Beyond Human Limits: The Introduction of Artificial Intelligence into Screening Mammography to Improve Breast Cancer Detection

by

Justin Kalman

Bachelor of Science in Business Administration, Binghamton University, 2021

Submitted to the Graduate Faculty of the

School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Health Administration

University of Pittsburgh

2024

UNIVERSITY OF PITTSBURGH

SCHOOL OF PUBLIC HEALTH

This essay is submitted

by

Justin Kalman

on

March 22, 2024

and approved by

Narayan Ramasubbu, PhD
Professor
Katz Graduate School of Business and College of Business Administration
University of Pittsburgh

Thesis Advisor: Kevin Broom, PhD, MBA
Vice Chair of Education and Director of MHA and MHA/MBA
School of Public Health
University of Pittsburgh

Beyond Human Limits: The Introduction of Artificial Intelligence Into Screening Mammography to Improve Breast Cancer Detection

Justin Kalman, MHA

University of Pittsburgh, 2024

Abstract

Radiologists can greatly improve their ability to detect breast cancer, with the integration of artificial intelligence (AI) into the breast cancer screening process. The use of AI within screening mammography can help minimize the frequency of false negative diagnoses of breast cancer, by calling out areas of concern that may otherwise be overlooked. By maximizing the ability to detect breast cancer at an earlier stage, the patient can maximize their ability to receive treatment at a time to control if not eradicate the breast cancer. The early detection of breast cancer can potentially reduce healthcare costs by treating patients at an early stage and avoiding the costs for treatment of cancer that has metastasized.

In the United States, breast cancer is the second leading cause of cancer death in women and is estimated to kill 42,250 women in just 2024 alone (Breast Cancer Statistics, n.d.). Breast cancer can be treated effectively when detected early on, which is why the American Cancer Society (ACS) recommends that women receive a screening mammogram each year. This recommendation has helped to reduce the breast cancer death rate 43% since 1989 (Breast Cancer Statistics, n.d.). Although this is a drastic improvement, not all women are able to receive the intended benefit of a yearly screening due to an epidemic of misdiagnosis.

Upwards of 30% of mammograms are misdiagnosed, causing women to forego treatment that may ultimately save their lives. 85% of the misdiagnoses can be attributed directly to human error, and could have been avoided (Ganesan, Karthikeyan, et al., 2013). In order to reduce human

error, we need to provide greater support to radiologists and improve screening mammography. AI has the ability to diagnose patients and provide explanations simple enough for both physicians and patients to understand. While this technology is new and currently still being tested, integration is imperative to ensure that women receive treatment for breast cancer as soon as it's detectable.

Table of Contents

1.0 Introduction

In the United States the average woman has a 13% chance of developing breast cancer while approximately 1 in 40 women will die from the disease (Breast Cancer Statistics, n.d.). The American Cancer Society (ACS) recommends that all woman aged 40 and above receive a screening mammogram each year to identify breast cancer at the earliest detectable stage. This recommendation in unison with awareness campaigns and improved treatment options have reduced the breast cancer death rate by 43% since 1989 but has slowed in recent years (Breast Cancer Statistics, n.d.). Despite significant advances in medical science and treatment options, breast cancer remains the second leading cause of cancer-related deaths among women in the United States (Breast Cancer Statistics, n.d.).

Cancer can take on many forms, with varying mortality rates. Cancer can be defined simply as an uncontrolled multiplication of cells. These cells divide rapidly, which may form a lump, microcalcifications or architectural disorders which are often called tumors. Thankfully breast cancer can be treated effectively if it is detected early, which translates to lower mortality rates. This is due to the development of better diagnostic facilities and effective treatments, however, there is still a need for further improvement (Castellino, 2005).

The main line of defense against breast cancer in the United States is screening mammography. This process involves a radiologist taking an X-ray of breast tissue and then visually examining the results with the assistance of computer-aided detection (CAD). CAD software is able to analyze the X-ray image and call out areas of concern, but it is up to the radiologist to determine if cancer is suspected and what further testing is needed. It is estimated that CAD assisted mammography has reduced breast cancer mortality by 28% (Weedon-Fekjaer

et al., 2014). While this is a drastic improvement, technology has failed to see any major advancements since its conception in 1998 (Lamb et al., 2022). With millions of women suffering from breast cancer, it is imperative to improve screening technology to catch the cancer before it becomes too late.

Despite worldwide use and general acceptance, mammography has many flaws due to human factors. Although mammography is cost-effective, there is great debate about its use due to false-positive findings, false-negative findings, and over diagnosis. In many European countries, a double reading by two radiologists is used to minimize human related errors (Hickman et al., 2022). However, in the United States double readings are not as common due to staffing and cost issues that are affecting the entire healthcare industry.

