to the rarity of some organ invasions.

The response to the additional findings section and the display of values in the lesion graphic was varied. According to P3, variables should only appear on the graphic if their presence or absence can impact the lesions. *(P3:) "There are cases where it is important to mention prostatitis on the form because it can put the lesion into a different context, but I don't need to write on the form that he has a BPH or atrophy."* Moreover, presenting the entire range of findings could potentially clutter the visual clarity of the graphic and divert attention from the most crucial information. P1 even suggested excluding this section entirely from the graphic. Nonetheless, systematically going through the list of potential additional findings and having them stored within the program was still perceived as advantageous. Consequently, the option to choose which variables to display on the graphic was viewed as valuable. P1 and P3 believed that such variables should be noted in the written report, which offers more comprehensive documentation. On the other hand, P2 took a different stance, supporting the inclusion of all additional values in the graphic as he aimed for a predominantly structured report format.

The feedback regarding the various final scores varied among participants. Each participant expressed different preferences regarding which scoring system to utilize. P2 favored using only the PI-RADS score, P3 preferred using both the PI-RADS score and the TNM formula, while P1, accustomed to using the PI-RADS and PRECISE scores, considered integrating the TNM formula. Despite the divergence in preferences for different scoring systems, all participants acknowledged the advantage of having the option to select the desired score. *(P3:) "PRECISE was not really a subject with us, I don't have much experience with it. It's mainly relevant for progress control and active surveillance. There we don't have that many. But it's good that it's included, especially for ambulances it will become increasingly relevant."* A point regarding the TNM formula that was discussed involved the need to clearly indicate that it is based on image data ("cTNM"), so as not to be mistaken for a pathological score ("pTNM").

Examining frameworks like the TNM formula was a scenario where the potential feature of info boxes was relevant. While working on the task, P1 encountered a situation where he wanted to select the correct TNM score but felt uncertain about which one to choose. This highlighted the need for additional information, as he mentioned that he would typically refer to external sources for clarification. *(P1:) "Especially for TNM, because those are things that I personally cannot remember by heart anymore. I always look that up."* Overall, the concept of the feature was well-received, as it could streamline the process of seeking information. Particularly, the idea

was deemed valuable for less experienced radiologists. Nonetheless, one participant highlighted the importance of maintaining the included information over time.

Regarding the prototype as a whole, participants expressed their endorsement of the concept. *(P1:) "I would take it how it is right now, with the points we discussed, the changes, improvements, etc. I would take it, let it be certified, and bring it to the market."* However, it was also emphasized that the solution's effectiveness hinges on the performance of the AI algorithm.

6.3 Discussion and Conclusion

The evaluation of the prototype has provided valuable insights into several dimensions of integrating AI-based solutions into the practice of prostate MRI diagnosis. Alongside supporting the efficacy of the design concept in meeting user needs, as evidenced by the overall positive feedback, the evaluation has also revealed multiple considerations essential for the development of AI solutions in this domain.

When evaluating the AI outcomes, a significant requirement that emerged was the need for verification. Despite the absence of specific XAI techniques like heat maps to enhance the understanding of AI outputs, participants still utilized available information to interpret the automatically generated results. For instance, the ability to navigate through images with the displayed prostate segmentation enabled participants to assess the accuracy of the given prostate volume. However, for other automatically computed metrics, such verification mechanisms were missed. For instance, participants wanted to determine where the maximum lesion diameter was measured or which group of pixels contributed to the ADC value. Providing the users with such information allows them to compare the measurements to their personal approach, empowering them to leverage their domain expertise. For instance, they can assess whether the lesion segmentation meets their expectations based on the image data.

Closely related to validating AI-generated results is the essential requirement to have the capability to rectify them. When users identify inaccuracies in the generated output, proceeding with the incorrect value is not a viable choice. Hence, enabling radiologists to rectify errors while minimizing disruption to the workflow holds the utmost significance. The prototype evaluation also highlighted the benefits of directly correcting the output within the image

data, such as adjusting the prostate delineation rather than overwriting the prostate volume. This approach would enable users to seamlessly continue working with the delineation instead of discarding it due to perceived flaws.

Alongside the capability to rectify the AI-generated results emerged a necessity to preserve the overwritten output and ensure its clear visual differentiation. The preservation of data enables users to make comparisons between their own efforts and the AI output. Additionally, participants recognized the potential to track discrepancies in numbers over time between AI and user assessments, indicating a desire for AI performance monitoring. In such instances, a specialized feature like a dedicated dashboard might be of value. Conversely, it is crucial to establish a noticeable visual contrast between rectified values and AI-generated output to prevent any potential confusion. This is especially important as radiologists might wish to indicate which results originated from AI and comprehend the process leading to a particular outcome in retrospect.

While the three participants shared the same opinion with most of the prototype's features, the discord on specific aspects exemplifies the variation between the values, preferences, and workflows of individual radiologists. For example, different opinions were given to the importance of the lesion volume, the information included in the report graphic, and the use of color to highlight elements. For the development of medical development, this represents a challenge since a design decision that is appreciated by some of its users can be an inconvenience to others. One aspect that supported a seamless interaction across participants was the possibility to omit values so that they would not appear in the graphic allowing the user to choose which values are relevant. Another approach that has been mentioned by the participants is to allow users to customize the available features in the settings.

