Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

Anchana Visvalingam Balasingham

Challenges of Developing and Implementing AI Solutions in the Norwegian Healthcare Sector

By Using Breast Cancer Detection in Mammograms as an Illustrative Example

Master's thesis in Computer Science
Supervisor: Patrick Mikalef
July 2023

NTNU
Norwegian University of Science and Technology

Anchana Visvalingam Balasingham

Challenges of Developing and Implementing AI Solutions in the Norwegian Healthcare Sector

By Using Breast Cancer Detection in Mammograms as an Illustrative Example

Master's thesis in Computer Science
Supervisor: Patrick Mikalef
July 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

ABSTRACT

Integrating artificial intelligence (AI) in healthcare holds great promise for enhancing diagnostics, prognosis, and patient care. However, there are several challenges that hinder the development and effective use of AI in this sector. This master's thesis aims to comprehensively investigate and analyze these challenges, focusing on breast cancer detection as an illustrative example. The goal is to identify these challenges and propose solutions to facilitate the advancement and successful implementation of AI in healthcare settings.

In this study, a research methodology was utilized, incorporating multiple case studies and interviews with four distinct stakeholder groups. To gain deeper insights into the research topic, semi-structured expert interviews were conducted with a total of eight participants. The interviews generated a significant amount of data, which was transcribed and systematically coded for analysis.

The interview findings highlight several key themes that impact the development and implementation of AI in healthcare. These themes include data-related challenges, difficulties in complying with regulations and laws, and a lack of communication between different academic disciplines. These challenges act as barriers to the progress and adoption of AI solutions in the healthcare sector.

To address these challenges, previous studies and conducted interviews have put forth various propositions. These propositions suggest technical and non-technical solutions that can help mitigate the identified challenges. The key propositions involve transforming data in different ways and utilizing AI models to improve data quality. Additionally, introducing middlemen as part of AI projects to enhance communication is proposed.

By tackling and resolving these challenges, the potential of AI in healthcare can be maximized, enabling significant benefits for patients and healthcare providers. The thesis is written to help inform decision-makers and stakeholders in the healthcare industry about the necessary measures to

promote the development, implementation, and utilization of AI in a way that effectively supports diagnosis, prognosis, and patient care.

SAMMENDRAG

Integrering av kunstig intelligens (KI) i helsevesenet har et stort potensial for å forbedre diagnostisering, prognostisering og pasientbehandling. Likevel er det flere utfordringer som begrenser utviklingen og den effektive bruken av KI i denne sektoren. Målet med denne masteroppgaven er å undersøke og analysere disse utfordringene, med deteksjon av brystkreft i mammogrammer som et illustrativt eksempel. Hensikten er å identifisere disse utfordringene og foreslå løsninger som kan lette fremgangen og en vellykket implementering av KI i helsevesenet.

I denne studien ble det benyttet en forskningsmetode som inkluderte flere case-studier og intervjuer med fire ulike interessentgrupper. For å få en dypere forståelse av forskningstemaet ble det gjennomført semistrukturerte ekspertintervjuer med totalt åtte deltakere. Intervjuene genererte mengder data som deretter ble transkribert og systematisk kodet for analyse.

Intervjufunnene fremhever flere sentrale temaer som påvirker utviklingen og implementeringen av KI i helsevesenet. Disse temaene inkluderer utfordringer knyttet til data, vanskeligheter med å overholde regelverk og lover, samt manglende kommunikasjon mellom ulike fagområder. Disse utfordringene fungerer som barrierer for fremskritt og implementering av KI-løsninger i helsesektoren.

For å håndtere disse utfordringene har tidligere studier og gjennomførte intervjuer foreslått ulike tiltak. Disse tiltakene innebærer både tekniske og ikke-tekniske løsninger som kan bidra til å håndtere de identifiserte utfordringene. Sentrale forslag inkluderer transformasjon av data på ulike måter og bruk av KI-modeller for å forbedre datakvaliteten. I tillegg blir det foreslått å introdusere mellommenn som en del av KI-prosjekter for å forbedre kommunikasjonen.

Ved å håndtere og løse disse utfordringene kan potensialet for KI i helsevesenet maksimeres, og det kan oppnås betydelige fordeler for pasienter og helsepersonell. Målet med oppgaven er å informere beslutningstakere og

interessenter i helsevesenet om nødvendige tiltak for å fremme utviklingen, implementeringen og bruk av KI på en måte som effektivt støtter diagnostisering, prognostisering og pasientbehandling.

PREFACE

This master's thesis was undertaken at the Norwegian University of Science and Technology in Trondheim, Norway, as part of the *TDT4900 Computer Science, Master's Thesis* course. The research was carried out during the Spring semester of 2023 for Information Systems and Software Engineering. This study builds upon and extends previous work conducted in the course *TDT4501 Computer Science, Specialization Project.*

I want to thank Professor Patrick Mikalef, the thesis supervisor, for his invaluable guidance and support throughout my specialization project and master's thesis. His guidance has been enjoyable and has greatly contributed to the success of this thesis. I want to thank Hemin Qadir for helping me understand GAN and ViT, it has been of enormous help. Lastly, I want to thank everyone at study hall 303 at Gamle Fysikk, for their invaluable support, discussions and laughter. Thank you for making my time at NTNU wonderful.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

List of all abbreviations in alphabetic order:

- **AI** Artificial Intelligence
- **DL** Deep Learning
- **EBP** Evidence-Based Practice
- **ML** Machine Learning
- **NTNU** Norwegian University of Science and Technology

CHAPTER

ONE

INTRODUCTION

Artificial Intelligence (AI) has witnessed remarkable advancements in recent years, revolutionizing numerous industries with innovative applications. AI has become integral to people's daily lives, from transportation and appliances to mobile phones. The healthcare sector has been particularly proactive in embracing AI, with a significant 167.1% increase in the healthcare market during the COVID-19 pandemic [1]. While the potential applications of AI in healthcare are extensive, it is crucial to recognize the challenges faced by the industry.

One area within healthcare that holds potential for AI utilization is radiology, particularly mammography [2]. In Norway, the national screening program for mammography, BreastScreen Norway, aims to screen women aged 50-69 years biennially, with the primary objective of reducing breast cancer mortality rates [3][4]. The radiology sector generates large volumes of images and experiences a heavy workload, making it an ideal candidate for AI implementation [5]. However, despite these opportunities, the widespread adoption of AI in mammography has not been enforced at the national level [6].

The research is motivated by the need to comprehensively understand the implementation and impact of AI in the Norwegian healthcare system. Although AI holds promising prospects for specialized health services, its practical application often falls short of initial expectations regarding time and resource requirements [6][7]. Therefore, it is important to study the factors contributing to this disparity between potential and reality, focusing on the challenges specific to breast cancer detection as an illustrative example.

Furthermore, there is a growing societal concern regarding the implications of AI in healthcare [8]. Questions arise regarding its effect on healthcare disparities, doctor-patient relationships, patient safety, and the evolving roles of medical professionals. Gaining a comprehensive understanding of the impact of AI on these stakeholders and identifying the factors that hinder AI adoption is necessary for informed decision-making and formulating effective policies [6][8].

Hence, the primary aim of this thesis is to enhance our comprehension of the challenges that arise during the development and implementation of AI in the healthcare sector. The selection of mammography as the primary example is attributed to various factors that underscore its significance.

Firstly, numerous countries, including Norway, have established national screening programs for breast cancer, leading to the generation of vast amounts of mammographic data [3]. This quantity of data provides an extensive source for analysis and investigation. Moreover, mammography and AI analysis are highly intriguing and have garnered significant research attention. The existing knowledge offers valuable insights and establishes a solid foundation for further investigation [5].

The study delves into various critical aspects by analyzing the challenges encountered in this research field, including technical limitations, ethical considerations, and practical implications. This analysis aims to drive progress in the field and facilitate the development of strategies that promote the effective and responsible utilization of AI technologies in the Norwegian healthcare system.

To accomplish these objectives, the research questions that drive this research are the following:

RQ1: What are the challenges of developing AI solutions for the healthcare sector?

RQ2: What are the challenges of implementing AI solutions in the healthcare sector?

RQ3: How can technical and non-technical solutions mitigate the challenges of developing and implementing AI solutions in the healthcare sector?

To address the research questions, this study employed a systematic approach. Firstly, a comprehensive review of relevant literature was conducted, as outlined in chapter 2. This literature review served as the foundation for understanding the research topic and identifying existing gaps and areas requiring further investigation. Subsequently, a well-defined research methodology was developed and implemented, as presented in chapter 3. A semi-structured expert method was utilized to gather data from knowledgeable individuals in the field. The data collection involved recording interviews, transcribing them, and applying coding techniques to analyze the information obtained. The research findings and supporting evidence are presented in chapter 4, shedding light on the various propositions derived from the analysis.

Furthermore, the study critically discusses its methodology, results, and implications in chapter 5. This section provides an in-depth exploration of the study, including identifying key findings, analysis of limitations encountered during the research process, and suggestions for future research directions. Finally, a comprehensive conclusion is presented in chapter 7, summarizing the key findings and their implications for the research topic. This section aims to provide a concise overview of the study's contributions, highlighting its significance and potential impact on the field.

CHAPTER

TWO

THEORETICAL BACKGROUND

This chapter explains the theoretical background essential to the research topic. It covers the application of AI in healthcare, the fundamentals of breast cancer, AI's contributions to breast cancer research, and various AI models and algorithms for detecting breast cancer in medical images. It discusses AI's advantages, limitations, and potential challenges in this context. The chapter prepares readers to comprehend the practical applications and findings presented in the following chapters.

2.1 Overview of AI in Healthcare

This section thoroughly introduces AI in healthcare, including its definition and diagnostic applications. It also investigates the enabling and inhibiting factors that influence AI adoption. This section aims to promote an understanding of AI's potential benefits and the factors that influence its successful integration into the healthcare sector.

2.1.1 Definition of AI

In a scientific context, the definition of AI was coined by John McCarthy in 1955 as *"the science and engineering of making intelligent things"* [9]. However, the definition has evolved alongside the technology it describes. To better understand how AI can be utilized in the healthcare sector and with breast cancer detection, it is essential to distinguish AI from other information technologies. Sampled definitions from various scientific papers are presented in Table 2.1.1.

These differing definitions of AI reflect the various perspectives and applications across different fields. Each definition provides a unique insight into how AI can solve problems and improve outcomes in different areas of

Table 2.1.1: Sample definitions of AI from various publications.

Title of Publication	**Definition of AI**
A Clinicians Guide to Artificial Intelligence (AI): Why and how primary care should lead the health care AI revolution [10]	There are many definitions of AI, but the simplest is just a process where a computer is trained to do a task in a way that mimics human behavior.
Artificial intelligence and big data in public health [2]	Artificial intelligence is a generic term for a machine (or process) that responds to environmental stimuli (or new data) and then modifies its operation to maximize a performance index.
Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges [5]	The concept of AI first emerged in 1956, the aim being to build machines that can think and reason over complex tasks just like human beings and thereby sharing the same essential cognitive characteristics.
Artificial intelligence in digital breast pathology: techniques and applications [7]	AI is a broad research field which aims at designing computer systems that simulate human intelligence.
Artificial intelligence in healthcare: A critical analysis of the legal and ethical implications [9]	The term ‘artificial intelligence’ was originally coined by John McCarthy and others defined as the ‘science and engineering of making intelligent machines. No universally accepted definition exists, though.
Artificial intelligence transforms the future of healthcare [11]	Artificial intelligence (AI) increases learning capacity and provides decision support system at scales that are transforming the future of health care.
Causability and explainability of artificial intelligence in medicine [12]	Artificial intelligence (AI) is perhaps the oldest field of computer science and very broad, dealing with all aspects of mimicking cognitive functions for real-world problem solving and building systems that learn and think like people.

study. For instance, the definition from "Artificial intelligence in healthcare: A critical analysis of the legal and ethical implications" explains that the original meaning of the term *artificial intelligence* was to describe the science and engineering of creating intelligent machines [9]. In contrast, "Artificial intelligence in digital breast pathology: techniques and applications" defines AI as a broad research field aimed at designing computer systems that simulate human intelligence [7].

The lack of a universally accepted definition for AI is also because the field encompasses various techniques and approaches, making it a constantly evolving study area. As new technologies and applications emerge, the definitions of AI will continue to evolve. Therefore, understanding AI's different definitions and perspectives is crucial for appreciating its full potential and impact across different fields.

2.1.2 Applications of AI in Healthcare

AI is increasingly used in medicine, particularly in preventive screening and treatment. In preventive screening, mammography is commonly used to detect breast cancer before symptoms are felt [13]. However, mammography can have a false-positive rate, and AI is being used to create software that helps radiologists improve accuracy and reduce false positives and negatives. With machine learning (ML) and deep learning (DL) algorithms, AI can train on clean labelled mammograms to identify patterns and improve its ability to read photos accurately [14].

In treatment, AI is used to help decide the best course of action for a patient based on available health data and medical journals [8]. For example, IBM has developed ML algorithms that can outperform doctors in finding the correct treatment for patients using a precision cohort treatment option (PCTO). PCTO combines clinical treatment guidelines with electronic health records (EHR) data to evaluate treatment options for specific patients. With PCTO, practitioners can incorporate real-world evidence into their medical decision-making and improve treatment outcomes [15].

AI can also monitor patients' conditions, such as predicting the risk of intraoperative hypotension (IOH) during surgery. The Acumen Hypotension Prediction Index (HPI) software by Edwards Lifesciences uses AI to inform clinicians of the risk that a patient will experience IOH, improving patient safety. AI can improve medical practices and outcomes, especially when combined with human expertise and experience [16][17].

2.1.3 Technical Enablers and Inhibitors

Efficiently incorporating AI into the healthcare industry necessitates adopting and accepting this technology by decision-makers, medical experts, and patients. To delve into this topic further, the following subsection examines several key factors that either facilitate or hinder the implementation of AI in healthcare.

2.1.3.1 Access to Data and Industry Collaboration

High-quality data are required to build AI systems that can be applied in healthcare. However, it is sometimes argued that "the race for AI is a race for data" [18]. The growth of global health data due to new medical devices, IoT, and EHRs means access to high-quality data is critical. In 2016, IBM obtained access to the health records of all 61 million Italian citizens in exchange for establishing an IBM Watson Health research centre in Milan. The absence of competition could lead to a monopolistic healthcare industry, which reduces the value AI offers to consumers and healthcare professionals. Therefore, stakeholders, healthcare professionals, and governments must consider the influence that data access, privacy, and confidentiality have on AI in healthcare [19].

2.1.3.2 Reducing Statistical and Social Bias

AI algorithms can be statistically biased if low-quality or improperly formatted patient data is entered. Statistical bias produces results that differ from the true underlying estimate [8]. For example, data sampling can be biased when data is primarily from patients of a particular ethnicity. AI-assisted algorithms can evaluate risks incorrectly for patients whose electronic health records (EHR) have missing data. For example, AI models based on genetic test results are more likely to underestimate the chance of developing breast cancer in black patients than in white patients. However, AI can assist in overcoming bias in healthcare if used appropriately. AI-based decision support systems could detect bias in real-time and alert doctors about potentially biased decisions. Using unbiased data could be a solution, and future AI technology may utilize this data as the benchmark to address bias and enable AI to realize its full potential [18].

2.1.3.3 Maintaining Privacy and Confidentiality

The General Data Protection Regulation (GDPR) safeguards individuals' right to privacy and self-determination regarding personal data. It impacts how AI models are trained using patient data. Patients must be informed of the GDPR's requirements for managing their data, and their consent must be obtained before using their data to train models. If not all relevant parties agree to provide their personal information, the models will be trained with biased data, making them biased. Researchers must explain the *black box* in AI to patients to access their health data and provide AI-assisted treatment, an inhibitor to implementing AI in healthcare [19].

2.2 Basics of Breast Cancer

Breast cancer is one of the most common cancer types affecting women worldwide, representing one in every eight cancer diagnoses. In 2021 in Norway, there were 4023 new breast cancer cases and 601 deaths. Figure 2.2.1 showcases the distribution of women diagnosed with breast cancer in Norway from 1981 to 2021. Over time, there has been a continuous rate for the 30-39 age range, with a minor increase in recent years. The incidence rates for people aged 40-49 have steadily risen. As a result of the progressive implementation of breast cancer screening through BreastScreen Norway, explained in section subsection 2.2.3, beginning in 1996, as well as a rise in hormone replacement treatment during menopause, there was a considerable increase in incidence for the age group of 50 to 69 years. For women aged 70+ years, there has also been an increase in recent years, with few researchers knowing why [20][3].

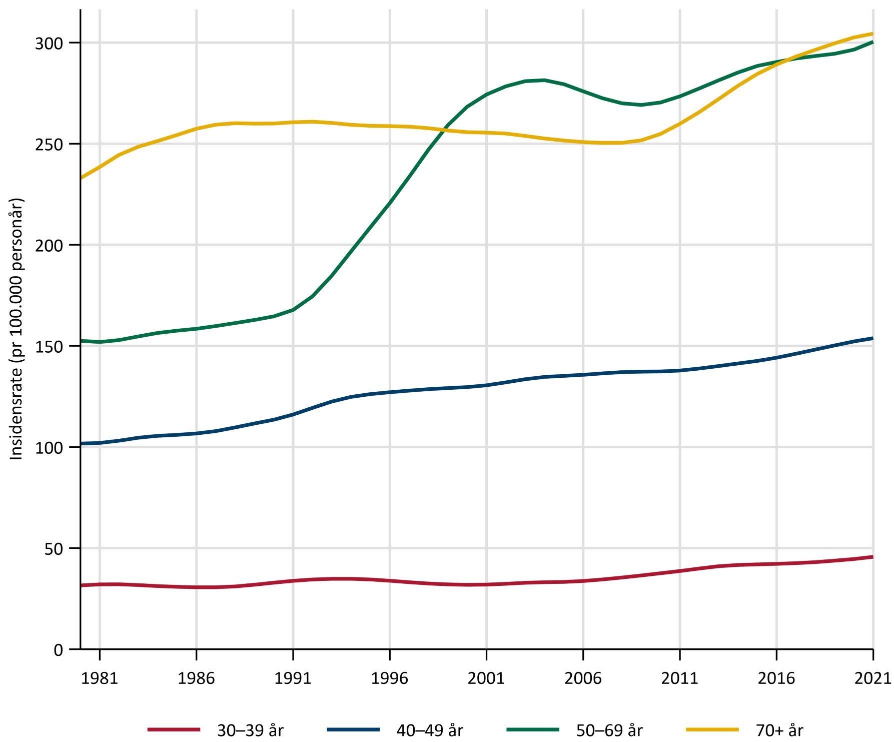

Figure 2.2.1: Age distribution of women in Norway diagnosed with breast cancer, from 2021. Created by The Norwegian Cancer registry [3].

Implementing a national breast cancer screening program in Norway has decreased breast cancer mortality rates, as visualized in Figure 2.2.2. According to data from the Cancer Registry of Norway, the age-standardized breast cancer mortality rate for women has decreased from 31.6 per 100000 in 1980 to 13.1 per $100,000$ in 2019 [3]. This decrease is primarily attributed to the early detection of breast cancer through mammography screening, which allows for identifying breast cancer at an earlier stage when it is more treatable [20].

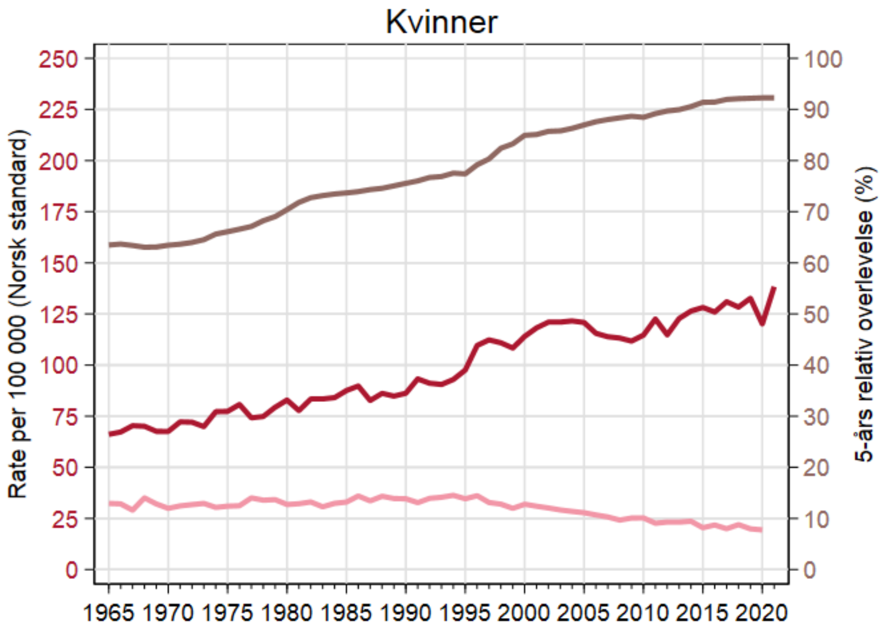

Figure 2.2.2: Trends in incidence (red), mortality (pink), and 5-year relative survival (brown) rates of breast cancer among women in Norway, from 2021 [3].

Detecting breast cancer early is essential for reducing the risk of death and enhancing the patient's quality of life. Early detection of breast cancer enables doctors to detect smaller tumours that have not spread to other body parts. This allows for less invasive treatments with fewer side effects and a shorter recovery time. Patients diagnosed early may also have a broader range of treatment options, resulting in a customized, less aggressive treatment plan. Early detection of breast cancer can also help reduce anxiety and stress for patients and their families, allowing for a faster start to treatment and a clearer understanding of their prognosis [20][3].

2.2.1 Definition of Breast Cancer

Breast cancer is characterized by the uncontrolled growth of cells in the breast tissue, resulting in lumps or masses. For women, there are three types of tissue in the breasts: glandular tissue, connective tissue, and fatty tissue fillers. The subtypes of breast cancer are differentiated by where in the breast the cancer originates. Common for all the subtypes is that the cancerous cells divide at a higher rate than healthy cells and can spread from the breast to other areas of the body through metastasis [20].

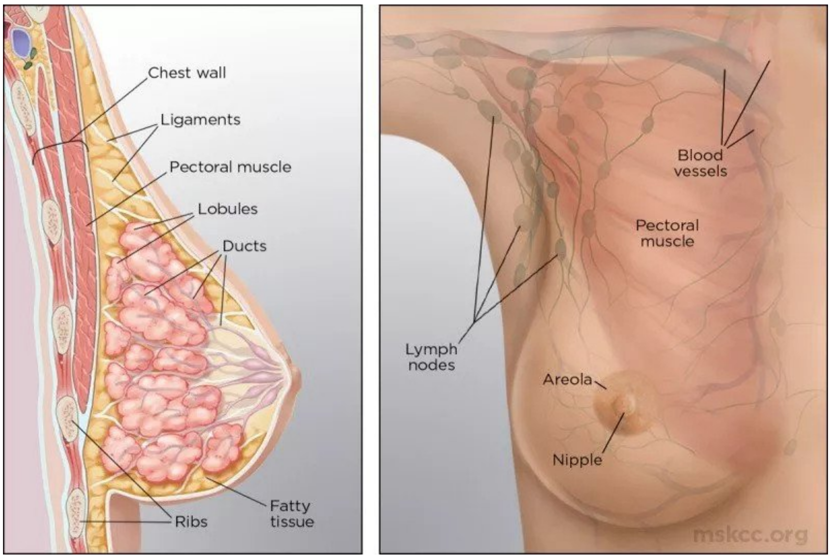

Figure 2.2.3: The female breast anatomy, illustrated by Memorial Sloan Kettering Cancer Center [21].

Figure 2.2.3 illustrates the breast anatomy from both the inside of the breast, the left side of the figure, and the outside of the breast, the right side. Breast anatomy is important to distinguish various subtypes of breast cancer from one another and to better understand the course of illness.

2.2.1.1 Subtypes of Breast cancer

Breast cancer has different subtypes distinguished by the type of cells involved, their appearance under a microscope, and the presence or absence of certain proteins or genes.

One important aspect of breast cancer is whether the cancer cells have receptors responsive to estrogen and progesterone. In hormone receptor-positive breast cancer, the cells use estrogen and progesterone to grow and divide. This means that hormonal therapies can block the effects of these hormones and stop the cancer cells from growing. Hormonal therapies work by either reducing the amount of estrogen in the body or blocking the receptors on the cancer cells that bind to estrogen and progesterone. This targeted approach is usually effective and has improved survival rates for patients with hormone receptor-positive breast cancer [3][20].

Obtaining information about cancer subtypes helps doctors determine the best treatment for each case of breast cancer and predict how the cancer may spread. Table 2.2.1 presents the most common breast cancer subtypes, what causes cancer, and where they typically originate in the breast.

Table 2.2.1: Common subtypes of breast cancer, based on findings from [20] and [3].

Subtype	**Description**
Ductal carcinoma in situ (DCIS)	A non-invasive type of breast cancer originating in the milk ducts, highlighted in Figure 2.2.3. Considered to be an early stage of breast cancer and can progress to invasive breast cancer if left untreated.
Invasive ductal carcinoma (IDC)	The most common type of breast cancer, accounting for about 80% of all cases. It begins in the milk ducts and invades the surrounding breast tissue. If left untreated, it can also spread to other body parts through lymph nodes.
Invasive lobular carcinoma (ILC)	A less common type of breast cancer, accounting for about 10% of all cases. It originates in the lobules, seen in Figure 2.2.3. Like IDC, ILC can spread to other body parts if left untreated.
Triple-negative breast cancer (TNBC)	An aggressive subtype of breast cancer that does not have estrogen, progesterone, or HER2 receptors. It is more aggressive and difficult to treat than other breast cancer subtypes. Researchers believe the BRCA1 gene is related to the cancer.
HER2-positive breast cancer	A subtype of breast cancer that overexpresses the HER2 protein. It is more aggressive than other subtypes of breast cancer.

