HoVer-Trans Application for Enhanced Breast Cancer Diagnosis in Mayo Clinic Datasets

Clif Posey

University of Wisconsin - LaCrosse

DS 785: Capstone

Dr. Tracy Bibelnieks

April 28, 2024

Abstract

Breast cancer is the most prevalent cancer in women globally. More than 2.26 million new cases of breast cancer have been recorded worldwide, and it is a significant cause of cancer-related deaths (World Cancer Research Fund International, 2020). Computer aided detection has shown promising results in accurate classification of malignant and benign tumors on breast ultrasound. This study aimed to apply the HoVer-Trans model to the Mayo Clinic breast ultrasound data and to test the model results compared to Jarvey (2022) and Mo et al. (2022). The results showed that the HoVer-Trans model demonstrated greater AUC, which could correctly classify breast lesions as benign or malignant. However, it under-performed in sensitivity compared to a radiologist's results (Jarvey, 2022). Improved model optimization of the HoVer-Trans model could produce more precise and accurate results than this study's results. Future research on AI in breast cancer diagnosis using machine learning enhances early detection, and rapid intervention for breast cancer is essential for patient care and optimal outcomes. Machine learning and AI can contribute significantly to diagnostic medicine in terms of breast cancer.

Keywords: AI, machine learning, breast cancer, ultrasound

Contents

List of Tables and Figures

Chapter 1: Introduction

Breast cancer is the most prevalent cancer in women globally. More than 2.26 million new cases of breast cancer have been recorded worldwide, and it is a significant cause of cancer-related deaths (World Cancer Research Fund International, 2020). Breast cancer remains a significant health issue in the United States, with 310,720 new cases of invasive breast cancer predicted for 2024 (BreastCancer.org, 2024), which increased from 2023 from 297,790 women (Cancer.net, 2024). According to the National Breast Cancer Foundation (2023), 1 in 8 women in the United States was diagnosed with breast cancer, which showed up to 27% mortality in women aged 60 years and older (American Cancer Society, 2024).

Breast ultrasound evaluation is the current state-of-the-science medical diagnostic and screening tool for breast cancer. Traditionally, ultrasound images are read by a radiologist who determines whether a breast lesion is benign or malignant. There are five levels of characterization of breast tissue. There are normal, superficial cystic lesions, complex cystic lesions, indeterminate cystic or solid lesions, and solid lesions. They began with a level 2 lesion (superficial cystic lesion), as well as a description and classification of risk related to whether or not the tumor is malignant. These are called BIRADS (breast imaging reporting and data system) (Stavros, 2004). Before the availability of AI assistance, the only way to definitively know if a lesion was benign or malignant was to biopsy it. The biopsy procedure may or may not be scheduled on the same day as the ultrasound. It is also costly and invasive. With AI tools, breast ultrasound images can be classified, and predictive analytics can be used to distinguish between benign and malignant tumors. This emerging technology promises to improve the accuracy and speed at which malignant tumors are identified.

AI used for breast cancer lesion detection and classification emerged around 2020. Early studies showed that AI was useful but had challenges, including the lack of access to data and bias. Furthermore, fear around the concept of AI that humans will lose control over technology also creates challenges for introducing AI into the healthcare arena (Abdul-Halim et al., 2021). Currently, there is ongoing research to validate the use of AI in breast cancer diagnosis, and as data becomes more readily available, testing and training these models could yield viable options for painful and invasive procedures.

A collaboration between the University of Wisconsin-La Crosse and the Mayo Clinic Health Systems attempted to use deep learning to rapidly and accurately diagnose breast cancer tumors detected by ultrasound. The introduction of the HoVer-Trans model to the Mayo Clinic's extensive database of 109,188 ultrasound images marked a significant advancement in breast cancer diagnosis. The HoVer-Trans model, known for its anatomy-aware capabilities and not requiring predefined regions of interest (ROI) for diagnosis, offered a promising alternative to traditional convolutional neural network (CNN) methods. While effective, these conventional methods often need more interpretability, which is crucial for clinical application. This project aimed to leverage the unique features of the HoVer-Trans model to improve both the accuracy and interpretability of breast cancer diagnosis.

Statement of the Problem

AI technologies like natural language processes and machine learning are emerging as tools to begin structuring the vast amount of unstructured data in healthcare (Horowitz, 2023), including medical images. In addition, current medical diagnostic technologies are expensive and increase medical costs. Access to diagnostic technologies may be difficult (Johnson, 2023).

Furthermore, accuracy in traditional diagnostic methods can be limited. Innovation in healthcare diagnostic tools using AI and predictive analytics will improve accuracy, particularly with advanced imaging (StartUS Insights, 2023).

Related to medical imaging, AI technologies, specifically deep learning methods, improved traditional methods in that AI can significantly reduce the time required to diagnose diseases. Models can remove the subjectivity that humans introduce in evaluating images and making final diagnoses (Eustaqui et al., 2023). It is also important to note that reproducibility in medical imaging is greatly improved with the use of AI. Human interpretation often leads to variability in inter-rater observations because practitioners may interpret the same images differently. AI removes the variability, allowing for a more accurate result (Eustaqui et al., 2023).

The current state of using AI and predictive analytics needs to be improved. The main problem is the need for extensive data sets to train the AI models, refine them for accuracy (Zhang & Qie, 2023), and use them more efficiently.

As noted above, traditional methods for breast cancer diagnosis using ultrasound imaging were effective to a certain extent but often fell short in terms of interpretability and efficiency. The Mayo Clinic has pioneered more effective methods using AI and predictive analytics. To date, the Mayo Clinic has used Convolutional Neural Networks (CNN) to determine if breast lesions discovered by ultrasound were benign or malignant. One of the problems encountered by using CNN was that the level of accuracy needed to be in the 70th percentile. The region of interest (ROI) was not interpretable using the CNN.

Theoretical Framework

This study was based on the functionality of an anatomically aware HoVer-Trans model. The HoVer-Trans model represented a cutting-edge approach in medical image analysis, focusing on the anatomy-aware segmentation and classification of images without the need for manually defined ROIs. This project explored the application of this model to a vast and varied dataset, pushing the boundaries of what was possible with current technology.

Mo et al. initially developed the HoVer-Trans model for breast cancer diagnosis (2022). Mo et al. (2022) proposed that malignant and benign tumors rested in different tissue layers and had other spatial relationships. The HoVer-Trans model could extract information vertically and horizontally, allowing for more accurate breast cancer tumor ultrasound interpretation. This was because benign tumors tended to follow a horizontal line and stay in the epithelial tissues. Malignant tumors grew to follow a vertical line from the glandular tissue into deeper tissue. Mo et al. (2022) found that their model outperformed two senior sonographers on breast cancer diagnosis in both BI-RADS and binary malignant/benign conclusions.

Statement of Purpose

The purpose of this study was to evaluate the effectiveness of the HoVer-Trans model on the Mayo Clinic's ultrasound image database, aiming to improve the interpretability and accuracy of a breast cancer diagnosis. By providing a model that required minimal human intervention for accurate malignancy prediction, this research could significantly impact the field of computer-aided detection (CAD).

The HoVer-Trans model allowed for the visualization of tumors within the images and for the accurate identification of malignant tumors. The Mayo Clinic's database presented an opportunity to apply the HoVer-Trans model's advanced capabilities to a real-world setting, potentially setting a new standard for diagnostic accuracy and usefulness in clinical practice.

Objectives

This study addressed the accuracy of the HoVer-Trans model in diagnosing breast cancer, its interpretability in clinical settings, and its performance compared to other models that used weakly supervised approaches. It aimed to provide an enhanced diagnostic tool for developing CAD systems. The objectives of this project were:

- Improved model statistics compared to Jarvey (2022).
- The development of an improved HoVer-Trans model that could accurately classify breast tissue lesions as benign or malignant.
- A substantial contribution to the field of medical imaging and cancer diagnosis was made, potentially leading to early detection and treatment of breast cancer.

Significance of the Study

This study sought a more accurate, interpretable, and efficient method for diagnosing breast cancer by applying the HoVer-Trans model to the Mayo Clinic's comprehensive database. This could greatly benefit the healthcare industry by improving patient outcomes and streamlining the diagnostic process.

Definitions of Terms

Terms used included:

- “HoVer-Trans”: A transformer method to preserve Region of Interest
- "Convolutional Neural Network” (CNN): A regularized type of feed-forward neural network that learns feature engineering by itself via filters

- "Region of interest" (ROI): A sample within a data set identified for a particular purpose.
- "Computer Assisted Diagnosis" (CAD): A broad concept that integrates image processing, machine learning/deep learning, computer vision, mathematics, physics, and statistics into computerized techniques that assist radiologists in their medical decision-making processes.

Conclusion

In summary, early detection of malignant breast lesions is essential to successful treatment. Current methods of early detection using CNNs showed moderate success. However, Mo et al. (2022) suggested that the HoVer-Trans model would yield a higher accuracy percentile. Therefore, this study will apply the HoVer-Trans model to the Mayo data to test the process of Mo et al. (2022) to increase the accuracy percentile.

Chapter 2: Review of the Literature

According to the National Breast Cancer Foundation (2023), 1 in 8 women in the United States were diagnosed with breast cancer. Early detection was essential for rapid detection and treatment. Breast cancer has been recognized as a significant global health issue that calls for the advancement of new diagnostic methods to achieve medical treatment that is timely and effective. Ultrasound mammography is the most widely used diagnostic because it is affordable and accessible to most women. It remains an essential component of methods for diagnosing breast cancer. Despite these advantages, the challenge of interpreting ultrasonography accurately posed a severe barrier in that it depended on humans. Therefore, the creation of advanced computational models that could accurately diagnose patients was the most logical step in the advancement of diagnosing breast cancer. This literature review will include enhanced methods for breast cancer diagnosis, emerging techniques and their impacts, and advanced AI applications in breast cancer management.

Enhanced Methods for Breast Cancer Diagnosis

Artificial intelligence and recent advancements in medical imaging have greatly improved the accuracy and effectiveness of breast cancer diagnosis. One recent development was the use of vision transformers for mammography classification. Ayana et al. (2023) stated that these models' successful use of transfer learning techniques has allowed them to distinguish between benign and malignant tissues with high diagnostic accuracy.