One interesting aspect that became apparent during the conduction of the task was how the availability of the data impacts the order of the workflow. Participants have mentioned that information about the prostate volume and PSA density, and in cases of a follow-up, the history is reviewed at the beginning of the diagnostic process. On the other hand, they preferred viewing the lesion detected by the AI after having examined the images independently during the session. Considering the possible processing time of the algorithm, it should be considered if certain information, such as the prostate segmentation, is given before the system has completed the detection process.

Moreover, the sessions highlighted the importance of supporting a smooth and efficient workflow. Features that supported a smooth work procedure were highly appreciated, such as the ability to quickly jump to the respective lesion delineation from the input form. On the other hand, usability issues, such as not being able to easily match different lesions registered in the input form to their respective delineation, led to disruption in the workflow. However, the most critical aspects that lead to a full stoppage of the process. For example, not being able to register multiple prostate regions for the location of the lesion in the case of a large tumor would have led to an abortion of the software since delivering erroneous or incomplete information would not be an option. This points to the importance of considering any possible cases that might need to be handled by the solution.

While the study has provided valuable insights into the research goal, it is essential to acknowledge its various limitations. The first limitation arises from the participant sample. As previously established in the CI study, the relatively modest size and inclination towards highly experienced radiologists might result in an incomplete representation of the diverse variations, perspectives, or nuances within the broader radiologist population. In addition, having sampled participants from the existing pool of radiologists associated with the PAIRADS research project could potentially incline them towards a lesser degree of skepticism towards AI solutions, on average. Moreover, reusing participants from the CI study in this research might have led to the prototype solution being more tailored to their specific needs, possibly yielding a more positive response compared to the inclusion of different radiologists.

An additional constraint arises due to the prototype's limited scope and the range of test cases involved. As evidenced by the CI study and the extensive list of circumstance statements, a substantial number of scenarios exist. Unfortunately, the scope of this research was insufficient to encompass all conceivable variations in the assessment. This suggests that potential insights specific to these cases may not have been fully captured. For instance, a scenario involving a more ambiguous case where lesion localization could be more challenging might have underscored the value of segmenting the PZ.

Regarding AI technology, certain limitations exist based on the variety of AI methods replicated within the prototype. Notably, a technique included in the PAIRADS concept, namely the assignment of a PI-RADS score to identified lesions, was omitted. Such AI methods are likely associated with unique requirements, particularly concerning aspects like explainability and

interpretability. These specific requirements warrant consideration for future developments.

Lastly, recognizing the constraints posed by the controlled environment in which the evaluation took place is crucial. Despite efforts to enhance authenticity, such as allowing participants to perform the study in their actual work settings, these conditions inherently deviate from actual practice. Unforeseen challenges and socio-technical influences may only manifest in real-world scenarios, warranting cautious interpretation of the insights gathered.

Nonetheless, this research approach has demonstrated its effectiveness in exploring additional user needs and system requirements for integrating AI into prostate MRI diagnosis practices. The prototype played a significant role in facilitating discussions about specific features, their alignment with user needs, and their influence on workflow. Furthermore, it stimulated participants to contribute their creative ideas. For instance, the proposal to add new lesions by directly drawing them onto the image was likely inspired by the specific structure and interactions of the prototype. These user suggestions hold particular significance as they stem from the practical expertise of the practitioners. In this way, the prototype not only aided in the research process but also fostered valuable insights from those immersed in the field.

Chapter 7

Conclusion

This thesis aimed to enhance the user-centered approach to the development of AI-driven medical solutions. To do so, the research was structured following the double diamond framework, first uncovering the problem space before then turning to the creation and evaluation of possible solutions. Initializing with an in-depth exploration of the objectives, contexts, and requirements within the work practice allowed for the informed creation of a solution prototype. Moreover, the subsequent evaluation with the end-users uncovered further aspects relevant to the development of such solutions. The investigation underscored the effectiveness of a design thinking approach for developing appropriate solutions.

Moreover, the distinct insights from each individual study bring their own unique contributions. The JTBD statements derived from the CI study provide a valuable understanding of the prostate MRI diagnosis practice and can aid in the design of further innovative solutions. Although these statements were employed to develop a specific AI-based prototype for evaluation purposes, the framework's solution-agnostic nature allows them to inform diverse potential solutions, including those not driven by AI technology.

The insights derived from the second study offer a focused contribution by offering a nuanced understanding of the potential opportunities and challenges associated with the targeted implementation of AI within the practice. By affording end-users the opportunity to engage with and articulate their perspectives on a specific design solution, the study provides specific insights into the ways particular design elements influence the practice. Through open exchanges with participants, the study unveiled underlying needs, mo-

tivations, concerns, and other factors underpinning their responses, thereby providing insights of broader applicability. Moreover, by enabling participants to share their individual ideas and recommendations for the solution, the study deepens comprehension of their perspectives on how they wish to be supported.

Looking forward, there are several promising directions for future research that can build upon the groundwork laid by this study. Although the research adhered to a design thinking approach, certain aspects could only be partially implemented. For example, in the context of actual product development, a greater number of iterations would be necessary. Subsequent studies could leverage the insights gained from this research to improve on the proposed design concept and reevaluate the improved version. This approach would not only foster more effective design solutions but also deepen the understanding of user needs and optimal ways to address them.

Another design thinking aspect that could be further expanded is the targeting of a broader range of stakeholders in the research process. While radiologists constitute a critical end-user group, other parties are also impacted, such as patients and referring physicians. The research highlighted how radiologists anticipate the needs of referring physicians, yet these needs were interpreted through the lens of radiologists rather than directly from the referring physicians themselves. Incorporating their perspectives in future studies could yield valuable insights into their specific needs, enhancing the value that a solution offers within the larger system. For a more comprehensive perspective on system adoption, the involvement of additional stakeholders like hospital administrators may be warranted.