2.2.2 Causes, Risk Factors, and Symptoms of Breast Cancer

To understand the significance of early detection of breast cancer and the role of AI in assisting with it, it is crucial to understand the diverse causes, risk factors, and symptoms associated with this disease, which will be elaborated in the following section.

2.2.2.1 Causes and Risk Factors of Breast Cancer

Table 2.2.2 presents the major risk factors that increase the chances of developing breast cancer. The table is derived from [22]. Age is a critical risk

Table 2.2.2: Risk factors for breast cancer, derived from [22].

Risk Factor	**Description of Risk Factor**
Gender	Women are at a higher risk of developing breast cancer than men.
Age	The risk of breast cancer increases with age.
Family history	Women with a family history of breast cancer are at a higher risk of developing the disease.
Inherited genetic mutations	Some inherited genetic mutations, such as BRCA1 and BRCA2, can increase the risk of breast cancer.
Personal history	Women with breast cancer in one breast are at an increased risk of developing cancer in the other breast.
Radiation exposure	Women who have had radiation therapy to the chest area before age 30 are at an increased risk of developing breast cancer.
Lifestyle factors	Factors such as alcohol consumption, being overweight, and lack of physical activity can increase the risk of breast cancer.
Hormonal factors	Hormonal factors, such as early onset of menstruation, late onset of menopause, and use of hormone replacement therapy (HRT), can increase the risk of breast cancer.

factor for breast cancer, as the likelihood of developing the disease increases as age increases. Women over 50 are at a higher risk of developing breast cancer than younger women, as seen in Figure 2.2.1. Therefore, routine screening is essential for early detection and treatment [3][22].

Further, a family history of breast cancer is a significant risk factor for developing the disease. Women with a first-degree relative with breast cancer are at an increased risk of developing the disease. Inherited genetic mutations, as BRCA1 and BRCA2, also increase the risk of breast cancer [20].

Personal history of breast cancer is another risk factor for developing the disease. Women with breast cancer in one breast are at an increased risk of developing cancer in the other breast. Furthermore, women who have had radiation therapy to the chest area before the age of 30 are at an increased risk of developing breast cancer [3][23].

Hormonal factors, such as early onset of menstruation, late onset of menopause, and hormone replacement therapy (HRT), can also increase the risk of breast cancer [23]. Women who start menstruating early or go through menopause late have a higher risk of developing the disease. Lifestyle factors such as alcohol consumption, being overweight or obese, and lack of physical activity have also been identified as risk factors for breast cancer [3]. Women who consume alcohol regularly have a higher risk of developing the disease. Therefore, adopting a healthy lifestyle can reduce the risk of developing breast cancer [20].

2.2.2.2 Symptoms of Breast Cancer

Breast cancer originates in breast tissue cells and primarily affects women worldwide [3]. Early detection and treatment of breast cancer can significantly increase the chances of survival, making it crucial to recognize and promptly address any potential symptoms [20].

The most typical sign of breast cancer is a lump or mass in the breast tissue, which may feel firm or hard and be painless or tender to the touch. While not all lumps are cancerous, it is important to seek medical evaluation for any new lump or changes in breast tissue [22]. Other possible symptoms of breast cancer include alterations in breast size or shape, skin dimpling or puckering, nipple changes such as inversion, redness or scaling of the breast or nipple skin, nipple discharge, or new or persistent breast or armpit pain.

Notably, some women with breast cancer may not present any symptoms, especially in the early stages of the disease. Hence, routine breast cancer screening, performed by BreastScreen Norway, is recommended for women over 50 and those at a higher risk of developing breast cancer. Early detection and treatment of breast cancer can significantly improve the chances of a positive outcome [4].

2.2.3 Preventive Screening: BreastScreen Norway

BreastScreen Norway, supervised by the Norwegian Cancer Registry, is a public health initiative to reduce breast cancer-related mortality. The program invites women aged 50 to 69 for a mammogram every two years. The women are invited back for further tests and examinations if anything cancerous or otherwise concerning is found in the mammograms. By detecting breast cancer at its earliest stages, BreastScreen Norway strives to minimize the number of women who succumb to this disease [13].

The national screening program plays a pivotal role in the early detection of breast cancer, which is important for various reasons. The American Cancer Society explained that the 5-year relative survival rate for localized breast cancer was 99%, compared to distant cancer, with a 30 % survival rate [24]. Therefore, preventive screening aids in discovering the disease early on, which gives the affected patients much better chances of survival.

For patients affected by the disease, the course of treatment drastically changes depending on when the cancer is discovered. If discovered early, the tumours are often smaller, which relates to better patient prognoses [24]. The treatment can be less invasive and diverse, giving patients multiple actions to consider when fighting the disease. The early detection of breast cancer can therefore be said to enrich the patient's quality of life.

2.2.4 Treatment Options

The Norwegian healthcare authorities have developed a national pathway for breast cancer to ensure that patients with suspicion of breast cancer receive a prompt and efficient investigation and treatment. The pathway includes the following steps, which are based on information from [4]:

1. **Referral:** The patient is referred by their general practitioner to a specialized healthcare service for investigation.

2. **Investigation:** The patient is examined and investigated by specialists, including radiologists, surgeons, oncologists, and others with specialized expertise in breast cancer. The investigation includes mammography, ultrasound, MRI, and possibly a biopsy to take breast tissue samples.

3. **Diagnosis:** Based on the investigation, the diagnosis and staging of cancer are determined.

4. **Treatment plan:** An individualized treatment plan is developed for the patient based on the diagnosis and staging. The treatment can include various measures to mitigate the disease, such as surgery, radiation therapy, chemotherapy, hormone therapy, or a combination.

5. **Treatment:** The patient receives the treatment recommended in the treatment plan, carefully monitored by healthcare personnel.

6. **Follow-up:** After the treatment, the patient is regularly followed up to monitor for possible relapse or side effects. The follow-up program varies depending on the patient's needs and treatment plan.

The breast cancer pathway is developed to ensure that patients receive a coordinated and effective treatment tailored to their needs. There is also a focus on ensuring good communication and information to the patient throughout the process so they can be confident that they are receiving the best possible treatment. Even though a pathway exists, much depends on how early the cancer is found and how invasive it is. If cancer is found after a routine screening, the patient will likely be called back for further examination and confirmation, for instance, with biopsies, after a few weeks [4].

2.3 Basics of AI

The subsequent section provides a comprehensive overview of AI, which includes concise explanations of machine learning and deep learning. Furthermore, a diverse range of AI models that hold significance for the remainder of the study is presented. The section aims to promote a basic understanding of AI in healthcare.

2.3.1 Overview of ML and DL

The impact of AI has been felt across many sectors, with the medical field being the most significant beneficiary of this technology [25]. From disease diagnosis to treatment optimization, AI has revolutionized every aspect of medical practice [11]. However, the development of novel AI technology is not merely dependent on technical advancements but also on ethical, social, and legal frameworks [8].

To fully comprehend the potential of AI in the medical field, one must have a coherent understanding of the terminologies used in various scientific papers. Hence, it is crucial to grasp the significance of AI, DL, and ML, as they form the foundation of AI technology. As depicted in Figure 2.3.1, understanding the context between AI, ML, and DL is essential for comprehending their practical implications.

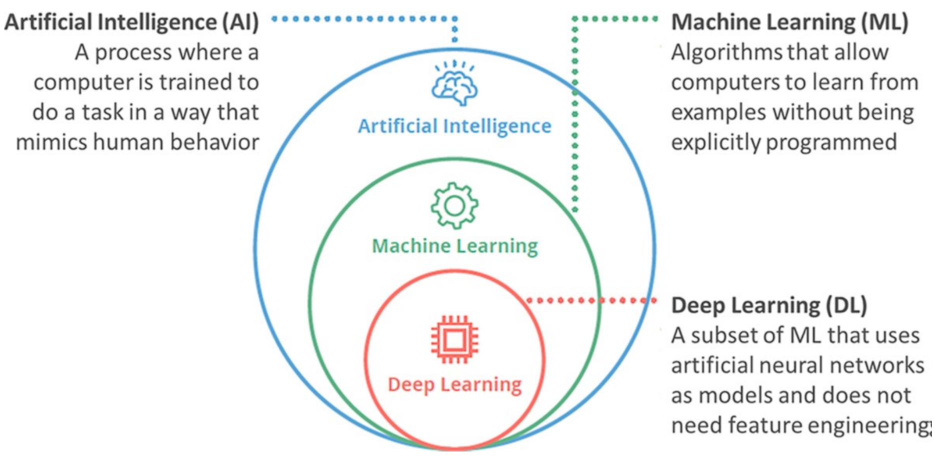

Figure 2.3.1: The context between AI, ML, and DL, showcased with definitions of each term [10].

2.3.1.1 Machine Learning (ML)

Machine learning, commonly abbreviated as ML, is an AI subfield that enables computer programs to make highly accurate predictions without being explicitly programmed. ML algorithms can accurately forecast new output values by analyzing historical and structured data as input. ML can be categorized into four distinct subcategories based on [26]:

Supervised learning is the first subcategory, where algorithms are provided with labelled training data and directed to seek correlations between specific variables. Describing the algorithm with its input and output makes it highly accurate in making predictions.

The second subcategory is *unsupervised learning*, where algorithms are trained on unlabeled data, and it is the algorithm's job to identify correlations between the given data set. Both the input data that algorithms use to train and the predictions or suggestions they produce are predefined, making it a highly effective way of discovering previously unknown patterns and associations.

Semi-supervised learning is the third subcategory, a mixture of supervised and unsupervised learning. Algorithms are fed labelled and unlabeled training data, but the algorithm is free to explore independently, allowing it to uncover previously unknown correlations.

The fourth and final subcategory is *reinforcement learning*, frequently used to train a system to complete a multi-step process with well-defined criteria. An algorithm is programmed to fulfil a goal and provides positive or negative feedback as it determines how to do so. However, the algorithm typically chooses the course of action independently, making it an efficient way to train machines to learn through experience.

2.3.1.2 Deep Learning (DL)

Deep learning (DL) is another AI subfield that utilizes complex and unstructured data to create predictions. DL algorithms can learn from large amounts of labelled and unlabeled data to perform complex tasks, such as image recognition, in detecting breast cancer. DL algorithms are often called *deep neural networks* because they utilize neural network topologies [26]. The term *deep* refers to the number of hidden layers within the neural network, differentiating them from traditional neural networks that typically have 2 or 3 hidden layers. In contrast, deep networks can have multiple layers, providing greater complexity and potential for learning [26][27].

An illustration of a layered neural network can be seen in Figure 2.3.2.

These DL models are trained using large sets of labelled data and neural network topologies that automatically extract features from the data [26]. This approach enables the models to learn and make accurate predictions based on patterns and relationships in the data.

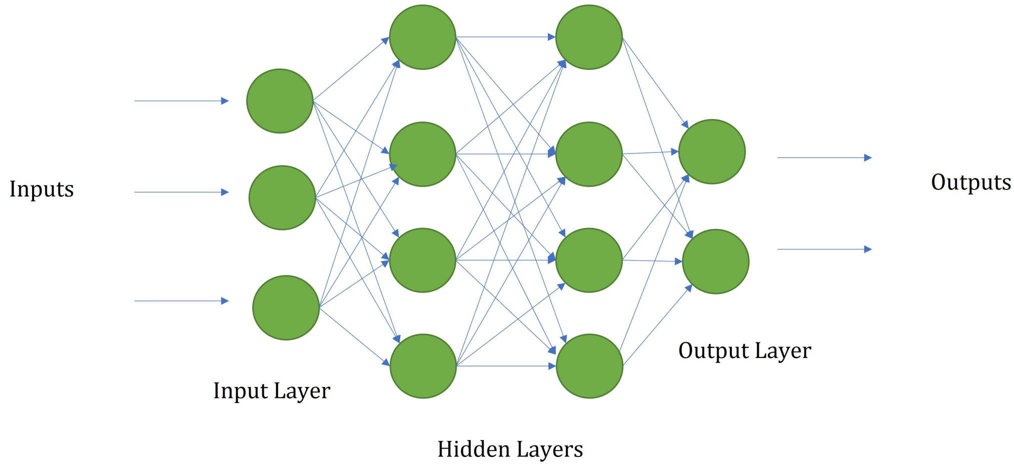

Figure 2.3.2: Layered neural networks are made up of a collection of interconnected nodes [26].

2.3.2 CNN, GAN and VIT

To comprehend the subsequent results and findings presented in chapter 4 and chapter 5, it is crucial to have a clear understanding of three distinct AI models. These models play a significant role in shaping the outcomes and insights discussed further in the study.

2.3.2.1 CNN

Convolutional Neural Networks (CNNs) have revolutionised healthcare by analysing medical images with advanced capabilities. CNNs use complex mathematical operations to extract pertinent features and make precise predictions.

CNNs consist of multiple layers, each designed to perform particular operations on the input data. The convolutional layer is the foundational layer of a CNN. It consists of filters (kernels) which are small matrices that perform convolution on the input image. The filter is convolved with the input image during convolution by element-wise multiplication and summation. This procedure emphasises distinct image patterns or characteristics, such as margins, textures, and shapes [28].

Following the convolutional layer are activation functions, typically Rectified Linear Units (ReLU), as seen in Figure 2.3.3, that introduce nonlinearities into the network. These activation functions facilitate the CNN model's learning of intricate relationships and enhance its expressiveness [28].

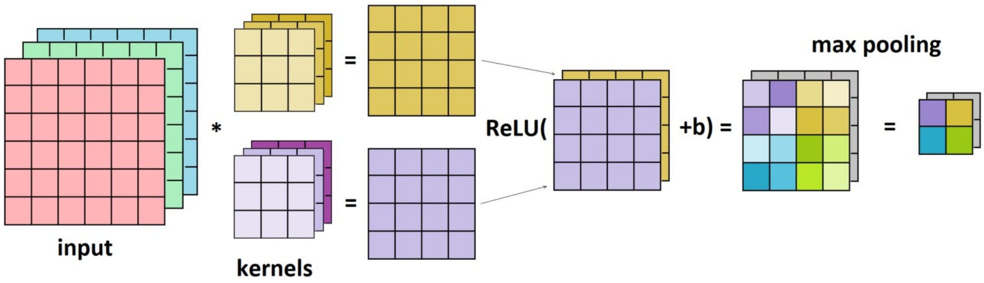

Figure 2.3.3: Figure illustrating an example of CNN architecture. Illustration created by [28].

After a series of convolutional and pooling layers, the extracted features are condensed and forwarded to fully connected layers. These layers connect every neuron in one layer to every neuron in the following layer, enabling the model to learn complex relationships between features. Fully connected layers enable the network to map extracted features to specific classes or predictions, such as disease or abnormality detection in medical images [28].

Once trained, CNNs can be utilised for medical image analysis. The input image is supplied into the network, and through convolutional layers, the model extracts pertinent features at various levels of abstraction. The fully interconnected layers then use these features to make predictions or classifications, providing insights into the patient's condition [28].

2.3.2.2 GAN

Generative adversarial networks, often known as GAN, are AI models composed of two main components: the generator and the discriminator. The generator network takes a random noise vector as input and transforms it into synthetic data samples. These samples are designed to resemble the real data from the training dataset. On the other hand, the discriminator network takes both real data samples from the training dataset and synthetic samples generated by the generator. Its task is to distinguish between real and fake data. The discriminator is trained to become more accurate in differentiating the two data types. In contrast, the generator is trained to improve its ability to generate synthetic samples that can fool the discriminator [29]. An example of a GAN system is presented in Figure 2.3.4.

The training process of GANs involves an adversarial competition between the generator and the discriminator. The generator aims to generate synthetic samples increasingly similar to real data, while the discriminator aims to better distinguish between real and fake data. Through iterations of this process, both networks continuously improve their performance. Ideally, the generator becomes proficient at generating synthetic data that is indistinguishable from real data, and the discriminator becomes skilled at accurately discerning between real and synthetic data [29].

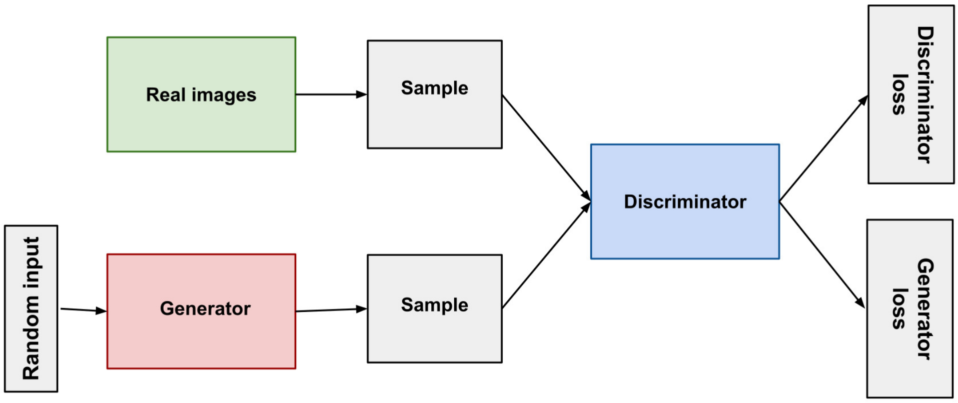

Figure 2.3.4: Figure illustrating an example of a GAN system. Illustration created by [29].

In the context of AI in healthcare, GANs offer several technical advantages. They can generate synthetic medical images that closely resemble real patient data. By doing so, GANs can overcome limitations such as a scarcity of real data or the need for additional data augmentation. The generated images can then be used for various medical image analysis tasks, including disease diagnosis, segmentation, and anomaly detection [29].

2.3.2.3 ViT

Vision Transformers (ViT) are deep learning models that apply the utility of transformer architecture to computer vision. They excel at image recognition and comprehension tasks and have sparked interest in various disciplines, including healthcare.

ViTs manipulate images by partitioning them into small patches, flattened and projecting to generate patch embeddings. These embeddings include both spatial and semantic data. Adding a positional embedding preserves spatial relationships. Ultimately, these patch and positional embeddings are input into a transformer encoder. This encoder utilises self-attention layers and feed-forward networks to capture the image's global context and derive intricate features. Using self-attention mechanisms, ViTs excel at capturing long-range dependencies, enabling them to analyse the entire image and comprehend intricate relationships between elements [30]. Figure 2.3.5 illustrates the architecture of a ViT system.

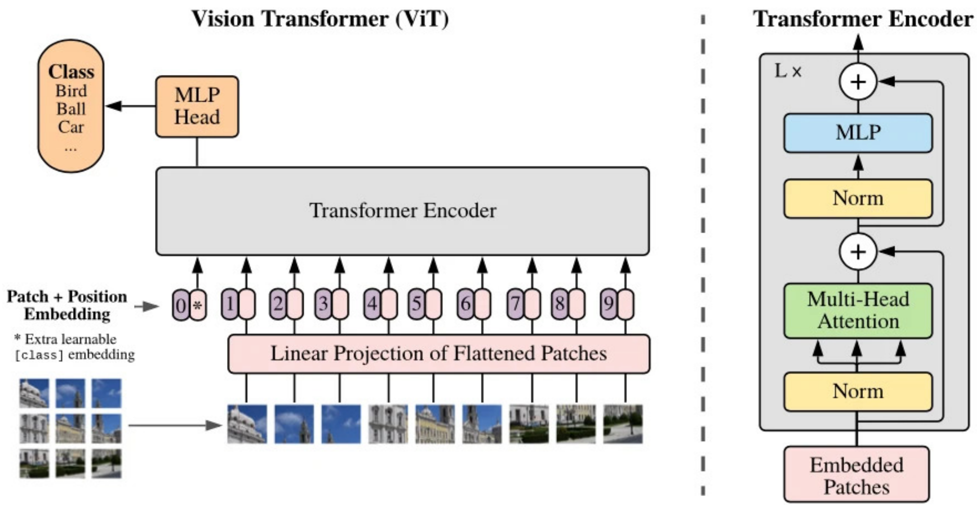

Figure 2.3.5: Figure illustrating an example of a ViT architecture. Illustration created by [30].

In the context of healthcare, ViTs offer numerous advantages. They can achieve cutting-edge performance in medical image analysis tasks such as disease detection, segmentation, and classification. By accurately documenting the global context and long-range dependencies within medical images, ViTs enable precise and exhaustive analysis, thereby assisting healthcare professionals in making informed diagnoses and treatment decisions [30].

2.4 AI in Breast Cancer Research

The research field of AI in healthcare, particularly in breast cancer detection, requires examining previous studies, highlighting AI's advantages, and addressing challenges. Prior research shows AI's potential to improve accuracy, speed up analysis, and enable early detection. However, challenges include data quality, algorithm interpretability, ethical considerations, regulatory compliance, and system integration. By understanding previous work, benefits, and challenges, a deeper comprehension of AI in healthcare, specifically breast cancer detection, can be achieved.

2.4.1 Previous Studies

The use of AI in breast cancer research is a growing area that has seen a significant increase in publications in recent years. A search on PubMed with the search string "artificial intelligence breast cancer" revealed that there are currently over 4,000 publications related to AI and breast cancer. The historical development is visualized in Figure 2.4.1

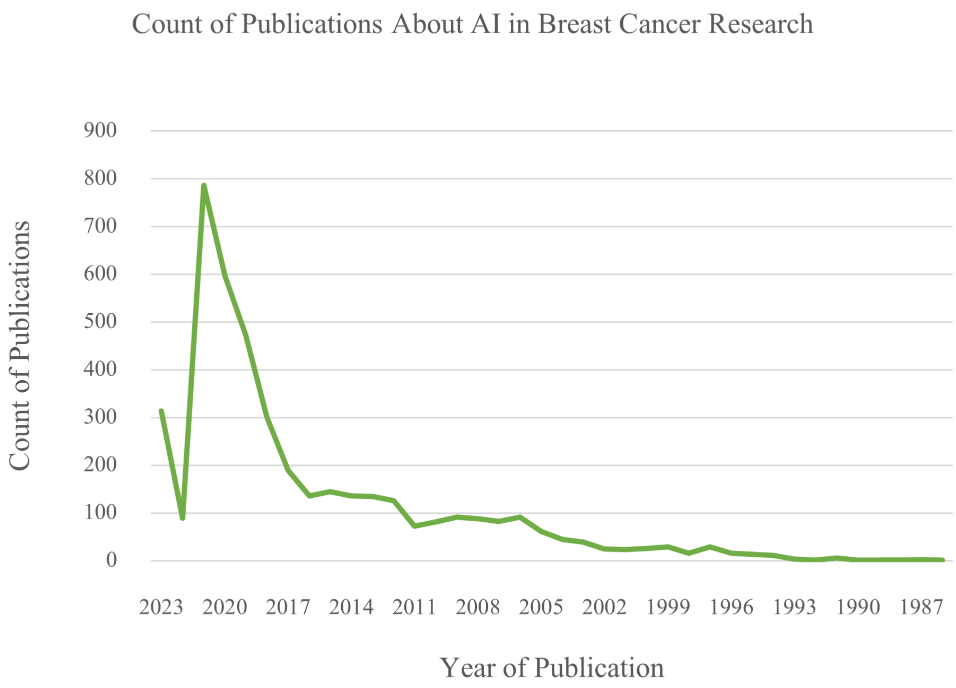

Figure 2.4.1: Count of publications about AI in breast cancer research from PubMed (08.05.2023).

Regular mammography, as revealed by Lehman et al. in their 2017 study "National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium," has a limitation of missing approximately one in eight cases of breast cancer [31]. This underscores the need for improved detection methods in breast cancer screening. Implementing AI shows promise in enhancing the efficiency of mammogram evaluation and screening processes. By harnessing AI technology, the aim is to enhance the identification of all instances of breast cancer, leading to improved outcomes in diagnosis and treatment [31][32].

In 2020, McKinney et al. conducted a study titled "International evaluation of an AI system for breast cancer screening," which brought together researchers from Google Health and Imperial College London [33]. The research team developed and trained a computer model using X-ray images from nearly 29,000 women. The AI algorithm outperformed six radiologists in interpreting mammograms. Moreover, the AI system demonstrated comparable accuracy to two collaborating doctors. The analytical capacity of AI sets it apart from humans, making it a promising technology for enhancing breast cancer detection capabilities [33][34].

Highlighting the potential of AI in breast cancer detection, the study titled "Artificial Intelligence Evaluation of 122,969 Mammography Examinations from a Population-based Screening Program" showcased impressive results [35]. The AI system assigned a score of 10 to 86.8% (653 out of 752) of screen-detected cancers and 44.9% (92 out of 205) of interval cancers when applying threshold 1. These findings highlighted the meaningful role of AI in the future of breast cancer screening [35].

While the studies conducted by Lehman et al., Larsen et al., and McKinney et al. shed light on the potential benefits of AI in breast cancer detection, it is crucial to acknowledge the challenges associated with implementing AI in the healthcare sector, particularly in the context of breast cancer detection.

One challenge is the need for large and diverse datasets to train AI models effectively. The availability of comprehensive and diverse datasets is essential to ensure that AI algorithms can accurately learn and generalize patterns related to breast cancer, according to Norori et al. in the research article "Addressing bias in big data and AI for health care: A call for open science" from 2021 [36]. However, acquiring such datasets can be complex due to privacy concerns, limited data sharing, and variability in data collection practices across different healthcare institutions [36].