The integration of more complex algorithms, utilizing AI, into the field of screening mammography will transform the idea of early detection. Radiologists are highly trained professionals who are using their vision to determine if breast cancer is present, but to be human is to err. While a radiologists' experience and training are able to give accurate readings most of the time, new AI tools are able view each case at a greater granularity than the human eye and reference a database of known breast cancer cases and predict with great certainty if cancer is present and the current stage (Lamy, Jean-Baptiste, et al., 2019). The integration of more advanced AI will enable radiologists to detect breast cancer earlier, have a second opinion to fortify their diagnosis, and create a new ability to challenge the diagnosis to ensure greater accuracy and save more lives.

This literature review will discuss the faults of current breast cancer detection and capabilities of AI to reduce, if not eliminate, screening errors. This paper will further fortify the need and benefits of new AI assisted screening mammography algorithms. Once the capabilities

and benefits are understood, it is essential to foster support from physicians and the larger healthcare organization to ensure the greatest benefit for patients and health systems is achieved. In a society obsessed with the newest and best technology, it is about time that the healthcare industry realized that AI assistance is the future of healthcare.

2.0 Methodology

The primary purpose of this literature review is to encourage further studies that focus on AI assisted screening mammography to improve the ability to detect breast cancer. A secondary goal of this literature review is to decrease the stigma and fear that comes with the integration of AI in a diagnostic setting. Literature searches were performed four times between August 4th, 2023, and November 16th, 2023, on Google Scholar and Google. The search terms on Google Scholar include "breast cancer", "breast cancer detection", "AI within healthcare", and "gaining physician buy-in". Each subsequent search I focused on filling in the gaps of this paper, without including phrasing that may lead to a potential bias. When selecting which sources to utilize, I focused on papers that were published between 2000 and 2024 to avoid any potential inaccuracies. When discussing AI, I only searched for papers published after 2015 to ensure that I had the most up to date information available. Each search had between 20,000 and 482,000 results. I manually examined the first 5 pages of results, which represent the top 50 most relevant articles based on my search criteria, to narrow down my sources and ensure the accuracy of this paper.

I additionally used Google to access articles that I have viewed during my coursework at the University of Pittsburgh. I explicitly searched for the articles that I used from Harvard Business Review, The American Cancer Association and Stat News. The articles that I used from Harvard Business Review and Stat News were used as references for recommended implementation practices and do not relate to AI or breast cancer directly. The American Cancer Association was used to provide statistics related to breast cancer and acted to provide more current statistics that older research papers were referencing.

3.0 Need for Improved Screening Methods for Breast Cancer

Breast cancer, along with all other forms of cancer, consists of out-of-control cells that rapidly replicate. It is imperative to diagnose and treat cancer as early as possible to prevent the cells from spreading to the rest of the body. When successful, a screening mammogram is able to help identify the presence of breast cancer and the appropriate interventions can be implemented by physicians. Unfortunately, current screening practices are riddled with shortcomings that lead to false negatives that delay treatment.

Interval Cancers (ICs) are a direct result of the current flaws in the screening process. ICs are detected in the interval after a planned screening mammogram that did not show signs of cancer. Some cases of ICs arise due to rapidly growing cancerous cells that were not visible on the prior screening but are then detected during the time of diagnosis. While screening mammography is unable to diagnose cancer before it is detectable, ICs are often present during the previous screening, but not seen due to poor technique or positioning (Gordon, 2022). In these cases, the cancer may have been masked by dense breast tissue or misinterpreted by the screening radiologist. The prevalence of human error makes it especially important to receive an annual screening mammography.

The need for annual breast cancer screenings and early intervention strategies are especially important for those with triple-negative breast cancer, which accounts for about 10%-15% of all breast cancers (Triple-Negative Breast Cancer, American Cancer Society, 2023). Triple-negative breast cancer refers to cancer cells that do not make any or too much of the protein HER2 and fail to have estrogen or progesterone receptors. Due to these traits, hormone therapy and other HER2 drugs are not viable treatments. Chemotherapy and surgery are often the only

options for women suffering with triple negative-breast cancer. This type of breast cancer tends to grow and spread faster than other types of breast cancer, which further increases the need for early detection (Triple-Negative Breast Cancer, American Cancer Society, 2023). The speed at which this form of cancer replicates and spreads makes it imperative to be diagnosed accurately the first time.

4.0 Detection of Breast Cancer

The leading method for diagnosing breast cancer is CAD assisted screening mammography. This method of detection involves using X-ray imaging of the breast of asymptomatic women to detect cancer as early as possible. The X-ray is then visually examined by the radiologist (Castellino, 2005). If breast cancer is detected, patients have multiple treatment options they can choose from. The options for patients become more limited as the cancer develops. Unfortunately, there is a human factor involved in the screening and detection process, which leads to a high degree of error.