Furthermore, future endeavors might also encompass greater stakeholder involvement in the creative process. Collaborative creation of solutions is a pivotal tenet of design thinking. Enlisting a broader spectrum of contributors, including developers, AI experts, and designers, can tap into a diverse array of viewpoints, fostering innovation and integrating those responsible for various aspects of product creation into the user-centered approach.

Not least, a critical necessity emerges to assess the impacts of the generated solution on the diagnostic performance within the real clinical setting. While this research places substantial importance on subjective user experience as an inherent component for the development of effective supportive tools, the ultimate goal remains the improvement of patient outcomes. This endeavor entails not only quantitative performance analysis but also a qualitative exploration of additional socio-technical elements influencing real-world perfor-

mance.

In conclusion, his research endeavor takes a small yet vital step toward more effective AI-driven medical systems by elevating the significance of user needs. While the intersection of AI and healthcare presents a large range of complex and diverse challenges, upholding the importance of a human-centered approach continues to be an indispensable component in creating innovative AI solutions that genuinely enhance clinical work practices and contribute to positive transformation.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (xai). *IEEE access*, *6*, 52138–52160.

Alberdi, E., Povyakalo, A., Strigini, L., & Ayton, P. (2004). Effects of incorrect computer-aided detection (cad) output on human decision-making in mammography. *Academic radiology*, *11*(8), 909–918.

Altman, M., Huang, T. T., & Breland, J. Y. (2018). Peer reviewed: Design thinking in health care. *Preventing chronic disease*, *15*.

American College of Radiology. (2019). *Prostate imaging – reporting and data system, version 2.1.* `https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/PI-RADS`. (Accessed: July 31, 2023)

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... others (2019). Guidelines for human-ai interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1–13).

Balint, B. J., Steenburg, S. D., Lin, H., Shen, C., Steele, J. L., & Gunderman, R. B. (2014). Do telephone call interruptions have an impact on radiology resident diagnostic accuracy? *Academic radiology*, *21*(12), 1623–1628.

Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., & Horvitz, E. (2019). Beyond accuracy: The role of mental models in human-ai team performance. In *Proceedings of the aaai conference on human computation and crowdsourcing* (Vol. 7, pp. 2–11).

Barrett, T., & Haider, M. A. (2017). The emerging role of mri in prostate cancer active surveillance and ongoing challenges. *American Journal of Roentgenology*, *208*(1), 131–139.

Bates, D. W., Kuperman, G. J., Wang, S., Gandhi, T., Kittler, A., Volk, L., ... Middleton, B. (2003). Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. *Journal of the American Medical Informatics Association*, *10*(6),

523–530.

Beede, E., Baylor, E., Hersch, F., Iurchenko, A., Wilcox, L., Ruamviboonsuk, P., & Vardoulakis, L. M. (2020). A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In *Proceedings of the 2020 chi conference on human factors in computing systems* (pp. 1–12).

Beyer, H., & Holtzblatt, K. (1999). Contextual design. *interactions*, *6*(1), 32–42.

Bijl, D., Blaumer, N., & Matuschek, D. (2022). Pairads: Interaction of humans and technology rethought. *Trust, Professional Vision, and Diagnostic Work..... 19*, 51.

Blezek, D. J., Olson-Williams, L., Missert, A., & Korfiatis, P. (2021). Ai integration in the clinical workflow. *Journal of Digital Imaging*, *34*, 1435–1446.

Brierley, J. D., Gospodarowicz, M. K., & Wittekind, C. (2017). *Tnm classification of malignant tumours*. John Wiley & Sons.

Cabitza, F. (2019). Biases affecting human decision making in ai-supported second opinion settings. In *Modeling decisions for artificial intelligence: 16th international conference, mdai 2019, milan, italy, september 4–6, 2019, proceedings 16* (pp. 283–294).

Cabitza, F., Campagner, A., & Balsano, C. (2020). Bridging the "last mile" gap between ai implementation and operation:"data awareness" that matters. *Annals of translational medicine*, *8*(7).

Cabitza, F., Campagner, A., & Sconfienza, L. M. (2021). Studying human-ai collaboration protocols: the case of the kasparov's law in radiological double reading. *Health information science and systems*, *9*, 1–20.

Cabrera, Á. A., Perer, A., & Hong, J. I. (2023). Improving human-ai collaboration with descriptions of ai behavior. *arXiv preprint arXiv:2301.06937*.

Cai, C. J., Reif, E., Hegde, N., Hipp, J., Kim, B., Smilkov, D., ... others (2019). Human-centered tools for coping with imperfect algorithms during medical decision-making. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1–14).

Cai, C. J., Winter, S., Steiner, D., Wilcox, L., & Terry, M. (2019). " hello ai": uncovering the onboarding needs of medical practitioners for human-ai collaborative decision-making. *Proceedings of the ACM on Human-computer Interaction*, *3*(CSCW), 1–24.

Calisto, F. M., Santiago, C., Nunes, N., & Nascimento, J. C. (2022). Breastscreening-ai: Evaluating medical intelligent agents for human-ai interactions. *Artificial Intelligence in Medicine*, *127*, 102285.

Choy, G., Khalilzadeh, O., Michalski, M., Do, S., Samir, A. E., Pianykh,

O. S., ... Dreyer, K. J. (2018). Current applications and future impact of machine learning in radiology. *Radiology*, *288*(2), 318–328.