Another challenge lies in the interpretability and transparency of AI algorithms. While AI models can demonstrate impressive performance, understanding these models' underlying reasoning and decision-making process can be challenging [36][37]. AI's "black-box" nature can hinder the trust and acceptance of these systems among healthcare professionals and patients. Ensuring interpretability and transparency in AI algorithms is vital to gain insights into how decisions are made and facilitating effective collaboration between AI and healthcare professionals.

Furthermore, integrating AI into existing clinical workflows poses practical challenges [37]. Adapting AI technologies within clinical settings requires careful consideration of system compatibility, workflow integration, and user experience. Effective implementation strategies and seamless integration into existing healthcare systems are crucial to ensure the successful adoption and utilization of AI tools for breast cancer detection [36].

2.4.2 Benefits of Using AI

AI in screening provides a significant advantage by enabling radiologists to concentrate on more challenging cases while accelerating the analysis of less complicated situations. ML and DL-assisted image classification can offer radiologists scores indicating the likelihood of cancer in an image, facilitating the identification of images that require closer examination. This approach can potentially reduce false positive and negative rates, resulting in less stress and unnecessary treatment for patients [14][38].

2.4.3 Challenges of Implementing AI

AI presents several challenges when applied to breast cancer screening, primarily due to the high rate of false positives resulting in increased costs, psychological distress for affected patients, and the need to ensure equitable access to diagnosis and treatment for diverse populations, as highlighted in [38] and [14]. To address these challenges, software developers must demonstrate their models' effectiveness across different populations and risk factors, requiring collaboration with hospitals and organizations with diverse imaging data. Additionally, the transition from 2D to 3D data, as depicted in Figure 2.4.2, poses challenges regarding computing resources and storage capabilities. Security issues related to the use of clinical data in AI models also require attention, and future research may face challenges in accessing the large quantities of mammograms and patient records needed to develop new AI technologies, as pointed out in [14].

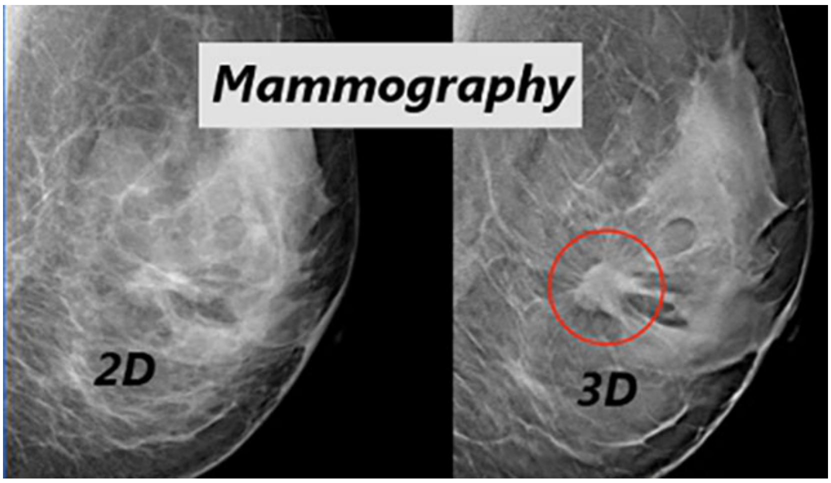

Figure 2.4.2: Comparison of 2D and 3D mammography [39].

CHAPTER

THREE

RESEARCH METHOD

This chapter explores the methods and tools used to conduct qualitative research on the research topic. The philosophical perspective and research approach adopted in the study are discussed, along with an explanation of the data collection method employed and its relevance to the relevant stakeholder groups. Moreover, detailed profiles of all interview participants are provided, alongside a description of the research setting and the methodology utilized for data analysis.

3.1 Philosophical Perspective: Interpretive Research

When exploring topics concerning Information Systems (IS), philosophical perspectives have a crucial impact on how researchers approach their chosen subject. These perspectives are based on diverse assumptions about knowledge, reality, and the connection between the researcher and the studied topic. Thus, understanding these perspectives is important for conducting insightful research [40]. According to Myers (2002), three main underlying philosophical assumptions are employed in qualitative research; *positivist research*, *interpretive research*, and *critical research*, presented in Figure 3.1.1 [40].

The positivist theory generally tries to test a theory, while critical theory attempts to disrupt the status quo by addressing oppositions and conflicts [41]. Interpretive research was deemed the most appropriate philosophical perspective for the research covered in this study. Interpretive research is detailed in the book *Researching Information Systems and Computing* (2006) by Oates [41] and is defined as,

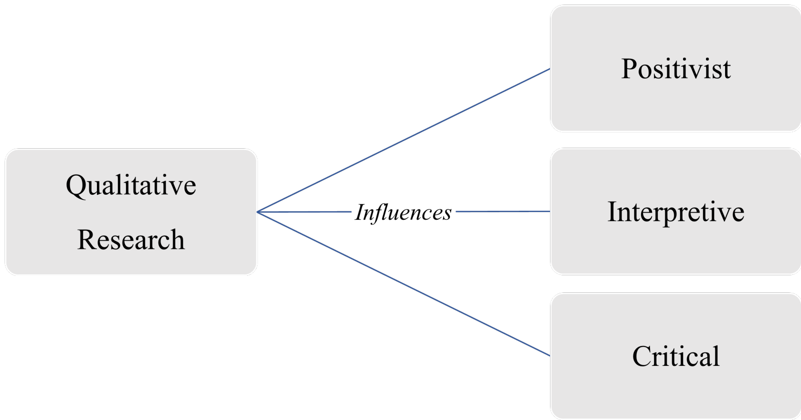

Figure 3.1.1: Philosophical perspectives employed in qualitative research, based on findings from [40].

> Interpretive research in IS and computing is concerned with understanding the social context of an information system: the social processes by which it is developed and construed by people and through which it influences, and is influenced by, it's social setting.
>
> - Oates (2006), [41, p. 292]

Interpretive research focuses on examining and explaining the elements of a setting, like an organization, that are connected and interdependent. As a result, it is possible to claim that interpretive research seeks to comprehend how human agents perceive the world and how this understanding alters and diverges across time, which fits the research topic [41].

An essential aspect of the interpretive perspective is that there is no truth. People can perceive the truth differently based on how they interpret the world. In the case of AI in breast cancer detection, doctors and developers may have different versions of the truth based on their knowledge and priorities, but both truths can be "correct." Another aspect to consider is the researcher's reflexivity, meaning that researchers are human beings and, thus, not neutral. The researcher's beliefs, values, priorities, and experiences can shape the research process, and interpretive research focuses on the researcher's reflexivity to better understand how their biases can influence the research and interpretation of the research topic [40][41].

3.2 Research Method: Multiple Case Studies

The multiple case studies method is employed in this study and is often used in interpretive research, presented in section 3.1. A case study usually focuses on one instance of the topic to be further researched: an organization, a department, an information system, a developer, and so on [41]. The chosen instance is studied deeply, and various data generation methods are used to understand all aspects. In this case, four instances, elicited in 3.4, with the main data generation method as semi-structured expert interviews, expounded in section 3.3.

Case study research is well-suited to IS research, mainly because the objective of the field is studying IS in organizations rather than only technical issues. As interpretive research acknowledges, the world is entangled and complex, meaning that IS research relies on technical and other aspects to be meaningful. The study conducted in this project used multiple sources to understand the theme of AI in breast cancer detection. Therefore, the study's approach falls under multiple case studies. The difference between a single case study and multiple case studies is that multiple cases are addressed in the latter, contrary to only one instance being addressed in the first. Further, the multiple case studies method was utilized as various groups of stakeholders were addressed in the data generation phase of the research, presented in section 3.4. According to Oates, multiple sources can aid in gaining multiple perspectives about the research topics and gaining access to both qualitative and quantitative data [40][41].

3.3 Data Collection Method: Semi-structured Expert Interviews

Semi-structured expert interviews were the data generation method used in the study and represent a dynamic and notably effective approach to data generation. Semi-structured refers to interview guidelines incorporating open-ended questions so that interviewees can provide extensive and candid responses. The flexible format ensures that conversations remain focused on intended topics while allowing for exploring uncharted areas of interest. Expert interviews involve individuals who are experts in their respective fields, which provides a specific focus and depth to the conversation [41].

The semi-structured approach empowers interviewees to share candid thoughts, emotions, and experiences, which can often prove challenging to capture through other data generation methods. By creating a relaxed and open environment, expert interviews encourage a free-flowing exchange of ideas and insights, fostering a deeper understanding of complex issues and facilitating valuable discoveries [41][40].

When deriving conclusions about an entire population or field, there is a drawback to using semi-structured expert interviews as a data generation method. While this disadvantage was considered carefully when selecting the appropriate research method, it did not present a significant obstacle in the context of the topic being studied. Moreover, the researcher's role and identity can also present a potential disadvantage, as they can significantly influence how interviewees perceive the questions and discussions. Age, gender, ethnicity, and professional role may impact the interviewer's perception. Although these factors cannot be changed, it is crucial that the interviewer is mindful of such biases and strives to maintain a professional demeanour, remaining polite, attentive, and genuinely interested in the interviewee's responses [41][40].

3.3.1 Planning the Interviews

The interviews were thoughtfully planned to leverage the numerous advantages of a semi-structured interview format while also attempting to minimize the potential for biased data. Following the advice in [41], an interview guideline was crafted. The final guidelines for researchers, developers, and data owners are presented in Table A.1 and medical professionals in Table A.2. The guidelines ensured that conversations flowed seamlessly throughout the interview process. The guidelines included carefully crafted questions that revolved around relevant themes and follow-up questions, as advised by Oates (2006) [41]. Furthermore, to prevent biased data and prepare all interviewees thoroughly, informative emails outlining the research project, the importance of the interviewee's expertise, and the careful handling of all generated data were exchanged [40].

Thorough preparation is crucial when conducting expert interviews. This involves conducting background research on the interviewee, their workplace, publications, and relevant research and considering other factors pertinent to the topic and interview. By putting in this effort, the interviewer establishes themselves as a professional, fostering a sense of trust and encouraging the interviewee to share their thoughts and experiences openly. Moreover, this preparation can help identify specific topics and aspects relevant to the chosen topic that can be referenced throughout the interview, facilitating a more engaging and productive conversation for all participants [40].

To enhance preparation for conducting the interviews, a valuable tool was presenting the questions to the project supervisor and other students for feedback, which was recommended in [41]. The interview questions were refined by sharing the questions with the supervisor. Additionally, presenting the questions to other students provided an opportunity to gain insights from peers who could offer diverse perspectives and constructive criticism [40].

Regulations enforced by Norwegian Centre for Research Data (NSD) state that interviewees must be granted the right to withdraw from the interview at any given time. This information was communicated to each participant at the interview's outset. It can also be referenced in the interview guidelines, as seen in Table A.1 and Table A.2. If an interviewee wished to have the video or automatic transcription removed, they were instructed to send an email to either the researcher or the project supervisor, a process that was thoroughly explained beforehand.

3.4 Stakeholders

The selection of interviewees was a critical consideration for the study, as it was the main data generation method for the multiple case studies. The appropriate selection was important when ensuring the insights collected were inclusive, in-depth, and diverse. Further, the research field is interdisciplinary and diverse. Hence the stakeholders were chosen to reflect that [41]. In light of this, a stakeholder analysis was conducted to identify and prioritize stakeholders likely to be impacted by the research, presented in Table 3.4.1.

The stakeholder analysis involved a multi-step process. Initially, identifying relevant stakeholders was performed by considering their potential influence on or impact on the research outcomes. Secondly, the stakeholders were mapped and prioritized based on their importance to the study. Finally, stakeholders were engaged through interviews, allowing them to share their thoughts, opinions, and experiences related to the research topic, which is elaborated in section 3.3.

The project's boundaries were established by considering only stakeholders with relevant technical or medical expertise. However, it is crucial to note that to create favourable solutions for all parties involved, other factors, such as social, economic, legal, and ethical considerations, must also be considered [9]. Patients affected by the research are also a significant stakeholder group, but their involvement in the project was not pursued to maintain focus on the technical aspect. Nevertheless, it is imperative to prioritize the patient's viewpoint in assessing solutions, as their well-being should be an objective of the project [41].

3.4.1 Final Pick of Stakeholders

The stakeholders in this project have been thoughtfully chosen, prioritizing the following order: researchers, developers, doctors, and data owners. Table 3.4.1 provides valuable insights into the significance of each stakeholder group and their respective impact on the research study. It's important to acknowledge that since this study focuses on the intersection of computer science and healthcare, the priority given to the stakeholder groups aligns with that perspective.

Table 3.4.1: Table showcasing the relevant stakeholder groups, interests, impact on the project, and priority.

Stakeholder Group	**Interests**	**Project Impact**	**Priority**
Researchers	Advancing scientific knowledge, developing new AI algorithms and techniques, and publishing research findings.	Drive innovation, contribute to developing AI solutions, evaluate and validate AI models, and provide insights into real-world applicability.	1
Developers	Building and maintaining AI systems, implementing algorithms and models, ensuring technical feasibility, and optimizing performance.	Translate research into practical applications, design and develop AI-based healthcare solutions, and ensure smooth functioning and usability of AI systems.	2
Doctors	Enhancing patient care, improving diagnostic accuracy, streamlining workflows, ensuring patient safety.	Utilize AI tools to aid in diagnosis and treatment decisions, integrating AI systems into clinical workflows, and providing feedback on the usefulness and effectiveness of AI applications.	3
Data Owners	Protecting patient privacy and confidentiality, maintaining data security, complying with regulations, and ensuring ethical use of data.	Provide access to relevant datasets for training and testing AI models, ensuring data quality and integrity, establishing data-sharing agreements, and addressing privacy and security concerns related to AI projects.	4

3.5 Presentation of Participants

In the project's research phase, a multiple case study was conducted, and the following subsections present all the interviewees. The study included eight interviewees who were interviewed in seven sessions, with each participant belonging to one of four stakeholder groups. The first case study consisted of the researchers, the second consisted of the developers, the third was made up of doctors, and the fourth and final case study involved data owners.

The participants for the multiple case study were carefully selected based on their relevant involvement and expertise in the respective stakeholder group under investigation. The researchers were chosen for their expertise in researching the specific topic, while the developers were selected for their experience in creating software or applications related to the research topic. The doctor was chosen based on their experience in treating patients associated with the research topic. The data owners were selected for their proficiency in managing and owning data relevant to the research. The study gathered diverse perspectives and insights from various stakeholder groups by selecting these interviewees, facilitating a comprehensive understanding of the research topic [41].

3.5.1 Researcher A

Researcher A is an accomplished Associate Professor at NTNU, specializing in medical imaging. They are involved in projects that aim to develop and apply ultrasound technology for patient monitoring and interventional procedures. Researcher A's research interests have expanded to include imaging modalities such as MRI and CT.

The researcher has been working on one of the longest-running projects at a heart clinic. This project aims to monitor the condition of the heart in patients undergoing surgery. Specifically, the project focuses on measuring the movement of the wall between the atrium and the ventricle on the left side of the heart. By monitoring this movement, the researchers hope to detect signs of deterioration earlier than traditional measures, such as heart pressure, with the aid of AI. This is important because the heart can compensate for some time before it crashes, and catching these signs early could allow corrective action to be taken during the surgery. This would be preferable to re-open the patient later to address any issues after the initial surgery.

Researcher A's involvement in the study is attributed to their knowledge of the healthcare industry, particularly in imaging and AI. They were specifically chosen for their extensive experience and in-depth understanding of the research process, from ideation to project development. As an expert in their respective field, they provided valuable insights and perspectives that helped shape the direction of the study. Their reflections, thoughts, and opinions were instrumental in making key decisions throughout the study.

3.5.2 Researcher B

Researcher B is a highly skilled expert in stochastic process control, statistical modelling, neural networks with reinforcement learning, and machine learning. With a PhD in neural networks, they have extensively researched practical medical statistics and modelling.

In a current project, Researcher B and colleagues are training ML models to identify cancer in mammograms using a large database of several million mammograms taken with X-rays. Most images in the database show no signs of cancer, with only 0.7% of them showing any signs. The ML models are trained using CNN, explained in subsubsection 2.3.2.1, and the final diagnosis of each image is used to provide feedback to improve the models' accuracy. The technical aspect of the project involves designing and implementing these neural networks to identify cancer in mammograms accurately. The goal of the project is to improve the accuracy of mammogram cancer detection, resulting in earlier and more accurate diagnoses, which could potentially save lives.

Researcher B was chosen to participate in the study for several reasons. Firstly, they possess impressive and wide-ranging experience working with AI, particularly in healthcare and other sectors. Secondly, the project that Researcher B is involved in is relevant to the research topic being studied in this project. Their proficiency in the technical aspects of the research field, familiarity with other research organizations and practices, and willingness to share their expertise have produced interesting results and solutions for this project.

3.5.3 Researcher C

Researcher C is a renowned cancer imaging researcher focusing on developing biomarkers for cancer diagnosis, patient stratification, and treatment monitoring. With an educational background in chemistry and a PhD, Researcher C began their research career working on MR imaging projects in Trondheim, Norway, where they gained extensive experience working in an interdisciplinary environment. Since the early 2000s, Researcher C has been a valuable member of their current research group, starting as a postdoctoral researcher and eventually becoming the group leader in 2012.

The current project Researcher C is working on began the initial work with MR imaging of prostate cancer, as in the early 2000s, there were no good imaging modalities available to diagnose prostate cancer patients. Biopsies were taken systematically across the prostate, which could miss cancerous areas or target non-aggressive cancerous areas. MR imaging has changed the diagnostic process for prostate cancer, as it provides better soft tissue contrast and complementary information about cell density. This enables them to determine the location of the cancerous area more accurately, allowing for more precise biopsy targeting. The project focuses on fusing real-time ultrasound with MR imaging to target biopsies more accurately for the detection of aggressive prostate cancer, which would heighten the patient's quality of life, as well as save hospitals millions in expenses for certain procedures.

Researcher C was invited to join the study to discuss their ongoing project, including the path from concept to implementation, challenges encountered, and the potential of AI in healthcare. Their candid and informative responses have proven invaluable in steering the study in the right direction.

3.5.4 Developer A

Developer A has a medical background with a focus on mathematics and biology. They began their medical training at the University of Zagreb in Croatia and later completed a clinical internship to acquire a medical license. Subsequently, Developer A obtained a Ph.D. in mathematical biology from the Max Planck Institute in Germany.

Developer A works as a scientist and leads the AI department in a private health technology company. They were asked to partake in the study due to their varied and extensive background, covering both medical and technical aspects. They are currently obtaining medical data to use in their company, which gives Developer A a unique outlook on the research subject. Additionally, they have recent and hands-on experience dealing with the complex rules and regulations related to AI in healthcare, which adds new insight to the study.

3.5.5 Developer B

Developer B works at a Norwegian startup that utilizes AI to monitor sleep. They joined the company early on and have participated in the shift from rule-based movement detection to AI-based detection. The interviewee possesses extensive knowledge about the technology employed by the company and the procedures involved in product development.

The study requested the involvement of Developer B to contribute a unique and innovative perspective, particularly given their company's recent acquisition of data and ongoing efforts to develop AI-based devices that can be utilized in clinical settings. With this relevant and up-to-date experience in the field of health technology, Developer B's insight is highly valuable and has the potential to enhance the study significantly.

3.5.6 Doctor A

Doctor A is a radiologist working as a section chief at a Breast Diagnostic Center. With over two decades of experience, they have expertise in clinical breast diagnostics and mammography screening and collaborating with other clinicians and developers in cross-disciplinary projects.

As a section chief, Doctor A has been integral to BreastScreen Norway for several years. With their extensive experience in preventative and clinical screening, diagnosis, and treatment options, Doctor A brings a wealth of knowledge that is crucial to include in the study. Their expertise is of the utmost importance in ensuring that the study accurately reflects the medical practice surrounding breast cancer.

In addition to their medical expertise, Doctor A's insights and ideas regarding interdisciplinary collaboration are invaluable. They can provide unique perspectives on how different academic disciplines can work together to tackle the challenges posed by breast cancer screening, diagnosis, and treatment. Furthermore, Doctor A's thoughts on the future of AI in the field are highly relevant to the study, given the significant impact this technology is poised to have on breast cancer management.

3.5.7 Data Owners A and B

Data Owners A and B are working at a government registry in Norway and have been actively engaged in several breast cancer-related projects, which is why they are involved in this study.

Data Owner A holds a Master's degree in political science and serves as a research administration advisor. Their involvement in AI-related projects provides a valuable perspective on the use of technology in healthcare. They can offer insights into AI's ethical, legal, and social implications in breast cancer detection and management. On the other hand, Data Owner B is a statistician with a strong academic background in physics and mathematics. They play a role in the registry's projects, analyzing data to draw insights and inform decision-making. Their ability to derive meaningful insights from complex datasets brings new insight to the study.

Data Owners A and B are working with a research group led by Researcher B and eight medical centres in Norway to develop an algorithm for detecting breast cancer using AI. Their combined research administration, statistics, and AI expertise are essential to the project's success. Further, as employees of a government registry, Data Owners A and B have extensive knowledge and experience in data management. This knowledge is invaluable in ensuring that the algorithm developed is accurate and reliable. Their understanding of how to combine multiple academic disciplines is also essential to the study's success, as it allows for a more comprehensive approach to algorithm development.

3.6 Research Setting

The research setting refers to the specific environment where the research occurs. The interviews were conducted in person or through video conferencing tools. When the participants were located in Trondheim, it was preferred to conduct in-person interviews, as it allowed for a more natural flow of conversation. However, Microsoft Teams or Zoom was utilized for participants elsewhere to conduct the interviews. These video conferencing tools added the benefit of automatic interview transcription, facilitating the transcription process. In cases where the interviews were conducted in person, the audio was recorded using a telephone or computer and later transcribed [41].

The interview appointments were carefully scheduled to align with the interviewees' calendars to ensure that the interviews did not disrupt their schedules. Most interviews were held in late February or early March. The interviewees were informed that the interviews would last approximately 45 minutes to 1 hour, allowing them to prepare accordingly. This approach helped to minimize any potential negative impact on their schedules and the interview itself [41]. Table 3.6.1 presents an overview of when, how, and for how long each interview was conducted.

3.7 Data Analysis: Hermeneutics

Hermeneutics, the practice of interpreting language, offers a valuable approach for analyzing textual data, particularly in interview analysis. This method allows researchers to delve deeper into the layers of meaning conveyed by interviewees, moving beyond surface-level content and uncovering implicit assumptions, underlying themes, and subjective perspectives. By employing hermeneutics, researchers can gain a more comprehensive understanding of the interview data, considering the broader context, historical background, and cultural influences that shape the narratives shared by participants [40].

The significance of hermeneutics in interview analysis is highlighted by

Table 3.6.1: Table showcasing the interviewee, interview setting, duration of the interview, recording tool, and interview month.

Interviewee	**Interview Setting**	**Duration**	**Recording Tool**	**Month**
Researcher A	In-person	60 min	Microsoft Teams	February
Researcher B	Remote	60 min	Microsoft Teams	March
Researcher C	Remote	60 min	Microsoft Teams	March
Developer A	Remote	75 min	Zoom	February
Developer B	In-person	60 min	Mobile Phone	March
Doctor A	Remote	50 min	Microsoft Teams 4	March
Data Owners A and B	Remote	60 min	Microsoft Teams	March

definition provided by Canadian philosopher Charles Taylor [40],

> Interpretation, in the sense relevant to hermeneutics, is an attempt to make clear, to make sense of an object of study. This object must, therefore, be a text, or a text analogue, which in some way is confused, incomplete, cloudy, seemingly contradictory – in one way or another, unclear. The interpretation aims to bring to light an underlying coherence or sense.

- Taylor (1976) [40]

According to Taylor, interpretation, as relevant to hermeneutics, seeks to bring clarity and make sense of an object of study, which in this case, is a text or a text analogue. These textual artefacts may be clouded, incomplete, confused, or seemingly contradictory, requiring interpretation to reveal an underlying coherence or sense [40].

By utilizing hermeneutics in interview analysis, researchers move beyond identifying surface-level patterns and themes. They strive to unravel the intricate layers of meaning in the interview data, exploring participants' subjective perspectives and experiences. This approach allows researchers to consider the broader context that shapes these narratives, including historical, cultural, and social factors. Furthermore, hermeneutics enables researchers to engage in a reflective and interpretive dialogue with the interview data, promoting a deeper understanding of the meanings expressed by participants [41][40].

CHAPTER

FOUR

RESULTS

The following section presents the results from the data generation phase of the study. As explained in the research methodology, chapter 3, all the conducted interviews were transcribed and analyzed to gain further insight from the interviews. The generated data were coded in Excel, classifying themes and elaborating them with quotes from the interviewees. Due to compliance with GDPR and NSD regulations, the coding is not included in the appendix. Table 4.0.1 presents all the relevant themes discussed in the interviews, with discussion points further elaborated in each subsection. The themes, with related discussion points, are based on the thoughts, feelings, and observations from the interviews.

Table 4.0.1: Table presenting all the relevant themes from the data collection conducted, highlighted with relevant bullet points to be further elaborated.

Themes	**Discussion Points**
Data-related challenges	Lack of image data, Adapting to changes in the healthcare sector, Varying data quality, Data bias, Data management systems
Complying with regulations and laws	Data access, Understanding Regulations
Trust and collaboration between academic disciplines and health regions	Lack of scientific evidence, Collaboration between health regions, Communication between academic disciplines
Hierarchical structure of healthcare	Education, Healthcare Systems
Ethical concerns	Concerns, Priorities
Future of AI in healthcare	Personalized treatment of breast cancer, Precise AI models, Ethical and systematic data collection, AI replacing clinicians

4.1 Data-Related Challenges

The following section addresses several key data-related challenges that the interviewees identified. Its purpose is to shed light on these challenges, explain their impact on the interviewees' work, and propose potential solutions to overcome them.