Mammograms do not provide great contrast between normal glandular and malignant tissues, which leads to a high error rate. Studies have shown that radiologists screening for cancer have an error rate between 10%-30%. Of those errors, 52% can be attributed to misinterpretation of breast cancer signs and 43% of the errors are caused due to overlooking signs in abnormal scans (Ganesan, Karthikeyan, et al., 2013). This means that 85% of all misdiagnoses are attributed to human error, which could be avoided. The high error rate leads to frequent biopsies performed on benign lesions, resulting in unnecessary testing. These procedures increase patient anxiety and medical expenses that could have been avoided if the X-ray was interpreted correctly the first time.

To combat misinterpretation of screenings, a second radiologist could independently read the mammograms to challenge or affirm the original readings. While having a second reading may be useful in well-equipped facilities, not every radiologist clinic has the resources to provide a second opinion in a timely manner. The use of CAD has been an acceptable replacement of a second opinion since its conception, but the addition of new AI could further improve the screening mammography process.

CAD was one of the first AI tools approved by the FDA in 1998 to help detect breast cancer and mitigate errors in the diagnosis process (Lamb et al., 2022). CAD analyzes a mammogram by processing the X-ray image into segments. The segmentation process separates the breast tissue from the pectoral muscle and surrounding regions to make it easier to visually extract the suspicious tissues from the breast tissue. The program is then able to isolate suspicious tissue and then classify it as normal, benign, or malignant. Once completed the CAD process highlights the areas it detected, and it is up to the radiologist to diagnose the patient.

There are many types of algorithms that CAD programs use, which all have varying rates of success between 72%-93% (Ganesan, Karthikeyan, et al., 2013). With such great variability, it is obvious that improvements can be made. In order to further reduce error and improve the early detection process, more advanced AI needs to be used.

5.0 Artificial Intelligence in Healthcare

In 1956, John McCarthy described the term AI as the science and engineering of intelligence machines. AI began as a series of "if, then rules" that have advanced into complex algorithms that perform similarly to a human's brain (Kaul, Enslin, Gross, 2020). AI can be broken into many subfields, each with its own specialty. The most common AI used in healthcare are machine learning (ML) and deep learning (DL). ML is able to identify patterns based on a set of particular traits and will "learn" from the new information, which is used in subsequent scenarios (Kaul, Enslin, Gross, 2020). Unlike traditional algorithms, ML is able to dynamically predict outcomes, that improve each time that it is used. ML has advanced so much since its conception that it has evolved into DL. DL is composed of algorithms that create an artificial neuro network that can learn and make decisions on its own, more similar to a human brain than ML (Kaul, Enslin, Gross, 2020).

ML and DL algorithms need to be trained to make accurate predictions. There are three types of learning algorithms that could be used, which include unsupervised, supervised, and reinforced learning. Unsupervised algorithms are able to find patterns, while supervised algorithms can classify data and make predictions based on previous examples. Reinforced learning algorithms use a sequence of rewards and punishments to form a strategy for a specific problem (Hamet, Pavel, Tremblay, 2017). The type of algorithm used is dependent on the desired use case.

Each type of algorithm can be used during the screening mammography process, with varying degrees of results. Unsupervised algorithms can be used within electronic health records to categorize patients based on their medical history. If patients have pre-existing conditions or have been to specific physicians after a screening mammogram, they could be flagged as at risk of

breast cancer or as potentially having breast cancer. This could help with automated information dissemination to help inform patients prior to an actual diagnosis. CAD is a form of supervised learning that is able to detect an anomaly present and determine if it is normal, benign, or malignant. A Reinforced AI tool used for screening mammography could perform the same as a CAD program, but with the added benefit of revising diagnosis based on a database of known cases of breast cancer.

The primary difference between newer DL-based AI algorithms and traditional CAD is the ability for DL-based AI to identify individual features of the image that are useful for analysis, where traditional CAD can only highlight areas for radiologists to diagnose. DL-based AI has the ability to uncover features too small for human perception and identify relationships between features that may not be obvious to an observer (Lamb et al., 2022). With the addition of a much more precise screening mammography analysis, breast cancer can be detected at a much smaller size, which allows for an earlier intervention when compared to current screening processes.

DL-based algorithms provide great opportunities for health systems to improve their diagnosis process, through the use of cognitive based reasoning (CBR). CBR is a form of analogical reasoning that uses known examples to produce a solution for the new case, also known as the query case. CBR is able to make predictions utilizing a four-phase cycle that retrieves old cases most similar to the query, reuses the embedded information and knowledge to produce a solution for the query case, revises the solution to better adapt it to the query case and then retains the query case and its solution to be included in the case database (Lamy, Jean-Baptiste, et al., 2019). CBR as a process can only improve over time as it grows its data repository, which provides increasing amounts of example cases to strengthen predictions.

A prime example of a CBR system that is currently being tested is DESIREE, which stands for Decision Support and Information Management System for Breast Cancer. DESIREE is a European project that is aimed at developing web-based services for the management of primary breast cancer. The CBR system is able to classify the query case using an automatic algorithm, which is combined with visual reasoning to determine if breast cancer is present and at what stage it is at (Lamy, Jean-Baptiste, et al., 2019).