Christensen, C. M., Hall, T., Dillon, K., & Duncan, D. S. (2016). Know your customers' jobs to be done. *Harvard business review*, *94*(9), 54–62.

Design Council. (2023). *The double diamond: A universally accepted depiction of the design process.* Website. Retrieved from `https://www.designcouncil.org.uk/our-resources/the-double-diamond/` (Accessed: July 31, 2023)

Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF. (2021). *Leitlinienprogramm onkologie: S3-leitlinie prostatakarzinom, langversion 6.2.* `https://www.leitlinienprogramm-onkologie.de/leitlinien/prostatakarzinom/`. (Accessed: July 31, 2023)

Dunn, M. W., & Kazer, M. W. (2011). Prostate cancer overview. In *Seminars in oncology nursing* (Vol. 27, pp. 241–250).

Eiband, M., Schneider, H., Bilandzic, M., Fazekas-Con, J., Haug, M., & Hussmann, H. (2018). Bringing transparency design into practice. In *23rd international conference on intelligent user interfaces* (pp. 211–223).

Erdmann, F., Spix, C., Katalinic, A., Christ, M., Folkerts, J., Hansmann, J., ... others (2021). Krebs in deutschland für 2017/2018.

Fogliato, R., Chappidi, S., Lungren, M., Fisher, P., Wilson, D., Fitzke, M., ... Nushi, B. (2022). Who goes first? influences of human-ai workflow on decision making in clinical imaging. In *Proceedings of the 2022 acm conference on fairness, accountability, and transparency* (pp. 1362–1374).

Gagnon, M.-P., Ngangue, P., Payne-Gagnon, J., & Desmartis, M. (2016). m-health adoption by healthcare professionals: a systematic review. *Journal of the American Medical Informatics Association*, *23*(1), 212–220.

Giganti, F., Kirkham, A., Kasivisvanathan, V., Papoutsaki, M.-V., Punwani, S., Emberton, M., ... Allen, C. (2021). Understanding pi-qual for prostate mri quality: a practical primer for radiologists. *Insights into Imaging*, *12*(1), 1–19.

Greenhalgh, T., Wherton, J., Papoutsi, C., Lynch, J., Hughes, G., Hinder, S., ... others (2018). Analysing the role of complexity in explaining the fortunes of technology programmes: empirical application of the nasss framework. *BMC medicine*, *16*(1), 1–15.

Grots, A., & Pratschke, M. (2009). Design thinking—kreativität als methode. *Marketing Review St. Gallen*, *26*, 18–23.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). Xai—explainable artificial intelligence. *Science robotics*, *4*(37), eaay7120.

Harder, F. N., Heming, C. A., & Haider, M. A. (2023). mpmri interpretation in active surveillance for prostate cancer—an overview of the precise score. *Abdominal Radiology*, 1–7.

Hartswood, M., Procter, R., Rouncefield, M., Slack, R., Soutter, J., & Voss, A. (2003). 'repairing'the machine: A case study of the evaluation of computer-aided detection tools in breast screening. In *Ecscw 2003: Proceedings of the eighth european conference on computer supported cooperative work 14–18 september 2003, helsinki, finland* (pp. 375–394).

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(4), e1312.

Hong, S. R., Hullman, J., & Bertini, E. (2020). Human factors in model interpretability: Industry practices, challenges, and needs. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW1), 1–26.

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer*, *18*(8), 500–510.

Houde, S., & Hill, C. (1997). What do prototypes prototype? In *Handbook of human-computer interaction* (pp. 367–381). Elsevier.

Jääskeläinen, R. (2010). Think-aloud protocol. *Handbook of translation studies*, *1*, 371–374.

John-Paul, J. Y., Kansagra, A. P., & Mongan, J. (2014). The radiologist's workflow environment: evaluation of disruptors and potential implications. *Journal of the American College of Radiology*, *11*(6), 589–593.

Jorritsma, W., Cnossen, F., & van Ooijen, P. M. (2015). Improving the radiologist–cad interaction: designing for appropriate trust. *Clinical radiology*, *70*(2), 115–122.

Kalbach, J. (2020). *The jobs to be done playbook: Align your markets, organization, and strategy around customer needs.* Rosenfeld Media.

Kocielnik, R., Amershi, S., & Bennett, P. N. (2019). Will you accept an imperfect ai? exploring designs for adjusting end-user expectations of ai systems. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1–14).

Ku, B., & Lupton, E. (2022). *Health design thinking: creating products and services for better health.* mit Press.

Larson, D. B., Froehle, C. M., Johnson, N. D., & Towbin, A. J. (2014). Communication in diagnostic radiology: meeting the challenges of complexity. *American Journal of Roentgenology*, *203*(5), 957–964.

Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the ai: informing

design practices for explainable ai user experiences. In *Proceedings of the 2020 chi conference on human factors in computing systems* (pp. 1–15).

Liao, Q. V., Pribić, M., Han, J., Miller, S., & Sow, D. (2021). Question-driven design process for explainable ai user experiences. *arXiv preprint arXiv:2104.03483*.

Liao, Q. V., & Varshney, K. R. (2021). Human-centered explainable ai (xai): From algorithms to user experiences. *arXiv preprint arXiv:2110.10790*.

Loeb, S., Bjurlin, M. A., Nicholson, J., Tammela, T. L., Penson, D. F., Carter, H. B., ... Etzioni, R. (2014). Overdiagnosis and overtreatment of prostate cancer. *European urology*, *65*(6), 1046–1055.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, *267*, 1–38.