4.1.1 Lack of Image Data

The availability of image data, a vital resource for training, validating, and testing AI algorithms and models, frequently posed challenges during participant interviews. The following subsection will explore various factors contributing to image data scarcity. This section emphasizes the participants' concern regarding the impact of the lack of data on their work.

Researcher A emphasizes that healthcare is among the most tightly regulated sectors worldwide. They highlight the extensive regulations, guidelines, and standards enforced by governments and regulatory bodies to ensure patient safety and prevent malpractice, negligence, and fraud. Violating these regulations can result in significant legal and financial repercussions, explained Researcher A. As a result, numerous considerations must be considered before testing, implementing, and utilizing AI in clinical settings. Researcher A cautions, stating, *"The consequences of AI can be devastating."*

Moreover, the stringent laws protecting healthcare data make it challenging for researchers and private companies to acquire such data.

Researcher B shared an insightful perspective during the discussion, emphasizing the importance of data in developing AI-powered algorithms. They stated, *"With ML of this type, data is gold,"* highlighting the significance of high-quality data for tasks such as cancer detection.

Furthermore, Researcher B discussed their ongoing research project on leveraging AI to detect breast cancer in mammograms. The project aims to combine established ML models with novel models developed by the project task force, enabling the differentiation between images with and without cancer. To achieve accurate results, the project relies on substantial data for training and validating the model.

Researcher B's team has collaborated with the national screening program BreastScreen Norway to obtain the necessary data. This partnership has facilitated access to cancerous images from the program. The researcher acknowledged the importance of such access in advancing their project. When discussing the number of cancerous images from the BreastScreen program, Researcher B explained,

> Only around 7 per mille have cancer in them [mammograms]. Luckily, of course, most of them [patients] are healthy.

- Researcher B

The lack of image data is even greater for diseases not regularly screened. The volume of image data is based on the number of patients with the disease and if they are screened. Therefore, most diseases have scarce image data, affecting *image diversity*. Image diversity refers to variation in data, which ensures that subtypes of the disease, genetic variation, and other factors are represented. *"You cannot have a good AI model if it is biased towards one subtype [of cancer],"* explained Researcher C. When asked about the capability of the models to distinguish between cancer and no cancer based on the various types of breast cancer, discussed in section 2.2.1, Researcher B explained that the project task force does not know yet about potential biases of the model towards one or more cancer types. The researcher elaborated with the following statement,

> Having a couple of million healthy images does not matter. We could have been fine with 1/10 of the images, which would not have altered our results. However, the critical number is those related to rare cases, with varying breast tissue, cancer type, etc.

- Researcher B

Developer A agrees with Researcher B and explains that their company is trying to obtain relevant necessary data with little success thus far. As an example, Developer A said, *"I don't have 2000 patients with Alzheimer's,"* when explaining that their company lacks access to data to train their models. Further, as a private company, there are more regulatory concerns to consider, as they will use the obtained data to create some profit. However, Developer A still wanted to highlight the fact that their company wanted to help patients, explaining,

> We want to sell you the test, but ultimately it's about helping people and making an earlier diagnosis, which we all strive for.

- Developer A

To go around the many regulatory challenges when obtaining relevant data, Developer B explained that the company *"Created our own data."* The company partnered with two institutions, the University of Bergen and a clinic in Oslo, to collect data on sleep patterns. Participants wore equipment that measured their sleep, including a movement sensor and a fully equipped sleep cap. Developer B explained that the data was used to train the ML model to interpret the movements and analyze the participant's sleep. By doing so, Developer B's company could quickly begin their development, thus avoiding the bottleneck of data access. However, it should be noted that the company is creating wellness products and other regulations and clinical requirements for products used in clinical settings.

4.1.2 Adapting To Changes in the Healthcare Sector

Throughout the interviews, adapting to various changes was problematized regarding technical and medical changes and regulations to comply with. The following subsection presents the case studies experts' insights about the topic in light of technical and medical advances.

Researchers A and B highlighted a medical advancement in early breast cancer detection, possibly changing the standard of mammograms. The most recent changes are expected from 2D to 3D images, with studies conducted by The Norwegian Cancer Registry in Bergen to assess the effect of tomosynthesis. From a medical perspective, Researcher B reveals that they are unsure of the effect of using 3D imaging,

> There is a lot of talk about tomosynthesis. In the research environment, it is said that it [tomosynthesis] finds more cancer but that it finds the least dangerous cancers.

- Researcher B

From a technical perspective, the change from 2D to 3D imaging will affect how AI models are developed. Researcher B explained,

> No, you need to retrain with separate models for that purpose. Currently, the models we are creating are using 2D X-ray images. However, the methods we are studying definitely have relevance for possible adaptions, but the model we are creating will not be used if we switch to tomosynthesis.

- Researcher B

Therefore, decisions and changes related to better medical performance also impact technical advancements in medical AI. If tomosynthesis becomes the new standard in Norwegian hospitals, Researcher B explains the current image database with millions of mammograms of healthy and unhealthy breasts becomes irrelevant. This is because the images are taken with the wrong equipment. Researcher B speculates that some parts of current models can be reused. Still, since the models need to be retrained, validated again, and retested with new image data, the process becomes costly and time-consuming. Generating new images will also be a bottleneck since models need to be trained on various data, which takes time to generate and collect, as most screenings are done on healthy women with no cancer. Researcher B explained,

> The incoming images have different characteristics, so we have prepared for the possibility that the results [of the AI model] may be poorer. However, we expect that if we receive large amounts of images from other providers, we can also train models that generalize well across manufacturers. But that's [the challenge] precisely part of it. That is the research element of this project. It's not just a development project that uses well-known technology to solve a well-known problem.

- Researcher B

Another medical advancement that may impact AI models is the possible use of *contrast mammography*, which Doctor A explained,

> It means that a contrast will be placed in the arm and that the mammogram is done with a technique where a regular and subtraction mammogram is produced. The subtraction mammogram showcases contrasts in the breast, which is more like MR images, making it easier for radiologists to detect cancer. It can be a very useful method.

- Doctor A

If contrast mammography becomes a new standard for mammograms, the AI models face the same obstacles as if the standard were to shift to tomosynthesis. A large, diverse, and relevant image database must be created for any medical changes, which is costly and time-consuming, explained Researcher C. Researcher A argues that AI models are vulnerable to change and that it will affect the availability of data needed to pursue novel AI technologies.

When AI models are used in the medical field, they have to go through several legal approvals before they can be implemented, which is further explained in section 4.2. Researcher A elaborated that the technology used in these models, devices, and other medical products has to remain the same as when it was initially approved. If the technology changes or improves, the models must go through the approval process again before being used in a medical setting. Researcher A highlights this point and emphasizes the importance of complying with legal regulations when using AI models in the medical field,

> A new round of approvals must be done for any technical improvements to the code, even if it is just done to make it [AI model] more effective. The process is to ensure patient safety, but it is still time-consuming. All documentation must be sent in again.

- Researcher A

4.1.3 Varying Data Quality

Another theme highlighted by several interviewees was the technical challenge of having varying data quality. The following section explains how varying data quality affects AI models, how researchers, developers, and doctors manage the challenge, and wishes for the future.

Several aspects must be considered when discussing good data quality, and regular data should differ from image data. Table 4.1.1 is derived from various interviews and research papers and explains the various attributes that support good data quality, Table 4.1.2 showcases important attributes necessary for good image quality.

The challenge of varying data quality affects most interviewees in various ways, especially considering that *"data is gold,"* a concept explained by Researcher B in subsection 4.1.1. Even though the researcher explained that

Table 4.1.1: Table presenting relevant attributes that describe data quality. Based on findings from [42].

Attribute	Description
Accuracy	Correctness of the data.
Completeness	Extent to which all the required data is present
Consistency	Absence of conflicts or contradictions in the data.
Timeliness	Freshness of the data.
Relevance	Usefulness of the data for the intended purpose.
Validity	The degree to which the data accurately represents the real.

Table 4.1.2: Table presenting relevant attributes that describe good image quality. Based on findings from [43].

Attribute	Description
Resolution	Level of detail in an image.
Contrast	Level of difference between an image's lightest and darkest areas.
Colour accuracy	How well the colours in an image match the original scene.
Sharpness	Clarity of the image.
Noise	Random variations in brightness and colour in an image.
Artefacts	Any visual anomalies in the image, such as distortion or blurring.

they have access to millions of mammograms, it is the image quality that matters,

> If we succeed in creating something that works better than what other groups can achieve, it may be because we have come up with some clever tricks in the algorithms. But the main point is simply the quality of the data.

- Researcher B

One of the challenges discussed is the random variations, explained in Table 4.1.2, of the images used for diagnosis due to the differences in the equipment used and the scanner's location. Researcher B explained,

> Mammography images are not just mammography images. They vary. The equipment suppliers, how they handle it technically, the type of technical settings they use, the procedures they have during image capture, and so on can vary greatly from center to center and country to country. Therefore, having a very large dataset from Norway could mean that our models may generalize better to the Norwegian population.

- Researcher B

Therefore, one can say that the random variations affect the good data quality researchers seek. *"All radiologists follow international guidelines,"* explained Doctor A when discussing the variation in images from various health regions in Norway. However, most interviewed researchers explained that they also experience large image variations. Researcher B speculated the following,

> Do we really believe it's mostly about the patient group? Maybe it has more to do with how the equipment is used or the technical image quality.

- Researcher B

During the discussion, Data Owners A and B highlighted that different health regions and hospitals within the country often utilize distinct equipment for data generation. The Data Owners explained that the variations in equipment usage are determined by legislation specific to each region, *"It is legislation from each region that decides these things"*.

Moreover, Data Owners A and B emphasized that the data obtained from BreastScreen Norway is subject to several biases that impact its quality. By acknowledging the heterogeneity in equipment usage and some biases within the BreastScreen Norway data, Data Owners A and B highlighted the complexities of working with healthcare data. They underscored the importance of understanding and mitigating such variations and biases when utilizing the data for analysis and research. To exemplify how the data quality is affected by the participants of the screening program, Data Owner B said the following,

> A relatively low proportion of women from Somalia attended and used the screening service offered. If you base a detection algorithm on that, it can have unfortunate results.

- Data Owner B

Researcher C expounded that their team has used open data, which can be found online, and local data, from hospitals in Norway to create the models they are currently using. The team received data from an Asian country but explained that it has been difficult to use because the images were obtained differently than how they were obtained in Norway. In the received images, MRI was not used as part of a package procedure to determine whether a patient should undergo a biopsy but rather as an image guide for surgical or radiation therapy. Therefore, Researcher C explained that the images showed a more widespread cancer than those from Norwegian patients earlier in the disease.

Further, the image quality varied regarding resolution, noise, and sharpness, as seen in Table 4.1.2, making using the images obtained difficult. Researcher C explained the challenge through the following quote,

> The variation in the quality and characteristics of the images makes it difficult to use them to develop reliable models.
>
> \- Researcher B

4.1.4 Data Bias in Healthcare

Bias in healthcare was a topic that emerged while conducting the interviews. All stakeholder groups expressed their knowledge and concern about the field. The subsection presents various perspectives about data bias and why it is important when creating robust, fair, and secure AI models.

Parikh et al.'s scientific article "Addressing Bias in Artificial Intelligence in Health Care" defines *social bias* in healthcare as the systematic delivery of inequitable care that leads to suboptimal outcomes for specific groups [18]. During discussions on this topic, Data Owners A and B expressed concerns about mitigating social bias, stating that it is a complex challenge. However, they proposed a potential solution for mitigating data bias in breast cancer detection algorithms by implementing a "safety net" model. This safety net consists of multiple interconnected AI models and algorithms, acting as a protective measure to minimize the impact of biased algorithms on the outcomes.

When Researcher B was asked about bias, they explained that the concept is genuine but not something they have been able to work with yet effectively. *"In the process we are in now, bias towards the patients is not something we have yet worked in, but statistical bias is something we have researched".* Parikh et al. describe *statistical bias* as an algorithm that produces a result that differs from the true underlying estimate.

Developer A explained that bias in any form is something that should raise ethical concerns amongst researchers and developers. When discussing how biased data impacts training, validating and testing, Developer A highlighted the impact on the accuracy of an AI model. They elaborated with the following statement,

> The ethics behind this [bias] is so important because we do not want to add to the problem with our technology. There need to be more discussions about it.
>
> - Developer A

"Data bias should raise ethical concern", said Researcher C and explained that the quantity and quality of data directly impact the AI model's performance. Researcher C explained that gaining access to diverse, rich and high-quality data is something that all researchers should strive to do. Further, they elaborated on the challenge of data bias as a legitimate concern for all people working with technology and called for better education on the topic.

4.1.5 Data Management Systems

During the interviews, a recurring topic that some participants mentioned was the need for improved data management systems (DMS). Essentially, DMS are tools that enable organizations to handle vast amounts of data in a structured and organized way. With the help of such systems, organizations can efficiently store, manage and use data, which is crucial in healthcare, where a large amount of data is to be processed. The following section presents some of the interviewees' opinions and concerns about using DMS in healthcare. This section aims to gain insights into how DMS can facilitate AI development in healthcare, ultimately leading to improved patient outcomes.

Researcher B described the challenges their team faced when dealing with a large amount of data they gained access to. Specifically, they mentioned that the volume of data they retrieved was substantial, *"When millions of images need to be stored securely and properly, there are several challenges."* Further, they had to spend significant time figuring out how to store it securely and properly, *"Storing millions of images is costly and technical; we spent much time figuring it out."*

Proper anonymisation is another important factor for all healthcare AI projects, especially concerning GDPR regulations. All researchers and developers explained various processes of anonymizing data, with some relying on third parties and others doing it themselves. Researcher B explained how each image is labelled with a unique serial number for accurate identification. Any inaccuracies or errors in the anonymization process can compromise the accuracy and validity of the AI model or make it unusable for further work.

Moreover, data privacy and security are critical concerns that must be addressed while integrating new systems into existing hospital infrastructure. Compliance with guidelines and regulations related to data protection and risk assessment is necessary. Researcher C explains that using ISO-certified cloud service providers like Hunt Cloud can ensure data security. The technical challenges of testing algorithms in clinical settings and integrating them into hospital systems, such as overcoming firewalls and establishing research practices within hospital systems, must also be addressed through national systems,

> However, when hospitals and universities collaborate, data must cross firewalls. It's quite complicated to have everything in place to make it work. It's also challenging now when we need to test in a clinical setting and create systems for the hospital. They don't want our algorithms on their regular servers, so they need to be created separately. The loop where data runs through our algorithms and then returns to the radiologist is still in progress. Our health region is lagging because we haven't established research practices here or packaged it into the archive system that many hospitals use for their images.

- Researcher C

Creating a separate research system can enable more efficient data collection and algorithm testing, which is essential for the success of AI in healthcare. Researcher C explained how hospitals and scientists in Bergen are collaborating and sharing data through a shared system, *"In Bergen, they have established a research park which is, in a way, a parallel line to the clinical healthcare system, so that they have greater opportunities both to extract data and test algorithms in a clinical setting."* The researcher further explains how proper data management practices and adherence to regulations ensure the accuracy and validity of AI models and protect patient privacy and security. Thus, they explain that it is crucial to prioritize and invest in robust DMS to facilitate the integration of AI into healthcare effectively and cost-optimized.

4.2 Complying With Regulations And Laws

The interviews revealed a prevalent challenge of obtaining the necessary approvals to use or market an AI model or product for a clinical environment.

Private companies faced a prolonged, expensive, and resource-demanding journey of gaining data access and validating the end product. On the other hand, researchers faced fewer obstacles as they obtained the required data under the pretext of research. The next subsection presents the perspectives, emotions, and viewpoints of developers, researchers, and data owners concerning this subject.

For private companies, where both Developers A and B work respectively, several laws, regulations, and processes must be followed precisely and correctly to obtain the necessary approvals to move forward with the desired product. Developer B describes the process as *"costly, time-consuming and boring,"*. Further, Developer B explained that companies needing access to large amounts of medical data often face resistance from patients and regulatory authorities concerned about privacy and data protection. Hence, they emphasized that companies need to be careful with how they collect and use patient data and ensure the accuracy and safety of their products. Additionally, obtaining FDA/EMA approval for AI-driven medical devices can be lengthy and costly, challenging startups in this space. Developer A explains,

> We should not be looked at as the enemies and leeches of data. We don't want to leech your data to profit; it's about helping people.

- Developer A

There are not only challenges with regulations in terms of gaining access to data, but there are also several strict regulations, laws, and processes to be followed to be able to validate an AI-based algorithm. In Norway, a company must comply with several examples of authorities and guidelines. From interviews with all researchers and developers, Table 4.2.1 was created to understand some of the regulatory instances one may face when validating an AI algorithm used in medical settings in Norway.

As Table 4.2.1 showcases, several regulatory instances must be researched, understood, and applied for a product to be able to enter the medical market. Developer B explained that,

> We need to hire outside counsel to understand how we can comply with all regulations and laws. This is an extremely costly and time-consuming process for us, and we expect it will take 1-2 years and around 50 MNOK to approve it [medical device].

- Developer B

Table 4.2.1: Regulatory bodies and their descriptions in Norway's healthcare system

Regulatory Bodies	**Description**
The Norwegian Data Protection Authority (Datatilsynet)	Enforces data protection legislation, including the General Data Protection Regulation (GDPR), to ensure that the processing of personal data complies when using an AI algorithm in the context of healthcare.
The Norwegian Directorate of Health (Helsedirektoratet)	Develops national guidelines and manuals for healthcare professionals and institutions. To have an AI algorithm validated for use in healthcare, one must follow the guidelines and manuals applicable to the relevant area of the healthcare system.
The Norwegian Medicines Agency (Legemiddelverket)	Approves medical products and equipment for use in the healthcare system. If the AI algorithm is part of a medical product or equipment, one must apply for approval from the agency before it can be used.
The Norwegian Medical Devices Act	Sets requirements for medical equipment sold and used in Norway. If the AI algorithm is part of medical equipment, one must ensure that the equipment meets the requirements specified in the law.
The Norwegian Health Personnel Act	Sets healthcare personnel's professional practice and competence requirements. Suppose the AI algorithm is to be used by healthcare personnel. In that case, one must ensure they have the necessary competence and knowledge to use the algorithm safely and appropriately.

Developers A and B highlight the cost and time factor of the regulatory process, with Developer A saying, *"This is what kills them [AI startups]."* Developer B explains, *"For smaller companies with great and potentially life-saving ideas, the hurdles of getting the data are enough to bankrupt them."* Developer A explains that they have estimated the cost of validating each algorithm, saying, *"It might cost us over 1 million euros to validate one algorithm"*. Further, both Developers A and B share a growing concern for the future of AI in healthcare and calls for better oversight from the Norwegian government,

> The authorities should understand that having so much bureaucracy kills the growth in our industry. Of course, we need the best quality systems used on real-life patients, but for a regular

> company and developer, all these rules are confusing and frustrating.
>
> - Developer A

The researchers had a different point of view regarding rules and regulations. They acknowledged that there were various regulations to follow, but they agreed with developers that it was challenging to understand and comply with all the different laws and standards. The researchers used data in a research context, meaning their primary goal was not to create a product that generates revenue. However, if the researchers were to develop a product that becomes successful and receives enough funding, they would have to adhere to the same regulations, laws, and standards as private companies. This means that even though their current work is not for profit, they would still have to meet the exact legal requirements as commercial organizations if they decide to create a product in the future.

Researcher C explains a project they are currently working on. It focuses on prostate cancer, generating good data and images, exploring new imaging modalities, and using AI for image interpretation. They explain that diagnosing prostate cancer is challenging due to the varying aggressiveness of tumours and the potential for overdiagnosis, leading to unnecessary treatments. Due to the complex nature of the project, they were fortunate enough to receive funding for a doctoral researcher to focus on making the project comply with all rules. Researcher C explains the role in the following manner,

> We have been lucky and received funding for a researcher to work with this [comply with laws]. The role is important for the team and lets other researchers focus on their work, which is much appreciated.
>
> - Developer A

4.3 Trust and Collaboration Between Academic Disciplines and Health Regions

Several interviewees have pointed out a lack of trust between academic disciplines, which they believe directly impacts the acceptance of AI technologies. The following section explores the thoughts and sentiments expressed by these interviewees regarding this issue, highlighting the significance of trust as a crucial factor in determining the future of AI in healthcare.

4.3.1 Lack of Scientific Evidence

During the interviews, it was brought to attention that clinicians exhibit a certain degree of mistrust towards researchers and developers involved in medical AI within the healthcare sector. Specifically, Researcher C and Developer A emphasized the absence of scientific evidence as a significant factor contributing to this lack of trust. This particular concern will be expounded upon in the following subsection, which aims to shed light on the impact of insufficient scientific evidence on trust towards medical AI and propose potential strategies to address these issues and rebuild trust.

"Sometimes it seems that clinicians shy away from technological solutions because there is no evidence supporting it", explained Researcher C. They highlighted the deficiency of scientific evidence in the field of medical AI. *Scientific evidence* can be described as the following, *"scientific evidence is the results of scientific tests used to prove or disprove a theory or hypothesis."*

To shed light on the prevailing lack of trust within the medical AI sector, Researcher C referred to an article titled "Artificial intelligence in radiology: 100 commercially available products and their scientific evidence" by van Leeuwen et al. (2021) [6]. This article presented several key points, which are summarized as follows:

1. AI in radiology is still in its infancy even though 100 CE-marked AI products are commercially available.
2. Only 36 out of 100 products have peer-reviewed evidence, of which most studies demonstrate lower levels of efficacy.
3. There is a wide variety of deployment strategies, pricing models, and CE marking classes of AI products for radiology.

Researcher C emphasized that the article is interesting and raises concerns for their research field. They were especially concerned by the following statement from the article, explaining that *"there might be something wrong with how we research based on it [results from the article]"*. Further, Researcher C speculates that there might be some lack of trust from clinicians when the scientific evidence is presented as low, demonstrating that the actual efficiency of various AI models implemented has little effect.

Developer A also speculates several reasons for the lack of trust between clinicians and developers but mainly points to the fact that AI solutions rarely bring new value to the medical field, often regarding the efficiency of disease detection. Developer A asked the rhetorical question, *"Why should they [clinicians] bother to use solutions that provide nothing?"*, when discussing the theme. Developer A also wondered about the scientific evidence, explaining that many AI solutions are "unproved" and, therefore, understands that medical professionals hesitate to utilize them with real-life patients.

4.3.2 Collaboration and Communication Between Health Regions

A theme that was discussed by several of the interviewees was the lack of communication and collaboration between the various health regions and hospitals in Norway. The following section investigates the variations between health regions and the challenges researchers, developers, doctors, and data owners face.

To understand the challenges related to cooperation and communication between health regions, it is important to understand how the health regions function in Norway. The four health regions that are Norway's healthcare system are each in charge of providing healthcare to the residents in their region. These areas are governed by regional health authorities, who also manage the finances and operations of the local hospitals and speciality medical services. Western Norway Regional Health Authority, Central Norway Regional Health Authority, Northern Norway Regional Health Authority, and Southern and Eastern Norway Regional Health Authority are the four regions depicted in Figure 4.3.1. The Norwegian government funds hospitals and clinics in each region as well as other healthcare institutions [44].

Figure 4.3.1: The four health regions in Norway. Illustration from [45].

The researchers interviewed all for better cooperation between the health regions regarding data management and sharing. Researcher A explains the repetitive process of obtaining data from various health regions, as they sometimes have different application processes and response times.

4.3.3 Communication Between Academic Disciplines

As AI advances and affects various fields, the need for collaboration between diverse academic disciplines becomes increasingly important. AI solutions that are robust, secure, and unbiased require input from experts across multiple fields, such as computer science and medicine. Unfortunately, communication barriers sometimes hinder this collaboration. The following section will explain various opinions and thoughts from the interviewees regarding the topic. The section aims to understand the importance of interdisciplinary collaboration and various challenges when developing AI solutions.

"The development of AI solutions that are robust, secure, and unbiased requires cooperation between diverse academic disciplines", explained Researcher C. Unfortunately, communication barriers can hinder this cooperation, as team members with different academic backgrounds sometimes struggle to understand each other's perspectives. To ensure that AI products meet the needs of various fields, interdisciplinary collaboration is crucial. Researcher B explained that input from computer scientists and medical professionals is necessary when creating health tech products. Researcher C explained it with the following statement,

> I think that it can be difficult to understand each other in such an interdisciplinary project. On the one hand, you have a computer scientist who knows what he/she is doing in that field. On the other hand, you have a doctor who is an expert on the medical side, and it is challenging to communicate so that you understand what is being done and end up with something good. It can be quite demanding, actually.

- Researcher C

The importance of collaboration becomes even more apparent when considering the challenges of introducing new technologies to medical professionals. Getting doctors on board to collaborate with AI companies in the health tech industry can be a significant hurdle. Developer A explained one scenario they experienced once: *"Once I reached out to a doctor, and he simply said that he was too busy. It is not uncommon [to experience]."* Developer A believes that busy schedules and scepticism about the new technology make gaining acceptance from medical professionals challenging.

"It is important to be aware of the challenge, but most doctors are willing to work with us," explained Researcher B and emphasized how various

medical professionals understand that AI will be a part of the medical world. Doctor A suggested the following,

> We are sometimes requested to participate in projects, which is exciting. I can only speak for my division, but I think everyone understands the importance of cooperation between academic disciplines, and they [doctors] are interested in this type of collaboration.