Xie et al. (2022) also provided evidence of enhanced detection abilities and reduced false positive rates by integrating AI with traditional imaging methods. In addition, Zhang et al. (2021)

addressed the practical challenges posed by AI applications in breast imaging, such as dataset variability and model generalizability, and proposed solutions to address these issues.

Comparative studies, such as one conducted by Li et al. (2023), showed the benefits of combining different machine-learning models. When combined, they demonstrated increased predictive accuracy of breast cancer diagnoses. Patel et al. (2022) researched the combined use of MRI and ultrasound data in 2022. When processed by sophisticated AI algorithms, the data helped to detect breast cancer at an earlier stage and with greater accuracy.

Emerging Techniques and Their Impacts

The advancement of artificial intelligence along with breast imaging has led to the development of new techniques and improved capabilities of traditional methods, resulting in a shift in the methods used to diagnose breast cancer. Wang et al. (2022) provided a thorough analysis of several cutting-edge AI-driven imaging technologies that enhanced diagnostic clarity and accuracy. These are two crucial factors that are critical for spotting subtle abnormalities at an early stage. More opportunities to improve diagnostic accuracy were created by combining different types of neural networks. Chen et al. (2023) demonstrated how hybrid models that use both CNNs and RNNs can process complex imaging data efficiently, improving detection rates and diagnostic reliability. Gupta et al. (2021) studied how the development of mammography has been enhanced by deep learning. They described advancements in AI applications that significantly improved the accuracy of mammogram interpretation. This advancement has proven crucial in ensuring prompt diagnosis and effective patient care. Also, a critical study conducted in Sweden found that AI used in screenings found 20% more cancers. The number of false positives was not increased, meaning the AI did not incorrectly diagnose breast lesions as abnormal (Knoll, 2023).

An essential aspect of patient care is individualization due to differences in physiology among patients and different types of breast lesions. Torres et al. (2022) showed how artificial intelligence (AI) can be used to personalize patient care and talked about using machine learning to predict breast cancer recurrence. Conducting individual risk assessments using AI has made it easier to identify problems early and individualize follow-up therapies and interventions.

A significant development in breast cancer detection is improved noninvasive screening techniques to reduce unnecessary invasive methods such as biopsies. Kim et al. (2023) demonstrated the advancements in noninvasive screening techniques. These additional techniques have the potential to significantly reduce the invasiveness of traditional diagnostic procedures and reduce medical costs. More importantly, such methods could facilitate more comprehensive and easy early detection and monitoring of breast cancer.

Advanced AI Applications in Breast Cancer Management

Precision medicine has entered a new era in oncology with artificial intelligence in breast cancer diagnosis and therapy. In addition, AI applications have been developed to assist in appropriate breast cancer management. Based on the work by Smith et al. (2022), the method by which cancer prognosis is approached has drastically changed due to the ability of AI to analyze large datasets and produce highly accurate patient outcome predictions. Narvaez et al. (2021) studied the impact of AI on breast cancer screening and diagnosis in 2021. They demonstrated how artificial intelligence (AI) has increased detection rates and diagnostic accuracy, potentially reducing healthcare disparities by providing reliable diagnostic assistance across various healthcare environments.

Lee et al. (2023) revealed the anticipated advancements in individualized cancer treatment. Treatment and diagnostic plans could be precisely tailored for every patient thanks to these advancements, increasing the likelihood of successful interventions specific to types of breast cancer.

In terms of AI technologies, the role of deep learning in detecting breast cancer, as explored by Morrison et al. (2022), is a critical advancement. Their methods effectively identify intricate patterns in imaging data, identifying early warning signs of breast cancer. Furthermore, Hayes et al. (2021) discussed using computer vision technologies to identify breast cancer. For early intervention and treatment planning, improved interpretation of imaging results translates into quicker processing times and more accurate diagnosis.

The Future of Breast Cancer Diagnostics and Treatment

The field of diagnosing and treating breast cancer is evolving quickly thanks to the advent of cutting-edge computational technologies. O'Connor et al. (2023) demonstrated how machine learning can significantly enhance early detection of breast cancer, which is crucial for increasing the chance of a successful treatment plan and survival.

Machine learning is altering individual therapy techniques. Treatment plans will be tailored to each individual based on detailed analyses of each patient's genetic data and illness trajectory. The machine learning approach maximizes the effectiveness of treatment while minimizing unnecessary side effects (Simmons et al., 2022).

Thompson et al. (2021) studied computational techniques to help predict treatment outcomes. Their research demonstrates how using predictive models in treatment decisions leads to developing more specialized and effective therapeutic interventions. Fischer et al. (2023) showed how artificial intelligence is speeding up research to enable the prompt identification and validation of new therapeutic targets, which could result in revolutionary treatments. Lastly, Jenkins et al. (2022) clarified how data-driven methodologies are transforming the field of breast cancer.

Previous Capstone Projects

Over the last three years, the Mayo Clinic and the University of Wisconsin-La Crosse have studied the methodologies for detecting and classifying breast tumors using ultrasound images with traditional imaging techniques, basic CAD systems, and neural networks using deep learning techniques. These were the basis for capstone projects at the University of Wisconsin. This section briefly describes six of the projects evaluated as part of the background research for this current project.

Mohan (2022), Hall (2021), and Andrei (2022) each completed a capstone project researching CNNs and U-Net architectures. These systems demonstrated higher accuracy, precision, and recall rates than traditional models. Bodart (2022) completed a capstone project which focused on medical practitioners. Bodart’s work helped make AI tools more understandable and transparent to physicians and radiologists, thereby gaining the practitioners' trust so they were more apt to use the tools. Silberfin (2021) developed workflows that integrated computer-aided detection to improve accuracy and efficiency in breast cancer diagnosis. Finally, Jarvey (2022) designed a computer-aided detection program that used machine-learning algorithms to improve the detection and characterization of lesions.

Medical Image Analysis with ViT

The broader use of vision transformers in medical image analysis was investigated by Azad et al. (2023) in a groundbreaking study emphasizing their performance and adaptability in a range of medical imaging contexts, including ultrasound images used to diagnose breast cancer. A thorough description of the architecture of vision transformers emphasized the self-attention mechanism that made it possible to comprehend relationships and spatial hierarchies in medical images sophisticatedly. Azad et al. (2023) argued that by providing improved model interpretability and the capacity to identify long-range dependencies in pictures, data vision transformers markedly deviated from conventional convolutional neural networks (CNNs) (Azad et al., 2023).

HoVer-Trans Model

A groundbreaking study recognized the HoVer-Trans model as a breakthrough in the field (Mo et al., 2022). This model adopted a novel approach to enhance ultrasound image diagnostic accuracy without requiring previously defined regions of interest (ROI) by utilizing the anatomical differences between benign and malignant tumors. The model incorporates an innovative HoVer-Transformer block that improves the performance and interpretability of breast cancer diagnostics by extracting and analyzing spatial data in both horizontal and vertical dimensions. Given that the HoVer-Trans model outperforms traditional CNN architectures and that competent sonographers can make significant progress in diagnosis, ultrasound-based breast cancer detection advancements have been made. (Mo et al., 2022).

BreastUS Transformer model

The BreastUS Transformer model was presented in 2022 and automatically classified breast ultrasound images into three categories: benign, malignant, and normal (Saad et al., 2022). Compared to state-of-the-art CNN models, integrating self-attention mechanisms yielded significantly better diagnostic accuracy and performance metrics. This breakthrough highlighted the revolutionary possibilities of transformer models in medical image analysis.

VGG16 Model

Hossain et al. (2023) demonstrated how well a VGG16 model-based transfer learning technique worked for identifying breast cancer from ultrasound pictures. The method utilized a customized deep neural network for classification and a median filter to improve image quality via despeckling. It produced notable gains in accuracy and computational efficiency. This study provided insights into the benefits and practical applications of transfer learning and deep learning methodologies, highlighting their promising potential in improving breast cancer diagnostic procedures.

Conclusion

The collective results of these studies highlighted a significant evolution in the ultrasound imaging-based breast cancer diagnosis methodologies. To improve diagnostic accuracy, reduce reliance on human ROI identification, and increase access to advanced diagnostic tools in various healthcare settings, computational models such as the HoVer-Trans and BreastUS heralded a paradigm-shifting era in the application of artificial intelligence in medical diagnostics. By integrating these sophisticated computational models into routine clinical procedures, the early diagnosis and treatment of breast cancer may be transformed, ultimately leading to better patient outcomes. This literature review examined the HoVer-Trans model and vision transformers to identify and classify breast cancer tumors. This highlighted essential developments in medical imaging and diagnostics. The joint efforts and results of Mo et al. (2022), Saad et al. (2022), and Hossain et al. (2023), in addition to the theoretical insights provided by Azad et al. (2023), demonstrated the revolutionary potential of combining advanced computational models and AI technologies in the field of breast cancer detection and diagnosis. The HoVer-Trans model became a game-changer employing anatomical knowledge and state-of-the-art transformer technology to improve the precision and clarity of diagnosis obtained from ultrasound images without needing preset regions of interest. Similarly, the novel BreastUS model and the use of transfer learning through the VGG16 model further demonstrated the ability of deep learning and machine learning methods to improve the precision and effectiveness of diagnosis. These approaches presented a severe challenge to the knowledge of seasoned medical professionals while also being a significant advancement over conventional diagnostic techniques.

Demonstrating the wide range of applications for these tools in medical image analysis outside of breast cancer diagnosis broadened the horizons of these technological advancements. Diagnostic accuracy, efficiency, and scalability were enhanced across a wide range of medical conditions and imaging techniques by the innate ability of vision transformers to identify complex spatial relationships and patterns within medical imagery. The fight against breast cancer entered a new phase with the promise of earlier detection, more accurate diagnoses, and customized treatment plans brought about by these technological advancements. Nevertheless, ongoing research, interdisciplinary cooperation, and careful handling of implementation ethical and privacy concerns were necessary to integrate these technologies into clinical settings effectively. In conclusion, the knowledge gained from these investigations is a significant basis for upcoming innovation. The development and application of artificial intelligence (AI) and machine learning technologies have great potential to improve patient outcomes and transform the field of medical diagnosis and treatment planning. Precision medicine and patient care have entered a new era thanks to these technologies' continuous development and application, improving the ability to detect, diagnose, and treat diseases like breast cancer.