Miller, T., Howe, P., & Sonenberg, L. (2017). Explainable ai: Beware of inmates running the asylum or: How i learnt to stop worrying and love the social and behavioural sciences. *arXiv preprint arXiv:1712.00547*.

Murphy, G., Haider, M., Ghai, S., & Sreeharsha, B. (2013). The expanding role of mri in prostate cancer. *AJR Am J Roentgenol*, *201*(6), 1229–1238.

Nishikawa, R. M., & Bae, K. T. (2018). Importance of better human-computer interaction in the era of deep learning: mammography computer-aided diagnosis as a use case. *Journal of the American College of Radiology*, *15*(1), 49–52.

Oliveira, M., Zancul, E., & Fleury, A. L. (2021). Design thinking as an approach for innovation in healthcare: systematic review and research avenues. *BMJ Innovations*, *7*(2).

Omoumi, P., Ducarouge, A., Tournier, A., Harvey, H., Kahn, C. E., Louvet-de Verchère, F., ... Richiardi, J. (2021). To buy or not to buy—evaluating commercial ai solutions in radiology (the eclair guidelines). *European radiology*, *31*, 3786–3796.

PAIRADS. (2023). *Pairads - künstliche intelligenz für die prostatakarzinomerkennung.* Website. Retrieved from `https://pairads.ai/` (Accessed: July 31, 2023)

Panzone, J., Byler, T., Bratslavsky, G., & Goldberg, H. (2022). Transrectal ultrasound in prostate cancer: current utilization, integration with mpmri, hifu and other emerging applications. *Cancer Management and Research*, 1209–1228.

Pezaro, C., Woo, H. H., & Davis, I. D. (2014). Prostate cancer: measuring psa. *Internal medicine journal*, *44*(5), 433–440.

Privitera, M. B. (2015). *Contextual inquiry for medical device design.* Academic Press.

Productboard. (2023). *A guide to using the "double diamond" framework to score product home runs.* Website. Retrieved from https://www.productboard.com/blog/double-diamond-framework-product-managemer (Accessed: July 31, 2023)

Ratwani, R. M., Savage, E., Will, A., Fong, A., Karavite, D., Muthu, N., ... others (2018). Identifying electronic health record usability and safety challenges in pediatric settings. *Health affairs*, *37*(11), 1752–1759.

Rawla, P. (2019). Epidemiology of prostate cancer. *World journal of oncology*, *10*(2), 63.

Reiner, B. I., Knight, N., & Siegel, E. L. (2007). Radiology reporting, past, present, and future: the radiologist's perspective. *Journal of the American College of Radiology*, *4*(5), 313–319.

Reiner, B. I., & Krupinski, E. (2012). The insidious problem of fatigue in medical imaging practice. *Journal of digital imaging*, *25*, 3–6.

Rosala, M. (2023). *Using "how might we" questions to ideate on the right problems.* Website. Retrieved from https://www.nngroup.com/articles/how-might-we-questions/ (Accessed: July 31, 2023)

Rubin, D. L. (2019). Artificial intelligence in imaging: the radiologist's role. *Journal of the American College of Radiology*, *16*(9), 1309–1317.

Schlemmer, H.-P. (2018). Multiparametrische prostata-mrt: Vorzüge und grenzen. *Deutsches Aerzteblatt Online*.

Sorantin, E., Grasser, M. G., Hemmelmayr, A., Tschauner, S., Hrzic, F., Weiss, V., ... Holzinger, A. (2021). The augmented radiologist: artificial intelligence in the practice of radiology. *Pediatric Radiology*, 1–13.

Strohm, L., Hehakaya, C., Ranschaert, E. R., Boon, W. P., & Moors, E. H. (2020). Implementation of artificial intelligence (ai) applications in radiology: hindering and facilitating factors. *European radiology*, *30*, 5525–5532.

Suresh, H., Gomez, S. R., Nam, K. K., & Satyanarayan, A. (2021). Beyond expertise and roles: A framework to characterize the stakeholders of interpretable machine learning and their needs. In *Proceedings of the 2021 chi conference on human factors in computing systems* (pp. 1–16).

Ulwick, A. W. (2017). Outcome-driven innovation®(odi): Jobs-to-be-done theory in practice. *Strategyn, LLC Whitepaper*.

Verstijnen, I. M., van Leeuwen, C., Goldschmidt, G., Hamel, R., & Hennessey, J. (1998). Sketching and creative discovery. *Design studies*, *19*(4), 519–546.

Wolniak, R. (2017). The design thinking method and its stages. *Systemy Wspomagania w Inżynierii Produkcji*, *6*(6), 247–255.

Wulf, V., Müller, C., Pipek, V., Randall, D., Rohde, M., & Stevens, G. (2015). Practice-based computing: Empirically grounded conceptualizations derived from design case studies. *Designing socially embedded technologies in the real-world*, 111–150.

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, *9*, 611–629.

Zheng, B., Ganott, M. A., Britton, C. A., Hakim, C. M., Hardesty, L. A., Chang, T. S., ... Gur, D. (2001). Soft-copy mammographic readings with different computer-assisted detection cuing environments: preliminary findings. *Radiology*, *221*(3), 633–640.

Zheng, Q., Yang, L., Zeng, B., Li, J., Guo, K., Liang, Y., & Liao, G. (2021). Artificial intelligence performance in detecting tumor metastasis from medical radiology imaging: A systematic review and meta-analysis. *EClinicalMedicine*, *31*.