- Doctor A

Data Owners A and B often work on multi-disciplinary projects and emphasize the importance of all parties understanding each other to ensure the projects are completed with a favourable result. They explained, *"Most people from various disciplines can understand each other, but of course, some struggle more than others."*. As a remedy, they describe project structures that allow for multi-disciplinary knowledge exchange, with meeting together in all phases of the project and a clear understanding that many, especially clinicians, seldom have the time for this kind of participation,

> We see that heavy involvement from clinicians often benefits projects we [government registry] fund, but time is often a struggle. They [clinicians] have very busy schedules, so sometimes, even though they want to participate, they can't.

- Data Owner B

4.4 Hierarchical Structure of Healthcare

In recent years, the development of AI has impacted the healthcare industry, attempting to find innovative ways of improving patient care and outcomes. However, the adoption of AI in healthcare is not uniform across different regions and countries. Developer A has reflected on the theme of hierarchies in healthcare and its potential impact on the development of AI in the industry. This section delves into Developer A's insights on how hierarchical structures within healthcare systems can challenge AI's further advancement and integration in healthcare.

Developer A explains that they think the reason for the difference in enthusiasm around AI in healthcare between the US and Europe can be attributed to historical, educational, and hierarchical factors, and explained it with the following statement,

> Doctors [in America] go to university, to med school [medical school], with a bachelor's and, most times, a master's degree from a technical background. And they are much more ready to accept your novelty AI and something technical and outside of their scope. So yeah, the reasons are historical, educational, hierarchical

- Developer A

As Developer A explained, clinicians in the US usually have a more diverse academic background. During their university education, they are exposed to technical subjects like mathematics, engineering, and computer science. In Developer A's opinion, this exposure makes them more open to embracing new AI technologies and their potential applications in healthcare. Developer A compares it to medical education in Europa and says, *"European clinicians usually follow a more traditional educational path"* since they often go directly to medical school after high school and specialise in a specific area of medicine, which may not include exposure to fields like AI or computer science.

Developer A speculates that the hierarchical structure of healthcare in Europe also could play a role in the slower adoption of AI in healthcare. According to Developer A, European clinicians are generally more specialized and may not have as much incentive or motivation to adopt new technologies outside their areas of expertise. Therefore, in contrast, healthcare professionals in the US are more likely to collaborate with researchers and adopt new technologies that could benefit their patients.

To create a research field that allows for more adoption of AI technology, Developer A wishes that governments and other regulatory parties allow for changes. *The change has to come from above*, explained Developer A in the interview's conclusion and added that they wished that the future would be more diverse and welcoming of AI technology in the healthcare sector.

4.5 Ethical Concerns

Addressing several ethical concerns is crucial as AI plays an increasingly important role in society and healthcare. Various interviewees have shared their perspectives on these concerns and how they impact their work. The following subsection discusses their thoughts, feelings, and insights on the topic, shedding light on the complex ethical challenges surrounding AI in healthcare.

"There are ethical considerations in everything when working with AI," explains Researcher A when discussing the theme. Researcher A continues

by explaining that there are ethical considerations throughout the development process of AI projects in the healthcare sector. Therefore, Researcher A emphasizes the importance of all involved in these projects gaining knowledge about the topic and being willing to address the ethical concerns to ensure projects that allow for the best possible results. Researcher A formulated the statement when discussing the theme,

> I think everyone should be involved in this [ethical considerations]. We will never create solutions that are good enough if not.

- Researcher A

One ethical concern when working on medical AI projects is the lack of available data. As explained in subsection 4.1.1, the shortage of data can come from various reasons. Still, mainly with a shortage of patients with the relevant disease or regulatory obstacles. From an ethical viewpoint, Developer A problematized the shortage of data for several reasons. Firstly, they emphasized the concept of *data colonialism* as a great ethical concern, explained the concept in the following manner,

> Have you heard of data colonialism? It's a term used to describe a significant challenge. Imagine you're trying to access readily available information, like data on white women from Norway. But what about data from Africa, Asia, or Brazil? It becomes increasingly difficult to obtain a comprehensive understanding across diverse regions. Solving this issue won't be easy, as accessing data from places like Africa poses a question of availability. How can we gather data from these regions?

- Developer A

4.6 Future of AI in Healthcare

In all of the conclusions of the interviews conducted respectively, the question *"How do you see the future of AI in healthcare?"* was asked. The question aimed at yielding interesting results and knowledge from the participants, but was also asked to find possible solutions and future work in the sector. The subsequent subsection explains the various insights from the interviews, focusing on the personalized treatment of patients, more precise AI models, and better regulatory framework and collaboration with hospitals.

4.6.1 Personalized Treatment of Breast Cancer

When Doctor A was asked the question regarding the future of AI in healthcare, they were excited to share their views. As a senior clinician, Doctor A explained that they had recently experienced several changes in the sector and that they are positive that more and more clinicians are willing and excited about how AI can aid in radiology and healthcare in general. *I can only speak for myself, but I am very excited to see where the technology takes us*, explained Doctor A.

As for the future, Doctor A is positive that the healthcare sector will see some changes with technology. For instance, Doctor A explained that AI technology would aid in more personalized patient treatment. As Doctor A works closely with women with breast cancer and the BreastScreen Norway program, they hope that the future with more technology allows for more screening of women with higher risk for breast cancer, and fewer than biennial screenings of women with lower risk, and explained it with the following statement,

> I think that in the future, maybe we [BreastScreen Norway] will have more personalized treatment, with women genetically disposed of having breast cancer or women with dense breast tissue being invited more often for regular screenings. I think we can use their health records to determine how often screenings are needed, and maybe AI can help here. As for women with less risk of breast cancer, maybe they don't need to be invited as much.

- Doctor A

Developer B emphasized the importance of personalized treatment and expressed their desire for their company's devices to enhance the differentiation between individual patients. They aimed to achieve more accurate results and extend assistance to a wider range of patients. By leveraging advanced technology, they intended to tailor treatments and outcomes to suit each patient's unique needs, ultimately improving overall healthcare outcomes.

4.6.2 Precise AI Models

When asked about the future of AI in breast cancer research and the healthcare sector in general, Researcher B highlighted various topics that could be of interest. When talking about their ongoing project, Researcher B emphasized that it is a research project with a strict deadline, but that they are hoping to dig deeper into differentiating between various subtypes of breast cancer, and explained it with the following statement,

> When we assess more data, maybe it becomes more apparent how we can differentiate between various breast cancer subtypes. We have it in the backlog. [...] Maybe GAN can aid in our future work.
>
> - Researcher B

The statement highlights the introduction of GAN, as mentioned in subsubsection 2.3.2.2, as a potential solution worth exploring when delving into the transformations within the healthcare sector. Researcher B expresses optimism, acknowledging that the sector is undergoing significant technical, regulatory, and medical changes. However, Researcher B believes that the researchers working in this field possess the ability to adapt and navigate through these changes successfully.

Another interesting aspect Researcher B hopes to research is how an AI model can identify signs of breast cancer earlier. During the interview, while brainstorming ideas, Researcher B discussed a solution that combines historical data with AI technology to generate predictions that could assist clinicians in identifying patients at risk at an earlier stage. Researcher B emphasizes the importance of reducing the mortality rate and avoiding the detection of conditions that may not be present and explains with the following statement, *"We want to decrease the mortality rate but not detect something that is not there. Over-diagnosing can become a more common issue."* They raise concerns about the potential increase in over-diagnosing and thus advocate for developing more precise and unbiased AI models to effectively support clinicians in their decision-making process. Researcher B envisions a future where such advanced AI models are crucial in improving healthcare outcomes by providing accurate and timely insights to healthcare professionals.

Researcher C aligns with Researcher B's perspectives and elaborates on their plan to incorporate ViT into their future research project, as described in subsubsection 2.3.2.3. Researcher C describes ViT as an emerging technology their research group believes holds great potential for enhancing their existing algorithms, coining it as *"is up and coming."* By harnessing the power of ViT, they aim to leverage its capabilities to improve their research outcomes significantly. Researcher C highlights the growing importance of ViT in the field and expresses confidence that integrating this innovative technology will contribute to the further advancement of their research field. With an optimistic outlook, Researcher C envisions a future where ViT can help optimize algorithms and achieve remarkable progress in its research project.

4.6.3 Ethical and Systematical Data Collection

Developer A envisions a future where AI assumes a prominent role in healthcare, significantly emphasising ethical data collection and ensuring patient awareness regarding data usage. They commend a healthcare institution that has initiated digitising all medical records and intends to share the data with partners for AI advancement, explaining that *"this is what the future of AI needs."* Developer A firmly believes that systematic data collection from healthcare institutions is pivotal for the future development of AI in healthcare. They further emphasize the need for large healthcare establishments to prioritize anonymization and secure storage of patient data in adherence to stringent regulations; otherwise, the progress of AI in healthcare may face substantial challenges.

Developer A appreciates the commendable efforts of the healthcare institution, predicting that it will surpass other establishments lacking similar resources and initiatives. They elaborate on how this institution has collaborated with a technology company to provide an API to the database, facilitating the development of AI algorithms by engineers. However, Developer A acknowledges that extensive data collection of this scale may be less likely in Europe due to the vast patient population and the associated costs of digitization. Consequently, they suggest that countries such as India and China, possessing vast data, will lead the charge in AI development within the healthcare domain.

4.6.4 AI Replacing Clinicians

The idea of AI-based solutions in healthcare rendering clinicians obsolete and *"replacing their roles"* has sparked speculation among developers and researchers [10]. When exploring this scenario, the stakeholder groups provided unique insights and perspectives. In the subsequent subsection, the diverse responses from the interviewees regarding this topic are presented, aiming to highlight key factors that could shape the future perception of healthcare.

"We will always need doctors, but maybe not in the traditional way", was the first response when Researcher C got asked the question. They are confident that nothing can make clinicians obsolete or redundant. However, they believe that the nature of their work might change to comply with the changes in the sector enforced by novel technology and that clinicians might be able to be more hands-on with their patients. *If technology can take over the routine jobs for clinicians, they might have more time to be there for their patients*, explained Researcher C.

Another researcher who shares the perspective of Researcher C is Doctor

A. When asked about the position of clinicians, they were certain that *"I don't think we will be replaced. People want to be treated by other people."* Further, Doctor A envisions a future where technology aids the clinician in their daily tasks, freeing time that can be spent with patients and doing other types of work, and explained the following, *"If we get more time, maybe we can collaborate more with researchers."*

Developer A comes in with another perspective and explains that even though it might be cheaper to "replace" doctors with AI models and other kinds of technology, it simply is not *"what the people want."* They emphasized that AI might outperform clinicians sometime in the future, but that is not the case now, so it makes sense to retain clinicians.

CHAPTER

FIVE

PROPOSITIONS

Based on in-depth interviews and previous research, a comprehensive set of propositions has been formulated to address the challenges highlighted in chapter 4. These propositions will be thoroughly examined in the upcoming chapter, providing specific solutions to the identified challenges and exploring their potential impact on the healthcare sector. These propositions specifically aim to address *RQ3*, presented in chapter 1, which explores how technical and non-technical approaches can help mitigate the challenges of developing and implementing AI solutions in healthcare. It is important to acknowledge that these propositions are built upon insights derived from interviews, existing publications, and theoretical frameworks and thus may not encompass every relevant perspective. While the primary focus of these propositions lies within the research sector, certain propositions may also have applicability to other sectors, and these will be discussed further in the corresponding subsections.

To summarize the discussion and proposed solutions, Table 5.0.1 provides a concise overview derived from the research findings presented in chapter 4 and discussed in chapter 5. The first proposition, P1, highlights the potential of leveraging synthetic data in medical AI models to enhance dataset diversity and improve model robustness and accuracy. The second proposition, P2, emphasizes the importance of implementing effective techniques to mitigate bias in medical AI models, aiming to reduce disparities and promote fairness. Additionally, incorporating fairness goals from the outset of a project is encouraged. The third proposition, P3, focuses on fostering improved communication between developers and clinicians, recognizing its significance in successfully implementing and adopting AI solutions in healthcare. Strategies such as introducing intermediaries, organizing conferences and knowledge-sharing events, and establishing dedicated communication management systems are proposed. Lastly, the fourth proposition, P4, emphasizes the utilization of technical frameworks, including Evidence-Based Practice (EBP), to build trust and recognition among clinicians. Clinicians can be more receptive to AI solutions by promoting scientific evidence in research projects, driving advancements and improving outcomes. Despite the study's limitations in sample size, variability, and time constraints, the propositions presented in the table offer valuable recommendations for advancing the integration of AI in medical settings, providing a roadmap for future research and development efforts.

Table 5.0.1: Summary of propositions presented in chapter 5, derived from the results presented in chapter 4.

Number	Proposition	Description
1	Explore the potential of synthetic data in medical AI models.	Leveraging synthetic medical data can offer a viable solution to enhance dataset diversity for training AI models, leading to the development of more robust and accurate models.
2	Implement techniques to mitigate bias.	Employ various strategies to effectively address bias in medical AI models, aiming to reduce disparities and promote fairness. Incorporate fairness goals from the outset of the project.
3	Foster improved communication between developers and clinicians.	Enhance communication and efficiency between developers and clinicians by introducing intermediaries, organizing more conferences and knowledge-sharing events, and establishing a dedicated communication management system between research facilities and hospitals.
4	Utilize technical frameworks in research projects to build trust and recognition among clinicians.	By employing technical frameworks such as Evidence-Based Practice (EBP) and promoting scientific evidence in research projects, clinicians can become more receptive to AI solutions in the healthcare sector, leading to growth and improved solutions.

5.1 P1: Synthetic Data to Train Medical AI Models

The challenges discussed in section 4.1 shed light on several data-related issues and how they make it more difficult to create novel AI solutions. During the interview, Researcher B suggested that their research group could explore GAN, described in subsubsection 2.3.2.2, as a potential solution to generate more images to train their breast cancer detection algorithms. Additionally, Researcher C expressed their interest in investigating ViT, described in subsubsection 2.3.2.3, as a promising DL image recognition and analysis model, saying,

> [..] We are actually exploring how vision transformers might function as a promising complementary choice in modernity, so perhaps we need to start thinking in completely new ways.
>
> - Researcher C

This section explores the potential of employing synthetic data to overcome various data-related obstacles that emerge during AI model development, generally in healthcare and within breast cancer detection. Synthetic data presents a viable solution to address data quality, quantity, and diversity issues. Furthermore, it offers an efficient strategy to adjust to changes in the dataset while ensuring adherence to regulatory and privacy standards. The discussion also aims to discover how synthetic data can be created with ViT and GAN to tackle challenges highlighted by professionals in the industry and relevant research.

5.1.1 GAN and ViT to Create Synthetic Image Data

Generating synthetic data, known as *synthesis*, can be accomplished through various techniques such as decision trees or DL algorithms. A promising approach in AI for generating synthetic images that closely resemble real ones is the combination of GANs with ViT models. This subsection elaborates on how these two DL models can be combined to generate synthetic medical images.

In their 2020 scientific paper, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Dosovitskiy et al. introduced ViT for image recognition purposes [46]. Originally developed for NLP and text recognition, ViT demonstrated effectiveness in image recognition. The authors' research unveiled a cost-effective and efficient approach to image classification by treating an image as a sequence of patches and leveraging a standard Transformer encoder commonly used in NLP tasks [46].

On the other hand, GANs were first introduced in 2014 in the scientific article "Generative Adversarial Networks" by Goodfellow et al. [47]. This research paper presented GANs as a framework that explores the interplay between two neural networks: a generator and a discriminator. The generator generates realistic data, while the discriminator distinguishes between real and fake data. The success of the generator can be measured by its ability to deceive the discriminator into perceiving a fake image as real [47].

To better understand how ViT and GAN can be combined to generate synthetic images, Figure 5.1.1 is presented.

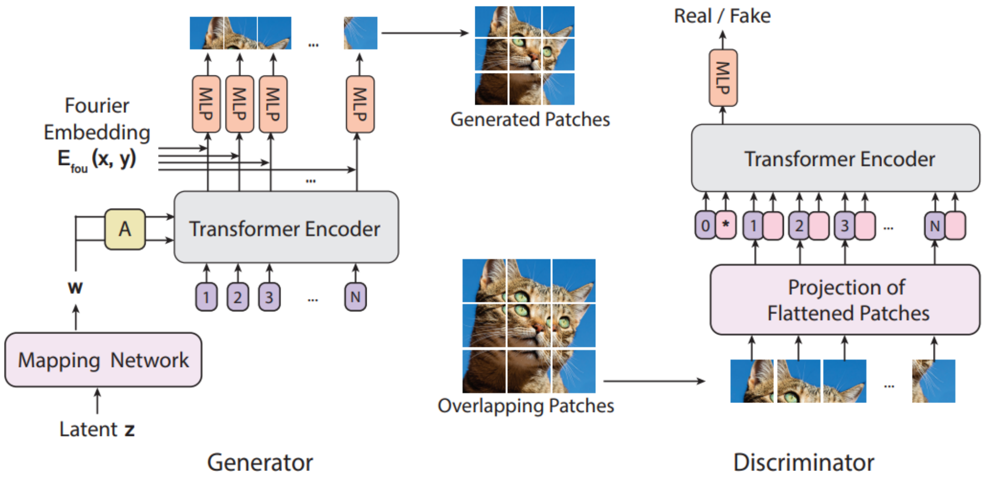

Figure 5.1.1: Figure illustrating how ViT and GAN can be combined, created by Demochkin (2021) [48].

One can combine GAN and ViT to create synthetic medical images in several ways. The following example is based on the theory presented in [46] and [47]. The explanation is simplified to fit the scope of the thesis.

The first step to combining ViT and GAN is to train a ViT model on a large dataset of real medical images, which can be obtained from both open and local sources that require regulatory approvals. The ViT model is trained to learn features from the images that can be used to generate new images. Next, the GAN model is trained using the learned features from the ViT model. The generator takes in a random noise vector and uses the features learned by the ViT model to create a synthetic medical image. The discriminator tries to differentiate between real and synthetic medical images, depicted as the [*] in Figure 5.1.1.

After the GAN model has been trained, the ViT model can be fine-tuned using the synthetic medical images generated by the GAN. The fine-tuning process can improve the ability of the ViT model to learn features from synthetic medical images. The GAN model generates the synthetic images using random noise vectors, a source of randomness, as input, and the ViT model can refine the images by learning features from them. The discriminator cannot differentiate between synthetic and real images if the images are good enough.

5.1.2 Results From Previous Studies

Previous studies highlight the possibilities of using synthetic images in a medical context. The following subsection presents findings from peer-reviewed studies published in various medical journals. All studies emphasize the benefits of using synthetic medical data in training AI models in medical settings and give context to understand further how synthetic data can be used in a medical context.

In the publication titled "SinGAN-Seg: Synthetic training data generation for medical image segmentation" by Thambawita et al., the authors delve into the realm of synthetic data for training models in medical image segmentation [49]. Their research introduces an innovative approach known as SinGAN-Seg, which combines the power of SinGAN with segmentation masks to generate synthetic medical images. These synthetic images are accompanied by ground truth segmentations provided by clinicians.

SinGAN-Seg distinguishes itself from traditional GANs' remarkable capability to generate synthetic data utilizing a single training image. This unique characteristic positions it as an ideal solution for constructing medical datasets. The study underscores synthetic data's numerous benefits, specifically in tackling the challenges associated with data annotation and labelling discussed in previous research. Moreover, SinGAN-Seg eliminates the need for expert annotation of the ground truth mask since each synthetic sample comprises a synthetic image and its corresponding ground truth mask. This approach streamlines the data generation process, reducing the burden of manual annotation. By leveraging synthetic data, SinGAN-Seg opens new avenues for enhancing medical image segmentation models and addresses critical challenges in the field. However, it is important to note that further investigation is still required to explore the potential of synthetic data in medical imaging fully [49].

However, while the potential of synthetic data in the medical imaging domain is promising, it is essential to acknowledge that its utilization requires thorough investigation. Privacy and confidentiality concerns must be handled with utmost care and caution. The authors of the study highlight the significance of these concerns and express the following perspective [49],

> For example, Norway should follow the rules given by the Norwegian data protection authority (NDPA) and enforce the personal data act, in addition to following the general data protection regulation 31 (GDPR) guidelines being the same for all European countries.

- Thambawita et al. (2022) [49]

The research presented in "Generation and evaluation of synthetic patient data" by Goncalves et al. focuses on generating and evaluating synthetic patient data in the medical field [50]. These studies shed light on the potential of synthetic data and its relevance in a medical context, providing valuable insights for understanding its applications.

The authors present a comparative analysis of different methods for generating synthetic data and highlight several important points for discussion. They emphasize the potential of synthetic medical data in accelerating research, as it can be generated without complying with strict data access regulations. This is particularly significant considering the challenges of limited data access discussed by multiple interviewees in chapter 4. Synthetic medical data could be a valuable resource to overcome these limitations and provide an advantage in the research sector.

The authors highlight an important aspect of using synthetic medical data in the healthcare sector, explaining,

> On one hand, the synthetic data must capture the relationships across the various features in the real population. On the other hand, the privacy of the subjects included in the real data must not be disclosed in the synthetic data.
>
> - Goncalves et al. (2020) [50]

The authors' recognition of these two contradictory aspects illustrates the delicate balance that must be achieved when generating synthetic medical data. It emphasises the significance of producing high-quality synthetic data that preserves real data's relationships and statistical properties while safeguarding individual privacy. By addressing both issues, synthetic medical data can serve as a valuable research resource, balancing data accessibility and privacy protection.

5.1.3 Synthetic Data to Mitigate Challenges

Synthetic data can be a helpful solution to data-related challenges discussed in chapter 4, such as data access, regulatory restrictions, and privacy concerns. Synthetic data involves generating artificial data that mimics the statistical properties of real data, which can help address data availability and privacy concerns.

Furthermore, in the medical field, there may be changes in standards and technologies that require adjustments to how data is processed and analyzed. For example, if the standard of mammography changes from 2D to 3D images, the existing database of images may not be sufficient for researchers and developers. By leveraging synthetic data, for instance, through ViT and GAN models, developers can scale data by generating their own, which may be comparable to, if not better, than real data [50]. This approach allows developers to stay efficient in building AI models without being constrained by regulatory restrictions or reluctance from governing bodies.

Synthetic data allows developers to comply with GDPR guidelines while obtaining accurate results from image data. In the case of using ViT and GAN to generate synthetic data, it is important to note that ViT uses a self-attention mechanism to process images and extract meaningful labels and features without exposing any real patient information [46]. By using only the learned representations from the ViT model based on real patient data, developers can ensure GDPR compliance. GAN can generate synthetic data resembling real patient data, which can be used to retrain the ViT model multiple times, improving its accuracy and performance. This way, developers can confidently use image data without violating patient privacy [49][50].

Another important aspect of synthetic medical data is that they can be trained to include a wider range of data, which can mitigate data bias present in medical data [50]. Regarding detecting breast cancer, Researcher B uttered concerns that the model their team is developing might not be reliable in identifying different types of breast cancer. However, using GAN and ViT to create synthetic data, a more diverse and extensive database could be used to train the models. This would potentially increase their ability to detect various cancer subtypes with different spreading methods, explained in chapter 2. For example, IDC is a type of breast cancer that develops in the milk ducts, while another less common subtype of cancer, ILC, spreads and develops in the lobules. Therefore, incorporating GAN and ViT can help improve the accuracy and reliability of breast cancer detection models.

It is worth noting that other AI models that can transform 2D images to 3D images exist, but their performance may vary. Therefore, developers and researchers should consider which approach best suits their use case. By adapting to changes in the medical sector and leveraging the latest technologies and methods, AI solutions can become more robust and effective.

5.1.4 Limitations of Using Synthetic Medical Data

When considering synthetic medical data for training AI models in healthcare, it is essential to address various concerns to ensure security, reliability, and value [46]. This section will explore the concerns of using synthetic medical data to create, train, and validate AI models.

One of the primary concerns involves ethical and legal considerations. Working with patient-sensitive medical data requires adherence to strict laws, regulations, and ethical guidelines. When using synthetic medical data, it is important to comply with these regulations. The European Data Protection Supervisor (EDPS) has identified three key factors that are affected when incorporating synthetic data [51].

Firstly, the issue of output control arises. Validating models trained with synthetic data requires comparing them to models trained with real-world medical data. This process can be complex, as ensuring the accuracy and consistency of the output necessitates careful examination and verification. The second concern highlighted by the EDPS is the difficulty in mapping outliers. Synthetic data is not an exact replica of real-world data, meaning certain outliers present in the original data may not be adequately represented in the synthetic data. This poses a challenge, as outliers can hold significant importance in certain healthcare applications and should not be overlooked. The third concern raised by the EDPS is the quality of the data. The quality of synthetic data is closely linked to the quality of the original data and the data generation model. Recognizing that synthetic data may inherit biases present in the original data, potentially introducing inaccuracies is crucial. Moreover, manipulating datasets to create fair synthetic data can further complicate matters and potentially lead to inaccurate representations [49][50][51].

5.2 P2: Ensure More Fairness in Medical AI

Fairness in research projects was consistently highlighted during the interviews, emphasizing the need for equitable outcomes and increased trust among clinicians and patients. The notion of fairness emerged as a crucial factor in determining the success of medical AI solutions. In response to *RQ3*, chapter 1, this section delves into the concept of fairness in algorithms and AI models. It explores the ramifications of lacking fairness and presents a range of technical solutions to mitigate different forms of bias and promote enhanced fairness. Addressing these issues aims to promote a deeper understanding of fairness and its significance within the context of medical AI research, ultimately facilitating more equitable and trustworthy outcomes.

5.2.1 Defining Fairness in Medical AI

Fairness in medical AI refers to the ethical principle and practice of promoting equitable and impartial outcomes for individuals and patient groups in the context of AI algorithms used in healthcare [52]. It strives to prevent discriminatory biases and ensure that the benefits and risks associated with AI technologies are distributed fairly among diverse populations. Fairness in medical AI is particularly important in addressing disparities in healthcare access, diagnosis, treatment, and patient outcomes, which can disproportionately affect marginalized or underrepresented groups. In the article "Ensuring Fairness in Machine Learning to Advance Health Equity" by Rajkomar et al. (2018), the authors explain how bias can be encompassed in ML models [53],

> Because machine-learning models learn from historically collected data, populations that have experienced human and structural biases in the past—called *protected groups*—are vulnerable to harm by incorrect predictions or withholding of resources.
>
> - Rajkomar et al. (2018) [53].

To underscore the significance of fairness in medical AI, it is crucial to delve into the root causes contributing to disparities in outcomes within different AI-based solutions utilized in healthcare. In the article "Addressing fairness in artificial intelligence for medical imaging" by Lara et al. (2022), the authors identified three primary sources of bias in AI models: data, people, and the model itself, as depicted in Figure 5.2.1 [54]. These three sources are influenced by the extent to which the available dataset accurately represents the overall population and how outcomes may be skewed in favour of underrepresented groups. Understanding and addressing these sources of bias is vital for advancing fairness in medical AI applications [53][54].

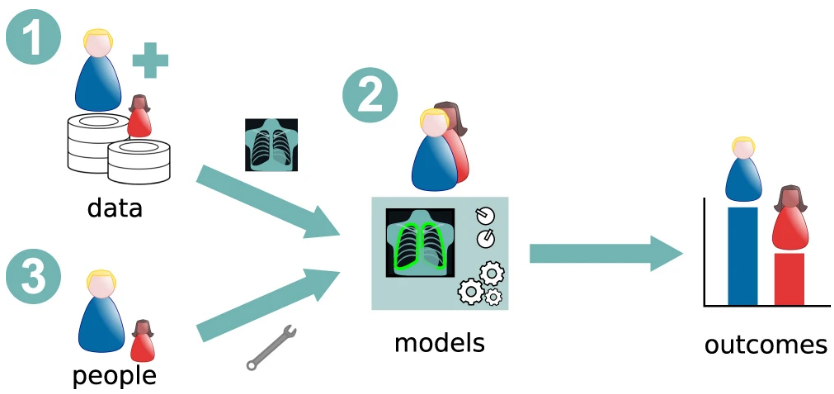

Figure 5.2.1: Figure illustrating three sources for disparities in outcomes from medical AI models, according to Lara et al. (2022) [53].

Rajkomar et al. highlighted various types of bias that may affect the fairness of a medical AI model, presented in Figure 5.2.2 [53]. The figure highlights the model development process, as it is essential to recognize that differences in feature distribution between protected and nonprotected groups can introduce biases, creating unfair outcomes. Furthermore, the figure highlights the risk of model deployment when the model is trained on data that may not accurately represent real-world scenarios. These model design and data biases impact patient outcomes as the model interacts with clinicians and patients, influencing decision-making and potentially perpetuating healthcare disparities [54][53].

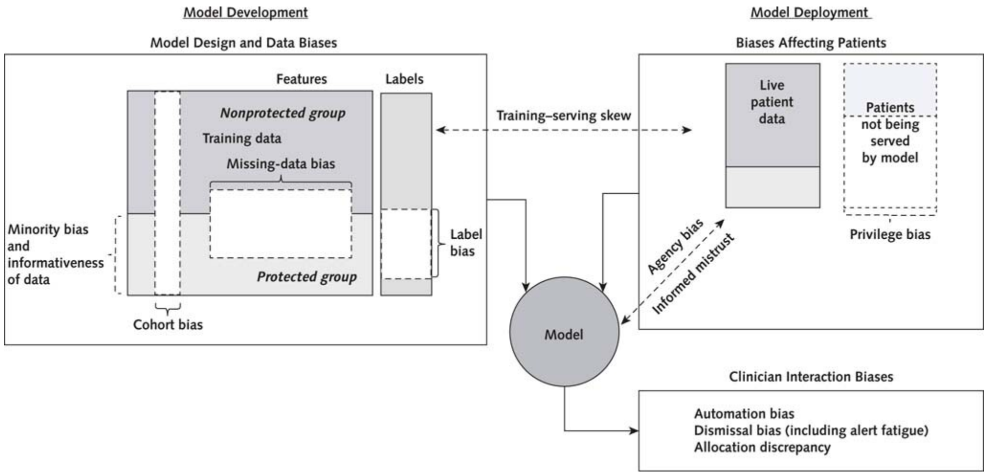

Figure 5.2.2: Figure created by Rajkomar et al. (2018) that highlights various forms of biases that impact a medical AI model [53].

The connection between algorithmic bias, ethical considerations, and fairness emphasises the need to develop and implement AI systems that are transparent, accountable, and unbiased. Researchers and healthcare practitioners recognize the importance of addressing bias and ensuring fairness in medical AI applications to prevent harm and foster health equity [54][53].

5.2.2 Propositions to Mitigate Consequences of Less Fairness

Researchers can employ various techniques to promote fairness to mitigate disparities in outcomes stemming from medical AI models. According to Lara et al., addressing algorithm bias can occur at three critical stages: before, during, and after training [54]. The following subsection explores various techniques that can be applied at different stages.

5.2.2.1 Before Training

Before training, researchers can modify the dataset, aiming to rebalance it and alleviate disparities. However, since the data used in AI models comes from patients, strict adherence to regulations and consultation with clinicians for accurate ground truth labelling is crucial. This ensures that ethical considerations and professional expertise are upheld throughout the process [54][53].

Promoting fairness in AI models requires collecting representative data, necessitating close collaboration with clinicians. By engaging with healthcare professionals, researchers can identify data gaps and explore ways to acquire the necessary information. This iterative process of data collection, guided by clinicians, plays a pivotal role in achieving a comprehensive and inclusive representation of diverse patient populations. By incorporating clinicians' perspectives, AI models can better address the challenges and biases within the dataset, resulting in more accurate and unbiased training. This collaboration leads to fairer outcomes in healthcare decision-making [54][53].

Recognizing the importance of clinicians' time, it is essential to streamline and optimize the data collection process, minimizing the burden on healthcare professionals, as it was introduced as a great challenge by multiple interviewees, presented in chapter 4. Clear communication, efficient data acquisition methods, and utilising existing resources effectively can help maximize the impact of clinicians' contributions without overburdening their already demanding schedules. By valuing and respecting their expertise, researchers can foster stronger collaborations and ensure that fairness remains a central focus in medical AI research [54][53].

Ensuring fairness in algorithms requires a proactive approach during the initial design phase of the project. According to Rajkomar et al., researchers and relevant stakeholders should comprehensively review the ML model's intended goal, emphasising including protected groups [53]. An essential aspect of this process is identifying the specific groups that should be protected and comprehending the overall implications for the project. Collaborating with clinicians and other stakeholders is crucial in making informed decisions and gaining contextual insights that can guide the inclusion of protected groups in a meaningful and equitable manner. By actively involving diverse perspectives, the design process can foster fairness and address potential biases from the outset, ultimately contributing to developing more just and reliable algorithms [53].

5.2.2.2 During Training

A technique known as *data augmentation* can address the imbalance in image-based datasets. Data augmentation involves artificially expanding the original dataset by generating additional data based on the existing samples. Typically implemented after preprocessing, this technique can impose greater diversity and variability into the dataset. Consequently, data augmentation has the potential to mitigate disparities within the original dataset by incorporating a wider range of instances, which can be helpful if the original dataset is small or lacking [55].

Synthetic data is another potential solution to mitigate bias and enhance datasets, as described in the referenced work, section 5.1. Synthetic data offers the advantage of augmenting the dataset by creating new data inputs that are not replicas of the original samples. Instead, synthetic data can be modelled based on the characteristics and patterns present in the original dataset. This approach allows for introducing diverse and representative data points, which can help address bias and promote fairness in machine learning models. By leveraging synthetic data, researchers can expand their analysis's scope and improve their algorithms' robustness and generalizability [50][49].

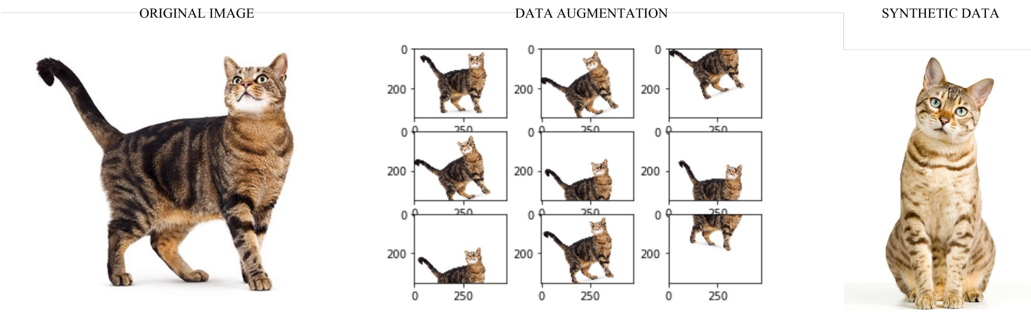

Figure 5.2.3: Figure showcasing the difference between data augmentation and synthetic image. The original image and data augmentation was created by Dwivedi, R. (2020) [56].

Bootstrapping is a powerful resampling technique that significantly enhances the accuracy and reliability of bias estimation. This approach treats the original dataset as the population and generates new datasets through random sampling. The key advantage of bootstrapping lies in its ability to assess the variability of bias estimates across these different samples. Through multiple iterations, new datasets of the same size as the original are generated, and bias is estimated for each. A more precise and robust estimation of bias is obtained by calculating the mean and variance of these bias estimates. This technique is particularly beneficial when working with small datasets or when the bias estimation method is sensitive to the choice of data samples and can potentially mitigate the effect of bias [57].

To ensure the integrity of the development process, it is vital for researchers and stakeholders to consistently uphold the fairness goals established during the project's preliminary phase. According to Rajkomar et al., this heightened awareness enables them to proactively address various aspects of data bias and implement effective mitigation strategies [53]. By focusing on fairness throughout the development process, researchers can work towards creating unbiased and ethically sound solutions that promote equitable outcomes.

5.2.2.3 After Training

Once the model development phase is complete, it is crucial to take several measurements to ensure its fairness. In particular, evaluating AI models, especially those deployed and utilized in a clinical environment, becomes critical. The evaluation process should focus on comparing the deployment and training data to ensure comparability and fairness [53].

Continuous monitoring and retraining of the model are crucial for its optimization. It is imperative to consistently evaluate the model's performance in real-world scenarios to identify and address any emerging biases, thus improving its fairness. Regular retraining, utilizing updated and diverse datasets, is an effective strategy to tackle biases that may have been overlooked during the initial training phase. In addition, involving human reviewers or domain experts in the validation process enhances the approach further. Their valuable perspectives enable the identification of biases that automated assessments might miss. Through assessing the model's predictions for fairness, these reviewers can detect potential issues and provide constructive feedback for improvement [58][59].

Transparent documentation and reporting of the model's development process are important. This includes documenting the datasets used, preprocessing techniques, and evaluation metrics employed. Transparent reporting ensures accountability, enables external audits, and builds trust among users and stakeholders [58][59].

5.2.3 Limitations of Ensuring More Fairness in Medical AI Models

When discussing fairness in medical AI solutions, it is important to understand potential limitations. To explore this issue further, the following section will discuss the findings from the article "Ethical limitations of algorithmic fairness solutions in health care machine learning" by McCradden et al. (2020) [60].

When discussing the concept of fairness, it is important to understand the many prerequisites needed before measures can be taken to advance further in the field. The following subsection explores various limitations of ensuring fairness in medical AI solutions.

While algorithmic fairness solutions have been developed to create neutral models that avoid discrimination, framing fairness as a purely technical problem can be problematic and potentially worsen harm to vulnerable groups. Recognizing the complex causal relationships between biological, environmental, and social factors that contribute to health differences across different identities is important. While incorporating differences between identities is justified in cases where causation is reasonable, including non-causative factors in recommendations can perpetuate unequal treatment and discrimination. Striking a balance between acknowledging differences and avoiding discrimination becomes a challenging task [60].

Another issue arises from the disconnection between a patient's clinical trajectory and fair predictions. If a model, adjusted for fairness, predicts that a patient will respond to treatment similarly to the reference group, discrepancies may arise if the patient's response differs. This discrepancy can impact the model's performance metrics and clinical utility, potentially obscuring more appropriate interventions and perpetuating health inequalities. Therefore, fairness, as measured solely by output metrics, is inadequate, and real-world consequences must be carefully considered [60].

Bias and ineffective algorithmic fairness solutions undermine the ethical obligation to avoid patient harm. Ensuring patient safety involves being aware of the limitations of AI models concerning protected identities and social determinants of health. To achieve this, transparency and accountability are crucial throughout machine learning models' development, testing, and implementation stages. In interpreting model outputs, it is essential to involve relevant stakeholders, including researchers, clinicians, patients, and auditors. Conducting clinical trials becomes imperative to understand the real-world implications, especially for vulnerable patients, when integrating machine learning into clinical decision-making [60].

5.3 P3: More Trust and Recognition in the Healthcare Sector

Participants in the interviews expressed concerns about the lack of robust scientific evidence, which contributes to clinicians' scepticism regarding AI solutions in healthcare. To address this challenge, a potential solution is adopting evidence-based practice. This theoretical framework will be explored further in the following section.

5.3.1 Evidence-Based Practice

Evidence-based practice (EBP) is a framework that can facilitate more scientific evidence in medical AI. The framework is widely used when implementing new medical procedures and has aspects that can be reused to ensure that AI solutions are suitable for the healthcare sector [61].

There are three main components to EBP, as depicted in Figure 5.3.1. In the book *Research for evidence-based practice in healthcare* by Newell and Burnard (2010), the following explanations are largely based on the definitions provided in the book [61].

EBP encompasses three main components, illustrated in Figure 5.3.1. The book titled *Research for evidence-based practice in healthcare* by Newell and Burnard (2010) serves as a primary reference for the subsequent explanations, which closely align with the definitions outlined within the book [61].

1. Research evidence encompasses scientifically rigorous findings from studies that systematically collect, analyze and interpret data to address research inquiries. It provides unbiased and reliable information about healthcare interventions' effectiveness, safety, and cost-effectiveness, facilitating informed decision-making by healthcare practitioners.

2. Clinical expertise represents the extensive knowledge, skills, and experience healthcare professionals possess in their specific fields. It involves proficiently integrating and applying research evidence into patient care, considering individual cases and making decisions based on a combination of the best available evidence, clinical judgment, and accumulated experience.

3. Patient preference recognizes each individual receiving healthcare's unique values, beliefs, expectations, and desires. It considers factors such as treatment goals, quality of life considerations, cultural and personal values, and the active involvement of patients in the decision-making process. Patient preference is seamlessly integrated with research evidence and clinical expertise to develop patient-centred care plans that align with their needs and preferences.

The framework of EBP holds great potential in fostering trust from clinicians towards AI solutions in healthcare. EBP emphasizes the significance of rigorous research design, ensuring medical AI studies' scientific validity and reliability. Adhering to principles such as appropriate sample sizes, control groups, and randomization, well-designed research studies provide a strong foundation for developing and validating AI algorithms and models [61].

Additionally, EBP encourages using systematic literature reviews to gather and evaluate existing research evidence. By conducting comprehensive reviews, relevant studies are assessed for their quality and synthesized to inform the development of AI algorithms. This emphasis on evidence synthesis helps clinicians understand the robustness of AI models, their data sources, and the benchmarks used for performance evaluation. By accessing well-evaluated evidence, clinicians can trust the reliability and applicability of AI solutions in their practice. Critical appraisal of research evidence is another crucial aspect of EBP. When applied to medical AI, this involves critically

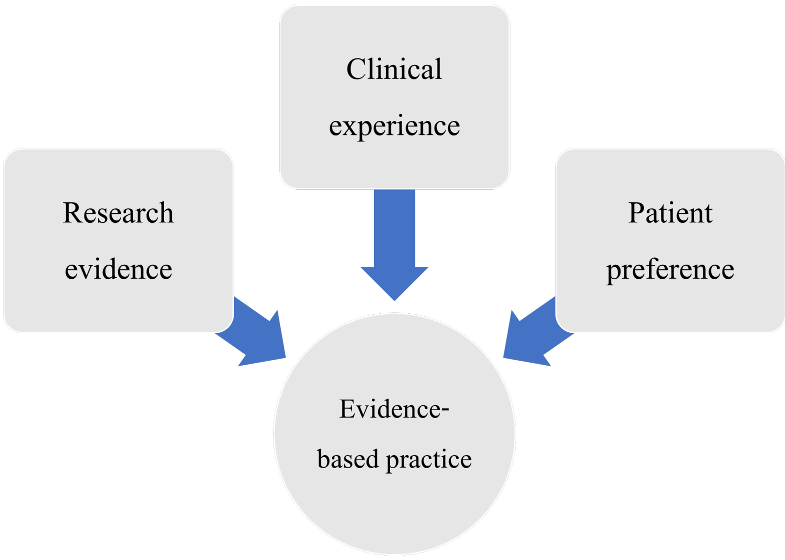

Figure 5.3.1: Figure illustrating the three main components of EBP [62].

evaluating studies on AI algorithms and examining methodology, biases, and limitations. By engaging in this critical evaluation, clinicians can assess the strengths and weaknesses of AI technologies, enhancing their confidence in the capabilities and limitations of these solutions. Trust is built when clinicians clearly understand the evidence supporting AI algorithms and can make informed decisions about their adoption and implementation [61].

Furthermore, EBP emphasizes translating research evidence into practice, including implementing AI algorithms in real healthcare settings. By systematically evaluating the impact of AI systems on patient outcomes, clinical workflows, and healthcare delivery, clinicians can witness firsthand the effectiveness and safety of these technologies. This practical evaluation helps build trust by demonstrating AI solutions' real-world value and benefits. The culture of continuous learning promoted by EBP ensures that clinicians stay updated with the latest research evidence in the rapidly evolving field of medical AI. Clinicians can effectively evaluate emerging AI technologies by staying informed about new discoveries, advancements, and best practices. This continuous learning fosters trust by enabling clinicians to make informed decisions about adopting and applying AI solutions, knowing they are keeping pace with the latest scientific evidence [61].

By incorporating the framework of EBP into medical AI research and practice, stakeholders can establish a strong foundation of scientific evidence and foster trust among clinicians. The emphasis on rigorous research design, evidence synthesis, critical appraisal, practical implementation, and continuous learning instils confidence in AI solutions' reliability, accuracy, and value, ultimately enhancing trust and acceptance in the healthcare community [61].

5.4 P4: Leverage Better Communication Between Developers and Clinicians

The importance of effective communication between clinicians and researchers was highlighted during the interviews as a significant challenge. This section delves into potential solutions, both technical and non-technical, to address this communication gap.

Improved communication between clinicians and developers is essential for leveraging better medical AI solutions. Proper communication helps understand healthcare data, which can be challenging to interpret without the right context. Metadata alone may not be sufficient, and effective communication helps comprehend the data's context and its effective utilization in training, testing, and validating AI models [10].

The collaboration between clinicians and developers offers significant benefits for medical AI development and application. Clinicians possess extensive domain expertise and knowledge about patient care, enabling them to share insights that help developers understand the complex clinical context. This understanding allows developers to create AI models that align with clinical workflows, address real-world healthcare challenges, and produce accurate and relevant solutions. Additionally, clinicians are crucial in guiding data collection and annotation processes. Their expertise helps identify relevant data sources, define inclusion/exclusion criteria, and label datasets for training AI models. This collaboration ensures that the collected data is comprehensive, representative, and clinically meaningful, resulting in the development of more robust AI algorithms [10].

Regular interactions and knowledge exchange between clinicians and developers facilitate AI solutions' iterative development and validation. Clinicians provide valuable feedback on initial prototypes, ensuring that the outputs of AI models are clinically relevant, interpretable, and aligned with evidence-based medicine. This iterative approach allows for algorithm refinement and continuous performance improvement over time. Moreover, clinicians' involvement in AI development helps address potential biases and ethical concerns. They can identify biases in training data or algorithmic outputs that may disproportionately impact specific patient populations. By working collaboratively, clinicians and developers can design fair and unbiased AI models, fostering trust and transparency in medical AI applications. This section will explore different strategies that can enhance communication between these two groups and examine their potential impact on the healthcare sector.

5.4.1 Using a Middleman in Medical AI Projects

One effective strategy to address communication challenges in the medical AI sector is implementing a *middleman* or intermediary facilitating improved communication. The concept of a middleman, widely utilized in large-scale projects across different industries, involves an individual or entity that acts as a liaison between the parties involved. Introducing a middleman ensures that developers and clinicians share a common understanding of the data, its context, and the objectives [10][63].

In medical AI, the middleman can play a crucial role in bridging the gap between developers and clinicians. Their primary responsibility is to facilitate clear and effective communication between these two groups. By leveraging the expertise of the middleman, potential misunderstandings or miscommunications can be minimized. They can actively facilitate discussions, mediate conflicts, and provide valuable insights to developers and clinicians. Furthermore, the middleman can aid in aligning the project's objectives with the practical requirements of medical professionals, thereby fostering a more collaborative and productive working environment [8].

Including a middleman in medical AI projects can effectively address a significant bottleneck: clinicians' time constraints. European clinicians typically possess specialized skills and primarily dedicate their time to patient care, as explored in chapter 4, leaving little room for involvement in additional research endeavours. By introducing a middleman, the burden on clinicians is significantly alleviated [10][63].

This approach not only streamlines the workflow but also maximizes the utilization of clinicians' time and expertise. Clinicians can focus on their essential role in patient care while the middleman effectively bridges the gap between them and the AI development team. Consequently, this collaboration ensures that the project benefits from the clinicians' valuable insights without imposing excessive demands on their limited time resources [10][63].

Having a middleman with knowledge in both domains can be beneficial to optimize research projects involving the medical field and AI. An ideal candidate for this role could be someone with a background in bioinformatics, medical data science, or a clinician with a strong foundation in computer science. By bringing together expertise from these diverse areas, the middleman can effectively bridge the gap between medical practitioners and AI developers, facilitating seamless communication and collaboration [8].

5.4.2 Areas for Sharing Knowledge

Knowledge sharing can further enhance better communication between developers and clinicians. Bridging the gap between the medical and data sectors requires understanding and familiarity with each other's work processes. One effective solution is to organize joint conferences that bring together medical researchers and clinicians, allowing them to share their knowledge and expertise while fostering a collaborative environment.

By bringing developers and clinicians together at these conferences, valuable insights can be exchanged, leading to a deeper understanding of each other's perspectives. Developers can gain firsthand knowledge of the challenges and intricacies of medical practice, while clinicians can learn about the potential of data-driven technologies and their applications in healthcare. These joint conferences serve as platforms for knowledge exchange, creating discussions and idea-generation opportunities. They encourage meaningful collaborations, where developers and clinicians can work together to address complex healthcare problems and develop innovative solutions [10].

In addition to knowledge sharing, these conferences can provide valuable networking opportunities. Developers and clinicians can establish connections and build relationships by bringing together professionals from both sectors. This networking aspect is crucial as it enables individuals to identify potential collaborators and experts to contact for further support, guidance, or collaboration in their respective fields. Creating a space for networking ensures that developers and clinicians are aware of the resources and expertise available to them. It promotes a sense of community and facilitates ongoing communication beyond the conference setting. By fostering these connections, developers and clinicians can leverage each other's strengths and experiences to address challenges and enhance their collaboration in the future [52].

To fully harness the potential of conferences and collaborative events, authorities must acknowledge their advantages and allocate adequate resources. While international conferences successfully promote communication and collaboration between clinicians and developers, it is important to recognize that not all healthcare regions in Norway have equal access to or participation in such events. Therefore, organizing local conferences that facilitate collaboration, meetings, and knowledge-sharing could greatly contribute to developing AI in healthcare within Norway. These conferences can foster innovation and drive progress in the field by providing a platform for local stakeholders to exchange ideas and expertise.

5.4.3 Communication Through DMS

In addition to conferences and collaborative events, another important factor that can enhance communication between clinicians and developers is the implementation of a data management system, DMS, between hospitals and affiliated research institutions.

The DMS can facilitate collaborative research projects by providing a secure environment for clinicians and developers to work together, exchange ideas, and share their expertise. Clinicians can contribute their understanding of patient care and clinical workflows, while developers can leverage their technical skills to create innovative AI solutions that address specific healthcare challenges. Furthermore, a well-implemented DMS ensures data privacy and security, complying with applicable regulations and ethical guidelines. This strengthens confidence among clinicians, patients, and researchers, promoting trust and encouraging participation in data-sharing initiatives [64].

Implementing and maintaining a DMS presents several challenges. The technical complexity of integrating systems, ensuring interoperability, and managing complex data formats requires expertise and dedicated resources. Additionally, ensuring data governance and privacy compliance in handling sensitive healthcare data is demanding. Interoperability and standardization are crucial for effective collaboration and data sharing, requiring harmonising data structures and terminologies. Introducing a DMS also involves cultural and organizational changes, necessitating efforts to overcome resistance, raise awareness, and secure stakeholder buy-in. Effective communication, training, and support are vital for fostering a collaborative culture and promoting data sharing within the DMS implementation [64].

Despite these challenges, organizations can overcome them by developing a comprehensive implementation strategy that involves all stakeholders. Proactive efforts to address technical, organizational, and cultural barriers are crucial. By doing so, a DMS can fully realize its potential in fostering effective communication and collaboration between clinicians and developers. This, in turn, can advance healthcare through AI and improve patient care and outcomes.