DESIREE was constructed on a small sample set of 315 patients and uses two complimentary approaches to diagnosing breast cancer. A polar-MDS scatter plot is used to visualize the level of similarity between the query case and similar cases within the database. Rainbow boxes are also used to visually show overlapping characteristics between the query cases and similar cases, while including other information outside of the X-ray, such as if the patient has received chemotherapy or surgery in the past. This process has accurately classified cancer 80.3% of the time, but with such a small dataset, any outliers can greatly affect the accuracy of the tool (Lamy, Jean-Baptiste, et al., 2019). If this program were to expand and retain each query case, accuracy would continually improve.

DESIREE also provides a level of explanation that CAD programs fail to have. Rather than seeing a box around suspicious tissue using CAD, physicians and patients can use DESIREE to see a visual similarity to known cases of breast cancer along with other factors that may relate to their diagnosis. Having a visual comparison will help patients with low physician trust to reconfirm their diagnosis through a direct comparison of their mammogram and the closest known example of breast cancer at the same stage.

As of December 2021, the FDA has approved 15 different models for screening mammography. The approved models can be used to detect and diagnose lesions, triage, density assessments, and risk assessments (Lamb et al., 2022). The approved models are similar to DESIREE in their processes, but with additional features that will aid in the dissemination of diagnoses, while providing evidence to support them. The approved models include triage tools, breast density analysis, and individual risk assessments.

Triage tools can be used to prioritize patients through the flagging of mammograms with one or more suspicious findings. By flagging suspicious screenings, radiologists can interpret batches of mammograms with call outs of specific features rather than looking at highlighted areas of general concern. Having direct call outs can help improve diagnosis speed and accuracy. After two years of clinical examinations with triage tools, the turnaround time from examination to reporting decreased from 6.9 days in 2019 down to 3.9 days in 2021 (Lamb et al., 2022). The three-day decrease in reporting time not only reduces the amount of patient stress felt while waiting for a potentially life changing diagnosis, but it will increase the volume of patients that radiologists could diagnose daily.

Breast density is a risk factor for breast cancer and can mask cancers on a mammogram. Breast density refers to the proportions of fat and fibro glandular tissue in the breast. Unlike other organs, normal breast tissue has variable densities that range from no dense tissue and mostly fat, to almost no fat and mostly dense tissue. The density of breast tissue is broken down by categories A, B, C, and D. Category A is almost entirely fat, category B has scattered fibro glandular densities, category C is heterogeneously dense and category D is extremely dense (Gordon, 2022). The performance of CAD assisted mammography is greatly dependent on tissue density. Mammographic sensitivity is directly correlated to the category of breast density, with 100%

sensitivity with category A, which is reduced to 83.9%,72.9%, and 50% for categories B, C, and D respectively (Gordon, 2022). This essentially means that 50% of cancers are left undiagnosed after a screening mammogram for women with category D breast tissue.

Although screening mammography fails to show cancer in category D breast tissue 50% of the time, breast density can be monitored to better alert radiologists that further testing may be needed. Women with category D breast density are 13-31 times more likely to develop breast cancer than those with category A (Gordon, 2022). Tracking all changes in breast tissue density can help to identify cancer that CAD assisted mammography fails to identify.

Individual risk assessments can also be used to determine which patients are at high risk of breast cancer. MRI screenings, chemoprevention or genetic testing are all beneficial to those at a high risk of developing breast cancer, but insurance will often not cover the examination without documentation stating that they have a risk level $\geq$ 20% (Gordon, 2022). Traditionally, the risk assessment models incorporate patient-specific features that are then calibrated at a population level rather than the individual level. Unfortunately, these population samples primarily consist of white women. One of the FDA approved models combines AI-based breast density, breast volume, age, and other risk factors to generate a risk prediction (Lamb et al., 2022). The use of this tool could better document women's risk factors, and aid in obtaining insurance approvals for an MRI.

The ability to detect and diagnose lesions, triage patients, perform density assessments, and generate risk assessments provides great value to health systems and their patients, but some systems may struggle to implement all features. When utilizing AI, the database needs to be stored in a secure server, which may have an additional cost component each time that it is queried. Additionally, the physical equipment at each health system may lack the resources to process information in an effective manner. The implementation of AI into the screening mammography

process will be dependent on the organizational structure, existing equipment, and cost structure. With 15 algorithms to choose from, each health system will be able to determine what features they want and the configuration that will integrate best into their current practices.

6.0 Considerations of AI used Within Patient Diagnosis

The CBR process used in the DESIREE program to diagnose breast cancer through a visual comparison of known cases could become the future of screening mammography. Physicians are able to confirm the output from the AI, through a visual and statistical comparison, that is simple enough to explain to most patients. This could prove to be extremely helpful for detecting breast cancer earlier, as a computer is able to analyze the X-rays with greater granularity than the human eye. The DESIREE program may also prove to be a viable solution for other cancers as well, but not all cancers are as easily diagnosable. Despite all of the benefits that this technology can bring, we need to ensure that peoples' medical information is protected and used appropriately.