.1 Appendix A: Consent Form for Contextual Inquiry

Informationen zum Datenschutz im Forschungsprojekt PAIRADS

Wir bedanken uns für Ihre Bereitschaft, das Forschungsprojekt PAIRADS bei der empirischen Forschung zu unterstützen.

Bei den geplanten Datenerhebungen wenden wir üblicherweise die Methoden „Contextual Inquiry" (Mischung aus Beobachtungen und Ad-hoc Interviews) sowie Interviews an. Die Sessions werden per Audio aufgezeichnet, sofern das der/die Gesprächspartner:in gestattet und zusätzlich protokolliert. Außerdem werden Evaluationen auf Audio und/oder Video aufgezeichnet. Zwecks Datenanalyse werden die Daten transkribiert. In der Contextual Inquiry Session werden wir Ihnen bei Ihren Arbeitstätigkeiten folgen, wo immer Sie es uns erlauben. Wir werden so unauffällig wie möglich sein, damit Sie Ihre Arbeit so ausführen können, wie Sie es gewohnt sind, und wir werden die Beobachtung in Feldnotizen festhalten. Möglicherweise bitten wir Sie auch um die Erlaubnis, Fotos und/oder Videos von dem Arbeitsplatz und/oder das Umfeld des Arbeitsplatzes mit Bezug zum Projektinteresse (verwendete Hard- und Software, Gegenstände oder andere Hilfsmittel, um ein besseres Verständnis über den Nutzungskontext zu erhalten und so zur Veranschaulichung oder zur Verdeutlichung der Studien nützlich sein können.

Erklärung zur Datenermittlung:

Wir verarbeiten Ihre personenbezogenen Daten auf Basis von Art. 6 Abs. 1 lit. a DSGVO. Das bedeutet, dass wir Ihre Daten aufgrund einer von Ihnen gegebenen Einwilligung verarbeiten

1) Die personenbezogenen Daten der Teilnehmer:innen und ggf. der Unternehmen / Krankenhäuser / Kliniken / Praxen, wie Namen und Kontaktdaten, werden für Forschungszwecke gespeichert und ausschließlich zum Zwecke der Analyse oder zur Kontaktierung verwendet. Die Forscher:innen verpflichten sich, die Daten vertraulich zu behandeln und nicht an Dritte weiter zu reichen.
2) Personenbezogene Daten der Teilnehmer:innen werden bis zum Projektende gespeichert, sofern der/die Teilnehmer:in nicht ausdrücklich die Löschung wünscht. Die DFG sieht vor, dass die Primärdaten 10 Jahre gespeichert werden. Die Frist für die Speicherung beginnt mit dem Datum der Herstellung des öffentlichen Zugangs. Anonymisierte Daten werden über das Projektende hinaus gespeichert.
3) Die durch die empirische Arbeit erzeugten Angaben und die im Projektkontext erhobenen Daten über den Arbeitsplatz werden für wissenschaftliche Zwecke gespeichert, im Rahmen der Forschungsarbeit analysiert und für Veröffentlichungen verwendet. Die Ergebnisanalyse erfolgt selbstverständlich anonymisiert, sodass das Datenmaterial keine Rückschlüsse auf einzelne Personen zulässt. Für wissenschaftliche Veröffentlichungen gelten die gleichen Regelungen.

4) Diese Forschung birgt keine potenziellen Risiken für die Teilnehmer:innen. Dennoch sollten Sie bedenken, dass Sie keine kriminellen Handlungen offenbaren sollten, da die Forscher:innen sonst möglicherweise die Vertraulichkeit und Anonymität außer Kraft setzen muss.
5) Nach Abschluss der Studien haben Sie die Möglichkeit, die erhobenen Daten zu überprüfen und ggf. Änderungen vorzunehmen. Sie können es auch verweigern bestimmte Fragen zu beantworten oder dass wir Ihnen in bestimmten Situationen folgen und/oder Fotos machen. Außerdem können Sie sich bis einen Monat nach der Datenerhebung ohne Angabe von Gründen von der Studie zurückziehen. Die Daten werden sicher auf dem Server der Universität Siegen gespeichert und können nur intern eingesehen werden. Darüber hinaus ist darauf hinzuweisen, dass diese Forschung den europäischen Datenschutzbestimmungen (GDPR) entspricht.

Zusammenfassung der Betroffenenrechte

Die DGVO sieht folgende Rechte für Betroffene vor:

1. Recht auf Auskunft (Art. 15 DSGVO)
2. Recht auf Berichtigung (Art. 16 DSGVO)
3. Recht auf Löschung („Recht auf Vergessenwerden“) (Art.17 DSGVO)
4. Recht auf Einschränkung der Datenverarbeitung (Art. 18 DSGVO)
5. Recht auf Datenübertragbarkeit (Art. 20 DSGVO)
6. Widerspruch gegen die Verarbeitung (Art. 21 DSGVO)

Sofern eine rechtliche Verpflichtung besteht, der der Verantwortliche unterliegt, sind die Rechte vom Betroffenen nur eingeschränkt wahrnehmbar.

Ihnen steht jederzeit das Recht auf Beschwerde bei einer Aufsichtsbehörde für Datenschutz zu.

LDI NRW
Kavallerierstr. 2-4
40213 Düsseldorf

Vertraulichkeitsvereinbarung für das Forschungsprojekt PAIRADS

Bitte füllen Sie dieses Formular aus, nachdem Sie das Teilnehmer:innen-Informationsblatt und die Informationen zum Datenschutz gelesen haben.