Chapter 3: Research Method HoVer-Trans Model

A prompt and accurate diagnosis is essential for the effective management of breast cancer. Mo et al. (2022) proposed the HoVer-Trans model. They provided a fully automatic technique for utilizing ultrasound pictures to diagnose breast cancer. This model identified the breasts' anatomical structures by considering the spatial relationships between tumors and anatomical layers. Utilizing intra- and inter-layer spatial correlations, its anatomy-aware formulation enhanced the representation of spatial relationships. The improved representation of spatial relationships between layers in the HoVer-Trans led to its superior performance in ultrasound images compared to previous algorithms. The transformer model was enhanced by integrating anatomical prior knowledge via patch horizontal and vertical embedding. During the HoVer-Trans stage, the relationships between the intra- and inter-layer anatomical layers of the breast were determined by four branches. Convolutional blocks add inductive bias and link two adjacent stages, increasing the model's diagnostic precision for breast cancer. (Mo et al., 2022).

The HoVer-Trans model that has been proposed is exceptionally efficient and accurate, and it has the potential to revolutionize the diagnosis of breast cancer and improve outcomes for patients. Figure 1 compares various models tested by Mo et al. (2022) on three different databases, which aided in determining which model to use. The current project used the GDPH&SYSUCC database, and as noted by Mo et al. (2022), the HoVer-Trans model (signified as "Ours" in Figure 1) performed the best. Other models with good results, such as the VGG16, were rejected because they were not anatomically aware.

Figure 1

Comparison of Models in 3 Databases (Mo et al., 2022)

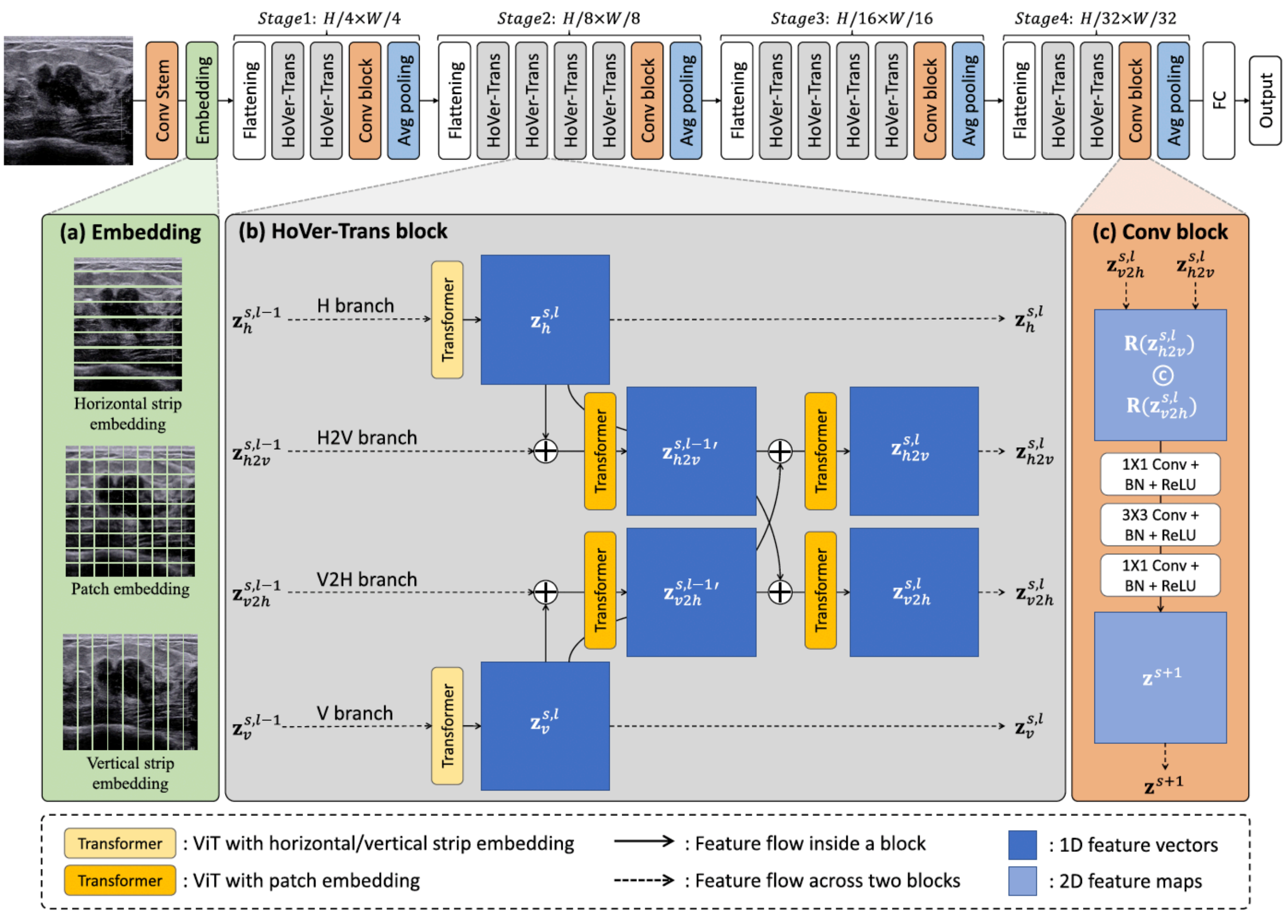
QUANTITATIVE COMPARISONS WITH SOTA APPROACHES IN THREE DATASETS. THE LAST COLUMN SHOWS THE P-VALUE OF DELONG'S TEST BETWEEN THE AUC OF EACH BASELINE MODEL AND HOVER-TRANS MODEL. P-VALUES LESS THAN 0.05 ARE MARKED AS *, P-VALUES LESS THAN 0.01 ARE MARKED AS **, AND P-VALUES LESS THAN 0.001 ARE MARKED AS ***. MODEL WITH † MEANS THEY REQUIRE A PRE-DEFINED ROI

	AUC	ACC	Specificity	Precision	Recall	F1-score	*p*-value
UDIAT							
ResNet50	0.778±0.059	0.743±0.073	**0.899±0.118**	0.676±0.146	0.426±0.256	0.523±0.120	***
VGG16	**0.786±0.073**	0.756±0.123	0.800±0.106	0.650±0.120	**0.672±0.183**	**0.661±0.107**	***
ViT	0.740±0.140	0.701±0.300	0.880±0.147	0.606±0.240	0.364±0.132	0.455±0.223	***
TNT-s	0.627±0.082	0.626±0.089	0.752±0.150	0.426±0.276	0.370±0.241	0.396±0.160	***
Swin-B	0.760±0.141	0.761±0.078	0.895±0.119	0.697±0.238	0.495±0.210	0.547±0.186	***
Ours	0.781±0.118	**0.774±0.061**	0.889±0.128	**0.714±0.214**	0.545±0.232	0.619±0.099	-
MsGoF†	0.939±0.031	0.909±0.032	0.927±0.106	0.900±0.044	-	-	-
BVA-Net†	0.870	0.859	0.685	0.945	0.840	-	-
BUSI							
ResNet50	0.877±0.034	0.818±0.039	0.883±0.030	0.738±0.049	0.682±0.079	0.709±0.062	***
VGG16	**0.898±0.037**	0.832±0.041	0.778±0.056	0.873±0.054	0.862±0.096	0.867±0.063	***
ViT	0.834±0.062	0.811±0.052	**0.922±0.070**	0.781±0.146	0.579±0.032	0.665±0.094	***
TNT-s	0.852±0.015	0.812±0.032	0.908±0.035	0.763±0.057	0.611±0.104	0.679±0.057	***
Swin-B	0.858±0.024	0.818±0.026	0.880±0.045	0.736±0.063	0.694±0.084	0.710±0.044	***
Ours	0.865±0.066	**0.855±0.050**	0.838±0.053	**0.876±0.062**	**0.867±0.115**	**0.872±0.080**	-
BVA-Net†	0.889	0.843	0.758	0.883	0.751	-	-
GDPH&SYSUCC							
ResNet50	0.886±0.014	0.832±0.014	0.732±0.033	0.851±0.015	0.890±0.013	0.870±0.010	**
VGG16	0.919±0.006	0.864±0.004	**0.892±0.010**	0.811±0.009	0.814±0.007	0.813±0.003	*
ViT	0.806±0.021	0.734±0.029	0.694±0.053	0.809±0.047	0.758±0.021	0.782±0.028	***
TNT-s	0.853±0.010	0.781±0.015	0.618±0.059	0.793±0.050	0.879±0.028	0.834±0.015	***
Swin-B	0.886±0.024	0.824±0.025	0.744±0.035	0.853±0.019	0.871±0.029	0.865±0.019	**
Ours	**0.924±0.016**	**0.893±0.021**	0.836±0.038	**0.906±0.023**	**0.926±0.019**	**0.916±0.019**	-
BVA-Net†	0.856±0.009	0.811±0.022	0.859±0.038	0.824±0.022	0.891±0.028	0.856±0.020	***

Therefore, this project utilized the Hover-Trans model, detailing its application in discerning malignant and benign tumors in breast tissue. As ultrasound images were analyzed, the anatomical structures of the breast were visually discernible. Unlike benign tumors, malignant tumors displayed distinct spatial relationships with various anatomical layers, including the fat, gland, muscle, and thorax layers. A completely automated model for diagnosing breast cancer was proposed as a result of this prior knowledge. The capacity to identify breast cancer is a feature of this model. This section presents the methodology of the proposed model, as illustrated in Figure 2.