CHAPTER

SIX

DISCUSSION

The following discussion section will explore the implications of the study's results, drawing connections between the findings and their broader significance. The study's limitations will also be thoroughly examined, shedding light on potential weaknesses or constraints that may have influenced the outcomes. Furthermore, a critical evaluation of the study's chosen methodology will be provided, carefully analyzing its strengths and weaknesses regarding reliability, validity, and generalizability. Additionally, roadmaps for future research will be explored, identifying areas that require further investigation and proposing potential directions to build upon the study's findings. This analysis aims to provide a well-rounded understanding of the study's implications, limitations, methodology, and future work.

6.1 Implication of Research

The following section elicits the research's various implications, both theoretical and practical. The research implications elaborate on how and why the conducted research is important for the research field and what it adds.

6.1.1 Theoretical Implications

Theoretical implications are essential in providing valuable insights into how the findings and propositions derived from a study align with existing ideas and concepts within the research field. The following subsection will explore the theoretical implications and their interconnectedness with other relevant theories and concepts in the research field of AI in the healthcare sector, with breast cancer detection as an illustrative example.

The study reveals a prevalent challenge interviewees face and corroborates previous research studies in identifying the difficulty of accessing and managing large volumes of data. This issue aligns with the findings of subsection 4.1.1, emphasising the substantial challenge of data access and volume in developing medical AI solutions. It is important to note that these challenges extend beyond practical limitations and carry significant theoretical implications. Thus, there is an urgent imperative for enhanced regulations and policies in this domain to foster more progress in medical AI research.

Research on data scarcity in mammography and breast cancer detection has contributed to the theoretical foundations of data analysis and access. By exploring this field, the study enhances the understanding of the challenges of limited data availability. The findings also provide valuable insights into developing more effective strategies and solutions. Two innovative approaches have been suggested for addressing data scarcity issues: synthetic data introduced in section 5.1 and exploring DMS between hospitals and research institutions discussed in section 5.4. These insights can guide the design and implementation of initiatives to bridge the gap between data availability and research requirements. Such steps will foster progress in the field, ensuring the availability of more scientific evidence.

Another focus point in the study was data bias in medical AI models, which emerged as a recurring theme during the interviews. This finding aligns with the research conducted by Wang and Preininger (2019), who also emphasized the significance of data bias when discussing the application of AI in healthcare [65]. By highlighting the prevalence of data bias and its impact on medical AI models, the study contributes to a broader understanding of the challenges associated with implementing AI in the healthcare sector. The study examines various techniques to mitigate data bias in various forms, aligning with findings from [52] and [53].

Furthermore, this study focuses primarily on the challenges encountered while implementing AI technology in the context of breast cancer detection. By centring on this specific aspect, the research contributes to a better understanding of the barriers that impede the widespread adoption of AI in healthcare. This finding aligns with previous studies, such as the work by van Leeuwen et al. (2021), which highlight that the challenges predominantly arise during the implementation phase rather than being inherent limitations of the technology itself [6].

However, it is important to emphasize that effective communication and collaboration between the AI and medical sectors are crucial for further advancements in the field. The study extends beyond the recognition of policy compliance and addresses the need for continuous development in the healthcare sector. By fostering a strong feedback loop and open communication channels, stakeholders can stay informed about new advancements, emerging challenges, and evolving needs in the medical field. This allows for the timely integration of AI solutions tailored to meet specific requirements and align with the dynamic nature of healthcare practices. Moreover, robust communication channels facilitate knowledge sharing, enabling researchers, developers, and healthcare professionals to exchange insights, best practices, and lessons learned. Possible solutions to enhance communication were introduced in section 5.4 and have theoretical implications for organising research projects. Further, the collaborative approach promotes synergies between technology and medicine, ultimately driving innovation, improving patient outcomes, and shaping the future of AI in healthcare.

One aspect of the study that unveils new research possibilities and holds substantial theoretical implications is its focus on the ability to adapt to changes in the healthcare sector. The ever-evolving nature of healthcare is continuous innovation and effective integration of emerging technologies. The study contributes to the theoretical understanding of technology adoption and integration within the healthcare domain by investigating how AI solutions can adapt to these changes. Specifically, the study highlights the importance of adaptability in AI solutions based on mammography, particularly in response to advancements in imaging techniques.

Constant advancements in imaging technologies and techniques occur in medical imaging, including mammography. Consequently, developing AI solutions that adapt effectively to these changes is crucial to maintain their relevance and efficacy. The study emphasizes the significance of adaptability by underscoring the need for AI systems to stay up-to-date with evolving imaging techniques in mammography. This recognition highlights the necessity for continuous research and development to enable AI solutions to process and analyze data from the latest imaging methodologies effectively.

6.1.2 Practical Implications

The study's propositions and findings have significant practical implications requiring further exploration and development. By elaborating on these implications, a clearer understanding emerges regarding the potential outcomes and applications that arise from implementing the study's recommendations.

One notable practical implication revolves around regulating AI usage in healthcare, particularly when synthetic data is employed to address the scarcity of real-world data. The study emphasizes the necessity of comprehensive regulatory frameworks that effectively address the challenges and considerations associated with implementing AI in healthcare settings. Such regulations should prioritize patient safety and confidentiality while fostering innovation. To achieve this, close collaboration between regulatory bodies, healthcare providers, and technology developers becomes essential in establishing clear guidelines and standards for developing, deploying, and monitoring AI systems in healthcare [8].

Another practical implication of this study is the importance of promoting knowledge-sharing between clinicians and developers. Effective collaboration between these two domains is crucial to ensure that AI systems are tailored to meet healthcare professionals' and patients' needs and requirements. Initiatives should be undertaken to facilitate communication, foster the exchange of expertise, and bridge the gap between clinical knowledge and technological advancements. By fostering such collaboration, the development and implementation of AI systems in healthcare can be enhanced, ultimately leading to improved patient outcomes [10].

Furthermore, this collaboration can potentially drive advancements in interdisciplinary research and collaborations. Breaking down the barriers between various disciplines allows for fully harnessing AI's potential in healthcare settings. Collaboration and integration among different fields, such as medicine, computer science, and data analytics, can lead to more effective and efficient utilization of AI technologies. A possible solution, introduced in section 5.4, is the introduction of a middleman to facilitate better communication and collaboration. The practical implications, therefore, revolve around how the middleman can be introduced in the sector and employed in various research projects. This interdisciplinary approach enables a holistic understanding of complex healthcare challenges and fosters innovative solutions that benefit patients and healthcare providers alike [10].

6.2 Results and Propositions

The propositions and presentation of results in this section directly address the research questions posed in chapter 1. The results shed light on the diverse challenges encountered in developing and implementing AI in the healthcare sector, including emerging challenges that have not yet had a significant impact. Primarily, the challenges identified in response to *RQ1* and *RQ2* pertain to data acquisition and meaningful utilization, as highlighted by multiple interviewees. Additionally, concerns regarding ethics and communication between different sectors impact the solutions proposed and the scientific evidence generated by AI tools. In response to *RQ3*, the propositions use synthetic data to address technical and ethical challenges. It should be noted that policies and regulations related to synthetic data are left as potential future work. Moreover, the propositions also explore non-technical solutions to mitigate technical challenges, which will be further discussed in the subsequent section.

The first proposition suggests exploring synthetic data in medical AI models. Synthetic data can help address the problem of limited diversity in medical datasets, which can affect the accuracy and reliability of AI models. Researchers can create more diverse and representative datasets for training AI models by generating synthetic data that resembles real patient data. However, there are challenges in accurately capturing the complexity of real-world data and avoiding introducing new biases during the data generation process.

The second proposition focuses on mitigating bias in medical AI models. Bias can arise from various sources, leading to unfair or inequitable healthcare outcomes. Addressing bias in AI models is crucial to ensure fairness. However, it's challenging to identify and mitigate biases, as they can be subtle and deeply ingrained in the data and algorithms. Developing effective strategies to detect and mitigate bias requires collaboration and ongoing monitoring.

The third proposition emphasizes the importance of improving communication between developers and clinicians. Effective communication is vital for successfully implementing and adopting AI solutions in healthcare. Enhancing communication channels and facilitating knowledge exchange allows for bridging the gap between technical expertise and clinical requirements. However, establishing effective communication mechanisms requires effort and a willingness to understand each other's perspectives.

The fourth proposition highlights using technical frameworks, such as Evidence-Based Practice (EBP), to build trust and recognition among clinicians. Clinicians often rely on scientific evidence and established frameworks to guide their decision-making. By incorporating technical frameworks into AI research projects and promoting the generation of scientific evidence, AI solutions can gain credibility and acceptance from the clinical community. However, translating technical advancements into practical applications aligned with established frameworks can be challenging.

While the propositions presented in Table 5.0.1 offer valuable recommendations for advancing the integration of AI in healthcare, it's important to acknowledge their limitations. The propositions may not apply universally to all healthcare domains and settings, and the specific strategies and solutions required to address the challenges may vary. Additionally, the propositions provide high-level recommendations and may not fully capture the complexity of implementing AI in healthcare. Healthcare can effectively leverage AI technologies to improve patient outcomes and transform healthcare delivery by addressing these limitations and refining the propositions through further research and collaboration.

6.3 Evaluation of Method

The selection of the research design in this study was carefully considered to ensure its alignment with the study's objectives. Using semi-structured expert interviews was a well-justified choice, as it provided valuable insights from knowledgeable individuals. The research recognized that this method would enable a deeper understanding of the specific domain under investigation. Alternative data collection methods were considered, but the semi-structured expert interviews were deemed the most appropriate and effective approach for gathering the desired insights.

Developing the interview guide was crucial to gain as much insight and knowledge from the chosen experts as possible. The study's interview guide was carefully designed to cover relevant topics and developed alongside the supervisor. The questions were well-structured and open-ended, allowing participants to provide in-depth responses. The interview guide managed to capture the nuances and complexities of the research topic, enabling the exploration of expert opinions and their respective views. The clarity, structure, and relevance of the interview guide contributed to the success of the data collection process.

The study aimed to carefully select participants with extensive expertise in the research field by leveraging the supervisor's network and the connections of other interviewees. This approach greatly facilitated the identification of suitable participants, enhancing the diversity and richness of the data. Efforts were made to include experts from various backgrounds, organizations, and geographic regions. However, some potential interviewees could not participate due to time constraints, suggesting the need to allocate more time to future recruitment processes. Ethical considerations were prioritized, with informed consent obtained from all participants and detailed information provided regarding the study's purpose, participants' roles, and time commitments. Participants were assured of their right to withdraw, demonstrating respect for their autonomy. These ethical aspects underscored the researchers' commitment to conducting the study with integrity and prioritizing the well-being and rights of the participants.

A total of seven interviews were conducted with carefully selected participants. Each interview lasted approximately one hour, comprehensively exploring the research topic and yielding intriguing findings. The insights obtained from these interviews were meticulously coded and analyzed to synthesize the findings and incorporate a deep understanding of the subject matter into the thesis. To ensure accuracy, all interviews were audio-recorded and subsequently transcribed. Additionally, interviewees who requested access to the transcriptions were provided with copies, further promoting transparency and inclusiveness in the research process.

Transcribing and coding the interviews required attention to detail to ensure precise execution. Any uncertainties were promptly addressed through email communication, allowing for timely clarification. In the subsequent data analysis phase, the foremost goal was to accurately identify and extract the relevant themes. They were discussed with the supervisor further to ensure the validity and significance of the themes. The methodology employed in this study successfully generated valuable insights and contributed to advancing knowledge in the field.

6.4 Limitations of Study

Conducting research, particularly in the context of a master's thesis, can present several limitations. The study can be further understood by addressing these limitations, and its results can be interpreted more precisely. The following section aims to identify the limitations associated with the study and how they might impact the research outcomes.

6.4.1 Limited Sample Size

An example of a challenge of the semi-structured expert interview method is the limited sample size. In this case, sample size refers to the number of participants in the study [41]. A small sample size can limit the scope of the research. It may result in a biased or incomplete understanding of the research topic, leading to challenges when drawing meaningful conclusions from the data. Further, the small sample size affects the generalizability of the study because the sample size is too narrow to portray the whole public [40].

When choosing the data generation method, acknowledging the implications of limited sample size on the research is crucial. However, it is important to note that the objective of this study is not to generalise findings to a broader population but rather to capture the unique insights and perspectives of the experts regarding the research topic [40]. Consequently, the collected data is characterised by depth and richness, enabling the research to be flexible and shaped by the knowledge and insight derived from the interviews. This approach allows for a more comprehensive understanding of complex topics that may not be achievable through generalised data [41].

6.4.2 Participant Bias

The first type of bias included in this study is *selection bias*, which refers to selecting participants for a study. As previously examined in subsection 3.4.1, selecting participants is essential in any research endeavour. In this particular study, the participants comprised three researchers, one doctor, two data owners, and two developers. However, it would have been advantageous to involve more doctors to obtain a more comprehensive perspective from the medical community regarding the research topic. Therefore, the selection bias is associated with the limited representation of doctors among the study participants.

In the book *Researching Information Systems and Computing* (2006) by Briony Oates, it was explained that *"Your age, sex, ethnic origin, accent and status can also all influence the respondents when they decide what information to give you"* [41, p. 188]. Even though some aspects of yourself are not changeable, it is important to be aware of such biases. The book also reflects upon the researcher's behaviour, to be patient, friendly and inviting when conducting the interviews and to facilitate an open and respectful environment. This can urge expert participants to share their knowledge, further elevating the research [40][41].

It is crucial to acknowledge that interviewees' responses in semi-structured expert interviews can be influenced by personal opinions or socially desirable perspectives, leading to *response bias*. This bias presents a significant limitation when using this interview method. Therefore, it is essential to approach data coding and analysis carefully, identifying recurring arguments or themes shared by multiple experts while considering perspectives that may differ from the research topic. However, it is important to note that valuable and insightful data should not be dismissed solely based on responses that do not align with the expert's opinion. Encouraging participants to share different perspectives and aspects of themselves can contribute to a more comprehensive understanding of the research topic [41].

6.4.3 Researcher Bias

One aspect to consider in research is the potential impact of the researcher's perspectives and opinions on the analysis, often referred to as *confirmation bias*. This bias can lead to favouring findings that align with the researcher's preconceived notions or desired outcomes, potentially hindering the exploration of new and diverse results. To address this limitation, it is imperative for researchers to be aware of their preconceived notions before conducting interviews and to approach the analysis phase with a clear and open mind. Researchers can promote objectivity and ensure a more comprehensive and unbiased data analysis by consciously setting aside personal biases and being receptive to alternative perspectives and unexpected findings. This approach enhances the rigour and validity of the research, allowing for a more nuanced understanding of the topic at hand [66][67].

Another crucial aspect that can influence the findings of a study is the presence of *question-order bias* and *leading questions and wording bias*. Question-order bias occurs when the sequence of questions influences respondents' answers, potentially altering the data. Leading questions and wording bias, on the other hand, occur when the phrasing or wording of questions subtly guide or manipulate respondents towards certain desired outcomes, thereby compromising the objectivity of the research [66][67].

To address these biases, it is important to carefully craft a comprehensive and well-suited interview guideline that remains consistent across all interviews. This includes avoiding leading questions that may predispose respondents to certain responses and being mindful of the wording used to ensure neutrality and impartiality. Researchers can promote more accurate and unbiased data collection by employing a standardized approach in the interview process. This, in turn, strengthens the validity and reliability of the study's findings and allows for a more trustworthy interpretation of the results [66][67].

6.4.4 Time Constraint

The study was conducted within a 20-week timeframe, which imposed time constraints on various aspects of the research, including data collection, analysis, and the final write-up. The limited duration of the master's thesis adds additional consideration regarding time management. Firstly, with a more extended timeline, additional semi-structured expert interviews could have been conducted, allowing for a broader range of perspectives and potentially enhancing the overall depth and comprehensiveness of the findings. Here, the focus could have been to involve more clinicians to get a broader perspective of the research field. This would have provided a richer and more nuanced understanding of the research topic [41][40].

Moreover, a longer duration would have allowed for a more comprehensive analysis of the gathered data. The transcription and coding process could have been conducted more thoroughly, ensuring no valuable insights or details were overlooked. This would have potentially yielded a more refined and detailed interpretation of the interview data. Additionally, with more time, there could have been an opportunity to incorporate supplementary research methods or approaches to validate further and triangulate the findings. This might have involved conducting surveys, focus groups, or exploring additional data generation methods to enhance the study's validity [41][40].

Furthermore, a more extended timeframe would have allowed for a more extensive literature review, enabling a deeper exploration and synthesis of existing theories, concepts, and empirical evidence. This would have provided a stronger theoretical foundation for the study and facilitated a more comprehensive research problem analysis. Lastly, with increased time availability, the write-up and revision process could have been more thorough, allowing for meticulous editing, proofreading, and thesis refinement. This would have resulted in a more polished and cohesive final document [41][40].

6.4.5 Transferability

When conducting qualitative research, transferability serves a similar purpose as external validity does in quantitative research. Transferability in qualitative research refers to the extent to which the findings could be applied to other contexts or settings beyond the immediate research environment. In this study, comprehensive information was provided about the characteristics of the participants, the context in which the interviews took place, and the criteria used for participant selection. By ensuring transparency and clarity, the relevance and applicability of the findings to other settings or populations were better assessed. While these details enhance the transparency and understanding of the study's findings within its specific context, they also limit the generalizability and transferability of the results to different settings or populations. Factors such as the unique characteristics of the participants and the specific context in which the interviews took place may restrict the applicability of the findings to other contexts [68].

Additionally, the study's methodology and analytical approach may present challenges to transferability. The interpretations and conclusions drawn from the qualitative data may be influenced by the researcher's subjectivity and context-specific factors. These subjective elements can make it difficult to directly transfer the findings to other settings or populations, as they may not fully capture the nuances and intricacies of different contexts [68].

6.4.6 Replicability

Replicability is important in establishing the reliability and consistency of research findings. In the context of this study, replicability involves thoroughly documenting the research process and methods employed, enabling other researchers to recreate or duplicate the study. This included providing explicit guidelines for conducting semi-structured expert interviews and outlining the steps of the coding and analysis of the interviews. Other researchers could replicate the study by offering a comprehensive and transparent account of the research process, thus validating and building upon the findings [69].

One challenge in achieving replicability within qualitative studies is the subjective nature of data collection and analysis. Qualitative research often involves interpretive and context-dependent elements that may vary between researchers. Factors such as the interviewer's style, participant responses, and the researcher's interpretation of the data can all influence the findings. As a result, replicating the study exactly as it was conducted becomes difficult, requiring capturing these subjective nuances accurately [41][69].

Additionally, qualitative studies focus on specific contexts, populations, or timeframes, limiting the findings' generalizability. Replicating the study in a different context or with different participants may yield different results, further complicating the replicability process [41][69].

6.4.7 Validation of Results

Validation of results is a crucial aspect of qualitative research to assess the accuracy, integrity, and trustworthiness of the findings. This study used various techniques to validate the results, including member checking and peer debriefing. Member checking involved allowing participants to review and confirm the accuracy of their responses, while peer debriefing sought insights and perspectives from other researchers to validate data interpretations and conclusions. These validation techniques ensured the rigour and credibility of the study's findings [41][40].

However, it is important to acknowledge that the nature of this study being part of a master's thesis and conducted within a limited timeframe poses challenges for directly validating the results. Due to the constraints of time and resources, it may be difficult to conduct extensive follow-up or verification processes to validate the findings in a traditional sense. Thus, while the study employed appropriate validation techniques within its limitations, the inability to directly validate the results is a notable constraint [41][40].

6.5 Evaluation of Future Work

Synthetic data in healthcare shows great potential to address privacy concerns and enhance data accessibility. However, it is crucial for future research to thoroughly examine the limitations and potential risks associated with synthetic data. This includes investigating its ability to accurately represent the complexity and nuances of real patient data while also addressing biases and distortions. Moreover, ethical considerations such as informed consent and patient privacy must be carefully examined within synthetic data usage [46][50].

Another critical aspect that needs attention is establishing a transparent legal framework governing AI in healthcare. As the field advances, it becomes increasingly important to establish clear guidelines and regulations for the development, deployment, and ethical use of AI technologies. Future research should focus on assessing the legal landscape, identifying gaps or ambiguities, and proposing strategies to enhance transparency and accountability. This includes data protection, algorithmic transparency, liability, and patient consent. Furthermore, exploring the international dimension of AI in healthcare and considering the harmonization of legal frameworks across jurisdictions is essential [8].

To successfully integrate AI into healthcare, improved collaboration between developers and clinicians is necessary. Close cooperation and communication between these two groups are vital to ensure that AI technologies meet the practical needs of healthcare professionals and align with patient-centred care. Future work should explore effective models of collaboration, mechanisms for fostering interdisciplinary understanding, and methods for incorporating clinician feedback into the development process. Additionally, research should address potential challenges and barriers to collaboration, such as language differences, knowledge gaps, and cultural variations between developers and clinicians [10][50].

Furthermore, future research should extend beyond technical aspects and encompass social considerations. Examining the legal frameworks and regulatory environments surrounding AI in healthcare is crucial, considering ethical, legal, and societal implications [8]. This includes investigating algorithmic fairness and bias issues and ensuring AI systems promote equity and do not perpetuate existing healthcare disparities. Additionally, understanding patient perspectives and building their trust in AI systems will be essential for successful adoption. This involves investigating factors contributing to patient trust, methods for enhancing transparency, and strategies for addressing concerns and building confidence in AI-enabled healthcare [10].

CHAPTER

SEVEN

CONCLUSIONS

This thesis aims to comprehensively understand the challenges in developing and implementing AI solutions in the healthcare sector, specifically focusing on breast cancer detection in mammograms. Throughout the research process, the author engaged in critical reflection and gained valuable insights, leading to a deeper comprehension of this complex subject.

The study employed multiple case studies to explore the complexity of AI development and implementation in healthcare. A total of seven interviews spread over four stakeholder groups, were held to collect data. These interviews provided diverse perspectives from key stakeholders, including researchers, developers, data owners, and clinicians, enabling a more holistic view of the challenges.

Several key themes emerged through the transcription and coding of the interviews, resonating throughout the discussions. These themes contain data-related challenges, regulatory compliance, the hierarchical structure of healthcare systems, trust between researchers and clinicians, and the interviewees' visions for the future of AI in healthcare. These findings expanded the study's knowledge and triggered critical reflection, deepening the understanding of the subject matter.

The study used breast cancer detection through mammograms as a concrete example to contextualize the challenges while recognizing their broader applicability within the healthcare sector. This analysis prompted thoughtful contemplation on how AI models can adapt to emerging technologies, such as the transition from 2D to 3D imaging. It stimulated consideration of the far-reaching implications of these challenges for the future of AI in healthcare.

Further, the propositions presented in this study were not abstract concepts but resulted from synthesising the interview findings, relevant literature, and existing works. These propositions emphasized the importance of adopting holistic approaches considering technical and non-technical solutions to promote sustainable development and responsible AI implementation in healthcare.

Four key propositions were formulated based on the research:

P1 Leveraging synthetic data: Harnessing the potential of synthetic data in medical AI models to enhance dataset diversity, improve model robustness, and increase accuracy.

P2 Mitigating bias: Implementing effective techniques to mitigate bias in medical AI models to reduce disparities and promote fairness. Emphasizing the integration of fairness goals from the project's inception is crucial.

P3 Communication and collaboration: Fostering improved communication between developers and clinicians to facilitate the successful implementation and adoption of AI solutions in healthcare. Strategies such as utilizing middlemen, organizing conferences and knowledge-sharing events, and establishing dedicated communication management systems are proposed.

P4 Trust-building through evidence-based practice: Utilizing technical frameworks, such as Evidence-Based Practice (EBP), to build trust and credibility among clinicians. By promoting scientific evidence in research projects and driving advancements, clinicians can become more receptive to AI solutions.

It is important to recognize and address the study's limitations and methodology to ensure the findings' credibility, validity, and replicability. This includes acknowledging the impact of the 20-week timeframe and limited resources on the study's comprehensiveness. Incorporating quantitative approaches would greatly enhance the validity of future results. Despite the study's limitations, such as small sample sizes, variability, and time constraints, the research propositions provide valuable recommendations for advancing the integration of AI in medical settings. These propositions serve as a roadmap for guiding future research and development efforts in AI in healthcare. Key areas that should be explored include expanding understanding of data accessibility and quality, developing a transparent legal framework, fostering collaboration between developers and clinicians, addressing limitations and ethical implications of synthetic data, and considering the social aspects surrounding AI adoption. Investigating these areas will help overcome challenges and ensure the effective integration of AI technologies, ultimately improving patient outcomes and healthcare delivery.

The study provides theoretical and practical implications for using AI in healthcare, specifically breast cancer detection. The theoretical implications shed light on aligning the study's findings with existing concepts in AI, addressing challenges such as data access, management, bias, and barriers to implementation. It contributes to the theoretical foundations by offering insights into strategies to overcome data scarcity, mitigate bias, and adapt to evolving imaging techniques. Additionally, the study emphasizes the importance of effective communication and collaboration between the AI and medical sectors for continuous development and innovation. On the practical side, the study underscores the need for comprehensive regulations, knowledge-sharing between clinicians and developers, and interdisciplinary collaborations to ensure responsible and ethical implementation of AI in healthcare. By embracing these implications, researchers, developers, and healthcare professionals can work together to shape the future of AI in healthcare, ultimately improving patient outcomes and advancing the field.

In conclusion, this research study has provided valuable insights into the challenges associated with AI in healthcare by using breast cancer detection as an illustrative example. The findings contribute to a better understanding of the complexities involved and will guide future efforts to advance AI solutions in the healthcare sector. This study adds to the ongoing conversation about the responsible, ethical, and sustainable development and implementation of AI in healthcare, driving progress and ensuring the well-being of patients.

REFERENCES

[1] *Artificial Intelligence In Healthcare Market Size, Share, And Trends Analysis Report By Component (Software Solutions, Hardware, Services), By Application (Virtual Assistants, Connected Machines), By Region, And Segment Forecasts, 2022 - 2030*. Engelsk. URL: https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-healthcare-market (visited on 09/29/2022).

[2] Kurt Benke and Geza Benke. "Artificial intelligence and big data in public health". In: *International journal of environmental research and public health* 15.12 (2018), p. 2796. ISSN: 1660-4601.

[3] *Brystkreft*. no. URL: https://www.kreftregisteret.no/Temasider/kreftformer/Brystkreft/ (visited on 04/27/2023).

[4] Helsedirektoratet. *Brystkreft*. no. URL: https://www.helsedirektoratet.no/nasjonale-forlop/brystkreft (visited on 06/10/2023).

[5] Shigao Huang et al. "Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges". In: *Cancer letters* 471 (2020), pp. 61–71. ISSN: 0304-3835.

[6] Kicky G. van Leeuwen et al. "Artificial intelligence in radiology: 100 commercially available products and their scientific evidence". eng. In: *European Radiology* 31.6 (June 2021), pp. 3797–3804. ISSN: 1432-1084. DOI: 10.1007/s00330-021-07892-z.

[7] Asmaa Ibrahim et al. "Artificial intelligence in digital breast pathology: techniques and applications". In: *The Breast* 49 (2020), pp. 267–273. ISSN: 0960-9776.

[8] Stacy M. Carter et al. "The ethical, legal and social implications of using artificial intelligence systems in breast cancer care". en. In: *The Breast* 49 (Feb. 2020), pp. 25–32. ISSN: 0960-9776. DOI: 10.1016/j.breast.2019.10.001. URL: https://www.sciencedirect.com/science/article/pii/S0960977619305648 (visited on 11/29/2022).

[9] Daniel Schönberger. "Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications". In: *International Journal of Law and Information Technology* 27.2 (2019), pp. 171–203. ISSN: 0967-0769.

[10] Steven Lin. "A Clinician's Guide to Artificial Intelligence (AI): why and how primary care should lead the health care AI revolution". In: *The Journal of the American Board of Family Medicine* 35.1 (2022), pp. 175–184.

[11] Nariman Noorbakhsh-Sabet et al. "Artificial intelligence transforms the future of health care". In: *The American journal of medicine* 132.7 (2019), pp. 795–801.

[12] Andreas Holzinger et al. "Causability and explainability of artificial intelligence in medicine". In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 9.4 (2019), e1312. ISSN: 1942-4787.

[13] *BreastScreen Norway*. en. URL: `https://www.kreftregisteret.no/en/screening/BreastScreen_Norway/breastscreen-norway/` (visited on 12/11/2022).

[14] Sai Batchu et al. "A review of applications of machine learning in mammography and future challenges". In: *Oncology* 99.8 (2021), pp. 483–490. ISSN: 0030-2414.

[15] Kenney Ng et al. "Personalized treatment options for chronic diseases using precision cohort analytics". en. In: *Scientific Reports* 11.1 (Jan. 2021). Number: 1 Publisher: Nature Publishing Group, p. 1139. ISSN: 2045-2322. DOI: `10.1038/s41598-021-80967-5`. URL: `https://www.nature.com/articles/s41598-021-80967-5` (visited on 12/11/2022).

[16] Stephanie Price. *How smart monitoring technologies and AI optimise patient safety*. en-GB. Section: Research & Innovation News. May 2021. URL: `https://www.healtheuropa.com/how-smart-monitoring-technologies-and-ai-optimise-patient-safety/107790/` (visited on 12/11/2022).

[17] Marina Tsoumpa et al. "The Use of the Hypotension Prediction Index Integrated in an Algorithm of Goal Directed Hemodynamic Treatment during Moderate and High-Risk Surgery". In: *Journal of Clinical Medicine* 10.24 (Dec. 2021), p. 5884. ISSN: 2077-0383. DOI: `10.3390/jcm10245884`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8707257/` (visited on 12/11/2022).

[18] Ravi B. Parikh, Stephanie Teeple, and Amol S. Navathe. "Addressing Bias in Artificial Intelligence in Health Care". In: *JAMA* 322.24 (Dec. 2019), pp. 2377–2378. ISSN: 0098-7484. DOI: `10.1001/jama.2019.18058`. URL: `https://doi.org/10.1001/jama.2019.18058` (visited on 06/09/2023).

[19] Annika M. Svensson and Fabrice Jotterand. "Doctor ex machina: A critical assessment of the use of artificial intelligence in health care". In: *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine*. Vol. 47. Oxford University Press US, 2022, pp. 155–178.

[20] *Breast Cancer Screening (PDQ®)–Patient Version - NCI*. en. pdq-CancerInfoSummary. Archive Location: nciglobal,ncienterprise. Oct. 2022. URL: https://www.cancer.gov/types/breast/patient/breast-screening-pdq (visited on 12/11/2022).

[21] *Anatomy of the Breast | Memorial Sloan Kettering Cancer Center*. en. URL: https://www.mskcc.org/cancer-care/types/breast/anatomy-breast (visited on 04/27/2023).

[22] *Risk Factors*. en-US. URL: https://www.nationalbreastcancer.org/breast-cancer-risk-factors (visited on 05/08/2023).

[23] *Breast Cancer Prevention: Aromatase Inhibitors*. en. URL: https://www.cancer.org/cancer/types/breast-cancer/risk-and-prevention/aromatase-inhibitors-for-lowering-breast-cancer-risk.html (visited on 06/06/2023).

[24] *Survival Rates for Breast Cancer*. en. URL: https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html (visited on 06/10/2023).

[25] Jacek Lorkowski, Oliwia Kolaszyńska, and Mieczysław Pokorski. "Artificial intelligence and precision medicine: A perspective". In: *Integrative Clinical Research*. Springer, 2021, pp. 1–11.

[26] *What Is Machine Learning and Why Is It Important?* en. URL: https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML (visited on 12/11/2022).

[27] Yogesh Kumar et al. "A Systematic Review of Artificial Intelligence Techniques in Cancer Prediction and Diagnosis". en. In: *Archives of Computational Methods in Engineering* 29.4 (June 2022), pp. 2043–2070. ISSN: 1886-1784. DOI: 10.1007/s11831-021-09648-w. URL: https://doi.org/10.1007/s11831-021-09648-w (visited on 11/29/2022).

[28] Enes Zvornicanin. *How to Design Deep Convolutional Neural Networks? | Baeldung on Computer Science*. en-US. Oct. 2021. URL: https://www.baeldung.com/cs/deep-cnn-design (visited on 06/06/2023).

[29] *Overview of GAN Structure | Machine Learning*. en. URL: https://developers.google.com/machine-learning/gan/gan_structure (visited on 06/06/2023).

[30] Gaudenz Boesch. *Vision Transformers (ViT) in Image Recognition - 2023 Guide*. en-US. Feb. 2023. URL: https://viso.ai/deep-learning/vision-transformer-vit/ (visited on 06/06/2023).

[31] Constance D. Lehman et al. "National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium". eng. In: *Radiology* 283.1 (Apr. 2017), pp. 49–58. ISSN: 1527-1315. DOI: 10.1148/radiol.2016161174.

[32] *AI Shows Potential in Breast Cancer Screening Programs*. URL: https://www.rsna.org/news/2022/march/AI-Potential-in-Breast-Screening (visited on 06/06/2023).

[33] Scott Mayer McKinney et al. "International evaluation of an AI system for breast cancer screening". en. In: *Nature* 577.7788 (Jan. 2020). Number: 7788 Publisher: Nature Publishing Group, pp. 89–94. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1799-6. URL: https://www.nature.com/articles/s41586-019-1799-6 (visited on 06/06/2023).

[34] "AI 'outperforms' doctors diagnosing breast cancer". en-GB. In: *BBC News* (Jan. 2020). URL: https://www.bbc.com/news/health-50857759 (visited on 06/06/2023).

[35] Marthe Larsen et al. "Artificial Intelligence Evaluation of 122969 Mammography Examinations from a Population-based Screening Program". In: *Radiology* 303.3 (June 2022). Publisher: Radiological Society of North America, pp. 502–511. ISSN: 0033-8419. DOI: 10.1148/radiol.212381. URL: https://pubs.rsna.org/doi/10.1148/radiol.212381 (visited on 06/06/2023).

[36] Natalia Norori et al. "Addressing bias in big data and AI for health care: A call for open science". en. In: *Patterns* 2.10 (Oct. 2021), p. 100347. ISSN: 2666-3899. DOI: 10.1016/j.patter.2021.100347. URL: https://www.sciencedirect.com/science/article/pii/S2666389921002026 (visited on 06/06/2023).

[37] Yuri Y. M. Aung, David C. S. Wong, and Daniel S. W. Ting. "The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare". eng. In: *British Medical Bulletin* 139.1 (Sept. 2021), pp. 4–15. ISSN: 1471-8391. DOI: 10.1093/bmb/ldab016.

[38] Li Shen et al. "Deep learning to improve breast cancer detection on screening mammography". In: *Scientific reports* 9.1 (2019), pp. 1–12. ISSN: 2045-2322.

[39] *Moffitt's Position on Breast Cancer Screening Guidelines*. en-US. URL: https://moffitt.org/for-healthcare-professionals/clinical-perspectives/clinical-perspectives-story-archive/moffitts-position-on-breast-cancer-screening-guidelines/ (visited on 12/11/2022).

[40] Michael D. Myers and David Avison. *Qualitative Research in Information Systems: A Reader*. en. Google-Books-ID: Oe9jkjrdFuoC. SAGE, July 2002. ISBN: 978-0-7619-6632-6.

[41] Briony J. Oates. *Researching Information Systems and Computing*. en. SAGE, 2006. ISBN: 978-1-4129-0224-3.

[42] *What is data quality? | IBM*. en-us. URL: https://www.ibm.com/topics/data-quality (visited on 04/21/2023).

[43] Aliaksei Mikhailiuk. *Deep Image Quality Assessment*. en. Sept. 2022. URL: https://towardsdatascience.com/deep-image-quality-assessment-30ad71641fac (visited on 06/21/2023).

[44] Helse-og omsorgsdepartementet. *NOU 2016: 25*. no. NOU. Publisher: regjeringen.no. Dec. 2016. URL: https://www.regjeringen.no/no/dokumenter/nou-2016-25/id2522062/ (visited on 06/09/2023).

[45] *Helse Norge (kart)*. nb-NO. May 2014. URL: https://behandlingshjelpemidler.no/norge/ (visited on 06/09/2023).

[46] Alexey Dosovitskiy et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". en. In: Jan. 2021. URL: https://openreview.net/forum?id=YicbFdNTTy (visited on 06/09/2023).

[47] Ian J. Goodfellow et al. *Generative Adversarial Networks*. arXiv:1406.2661 [cs, stat]. June 2014. URL: http://arxiv.org/abs/1406.2661 (visited on 06/09/2023).

[48] Kirill Demochkin. *Casual GAN Papers: ViTGAN Explained*. July 2021. URL: https://www.casualganpapers.com/gan-based-self-attention-vision-transformers/ViTGAN-explained.html (visited on 06/09/2023).

[49] Vajira Thambawita et al. "SinGAN-Seg: Synthetic training data generation for medical image segmentation". en. In: *PLOS ONE* 17.5 (2022). Publisher: Public Library of Science, e0267976. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0267976. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0267976 (visited on 06/09/2023).

[50] Andre Goncalves et al. "Generation and evaluation of synthetic patient data". In: *BMC Medical Research Methodology* 20.1 (May 2020), p. 108. ISSN: 1471-2288. DOI: 10.1186/s12874-020-00977-1. URL: https://doi.org/10.1186/s12874-020-00977-1 (visited on 06/09/2023).

[51] *Synthetic Data | European Data Protection Supervisor*. en. June 2023. URL: https://edps.europa.eu/press-publications/publications/techsonar/synthetic-data (visited on 06/09/2023).

[52] Jie Zhang and Zong-ming Zhang. "Ethics and governance of trustworthy medical artificial intelligence". In: *BMC Medical Informatics and Decision Making* 23 (Jan. 2023), p. 7. ISSN: 1472-6947. DOI: 10.1186/s12911-023-02103-9. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9840286/ (visited on 06/09/2023).

[53] Alvin Rajkomar et al. "Ensuring Fairness in Machine Learning to Advance Health Equity". In: *Annals of Internal Medicine* 169.12 (Dec. 2018). Publisher: American College of Physicians, pp. 866–872. ISSN: 0003-4819. DOI: 10.7326/M18-1990. URL: https://www.acpjournals.org/doi/10.7326/M18-1990 (visited on 05/30/2023).

[54] María Agustina Ricci Lara, Rodrigo Echeveste, and Enzo Ferrante. *Addressing fairness in artificial intelligence for medical imaging | Nature Communications*. June 2022. URL: https://www.nature.com/articles/s41467-022-32186-3 (visited on 04/09/2023).

[55] *The Essential Guide to Data Augmentation in Deep Learning*. en. URL: https://www.v7labs.com/blog/data-augmentation-guide,%20https://www.v7labs.com/blog/data-augmentation-guide (visited on 06/09/2023).

[56] Rohit Dwivedi. *How Data Augmentation Impacts Performance Of Image Classification, With Codes*. en-US. Aug. 2020. URL: https://analyticsindiamag.com/image-data-augmentation-impacts-performance-of-image-classification-with-codes/ (visited on 05/30/2023).

[57] Jason Brownlee. *A Gentle Introduction to the Bootstrap Method*. en-US. May 2018. URL: https://machinelearningmastery.com/a-gentle-introduction-to-the-bootstrap-method/ (visited on 06/09/2023).

[58] Andrea Gao (she/her). *Data Bias Identification and Mitigation: Methods and Practice*. en. June 2021. URL: https://medium.com/bcggamma/data-bias-identification-and-mitigation-methods-and-practice-c0640f35ff30 (visited on 06/09/2023).

[59] Sourabh Mehta. *A guide to different bias mitigation techniques in machine learning*. en-US. Apr. 2022. URL: https://analyticsindiamag.com/a-guide-to-different-bias-mitigation-techniques-in-machine-learning/ (visited on 06/09/2023).

[60] Melissa D. McCradden et al. "Ethical limitations of algorithmic fairness solutions in health care machine learning". English. In: *The Lancet Digital Health* 2.5 (May 2020). Publisher: Elsevier, e221–e223. ISSN: 2589-7500. DOI: 10.1016/S2589-7500(20)30065-0. URL: https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30065-0/fulltext (visited on 06/09/2023).

[61] Robert Newell and Philip Burnard. *Research for Evidence-Based Practice in Healthcare*. en. John Wiley & Sons, Nov. 2010. ISBN: 978-1-4443-3112-7.

[62] Peter Davis and Mark Limb. *The knowledge and evidence base for practice | Clinical Gate*. Oct. 2016. URL: https://clinicalgate.com/the-knowledge-and-evidence-base-for-practice/ (visited on 06/11/2023).

[63] *Middleman*. en-US. URL: https://corporatefinanceinstitute.com/resources/valuation/middleman/ (visited on 06/09/2023).

[64] Mary H. Stanfill and David T. Marc. "Health Information Management: Implications of Artificial Intelligence on Healthcare Data and Information Management". en. In: *Yearbook of Medical Informatics* 28.1 (Aug. 2019). Publisher: Georg Thieme Verlag KG, pp. 56–64. ISSN: 0943-4747, 2364-0502. DOI: 10.1055/s-0039-1677913. URL: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0039-1677913 (visited on 07/02/2023).

[65] Fei Wang and Anita Preininger. "AI in Health: State of the Art, Challenges, and Future Directions". en. In: *Yearbook of Medical Informatics* 28.1 (Aug. 2019). Publisher: Georg Thieme Verlag KG, pp. 16–26. ISSN: 0943-4747, 2364-0502. DOI: 10.1055/s-0039-1677908. URL: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0039-1677908 (visited on 05/30/2023).

[66] Christopher J. Pannucci and Edwin G. Wilkins. "Identifying and Avoiding Bias in Research". In: *Plastic and reconstructive surgery* 126.2 (Aug. 2010), pp. 619–625. ISSN: 0032-1052. DOI: 10.1097/PRS.0b013e3181de24bc. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2917255/ (visited on 06/13/2023).

[67] Parvathi Vijayamohan. *Biased Survey Questions: 7 Examples & How to Minimize Them*. en. July 2022. URL: https://surveysparrow.com/blog/biased-survey-questions/ (visited on 06/13/2023).

[68] Melissa Moran. *What is transferability in qualitative research and how do we establish it?* en-US. May 2017. URL: https://www.statisticssolutions.com/what-is-transferability-in-qualitative-research-and-how-do-we-establish-it/ (visited on 06/13/2023).

[69] *Replicability - Reproducibility and Replicability in Science - NCBI Bookshelf*. URL: https://www.ncbi.nlm.nih.gov/books/NBK547524/ (visited on 06/13/2023).

APPENDICES

A - INTERVIEW GUIDELINES

A1 - Interview Guidelines for Data Owners, Developers, and Researchers

Table A.1: Interview guidelines for interviewing with data owners, developers and researchers. Used as a basis for semi-structured interviews.

Introduction	
Introduction of myself	For my master thesis, I am working on a project regarding AI in the healthcare sector, focusing on detecting breast cancer. The project aims to identify inhibitors of AI use in the field and investigate the bias used in the detection methods used today. You have been asked to participate in the study as I am interested in your experience and professional knowledge regarding the healthcare sector and/or AI technology.
Practical information	The interview will last around 45 minutes and will be recorded. The interview will later be transcribed, and the recording will be deleted. The transcription will only be visible to my supervisor and me, and you can ask to delete the recording at any time.
Consent	The consent form must be signed before the interview, and the consent can be revoked by emailing me or my supervisor.
Main part of interview	

Introduction of interviewee	Can you describe your work with AI in the healthcare sector? **Keywords:** position or roles, duration of projects, type of projects/tools, areas of application, type of technologies, software/hardware, stage of development.
Development process	Can you describe the development process of the applications/tools? **Keywords:** process from idea-development-prototype-launch, other stages of development, kind of expertise used in each stage, project duration, challenges.
Relevance of project	Can you describe the project's relevance? **Keywords:** gap in knowledge, relevance, potential disruption of the sector.
Role of medical staff	Can you describe how the application/tool affects the work life and practice of medical workers? **Keywords:** impact on medical workers, advised in the development process, communication.
Identify added value	Can you describe how the application/tool adds value to the patients, medical workers, and healthcare sector? **Keywords:** illness detection accuracy, effect on patients, quality of life, resource allocation.
Identify challenges	Can you describe challenges the application/-tool faces in today's development process? **Keywords:** examples, scenarios, hospital/medical staff/patients, efforts to tackle the challenges.
Final part of interview	
Summary	Ask to clarify, if necessary, summarize the main points.
Ending	Ask if they have any questions, remarks about what I am focusing on, or contacts I can contact.

A2 - Interview Guideline for Medical Professional

Table A.2: Interview guidelines for interviewing with a medical professional. Used as a basis for semi-structured interviews.

Introduction	
Introduction of myself	In the context of my research project on advancements in medical technology, specifically AI applications in healthcare, I am conducting interviews to gather insights from experienced medical professionals like yourself. This project explores the impact and potential biases associated with AI-based breast cancer detection methods. Your expertise and professional knowledge are highly valued, and your participation in this study is greatly appreciated.
Practical information	The interview will take approximately 45 minutes and be recorded for accurate transcription. Rest assured that the recording will be deleted after transcription. The transcript will only be accessible to me and my supervisor. If you wish to have the recording deleted, please inform us at any point during or after the interview.
Consent	Please sign the consent form before proceeding with the interview. You can also revoke your consent by emailing me or my supervisor.
Main part of interview	
Introduction of interviewee	Could you please provide an overview of your involvement with AI in the healthcare sector? **Keywords:** professional role, duration of experience, specific projects or tools, application areas, utilized technologies (software/hardware), and current stage of development.
Development process	Can you describe the process involved in developing AI applications/tools within your work? **Keywords:** stages from idea conception to development, prototyping, and deployment; types of expertise required at each stage; project duration; notable challenges encountered.

Relevance of project	Please elaborate on the relevance of your project in the healthcare sector. **Keywords:** identified knowledge gaps, overall relevance, and potentially disruptive impact on the sector.
Role of medical staff	How does the application/tool you work with affect the daily work and practices of medical professionals? **Keywords:** impact on medical professionals, involvement in the development process, communication strategies.
Identifying added value	Can you explain how the application/tool you work with brings value to patients, medical professionals, and the healthcare sector? **Keywords:** accuracy in illness detection, impact on patients, improvements in quality of life, resource allocation optimization.
Identifying challenges	Please discuss the challenges currently faced by the application/tool in its development process. **Keywords:** specific examples or scenarios, impact on hospitals/medical staff/patients, ongoing efforts to address these challenges.
Final part of interview	
Summary	Seek clarification if needed and concisely summarise the key points discussed during the interview.
Ending	Ask if they have any questions, remarks about what I am focusing on, or contacts I can contact.

B - PROJECT INFORMATION

Vil du delta i forskningsprosjektet

Investigating bias in breast cancer detection methods

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å undersøke bias i deteksjonsalgoritmer og kunstig intelligens brukt for å detektere brystkreft. I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål
Prosjektet skrives i samarbeid med en masteroppgave ved NTNU i Trondheim, fakultet for informasjonsteknologi og elektronikk (IE), institutt for datateknologi og informatikk (IDI). Formålet med prosjektet vil være å få en bedre undersøke hvilke typer data som brukes i forbindelse med detektering av brystkreft av kvinner i Norge i dag, samt undersøke mulig «bias» blant disse metodene. Videre vil prosjektet undersøke hvordan implikasjoner av mulig bias påvirker pasientenes livskvalitet.

Hvem er ansvarlig for forskningsprosjektet?
Professor Patrick Mikalef ved IDI er ansvarlig for prosjektet.

Hvorfor får du spørsmål om å delta?
Du får spørsmål om å delta i masterprosjektet, fordi du befinner deg i nettverket til en av prosjektets deltakere, eller fordi du sitter på relevant kunnskap og erfaring som kan bidra til masterprosjektet.

Hva innebærer det for deg å delta?
Ved å delta i prosjektet samtykker du til å delta på et intervju, som varer i ca. 45 minutter. Intervjuet vil handle om din fagkunnskap rettet til prosjektet, og hvordan ditt fagfelt jobber med tematikken. Lyd fra intervjuet vil bli tatt opp og brukt til å transkribere intervjuet. Det vil være mulig å be om at lyd ikke tas opp under intervjuet, dersom intervjuobjektet ønsker det. Det vil til enhver tid være mulig å be om at data fra intervjuet slettes.

Det er frivillig å delta
Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle dine personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger
Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket. Det er forfatter av masterprosjektet og veileder som vil ha tilgang til dataene. Det vil ikke være mulig å gjenkjenne deltakerne i publikasjonen.

For datalagring benyttes SharePoint fra Microsoft. Informasjonen vil ligger under NTNU som organisasjon og derfor adgangsbegrenses. Mer om Microsoft sin behandling av data finnes her: https://learn.microsoft.com/en-us/sharepoint/safeguarding-your-data.

Persondata som lagres under prosjektet er:
- Navn
- Lydopptak fra intervjuet

På grunn av lydopptaket vil det bli informert om at sensitive opplysninger ikke skal deles. Lydopptaket vil bli gjort på en privat enhet og transkribert, slik at lydopptaket kan slettes. Transkriberingen vil lagres på SharePoint, som er adgangsbegrenset.

Hva skjer med personopplysningene dine når forskningsprosjektet avsluttes?

Prosjektet vil etter planen avsluttes når oppgaven avsluttes eller oppgaven godkjennes, som er planlagt til å være senest 26 juni. Etter prosjektslutt vil datamaterialet med dine personopplysninger slettes. Dette inkluderer lydopptak

Hva gir oss rett til å behandle personopplysninger om deg?
Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra IDI har Sikt – Kunnskapssektorens tjenesteleverandør vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Dine rettigheter
Så lenge du kan identifiseres i datamaterialet, har du rett til:
- innsyn i hvilke opplysninger vi behandler om deg, og å få utlevert en kopi av opplysningene
- å få rettet opplysninger om deg som er feil eller misvisende
- å få slettet personopplysninger om deg
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger

Hvis du har spørsmål til studien, eller ønsker å vite mer om eller benytte deg av dine rettigheter, ta kontakt med:
- Forfatter av masteroppgaven: Anchana Balasingham – anchanab@ntnu.no
- Veileder: Patrick Mikalef – patrick.mikalef@ntnu.no
- Vårt personvernombud: https://i.ntnu.no/wiki/-/wiki/Norsk/Personvernombud+NTNU

Hvis du har spørsmål knyttet til vurderingen som er gjort av personverntjenestene fra Sikt, kan du ta kontakt via:
- Epost: personverntjenester@sikt.no eller telefon: 73 98 40 40.

Med vennlig hilsen,

Patrick Mikalef (Forsker/veileder)   Anchana Balasingham

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet *Investigating bias in breast cancer detection methods* , og har fått anledning til å stille spørsmål. Jeg samtykker til:

- ☐ å delta i delta i et intervju

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet

----
(Signert av prosjektdeltaker, dato)

NTNU

Norwegian University of
Science and Technology