Bitte kreuzen Sie die entsprechenden Felder an!	**Ja**	**Nein**
Ich bin 18 Jahre alt oder älter.		
Ich habe das Projektinformationsblatt gelesen und verstanden.		
Ich hatte die Möglichkeit Fragen zum Projekt zu stellen.		
Mir ist bewusst, dass meine Teilnahme freiwillig ist.		
Ich bin damit einverstanden, dass die Beobachtungen (bzw. Hospitation) sowie die Interviews per Audio oder Video aufgezeichnet wird (Bitte streichen Sie ggf. eine Modalität).		
Mir ist bekannt, dass ich die Beantwortung von Fragen verweigern kann und/oder Aufzeichnungen bestimmter Situationen verweigern kann.		
Ich kann innerhalb eines Monats nach der Datenerhebung von der Studie zurücktreten, indem ich den/die beteiligten Forscher in davon in Kenntnis setze ohne Gründe angeben zu müssen.		
Mir ist bekannt, dass die über mich gesammelten persönlichen Daten, die mich identifizieren können, wie z. B. mein Name, mein Arbeitsort oder meine Kontaktdaten, nicht über das Forschungsteam hinaus weitergegeben werden.		
Mir ist bekannt, dass die von mir angegebenen Daten sicher gespeichert und anonymisiert werden, sodass ich nicht identifiziert werden kann.		
Mir ist bekannt, dass die von mir gemachten Angaben in Veröffentlichungen verwendet werden.		
Ich möchte die transkribierten Daten überprüfen, um sicherzustellen, dass der/die Forscher:in meine Identität in zufriedenstellender Weise verborgen gehalten und alle unwesentlichen Informationen herausgenommen hat. Ich erkläre mich bereit, die transkribierten Daten ab dem Tag der Erhaltung innerhalb von 2 Wochen zu überprüfen.		
Ich erkläre mich damit einverstanden, auf das Urheberrecht und andere Rechte an geistigem Eigentum an dem Material, das ich zu dem Projekt beisteuere, zu verzichten.		

Kontaktmöglichkeiten

Name	Email-Adresse	Telefonnummer
Kieu Tran	kieu.tran(at)student.uni-siegen de	+49 (0) 173/ 901 - 9594
Sheree May Saßmannshausen	sheree.sassmannshausen(at)uni-siegen.de	+49 (0) 271/ 740 – 2395
Prof. Dr. Volkmar Pipek	volkmar.pipek(at)uni-siegen.de	+49 (0) 271/ 740 – 4068
Nazmun Nisat Ontika	nazmun ontika(at)uni-siegen.de	+49 (0) 271/ 740 – 2395

Erklärung der/des Forschungsteilnehmerin/Forschungsteilnehmers

Name der/des Teilnehmerin/Teilnehmers	Unterschrift	Datum

Ich habe der/dem potenziellen Teilnehmer in das Informationsblatt ausgehändigt und mich nach bestem Wissen und Gewissen vergewissert, dass die/der Teilnehmer in versteht, was sie/er freiwillig zustimmt.

Name der/des Projektmitarbeiterin/Projektmitarbeiters	Unterschrift	Datum

.2 Appendix B: Consent Form for Prototype Evluation

UNIVERSITÄT SIEGEN

Einwilligungserklärung

Bei der folgenden Studie handelt es sich um die Evaluation eines digitalen Prototyps. Die Teilnahme erfolgt über eine Videokommunikationsplattfom und beinhaltet die Durchführung von Aufgaben sowie ein Interview. Zur besseren Nachvollziehung der Interaktion mit dem Prototypen, wird der Bildschirm der teilnehmenden Person angezeigt.

Die Studie wird in Bild und Ton aufgezeichnet. Zum Zwecke der Datenanalyse werden die mündlich erhobenen Daten verschriftlicht (Transkription), wobei die Daten anonymisiert werden. Die Aufzeichnungen werden mit Abschluss des Projekts und spätestens nach 5 Jahren gelöscht.

Der Speicherung der personenbezogenen Daten zu Dokumentationszwecken kann durch die interviewte Person jederzeit widersprochen werden. Die Teilnahme an der Studie erfolgt freiwillig. Das Gespräch kann zu jedem Zeitpunkt abgebrochen werden. Das Einverständnis zur Aufzeichnung und Weiterverwendung der Daten kann jederzeit widerrufen werden.

Ich erkläre hiermit mein Einverständnis zur Nutzung der personenbezogenen Daten, die im Rahmen der Studie erhoben werden.