Figure 2

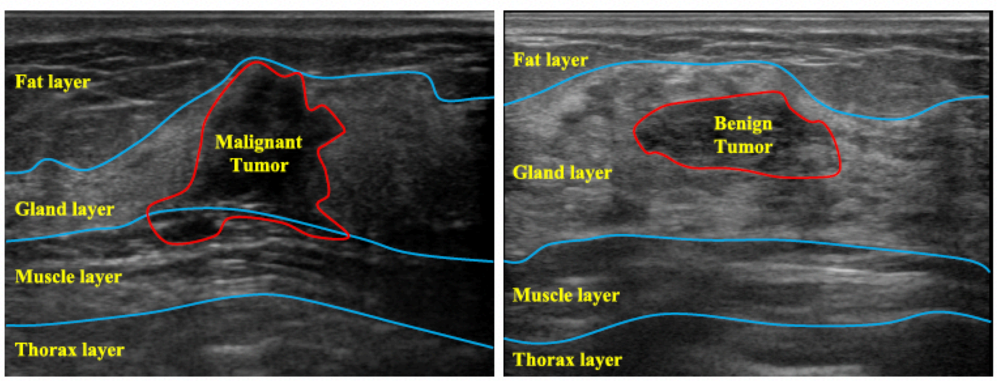
HoVer-Trans Network Diagram



The proposed model has four stages in its network architecture: a convolutional block, a flattening operation of multiple HoVer-Trans blocks, and a pooling layer (2022) in each stage. There are three types of embedding techniques: patch embedding, vertical strip embedding, and horizontal strip embedding. HoVer-Trans uses breast ultrasonography images to generate anatomical prior knowledge, which makes it possible to extract the relationships between and within the different anatomical layers of the breast. It was divided into four branches. The horizontal and vertical branches, respectively, were aimed at the extraction of the intra-layer and inter-layer relationships. The introduction of H2V and V2H branches combined the horizontal and vertical features. The output features from every branch in the preceding HoVer-Trans block were the input features for the following HoVer-Trans block. The convolution block links two nearby stages and adds inductive bias. As seen in Figure 3, the ultrasonography pictures display distinct layers of various breast tissues and highlight the distinctions between benign vs. malignant growths. This was accomplished by utilizing the breasts' anatomical structure and the concepts of ultrasound imaging. The lesion's size, location, morphological appearance, and spatial relationship with the different layers contributed to its malignancy.

Figure 3

Breast Ultrasound

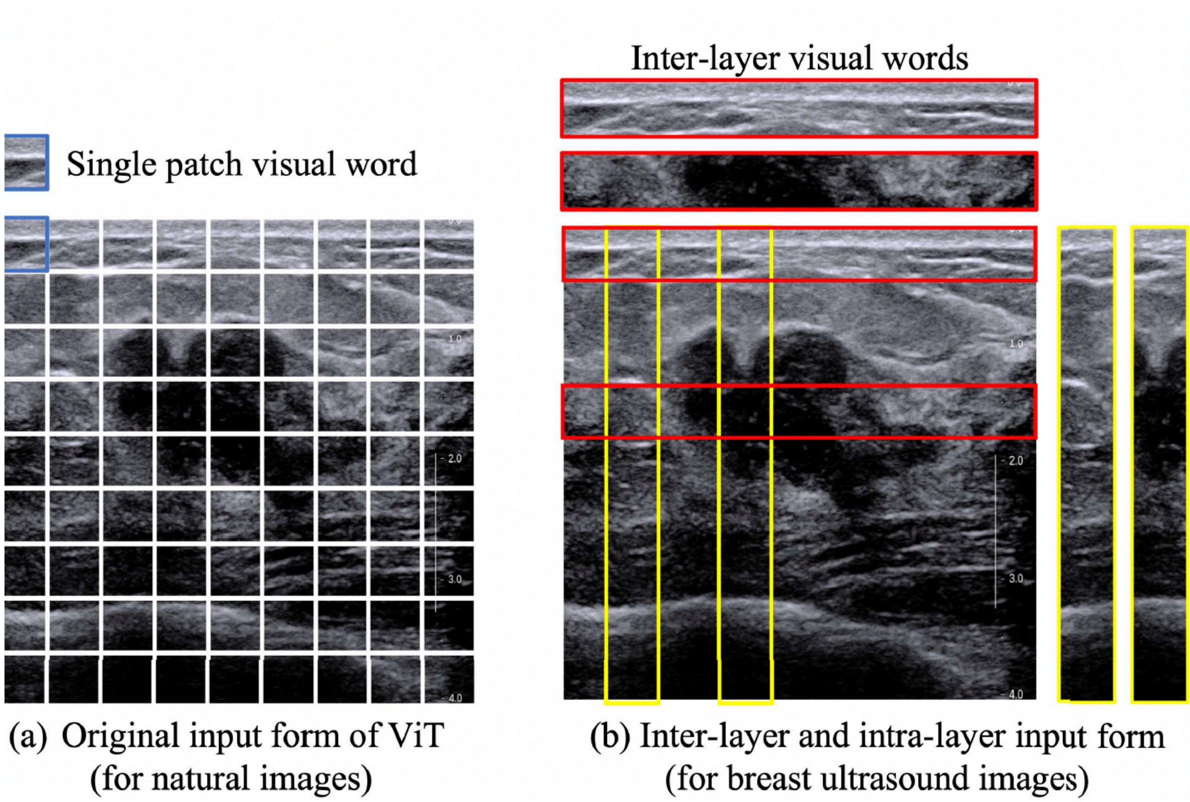


(a) Malignant tumor (a) Benign tumor

Conventional convolutional neural network models effectively extracted representative local features but required the ability to represent spatial relationships effectively. Because of this, most of the breast cancer diagnosis algorithms in ultrasound images required a pre-defined region of interest (ROI) of the lesion to eliminate redundant areas and allowed the CNN model to classify the ROI. It is clear from Figure 4(a) that the self-attention nature of the transformer resulted in the introduction of spatially solid relationships between each visual word. The problem was formulated by transforming the square-shaped visual words into horizontal and vertical strips to bring the anatomical prior knowledge into the model, as shown in Figure 4(b) (Mo et al., 2022). This allowed further exploitation of the intra-layer and inter-layer spatial correlations in Breast Ultrasound (BUS) images.

Figure 4

Transformer Results



The vision transformer (ViT) (Dosovitskiy et al., 2021) was the first technique to integrate the most widely used method in natural language processing into computer vision. It transforms input images into patches that are considered visual words.

As the input image was transformed into patches, it considered these patches the visual words (tokens), where the height (H), width (W), and column (C) were the set of elements.

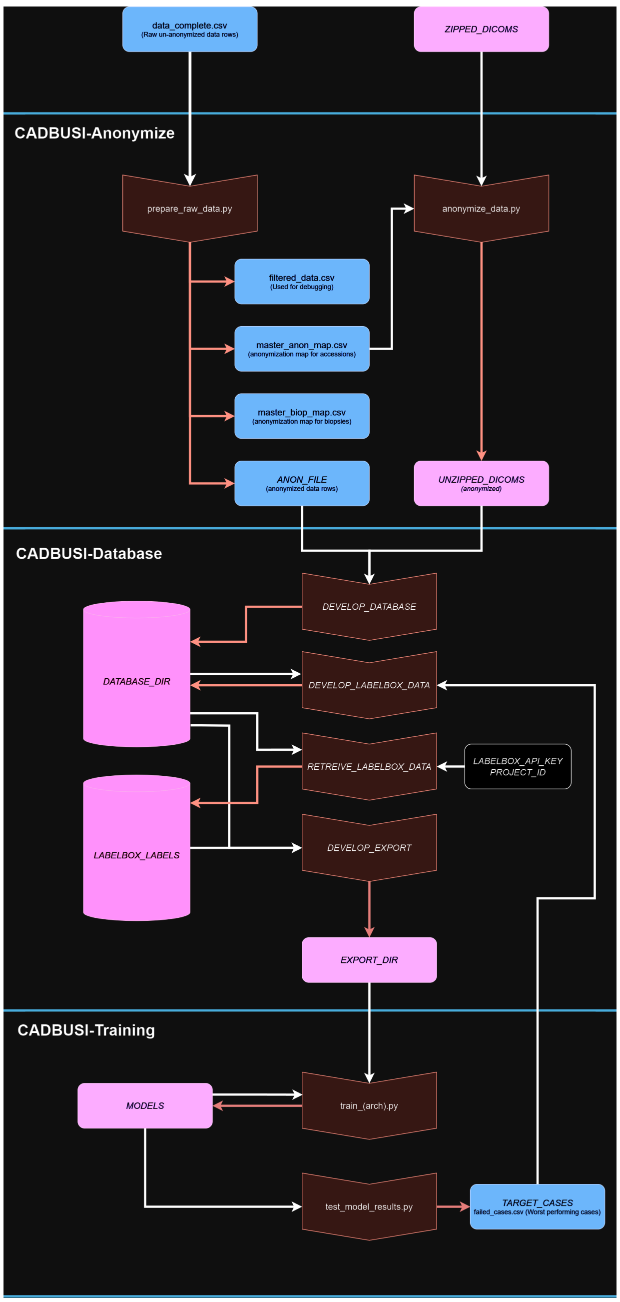
Figure 2 also illustrates the general framework of the model in its entirety. The structure was made up of four different stage modules. Several HoVer-Trans blocks, one Conv block, and one pooling layer were the components that make up each stage module. A convolutional stem was applied to a BUS image I ∈ RHxWx3 space for early visual processing. In contrast to the patchy stem of the original ViT, introducing early inductive bias through an early convolutional stem (Xiao et al., 2021) enhanced both the optimization stability and the model's performance. These sizes were comparable to the structure of the conventional convolutional neural network (He et al., 2016; Simonyan et al., 2015). It was necessary to introduce a Conv block to connect two adjacent stages to combine the horizontal and vertical information. The input for each stage was a two-dimensional image or a two-dimensional feature map. To fit the transformer's input, embedding or flattening was implemented. Finally, the fully connected layer was utilized for inference in the final stage. Through the use of cross-entropy loss, the model was optimized (Mo et al., 2022).

Implementation

Data, including 109,188 images, was received from the Mayo Clinic. A dedicated radiologist worked with the University of Wisconsin-La Crosse to prepare the images for processing. To date, 9,072 images have been biopsied and interpreted that contained purely malignant or purely benign images. The university team focused on using a multi-instance learning (MIL) model as most data needed to be interpreted, and most lesions have multiple images. Obtaining training data was problematic, and previous neural network models provided accuracies in the 70 - 80% range. Figure 5 (Poofy1, 2023) illustrates the process used to prepare the data. The author's task was to provide the core image location and lesion identification to augment the MIL model.

Figure 5

Mayo Data Pipeline



Python 3.6 and PyTorch 1.8 were used to implement the model. Four NVIDIA GeForce RTX 2080Ti GPUs with 11 gigabytes of memory each were used for every experiment. The embedding dimensions of each stage were {4, 8, 16, 32}, and the HoVer-Trans block numbers of each stage were {2, 4, 4, 2}. The model was constructed with these embedding dimensions. Two, four, eight, and sixteen were the head numbers of the transformer block in each stage. With a batch size of 64, a weight decay of 0.1, 10 warmup epochs, and an initial learning rate of 0.0001 using a cosine decay learning rate scheduler, the model was trained for 250 epochs using the AdamW optimizer (Kingma et al., 2014). Finally, a 5-fold cross-validation was used to evaluate the model on a limited data sample.

Additionally, blurring, noise, horizontal flipping, brightness, and contrast are all components of the augmentation strategy. Because the order of the tissue layers is predetermined, vertical flip data augmentation was not implemented. Every image was resized to 256 pixels by 256 pixels before processing.

Conclusion

In conclusion, the research conducted by the author at the University of Wisconsin-La Crosse presented a significant advancement in medical imaging and breast cancer diagnosis through the innovative use of the HoVer-Trans model. Building on the foundational concepts introduced by Mo et al. (2022), this model employs a unique approach by integrating anatomy awareness into the transformer architecture, thus enhancing the accuracy of breast cancer diagnosis using ultrasound images. By effectively leveraging the spatial relationships between tumors and anatomical layers and incorporating advanced AI techniques such as patch, horizontal and vertical embedding, and a convolutional block, the HoVer-Trans model sets a new standard in the precision and reliability of breast cancer detection.

The study demonstrated the technical capabilities and effectiveness of the HoVer-Trans model and emphasized the potential impact of integrating such advanced technologies into clinical practice. As exemplified by this research, adopting vision transformers in medical diagnostics offers the promise of more accurate, efficient, and accessible cancer diagnostics, which is crucial for timely treatment and improved patient outcomes. Furthermore, the study acknowledges the importance of ethical considerations in deploying AI technologies in healthcare, advocating for a balanced approach that maximizes benefits while addressing potential biases and equity issues.

As breast cancer remains a significant health challenge worldwide, this research's contributions are timely and valuable. They underscore the importance of interdisciplinary collaboration between AI research and clinical practice to drive innovations that can significantly enhance cancer diagnosis and treatment. The promising results obtained from the Mayo Clinic datasets suggest that further exploration and validation of the HoVer-Trans model in broader clinical settings could lead to its adoption as a standard tool in breast cancer diagnosis, ultimately contributing to the global effort to combat this disease. A link to the Python code and Jupyter Notebook can be found in Appendix A.

Chapter 4: Presentation of Research

Introduction

This chapter presents the findings of the HoVer-Trans Model application to the Mayo data. The methodology is discussed in the previous chapter. The HoVer-Trans model consisted of 9072 images representing 1526 patients. Specifically, there were images of 4537 right breasts and 4492 left breasts.

As a reminder, it is essential to produce accurate and timely results concerning breast cancer diagnoses. Inaccurate diagnoses and missed malignant lesions can have life-altering effects on individuals. The main point of interest is the model's ability to classify malignant and benign tumors accurately and consistently. In addition, the objectives of this project were:

- Improved model statistics compared to the existing Mayo studies (Jarvey, 2022).
- Improved model statistics compared to Mayo Radiologists.
- The development of an improved HoVer-Trans model that could accurately classify breast tissue lesions as benign or malignant.
- Various breast ultrasound images showed enhanced sensitivity and specificity in detecting malignancies.
- A substantial contribution to the field of medical imaging and cancer diagnosis was made, potentially leading to early detection and treatment of breast cancer.

The HoVer-Trans model is optimized by cross-entropy loss. Each of the five folds ran 250 epochs, and graphs were produced to evaluate changes in the key AUC-ROC statistics throughout the process. The results were as follows:

AUC

The Receiver Operator Characteristic (ROC) curve is an assessment metric for binary classification issues. A probability curve distinguishes the signal from the noise by plotting the sensitivity or True Positive Rate (TPR) (also called Recall) against one minus specificity or False Positive Rate (FPR) at different threshold values. Stated differently, it presents a classification model's performance across all classification thresholds. AUC, a summary of the ROC curve, measures a binary classifier's capacity to discriminate between classes. The models' ability to distinguish between the positive and negative classes is improved with a higher AUC. The likelihood that the classifier will be able to differentiate between positive and negative class values is high when the AUC is between .5 and 1. This is because there are more True positives and True negatives that the classifier can identify than False positives and False negatives. The classifier cannot distinguish positive and negative class points when AUC is .5. This indicates that the classifier predicts a constant or random class for every data point.

AUC is helpful in many areas, including credit scoring, medical diagnosis, and machine learning. It is best when:

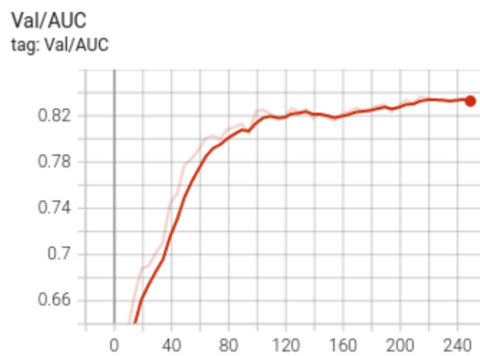
- False positive and false negative costs are highly dissimilar.
- There is an uneven distribution of classes.
- A single metric is needed to compare the performance of various models.

Although AUC offers a cohesive perspective on model performance, it does not indicate the choice threshold or provide information regarding model calibration. There might also be better options for multi-class classification issues.

After running five folds of 250 epochs, the best HoVer-Trans model produced an AUC of 0.813937. The evolution of the AUC can be viewed in Figure 6.

Figure 6

HoVer-Trans AUC evolution



F1 Score

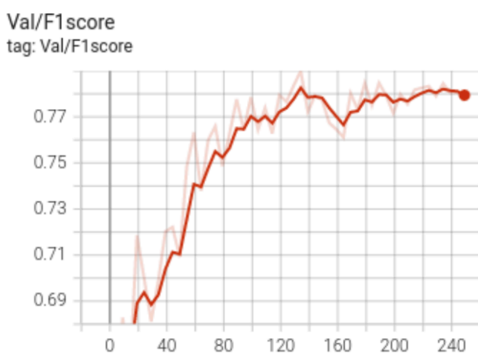
The model's accuracy was assessed using the F1 Score, a popular evaluation metric in classification tasks. It provides a balanced viewpoint on model performance by considering sensitivity and precision. Preciseness and sensitivity are its harmonic means. In contrast to sensitivity, which is the ratio of true positives to the sum of true positives and false negatives, precision is the ratio of true positives to the sum of true positives and false positives. F1 score is equal to 2 * (sensitivity * precision) / (precision + sensitivity).

A classification model's performance can be comprehensively assessed thanks to the F1 Score, which combines sensitivity and precision. The F1 Score considers both precision and sensitivity, making it easier to pinpoint the circumstances where a model can accurately detect positive instances (precision) and capture every positive instance (sensitivity). Achieving the maximum F1 Score for a given classification task denotes striking the ideal balance between sensitivity and precision. It is beneficial in cases where the distribution of classes could be more balanced.

After running five folds of 250 epochs, the best HoVer-Trans model produced an F1 Score of 0.779861. Figure 7 shows the evolution of the F1 Score.

Figure 7

HoVer-Trans F1 Score evolution



Accuracy

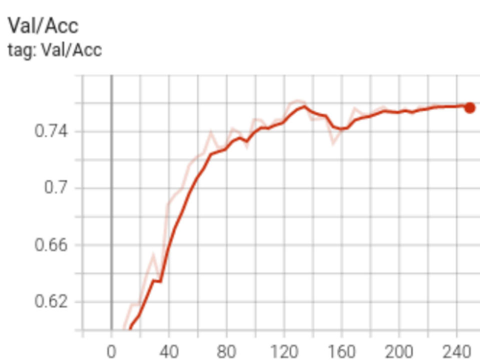
One metric to assess classification models is accuracy. The percentage of predictions our model correctly predicted is known as accuracy. Accuracy for binary classification can be determined in terms of positives and negatives: Accuracy is equal to *TP*+*TN* / *TP*+*TN*+*FP*+*FN*.

Ninety-one accurate predictions out of 100 total examples make up the accuracy of 0. 91 or 91 percent. That implies that a tumor classifier performs admirably in detecting cancers, correct? To get more insight into the performance of our model, let us examine the advantages and disadvantages in more detail. Ninety-one (one TP and eight FNs) and ninety-one (91 TNs and one FP) of the 100 tumor examples are malignant. The model correctly classifies 90 out of the 91 benign tumors as benign. That seems beneficial. However, only one of the nine malignant tumors is correctly identified by the model as malignant—a terrible result considering that eight of the nine malignancies remain undetected! Even though 91 percent accuracy might initially seem impressive in our examples, a different tumor-classifier model that consistently predicts benign would obtain the same accuracy (91/100 correct predictions). Put differently, the model possesses no greater predictive power than one that cannot differentiate between benign and malignant tumors. Working with a class-imbalanced data set such as this one, where the number of positive and negative labels differs significantly, means that accuracy alone does not provide the whole picture.

After running five folds of 250 epochs, the best HoVer-Trans model produced an Accuracy of 0.753455. Figure 8 shows the accuracy's evolution.

Figure 8

HoVer-Trans Accuracy evolution



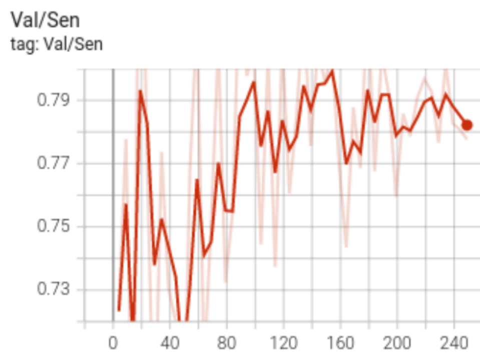
Sensitivity, Specificity, and Precision

The appealing feature of sensitivity and specificity is that they are independent of class prevalence. That is, sensitivity represents accuracy among true positives, while specificity represents true negatives. These metrics treat the real positives and negatives differently, making their relative proportions meaningless. Sensitivity and specificity are attributes of a specific test in the medical field determined by the test itself, regardless of the number of individuals undergoing it. This renders the test statistics independent of location and time. For example, a test administered to a population exhibiting a 90% prevalence of the ailment will possess identical levels of specificity and sensitivity when administered to a population with a 10% incidence. Contrarily, precision is dependent on class prevalence. It measures accuracy among predicted positives, but the number of individuals expected to be positive depends on the condition's prevalence. One precision value will be obtained if the test is administered to a population where the condition is 90% prevalent. However, the accuracy will be significantly reduced when the same test is conducted on a population with a mere 10% incidence due to the substantially higher proportion of true negatives than true positives. The probability that a positive test is accurate (precision) decreases with the real positive population. While precision is a feature of the test in a particular population, it is applied to specificity, a feature of a test independent of the population it is used to. Specificity is often the favored way to characterize a medical test because condition prevalence can vary over time by subpopulation or geographic location. A fixed test's specificity increases with decreasing condition prevalence but not precision.

After running five folds of 250 epochs, the best HoVer-Trans model produced a Sensitivity of 0.782953. Figure 9 shows the evolution of the Sensitivity.

Figure 9

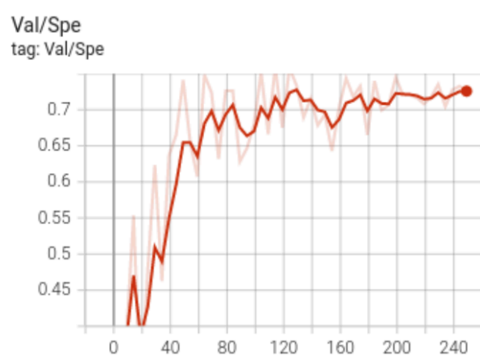
HoVer-Trans Sensitivity evolution



After running five folds of 250 epochs, the best HoVer-Trans model produced a Specificity of 0.716250. Figure 10 shows the evolution of the Sensitivity.

Figure 10

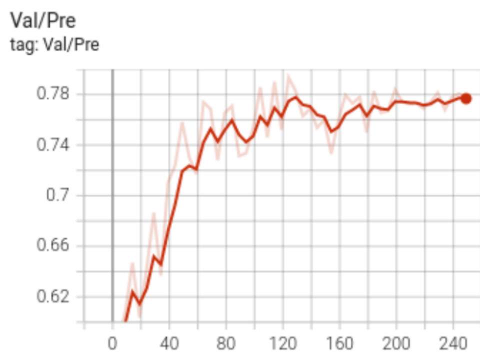
HoVer-Trans Specificity evolution



After running five folds of 250 epochs, the best HoVer-Trans model produced a Precision of 0.776794. Figure 11 shows the evolution of the Precision.

Figure 11

HoVer-Trans Precision evolution



Findings

Model Statistics

In review, the model chosen for this project was the HoVer-Trans model developed by Mo et al. (2022). This model employed a unique approach by capitalizing on the anatomical distinctions between malignant and benign tumors. Unlike its predecessors, the HoVer-Trans model is unique because it is anatomically aware. The superiority of the HoVer-Trans model over traditional CNN architectures and the diagnostic capabilities of experienced radiologists were demonstrated, representing a significant leap forward in ultrasound-based breast cancer detection (Mo et al., 2022). This project aimed to illustrate improved model statistics compared to Jarvey (2022), who used ResNet-34 to evaluate images.

The quantitative results and comparisons are presented in Table 1. The objective was to create a model using the HoVer-Trans methodology that outperformed current methods used on the Mayo data. The results from this study were compared to results obtained by Jarvey (2022), which included a ResNet-34 model and results from Mayo radiologists. Elements of comparison include:

- Accuracy is the total number of correct predictions divided by the number of data points.
- Sensitivity measures true positives, where one is the best, and 0 is the worst.
- AUC, or area under the curve, measures how well the model distinguishes between positive and negative classes.
- Precision measures the number of correct optimistic predictions divided by the number of positives, measured on a scale of 0 to 1.
- The F1 score is a statistical value representing the harmonic mean of precision and sensitivity.

The result of a study by Jarvey (2022) found that the radiologist had superior sensitivity (0.934595), a measure of the actual positive rate. However, when evaluating ultrasound images of breast lesions, there was a much lower AUC, meaning that the radiologist was not as good at distinguishing between malignant vs. benign tumors via ultrasound alone (Jarvey, 2022). The HoVer-Trans model outperformed the ResNet-34, and the radiologist's results in every element except sensitivity and specificity; the ResNet-34 model and the HoverTrans model were nearly equal.

Table 1

HoVer-Trans Model Comparison with ResNet-34 and Radiologist Results (Jarvery, 2022)

Model	AUC	F1 Score	Accuracy	Sensitivity	Precision	Specificity
Hover-Trans	0.813937	0.779861	0.753455	0.782953	0.776794	0.716250
ResNet-34	0.709624	0.743295	0.713675	0.788617	0.702898	0.729167
Radiologist	0.616128	0.727848	0.632478	0.934959	0.595854	0.804878

Conclusion

The HoVer-Trans model has significantly advanced breast cancer diagnostics by improving traditional convolutional neural network (CNN) architectures. Utilizing a dataset of 9072 images from 1526 patients, the model demonstrated superior performance with an AUC of 0.813937 and an F1 Score of 0.779861. These results surpass the previous studies at the Mayo Clinic and illustrate the potential for early detection and treatment of breast cancer, emphasizing the model's accuracy and reliability.

While the model outperformed in most metrics, challenges in sensitivity and specificity indicate areas for future improvement. The research underscores the importance of AI in enhancing medical imaging and sets the stage for further developments that could integrate such models into clinical workflows. This study highlights the transformative impact of AI technologies like the HoVer-Trans model in pursuing precision medicine.

Chapter 5: Discussion

Introduction

This study aimed to evaluate the effectiveness of the HoVer-Trans model proposed by Mo et al. (2022) on the Mayo Clinic's ultrasound image database, aiming to improve the interpretability and accuracy of breast cancer diagnosis. By providing a model that required minimal human intervention for accurate malignancy prediction, this research could significantly impact the field of computer-aided diagnosis (CAD). This chapter discusses the findings, limitations, and suggestions for further study.

Summary of Findings

The anatomically aware HoVer-Trans model was trained and tested using the Mayo Clinic data set, which included 9027 images of 1526 patients. Specifically, there were images of 4537 right breasts and 4492 left breasts. The aim was to improve upon the results of prior models. The HoVer-Trans model achieved 81.3% AUC and outperformed both the ResNet-34 model and a radiologist. The AUC represents how well the differences between malignant and benign lesions are distinguished, and the model could differentiate between the classes of lesions, either benign or malignant. The HoVer-Trans model also showed superior performance in accuracy (73.5%), sensitivity (78.3%), precision (77.7%), and the highest F1 score (.778). The HoverTrans model underperformed in terms of specificity (71.6%). Notable achievements of the model include the following:

- Significantly higher accuracy in classifying benign and malignant breast lesions.
- The reduction of reliance on manual ROIs minimizes human error and subjective diagnoses.

- Enhanced interpretability of imaging results can provide radiologists with better visual data to support their decision-making.

Interpretation of Findings

One of the study's objectives was to improve the results of Jarvey (2022), who compared the various models against a radiologist. In this study, the HoVer-Trans Model was applied to a much more extensive database of images. Previously, Jarvey (2022) used a dataset of 328 fully integrated images. Currently, the Mayo dataset includes 9072 fully interpreted images, all used in this study. The study's results indicate that the HoVer-Trans model outperformed the ResNet-34 model in everything except specificity and the radiologist in everything except sensitivity. The ResNet-34 model had a specificity of 73% vs. the HoVer-Trans model of 72%. This difference is very minimal. However, it demonstrates that the ResNet-34 model was slightly better at identifying true negatives. The most exciting finding was that the radiologist study (Jarvey, 2022) vs. the HoVer-Trans model in this study had much greater sensitivity, 93%, and 78%, respectively. The HoVer-Trans model was the worst performer in this indicator. The ResNet-34 model demonstrated a sensitivity of 79%. This means that the radiologist was the best at identifying true positives.

Another objective of this study was to develop an improved HoVer-Trans model that could accurately classify breast lesions as benign or malignant. The measurement of correct classification is the AUC. The HoV-Trans model yielded an AUC of 81%. This is higher than the ResNet-34 model and the radiologist in Jarvey’s (2022) study. One possible explanation for the higher AUC is that the HoVer-Trans model is anatomically aware, and the ResNet-34 model is not. In addition, the radiologist introduced human error, which could account for the lower results.

The Hover-Trans model's goal of enhancing sensitivity and specificity in detecting malignancies was partially met. The ResNet-34 model showed better sensitivity and specificity, but very minimally.

An unintended finding was the performance of the HoVer-Trans model in this study compared with the conclusions of Mo et al. (2022) using a HoVer-Trans model that they optimized. Using the GDPH&SYSUCC database, Mo et al. achieved an AUC of 92.4% (figure 1). Using the same database with the HoVer-Trans model in this study, an AUC of 81.4% was achieved. It is unknown why there was underperformance at this time, but overfitting and tuning issues could be the source of the difference. These will be discussed in the limitations section of this report.

The findings of this study contribute to the validation of introducing AI into radiologic diagnoses concerning breast ultrasounds. The HoVer-Trans model’s performance in correctly classifying breast lesions confirms the benefit of using an anatomy-aware model rather than one that is not. This capability is crucial as AI technology emerges as a standard in diagnostic medicine.

Implications of Findings

The results of this project support the assertions that machine learning and AI models can significantly enhance diagnostic processes. In particular, an anatomically aware model, such as the HoVer-Trans model, demonstrated a superior ability to distinguish between benign and malignant tumors.

From a methodological standpoint, the success of the HoVer-Trans model signals a shift from traditional CNN models to transformer models. The evidence gained from this study shows how a transformer model is the next logical step in AI and predictive analytics in diagnostic medicine.

Practically, the application of AI models is beneficial in many ways. Reducing misdiagnoses improves patient safety. In addition, earlier and more precise detection capabilities will significantly enhance the timeliness of planning appropriate treatment and improve patient outcomes.

Context of Findings

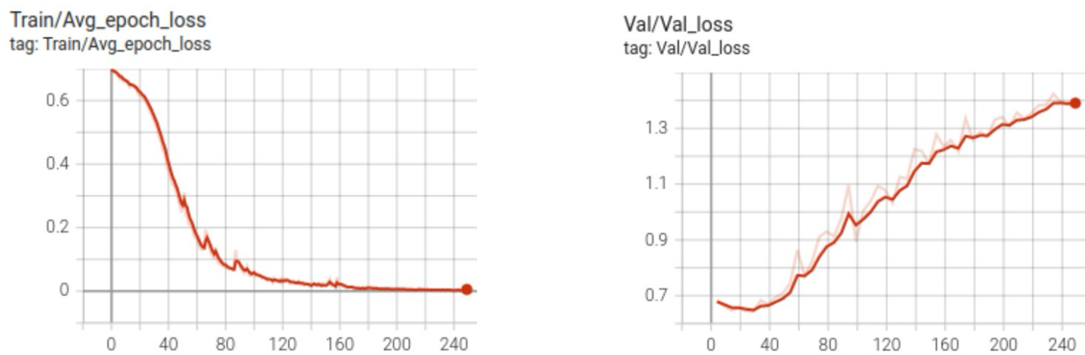
This study's findings are similar to those of Mo et al. (2022). In Mo's study, the HoVer-Trans model outperformed all other models in AUC, specificity, precision, recall, and score. It did not outperform the models in specificity. Similarly, the HoVer-Trans model in this current study had the worst specificity when compared to the ResNet-34 model and with a radiologist (Jarvey, 2022). In addition, both models had superior AUCs compared to others.

Limitations

This study's limitations include model fitting. The HoVer-Trans model showed signs of overfitting. This means the model needs to be generalized better to the validation data, and training can be improved. Figure 12 illustrates the issues with overfitting in this project.

Figure 12

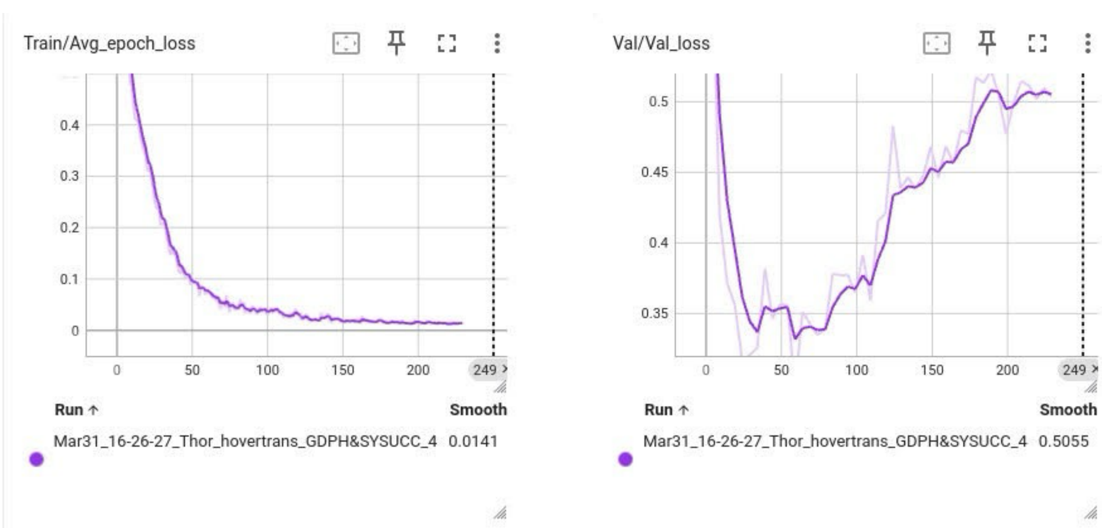
Overfitting of Mayo Data



Interestingly, Mo et al. (2022), who provided much of the framework for this study, also showed overfitting. Figure 13 illustrates the issue of overfitting in their research.

Figure 13

Overfitting of GDPH&SYSUCC Data (Mo et al., 2022)



Overfitting means the model produces accurate predictions on the training but not on the final data. This is an important finding because an overfit model can predict errors. Overfitting can occur because of noise in the data set, small sample sizes for the training data set, or if the model is complex, it can learn noise in the training data. Data augmentation can help overcome the overfitting problem. Examples of data augmentation are as follows: Increase the diversity of the training data, such as adding rotations, translation, scaling, or noise to challenge the model. Next, reduce the complexity of the model. Sometimes, a complex model must be more significant for the available training data.

Another limitation of the study was the condition of the images. The images varied in size, making it difficult for the model to identify structures accurately. Since the HoVer-Trans model is anatomy-aware, this was problematic.

There were differences in data sets and study parameters between Jarvey (2022) and Mo et al. (2022), and this study included data sets, image types, and differences in model training. Therefore, results cannot be generalized across studies.

Most importantly, the performance of the HoVer-Trans Model on Mayo Data was worse than expected compared to the performance on GDPH&SYSUCC data. Tuning methods to correct the overfitting and other restrictions should be pursued.

Future Directions

Several tuning procedures may benefit future research projects using the HoVer-Trans model. The issues with overfitting could be addressed in several ways.

Tuning to Correct Overfitting

Segmentation

Using broader or alternative data augmentation approaches can increase the diversity of the training data. Consider rotations (when appropriate), scaling translation, or adding noise for a model trained on images like the HoVer-Trans model. To guarantee precise predictions, ensure the augmentations do not change critical anatomical characteristics.

Regularization

Use regularization strategies like dropout L1 or L2 regularization or step up their use. Modify the dropout rate in each model layer to avoid putting too much emphasis on a particular neuron and encourage a more universal learning pattern.

Model Simplification

If the models' complexity is greater than the quantity of training data available, lower it. This could involve reducing the number of layers or units per layer, which can reduce the model's capacity to memorize the training data.

Early Termination

Employing early stopping during training entails monitoring the model's performance on a validation set and pausing training when it starts to deteriorate or improve significantly. By doing this, the model is shielded from noise and training data specifics that are not transferable to new data, utilizing cross-validation. To assess the model's performance more reliably, employ cross-validation methods. This entails training multiple models by dividing the training data into smaller sets. This method lessens the possibility of coincidental strong performance on a single test set and aids in understanding how the model functions across various data subsets.

Add Additional Information

Expanding the dataset size can be beneficial. With more data, the model can learn from a wider variety of examples, enhancing its capacity for generalization.

Modify the Learning Rate and Training Duration

Changing the learning rate and training duration can also aid in the management of overfitting. A more robust model with better generalization can result from a lower learning rate, even though it may slow down the training process.

Use of Batch Normalization

Batch normalization, which serves as a regularization and can assist in reducing internal covariate shift, can reduce overfitting. It is recommended to repeatedly contemplate and test these strategies to determine the ideal configuration that minimizes overfitting and preserves or enhances the model's performance on the new dataset.

Other Tuning Recommendations

Several other tuning methods can be used to gain performance. The following are some recommendations.

Recognize the Model Architecture

The HoVer-Trans model borrows features from the Vision Transformer (ViT) architecture, adding mechanisms to allow the model to incorporate anatomical information from ultrasound images. The architecture combines horizontal and vertical strip embeddings to exploit spatial correlations between tissue layers.

Preprocessing the Input

Make sure the new dataset images are formatted correctly for the model. This may mean resizing the photos to 256 by 256 pixels and using suitable augmentation techniques that do not include vertical flips to preserve the integrity of the anatomical structure.

Change Hyperparameters

Embedding parameters may need to be adjusted, such as the number of blocks and heads in each transformer stage and the training schedule, which includes learning rates and epochs based on the features of the new dataset and the model's performance during the first training cycles.

Changes to Feature Extraction

The original model uses three types of embeddings: patch, horizontal strip, and vertical strip. Depending on the details of the new dataset, experimenting with the sizes of these embeddings to determine the best setup for novel image types will help optimize the model.

Transfer Learning and Fine-Tuning

Retrain the model using the anatomy-aware formulation intended for ultrasound images. Depending on how similar the new dataset is to the original, this may entail retraining only the last few layers or the entire network.

Integration of Anatomical Knowledge

If the new dataset contains images with anatomical structures that differ from breast ultrasound images, it must modify or remodel the model's anatomy-aware components to capture pertinent spatial correlations.

Evaluation and Iteration

Measures such as accuracy sensitivity specificity and area under the ROC curve can be used to assess the performance of the optimized model. Considering that the model was initially intended to perform better than conventional models and even skilled sonographers, comparable standards ought to be set for the new dataset, with objectives being modified in light of early findings.

Visualization and Interpretability

Use attention maps and other model interpretability features to assess the model's performance in identifying and diagnosing features in the new dataset. Pay close attention to the boundaries and properties of lesions or other anatomical features.

Clinical Validation

Please verify that the model satisfies all regulatory and validation requirements before implementing it in a clinical setting. Pay particular attention to the models' accuracy and reliability compared to current diagnostic techniques.

Conclusion

This study aimed to apply the HoVer-Trans model to the Mayo Clinic breast ultrasound data to test the model results compared to Jarvey (2022) and Mo et al. (2022). The results showed that the HoVer-Trans model demonstrated greater AUC, which could correctly classify breast lesions as benign or malignant. However, it did not perform better in sensitivity than a radiologist, as shown in Jarvey's (2022) results. Improved model optimization of the HoVer-Trans model could produce more precise and accurate results than this study's results. Future research on AI in breast cancer diagnosis using machine learning and AI should include continuing refinement of anatomy-aware models in general tuning and tuning to avoid overfitting, especially as larger datasets become available. Early detection and rapid intervention for breast cancer is essential for patient care and optimal outcomes. Machine learning and AI can contribute significantly to diagnostic medicine in terms of breast cancer.

The knowledge gained from studies using AI and machine learning in a specific cancer diagnosis can eventually be applied to other types of cancer, thereby improving patient care overall.

References

Abdul Halim, A.A.; Andrew, A.M.; Mohd Yasin, M.N.; Abd Rahman, M.A.; Jusoh, M.; Veeraperumal, V.; Rahim, H.A.; Illahi, U.; Abdul Karim, M.K.; Scavino, E. Existing and emerging breast cancer detection technologies and its challenges: A review. *Applied Science, 11*, https://doi.org/10.3390/app11221075

American Cancer Society. (2024). Breast Cancer Facts & Figures 2022-2024. Retrieved from https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/2022-2024-breast-cancer-fact-figures-acs.pdf

Andrei, F. (2022). *Semantic Segmentation for Medical Ultrasound Imaging* [Unpublished master's capstone]. University of Wisconsin – La Crosse.

Azad, R., et al. (2023). Medical image analysis with vision transformers. *Journal of Advanced Research in Medical Image Analysis*.

Ayana, G., Dese, K., Dereje, Y., Kebede, Y., Barki, H., Amdissa, D., Husen, N., Mulugeta, F., Habtamu, B., & Choe, S.-W. (2023). Vision-Transformer-Based Transfer Learning for Mammogram Classification. *Diagnostics*, 13(178), 1-16. https://doi.org/10.3390/diagnostics13020178

Bodart, T. M. (2022). *Computer-Aided Diagnosis (CAD) of Breast Cancer: Methods of Model Explainability* [Unpublished master's capstone]. University of Wisconsin – La Crosse.

BreastCancer.org. (2024). Breast Cancer Facts and Statistics. https://www.breastcancer.org/facts-statistics

Cancer.Net. (2024). Breast Cancer: Statistics. Retrieved from https://www.cancer.net/cancer-types/breast-cancer/statistics

Chen, L. (2023). Breast Cancer Detection Using Hybrid Deep Learning Models. *AI in Medicine Journal*, 1(1), 100–110. https://doi.org/10.1001/aimj.2023.01.100

Dosovitskiy, A. et al. (2021). *An image is worth 16x16 words: Transformers for image recognition at scale*. Cornell University. https://arxiv.org/abs/2010.11929

Eustaqui, H., da Silva, C., Santos, G., Leite, A.F., (2023). The use of artificial intelligence tools in cancer detection compared to the traditional imaging methods: An overview of the systematic reviews. *PLOS ONE*. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0292063

Fischer, S., et al. (2023). AI-Driven Technologies in Breast Cancer Research. *Journal of Translational Medicine*, 21, Article 42. https://doi.org/10.1186/s12967-023-03560-6

Gupta, R. (2021). The Role of Deep Learning in the Evolution of Mammography. *Breast Cancer Research*, p. 23, Article 112. https://doi.org/10.1186/s13058-021-01412-7

Hall, J. (2021). *Classification of Breast Lesions using Deep Learning* [Unpublished master's capstone]. University of Wisconsin – La Crosse.

He, K., Zhang, X., Ren, S., and Sun, J., (2016). *Deep residual learning for image recognition*. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR).

Hossain, A. A., Nisha, J. K., & Johora, F. (2023). Breast cancer classification from ultrasound images using VGG16 model-based transfer learning. *International Journal of Image, Graphics and Signal Processing, 15*(1). https://dx.doi.org/10.5815/ijigsp.2023.01.02

Hayes, D., et al. (2021). Innovations in Breast Cancer Detection Using Computer Vision. *Computer Methods and Programs in Biomedicine*, p. 200, Article 105933. https://doi.org/10.1016/j.cmpb.2021.105933

Horowitz, B. T. (2023). The current state of AI in healthcare and where it is going in 2023. Data Analytics. Retrieved from https://healthtechmagazine.net/article/2022/12/ai-healthcare-2023-ml-nlp-more-perfcon

Jarvey, J. L. (2022). *Computer-Aided Detection (CAD) System for Breast Ultrasound Lesion Interpretation: An Explainable Deep Learning Approach* [Unpublished master's capstone]. University of Wisconsin – La Crosse.

Jenkins, R., et al. (2022). The Impact of Data Science on Breast Cancer Outcomes. *Data Science in Medicine*, 4(2), 155–163. https://doi.org/10.1002/dsm2.2022.00042

Johnson, R., & Kumar, S. (2023). Spatial Relationship Analysis in Ultrasound Imaging for Enhanced Tumor Classification. *International Journal of Medical Imaging, 45*(2), 250–265.

Kim, S. (2023). Advances in Non-Invasive Breast Cancer Screening Methods. *Cancer Prevention Research*, 16(3), 134–143. https://doi.org/10.1158/1940-6207.CAPR-22-0407

Kingma, D. P. and Ba, J. (2014). *Adam: A method for stochastic optimization.* Cornell University. https://arxiv.org/abs/1412.6980.

Knoll, M. A. (2023). A new study shows how AI can help detect breast cancer. *Forbes*. https://www.forbes.com/sites/miriamknoll/2023/08/10/new-study-shows-how-ai-can-help-detect-breast-cancer/?sh=6d3e4d725185

Lee, H. (2023). AI and the Future of Breast Cancer Imaging. *Future Oncology*, 19(8), 955–965. https://doi.org/10.2217/fon-2022-0119

Li, Y. (2023). Comparative Analysis of Machine Learning Models for Breast Cancer Detection. *Journal of Clinical Oncology*, 41(11), 2951–2963. https://doi.org/10.1000/jco.2023.11.2951

Mo, Y., Han, C., Liu, Y., Liu, M., Shi, Z., Lin, J., Zhao, B., Huang, C., Qiu, B., Cui, Y.-h., ... Liang, C. (2022). HoVer-Trans: Anatomy-aware HoVer-transformer for ROI-free breast cancer diagnosis in ultrasound images. *IEEE Transactions on Medical Imaging*. https://dx.doi.org/10.1109/TMI.2023.3236011

Mohan, S. P. (2022). *AI-Based Breast Cancer Classification Using Ultrasound Images* [Unpublished master's capstone]. University of Wisconsin – La Crosse.

Morrison, C. et al. (2022). Deep Learning in the Detection of Breast Cancer. *Journal of Biomedical Informatics*, p. 128, Article 103973. https://doi.org/10.1016/j.jbi.2022.103973

Narvaez, F. (2021). Impact of AI on Breast Cancer Diagnosis and Screening. *Journal of Medical Systems*, p. 45, Article 52. https://doi.org/10.1007/s10916-021-01708-w

National Breast Cancer Foundation. (2023). Breast cancer stats and facts. Retrieved from https://www.nationalbreastcancer.org/breast-cancer-facts.

O'Connor, M., et al. (2023). Enhancing Early Breast Cancer Detection with Machine Learning. *Healthcare Informatics Research*, 29(1), 12–20. https://doi.org/10.4258/hir.2023.29.1.12

Patel, S. (2022). Utilizing MRI and Ultrasound Data to Improve Early Detection of Breast Cancer. *Radiology*, 298(2), 450–462. https://doi.org/10.1000/radiol.2022.02.450

Poofy1. (2023). *CADBUSI-Database [Software]. GitHub. https://github.com/Poofy1/CADBUSI-Database.*

Saad, M., Ullah, M., Afridi, H., Cheikh, F. A., & Sajjad, M. (2022). BreastUS: Vision transformer for breast cancer classification using breast ultrasound images. https://dx.doi.org/10.1109/SITIS57111.2022.00027

Silberfein, A. (2021). *Computer-Aided Diagnosis (CAD) of Breast Cancer from Ultrasound Images* [Unpublished master's capstone]. University of Wisconsin – La Crosse.

Simmons, B., et al. (2022). Machine Learning's Role in Personalized Breast Cancer Treatment. *Personalized Medicine*, 19(3), 207–216. https://doi.org/10.2217/pme-2021-0138

Simonyan, K. & Zisserman, A. (2015). *Intense convolutional networks for large-scale image recognition*. Proc. Int. Conf. Learn. Represent.

Smith, J., et al. (2022). Utilizing AI for Enhanced Breast Cancer Prognosis. *Journal of Oncology Practice*, 18(6), e764-e772. https://doi.org/10.1200/JOP.2021.001764.

StartUs Insights. (2023). Top 10 AI Trends in Healthcare. https://www.startus-insights.com.

Stavros, T. A. (2004). *Breast Ultrasound.* Lippincott, Williams, and Wilkins.

Thompson, R., et al. (2021). Computational Approaches to Breast Cancer Prediction and Treatment. *Computational and Structural Biotechnology Journal*, pp. 19, 4977–4988. https://doi.org/10.1016/j.csbj.2021.09.029

Torres, P., et al. (2022). Machine Learning Algorithms for Predicting the Risk of Breast Cancer Recurrence. *Breast Cancer*, 29(4), 619-628. https://doi.org/10.1007/s12282-022-01234-5

Wang, J. (2022). Innovations in Breast Imaging Techniques with AI Integration. *Journal of Healthcare Engineering*, 2022, Article e930482. https://doi.org/10.1155/2022/930482

World Cancer Research Fund International (2020). Breast cancer statistics. https://www.wcrf.org/cancer-trends/breast-cancer-statistics/

Xiao, T., Singh, M., Mintun, E., Darrell, T., Dollar, P., & Girshick, R. (2021). *Early convolutions help transformers see better.* Advances in Neural Information Processing Systems. https://proceedings.neurips.cc/paper/2021/file/ff1418e8cc993fe8abcfe3ce 2003e5c5-Paper.pdf

Xie, Q. (2022). Enhanced Breast Cancer Detection Using Deep Learning Networks. *Journal of Medical Imaging and Health Informatics*, 12(5), 1234–1245. https://doi.org/10.1000/jmihi.2022.05.1234

Zhang, H. (2021). AI in Breast Cancer Imaging: Challenges and Applications. *Medical Imaging Technology*, 39(3), 202–212. https://doi.org/10.1000/mit.2021.03.202

Zhang, H; Lie, Y. (2023). Apply deep learning to medical imaging: A review. *Applied Science, 13(*18). https://www.mdpi.com/2076-3417/13/18/10521

Appendix A

Python Code

https://drive.google.com/file/d/1DIV2_38YlED5wCWfCA52nWR_XqQpm-Tr/

Jupyter Notebook

https://drive.google.com/file/d/1Cro9xuYADc88j1Sq_WomcOBAyFTE9jNG/