AI is often viewed as a black box of information, that is difficult to interpret and explain. This creates an ethical dilemma that may slow the growth of the technology's adoption. Without the ability for physicians to interpret or explain AI diagnoses, the accuracy may be difficult to prove, especially in the eyes of patients. For AI to be successfully implemented into healthcare, there needs to be provisions set in place to maintain physician and patient confidence in the technology. To ensure confidence of AI diagnoses, we need to guarantee the proper use of the data, minimize any potential biases, allow patients to contest the diagnosis, and provide a second opinion in the form of a physician opinion (Jobin et al., 2019). These provisions will ensure that AI is restricted to guidelines that maintain high accuracy, while providing options for both patients and physicians to confirm the validity of each diagnosis.

In all patient settings, it is essential to ask for consent prior to any examination. The need for consent will need to carry over to AI driven diagnostics. The accuracy of AI depends on the validity and quantity of data it has to analyze. As we are only in the initial stages of AI integrated

healthcare, it may be difficult to find volunteers to become part of base data sets, but as the technology is proven, trust in the process will hopefully improve. Additionally, people who are undergoing a screening mammogram should be given the ability to opt out of having their screening in the learning dataset. Over time, the need to retain screenings will diminish as more examples are collected. Eventually the need to retain screening may only be useful to test a new type of screening technology.

After gaining consent to utilize individuals' healthcare data, AI may be able to accurately diagnose patients, but this does not necessitate patient trust. Although large amounts of data are used to train AI, it is impossible to have an exact reference case for each new patient. There may be underlying biases that have unintentionally been included within the data (Jobin et al., 2019). Patients and physicians may be unaware of bias due to the similarity of most patients to the training data. To protect against biases, AI tools need to provide an alert to physicians if a patient has a characteristic that is not common and could provide potential inaccuracy. Physicians can then determine if the alert has an effect on the condition being screened for. The potential of unique patient traits will diminish as the technology's use is expanded and more cases are included in the training data set.

Patients need the ability to have a conversation about their diagnoses and question their validity. Physicians act as the liaison between patients and their healthcare journey and help to ease the burden of new diagnoses. Physicians cannot simply diagnose patients based on a computer output. AI models need to be explainable to confirm accuracy and provide patients with confirmation of their diagnoses. The DESIREE program has a high level of explanation due to the presence of image comparison and characteristic similarities between the patient and query cases.

The output will help physicians inform their patients about their diagnosis and help them understand the reasoning to trust the software.

7.0 Physician Buy-In

When introducing disruptive technology such as DL algorithms into screening mammography, there is bound to be pushback from those who do not understand the technology. The integration of increasingly complex AI has just begun in the healthcare industry and will continue to improve, but until there is an organizational transformation, acceptance may be challenging. A report by Harvard Business Review showed that 78% of companies failed to shift the way that they did business in 2021. The 28 most successful transformers were able to gain employee buy-in through employee compensation packages, fostering employee satisfaction, prioritizing diversity and inclusivity in their hiring processes and maintaining a higher representation of women in both managerial positions and their overall workforce (Argenti et al., 2021). Not every health system will need to use all of these tactics to gain physician buy-in to AI assisted mammography, but one or a combination of several of these methods could ease in a more company culture that embraces AI based innovations.

The difficulty of the buy-in process is determined by the current company culture. Highmark Health is a prime example of a positive company culture. Highmark Health starts to build their culture before each potential employee interviews with their team. During my application process, I was screened by human resources and then given a leadership assessment to ensure that I could further their mission to create a remarkable health experience, freeing people to be their best. During the interview I was not only asked about my skills, but how I could further improve the team and the company as a whole. Highmark Health ensures that all employees buy into their culture prior to extending an offer. Since I started, I have received numerous invitations to high level talks about AI, cyber security, and the future of healthcare - not because it adds value

currently, but because it encourages employees to take risks and try new things to further the company's goals. If the current company culture is poor, it will be difficult to integrate disruptive technologies because innovation is not part of the culture.

You cannot force culture, but you can incentivize innovation. The introduction of value-based reimbursement was aimed at improving health outcomes by focusing on data driven prevention, management of at-risk populations, limiting unnecessary hospital admission and re-admission, reduce costs, and empowerment and prioritization of primary care (Millstein, 2023). AI based screening mammography directly manages risk populations driven by data, making it a prime candidate for value-based reimbursement. Payers can help encourage adoption of AI in screening mammography through reimbursement structure that pays more when the tool used. Unfortunately, value-based reimbursement provides greater value to the health systems than the providers using the tool. This is why gainsharing may be the best way to foster physician acceptance of DL based AI into screening mammography.

Gainsharing is an internal program that is designed to improve productivity, enhance quality, and reduce costs by providing those involved with a percentage of the cost savings. In 1988, the Health Care Financing Administration (HCFA), now Centers for Medicare and Medicaid Services (CMS), tested gainsharing for Medicare participating Heart Bypass Center Demonstrations. This initiative assessed the financial feasibility of bundle payments and gainsharing principles within healthcare. After its initial use within healthcare, gainsharing has been used with varying restrictions (Anoushiravani & Nunley, 2017). By providing a financial incentive, those who are unsure of adopting the new technology may be less resistant to learning and implementing something new.

The ability for physicians to profit off of savings can become tricky without strict guidelines. There needs to be a set of rules to ensure that patient safety and health outcomes remain the main focus of physicians. To be successful, gainsharing must be provisional and have specific objective goals. The goals should be based on the current financial situation of the healthcare organization and target high-volume areas with high costs. A physician champion can help integrate the program and act as the voice of physicians to ensure that both sides are happy and in agreement with policies and procedures. The physician champion will need to be an individual who is mission focused and not driven purely by financial incentive (Anoushiravani & Nunley, 2017). If the wrong champion is chosen, the entire program can fail.

When establishing the gainsharing program, all physicians within the targeted specialty should be included. In practice, not all physicians will participate due to personal reasons and an unwillingness to change (Anoushiravani & Nunley, 2017). Over time, physicians will choose to adopt the better practices or seek employment at an organization that better aligns with their values. Although some physicians may leave, the opportunity to participate in a gainsharing program will help attract physicians who are more aligned with the health systems culture. In competitive markets, heath systems may need to implement a gainsharing system to stand out and attract top physician talent.

8.0 Integration of Gainsharing

The introduction of gainsharing could lead to faster adoption of DL AI assisted screening mammography. The stage at which breast cancer is detected greatly determines which treatments are used and the costs associated with care. The earlier breast cancer is detected, the cheaper and easier it is to treat. The average cost per patient allowed by insurance in the year after diagnosis in 2016 were $60,637, $82,121, $129,387, and $134,682 for stages 0, I/II, III, and IV respectively. The cost differences of each diagnosis are largely driven by the cost of chemotherapy and cancer treatments (Blumen et al., 2016). There are major cost savings associated with earlier detection of breast cancer and if some of those savings were passed onto the physicians using the new technology, physicians will be more willing to learn how to use and then integrate the technology into their practice.

When creating a gainsharing program, data capabilities should be one of the primary considerations. Before any conversations negotiating physician compensation from gainsharing, there needs to be an established benchmark of the current costs and resource consumption. The costs and consumption of resources should be measured for at least one year to reduce the impact of variability and to determine if there are any obvious ways to reduce costs. The data collected would need to be tracked at the physician level to create individualized metrics, which would also help to identify a physician champion to act as liaison between physicians and administration. Physician metrics will then need to be referenced against national and internal standards pertaining to product selection, resource usage, and best practices (Anoushiravani & Nunley, 2017). Each health system that implements a gainsharing system will need to determine other determinants of success with varying weights to determine the larger impact of the program.

For the integration of DL algorithms within Screening Mammography, there are many metrics other than cost that will affect the success of the gainsharing program. The increased accuracy of the AI will help to reduce the rate of misdiagnosis and allow for earlier detection, which can be the main indicators of success. Screening mammography has an average rate of error between 10%-30%, with 52% of the errors stemming from misinterpretation of the mammogram, and 43% if errors attributed to overlooking signs in abnormal scans. The integration of any of the 15 FDA approved algorithms within screening mammography will improve the rate of error, so that it is between 8% and 28% depending on the algorithm being implemented (Weedon-Fekjaer et al., 2014). This error rate will continue to improve as the algorithms are tested in larger populations and more diverse sample sets are collected.

In addition to an improved error rate, we would expect a decrease in turnaround time between examination and diagnosis. One of the FDA approved AI solutions was able to reduce the time it took physicians to diagnose mammograms by three days over the course of a two-year study (Lamb et al., 2022). There are many factors that affect the speed at which mammograms can be reviewed, so an internal benchmark should be weighted more than the average turnaround time as patient volume, staffing ratios, and equipment availability can vary greatly within different geographic areas.

The prevalence of each error would need to be tracked and reported. The accuracy of screening mammography should improve once integrated with better AI, but with new technology comes new problems that need to be addressed as they arise to prevent any negative impact on future patients. The current CAD software should remain in use while physicians are learning how to use and interpret the new technology. By running both programs simultaneously, physicians

will have a direct comparison to foster confidence and allow physicians to become comfortable without putting any patient at risk.

If acceptance of the technology remains an obstacle to adoption, a value-based reimbursement model may be another feasible option. Health insurance providers can negotiate higher rates paid out to health systems that implement AI into their mammography screening process. Health systems who participate in this type of value-based reimbursement will be able to earn greater revenue for the same process, while simultaneously diagnosing breast cancer earlier, which will help in lowering costs and improve patient outcomes.

9.0 Discussion

Nearly 13% of American women will have breast cancer at some point in their lives and without any improvement, the current standard of detection can greatly limit the effectiveness of treatment (Breast Cancer Statistics, n.d.). CAD, the standard program used in screening mammography, has remained relatively unchanged since its conception in 1998 (Lamb et al., 2022). With over 20 years of technological advancements and stalling improvements in the early detection of breast cancer, it is past due for some major upgrades.

When women receive a screening mammogram, they may receive news that will greatly affect their life, but upwards of 30% of these women will receive a misdiagnosis (Ganesan, Karthikeyan, et al., 2013). For the 30% of women who receive a misdiagnosis, breast cancer may be present, but they may have to wait until their next yearly screening to receive a diagnosis if they remain healthy. Women who do not have access to yearly screening services may wait even longer to be diagnosed, which gives the breast cancer a greater chance of spreading. Fortunately, 85% of the errors that led to misdiagnoses are attributed to human error and can be prevented (Ganesan, Karthikeyan, et al., 2013). It is way past the expiration of CAD software, and it needs to be replaced with newer, more precise algorithms.

The introduction of sophisticated AI models can be the solution to the current shortcomings of screening mammography. At the end of 2021, the FDA had approved 15 different AI models that focus on improving various aspects of screening mammography (Lamb et al., 2022). These models have greater capabilities than the current CAD software, including the ability to detect and diagnose lesions, triage patients, analyze breast density, and generate risk assessments. Radiologists using one of the newer algorithms will be able to review batches of screenings with

specific features called out and confirm or contest the diagnosis based on patient similarity to the reference cases.

With 15 FDA approved algorithms that provide a similar service, the decision of which algorithm to use will greatly depend on the health systems implementing them. The configuration of each algorithm will vary greatly depending on the organizational structure, existing equipment, and cost structure. Health systems with smaller budgets may seek to find a solution focused on lower costs that may include on premises data storage, while cash abundant systems may search for a solution with more features such as cloud-based storage.

While these improvements are able to better manage patients, what truly makes AI algorithms superior to CAD is the ability to explain findings to patients. Each algorithm utilizes previous cases to base its findings on, which can also be used as visual example to help patients better understand their diagnosis. Patients will gain the ability to compare their screening to a similar example of breast cancer and confirm the physician diagnosis. This will be especially beneficial to populations that have low levels of physician trust.

The largest obstacle in acceptance is physician buy in. Physicians are often reluctant to try new technology for a variety of personal reasons, so an incentive program will likely be needed for initial introduction. The cost of treating cancer increases as each stage advances, so early detection will not only improve outcomes but will also reduce overall costs of cancer treatment. Health systems will be able to share some of the cost savings with physicians using the new algorithms through gainsharing. The gainsharing program will compensate physicians with a percentage of cost savings based on a benchmark value determined prior to the start of the program. By providing a financial incentive, more physicians will opt to use the new algorithms for screening mammograms. Additionally, as the algorithm is used more frequently, the diversity and

quantity of reference cases will grow, which in turn will further improve the accuracy of the algorithm.

While not all health systems have the culture to foster the implementation of AI into their screening mammography process, a value-based reimbursement approach could also help further adoption. Where physicians may be unwilling to learn a new technology, the added revenue from value-based reimbursement can encourage health systems to utilize the technology. Once integrated, the detection accuracy and overall cost savings should be more than enough to encourage greater physician use, especially if a gainsharing program is in place. Despite being a viable alternative, value-based reimbursement will not be a guaranteed outcome of implementation and will need to be discussed between each health system and their payers.

The prevalence of AI within screening mammography will depend greatly on early adopters of the technology. While there are promising results from this type of technology, large scale studies are needed to confirm accuracy among diverse populations across the United States. The current success of DESIREE and the FDA approval of 15 AI algorithms shows a promising future for AI based screening mammography. It will be up to individual health systems to deploy these algorithms in diverse areas that could potentially create inaccuracy prior to any national deployment.

10.0 Conclusion

In this paper, the accuracy of CAD used in screening mammography has been challenged by human factors that can only be improved through better training and second opinions. Due to current challenges faced by the healthcare industry in the United States, supplementary trainings, and the availability of a second radiologist to confirm diagnoses may be impossible for many screening sites. AI algorithms need to be introduced into the screening mammography process to act as the second opinion and improve the rate of misdiagnosis.

Physicians get fatigued, experience stress, and get distracted by personal matters just like everyone else. A mistake for most professions may have economic ramifications, but if breast cancer is misdiagnosed, it may cost someone their life. Human error is the main cause of breast cancer misdiagnosis and AI could be the solution to mitigate this issue.

Replacing CAD with newer algorithms will fortify the diagnosis process and limit human related errors. The boxes that CAD places on abnormalities may work a majority of the time, but it is up to the physician to diagnose the patient. While CAD points out areas of concern, radiologists should be capable of detecting the same abnormality without CAD due to their extensive training. AI, unlike CAD, has a greater capability to point out specific abnormalities with visual comparisons and statistical similarities, which provides a much clearer picture to radiologists. With greater amounts of information at their fingertips, radiologists can diagnose patients quicker and with greater certainty.

This literature review has focused on the findings of multiple studies regarding breast cancer detection statistics, with a focus on early detection. It is clear that women who receive treatment for breast cancer as soon as it presents itself have a much greater treatment success rate.

While CAD based screening mammography has been widely accepted for over 20 years, there are still a large number of women who are negatively affected by the shortcomings of those reading their screenings. Human error should be limited at an institutional level, which cannot happen until CAD is replaced with a more accurate AI algorithm. While some facilities are implementing the new technology in their screening process, it is up to physicians and their health systems to adopt the new technology.

It is imperative to obtain physician buy in to successfully integrate any form of AI into screening mammography. While each of the 15 approved algorithms have proven to the FDA that they hold medical significance, without physicians willing to use the technology it is worthless. To further the use of AI in screening mammography, larger studies need to be conducted with respected health systems that operate in diverse communities. If implemented successfully, AI assisted screening mammography will eliminate misdiagnoses due to human error and allow for more women to seek treatment as soon as breast cancer is present.

Bibliography

Anoushiravani, Afshin A., and Ryan M. Nunley. "Gainsharing strategies, physician champions, getting physician buy in." *The Journal of Arthroplasty*, vol. 32, no. 6, 2017, pp. 1723–1727, https://doi.org/10.1016/j.arth.2017.02.011.

Argenti, Paul A., et al. "The Secret behind Successful Corporate Transformations." *Harvard Business Review*, 29 Nov. 2021, hbr.org/2021/09/the-secret-behind-successful-corporate-transformations.

Blumen, H., Fitch, K., & Polkus, V. (2016). Comparison of Treatment Costs for Breast Cancer, by Tumor Stage and Type of Service. American health & drug benefits, 9(1), 23–32.

"Breast Cancer Statistics: How Common Is Breast Cancer?" *Breast Cancer Statistics | How Common Is Breast Cancer? | American Cancer Society*, American Cancer Society, 14 Sept. 2023, www.cancer.org/cancer/types/breast-cancer/about/how-common-is-breast-cancer.html#:~:text=It%20is%20about%2030%25%20(or,(DCIS)%20will%20be%20diagnosed.

Castellino, Ronald A. "Computer Aided Detection (CAD): An overview." *Cancer Imaging*, vol. 5, no. 1, 13 Apr. 2005, pp. 17–19, https://doi.org/10.1102/1470-7330.2005.0018.

Ganesan, Karthikeyan, et al. "Computer-aided breast cancer detection using mammograms: A Review." *IEEE Reviews in Biomedical Engineering*, vol. 6, 2013, pp. 77–98, https://doi.org/10.1109/rbme.2012.2232289.

Gordon, Paula B. "The impact of dense breasts on the stage of breast cancer at diagnosis: A review and Options for Supplemental Screening." *Current Oncology*, vol. 29, no. 5, 17 May 2022, pp. 3595–3636, https://doi.org/10.3390/curroncol29050291.

Hickman, Sarah E., et al. "Machine Learning for Workflow Applications in screening mammography: systematic review and meta-analysis." *Radiology*, vol. 302, no. 1, 2022, pp. 88–104, https://doi.org/10.1148/radiol.2021210391.

Jobin, Anna, et al. "The global landscape of AI ethics guidelines." *Nature Machine Intelligence*, vol. 1, no. 9, 2 Sept. 2019, pp. 389–399, https://doi.org/10.1038/s42256-019-0088-2.

Lamb, Leslie R., et al. "Artificial Intelligence (AI) for Screening Mammography, from the *ajr* special series on Ai Applications." *American Journal of Roentgenology*, vol. 219, no. 3, 2022, pp. 369–380, https://doi.org/10.2214/ajr.21.27071.

Lamy, Jean-Baptiste, et al. "Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach." *Artificial Intelligence in Medicine*, vol. 94, 2019, pp. 42–53, https://doi.org/10.1016/j.artmed.2019.01.001.

Millstein, Jeffrey. "Value-Based Payments Are Making It Harder to See Your Primary Care Doctor on Short Notice." *STAT*, 19 Sept. 2023, www.statnews.com/2023/09/20/value-based-payments-primary-care-physicians-appointment-wait-times/.

"Triple-Negative Breast Cancer: Details, Diagnosis, and Signs." *Details, Diagnosis, and Signs | American Cancer Society*, American Cancer Society, 1 Mar. 2023, www.cancer.org/cancer/types/breast-cancer/about/types-of-breast-cancer/triple-negative.html.

Weedon-Fekjaer, H., et al. "Modern mammography screening and breast cancer mortality: Population study." *BMJ*, vol. 348, no. jun17 17, 2014, https://doi.org/10.1136/bmj.g3701