Name:	Datum:	Unterschrift
_____	_____	_____

.3 Appendix C: Interview Questionnaires for Contextual Inquiry

Interview Fragebogen - P03

- ☐ **Demoprafische Daten**
  - ☐ Wie alt sind Sie?
  - ☐ Wie lange arbeiten Sie bereits in der Radiologie?
  - ☐ Wie lange machen Sie bereits Prostata MRTs?
- ☐ **Befundung Allgemein**
  - ☐ Woher kommt der PSA Wert bei stationären Patienten?
  - ☐ Wie unterscheidet sich der Prozess bei hrem Chef?
  - ☐ Sind Grafiken/Tabellen über PI-RADS für Anfänger hilfreich (z B. Berechnung des PI-RADS Scores, Richtline für PSA-Dichte)
  - ☐ Gibt es Unterschiede zwischen Ihrem letzten Prozess und dem, als Sie angefangen haben? Welche?
  - ☐ Gibt es Tricks, die Sie für den Prozess gelernt haben?
  - ☐ Bekommt der Zuweiser die technischen Daten der MRT?
  - ☐ Bekommt der Zuweiser Zugriff auf die Bilder und den Bericht separat?
  - ☐ Wie werden Läsionen in der CZ oder AFS klassifiziert?
  - ☐ Gibt es Fälle, bei denen der PI-RADS Score nicht mit der eigenen Empfindung übereinstimmt?
  - ☐ Was am Prozess dauert am längsten bzw. ist am frustrierendsten?
  - ☐ Was macht am Prozess Freude?
- ☐ **Beurteilung Tabelle**
  - ☐ Wie wird die Index Läsion markiert?
  - ☐ Kann es zwei Läsionen geben, die beide die Indexläsion sein könnten?
  - ☐ Wird bei einer Läsion in der PZ auch die T2 Sequenzen bewertet und eingetragen?
  - ☐ Wird eine PI-RADS 2 Läsion eingetragen, wenn es eine PI-RADS 5 Läsion gibt?
  - ☐ Wie schreibt man eine Referenz zu einem Bild?
  - ☐ Was ist wichtig für einen guten Bericht?
- ☐ **KI-Software**
  - ☐ Was sind die Informationen, die besonders wichtig sind?
  - ☐ Wie könnte eine KI den Prozess unterstützen?
  - ☐ Würden Sie die Volumetrie-Berechnung nochmal überprüfen?
  - ☐ Wäre es hilfreich, die von der KI bearbeiteten Bilder im DICOM-Viewer sehen zu können?
  - ☐ Kann man auswählen, welche Bilder die Zuweiser*innen sehen können?
  - ☐ Was würden Sie sich für eine KI Anwendung noch wünschen?
  - ☐ Wann würden Sie dem Ergebnis einer KI trauen?

Interview Fragebogen - P01

- [ ] **Demoprafische Daten**
  - [ ] Wie alt sind Sie?
  - [ ] Wie lange arbeiten Sie bereits in der Radiologie?
  - [ ] Wie lange machen Sie bereits Prostata MRTs?
- [ ] **Gesamtprozess**
  - [ ] Wo im derzeitigen Prozess sehen Sie Probleme?
  - [ ] Wie lange dauern die einzelnen Etappen im Prozess?
  - [ ] Was bereitet Schwierigkeiten bei Grenzfällen?
- [ ] **Befundung Allgemein**
  - [ ] Welche MRT-Sequenzen werden gemacht?
  - [ ] Werden immer die gleichen MRT-Sequenzen gemacht?
  - [ ] Können Sie die Software-Anwendungen auflisten, die Sie nutzen?
  - [ ] Ist die Reihenfolge im Befundbericht wichtig?
  - [ ] Ist die Nutzung mit einer Hand angenehmer?
  - [ ] Beschreiben Sie die Befunde während der Gesamtbeurteilung aus dem Gedächtnis?
  - [ ] Gab es einen Grund für den Nutzen der Sprachassistentin statt des Taschenrechners?
  - [ ] Kann eingeschränkte Bildqualität einzelne Sequenzen betreffen?
  - [ ] Wie werden Läsionen in der CZ oder AFS klassifiziert?
  - [ ] Beschreiben Sie die technischen MRT Einstellungen im Bericht?
  - [ ] Dauert die Befundung von schwierigen Fällen deutlich länger?
- [ ] **Bearbeitung der Datei**
  - [ ] Wo werden die Annotationen im Bild gespeichert?
  - [ ] Warum ist die Erwähnung des Bildes (z B. T2) wichtig für den Bericht bei der Beschreibung eines Herds
  - [ ] Wie können Radiolog*innen die Referenzen zu den Bildern nutzen?
  - [ ] In welchem Format schicken Sie den Befund?
  - [ ] Wie kann man im DICOM Viewer zu den Markierungen springen?
  - [ ] Haben Sie schon mal Feedback zu der Struktur Ihrer Berichte bekommen?
- [ ] **Kartierung**
  - [ ] Muss nur die Region markiert werden oder ist Größe und genau Stelle wichtig?
  - [ ] Aus welchem Grund haben Sie die Farbe der Markierungen angepasst?
  - [ ] Wird jede Läsion eingetragen?
  - [ ] Wie kann der Zuweiser die Läsionen zuordnen?
- [ ] **KI Anwendung**
  - [ ] Wenn die Volumetrie automatisch berechnet werden würde, würden Sie das Ergebnis nochmal überprüfen?
  - [ ] Würden Sie KI Output in einem separaten Fenster oder direkt im Viewport präferieren?
  - [ ] Ab wann wäre der Aufwand den Output der KI zu überprüfen zu aufwändig
  - [ ] Könnten Sie sich vorstellen, das Ergebnis der KI zu übernehmen, ohne dieses vorher zu überprüfen?
  - [ ] Gibt es noch etwas, das Sie sich bei einer KI Anwendung wünschen würden?
  - [ ] Bei schwierigen Fällen, was müsste gegeben sein, damit Sie das Ergebnis berücksichtigen?
  - [ ] Gibt es einen Teil des Prozesses, den Sie komplett einer KI überlassen würden?
  - [ ] Helfen bereits eingesetzte KI-Applikationen bei der Qualität oder Geschwindigkeit?

Declaration of Academic Integrity

I declare that I have authored this thesis independently and have not used other sources and resources than the ones declared within. I have explicitly marked all material that has been quoted either literally or by content from the used sources.

Date: 16.08.2023 Signature: