T.C.

TOKAT GAZİOSMANPAŞA UNIVERSITY

GRADUATE EDUCATION INSTITUTE

DEPARTMENT OF COMPUTER ENGINEERING

M.SC RESEARCH

BREAST CANCER DETECTION USING CONVOLUTIONAL NEURAL NETWORK

Prepared by

Ragheed Idwer Bahjat Mokhtar

Department of Computer Engineering

Master Thesis

Advisor: Assoc. Prof. Dr. Mahir KAYA

TOKAT
2025

ETHICAL AGREEMENT

I hereby state that this Master's thesis entitled "Breast Cancer Detection Using Convolutional Neural Network" was submitted in accordance with the advice and principles of Tokat Gaziosmanpaşa University Graduate Education Institute during the thesis writing process under the esteemed supervision of Assoc. Prof. Dr. Mahir KAYA is my original work, conducted fully in compliance with the principles of scientific ethics and integrity. I hereby declare that if this work is found to violate any of these ethical standards, I am fully prepared to accept whatever legal or academic ramifications such a determination may entail.

…./…./…….

Ragheed Idwer Bahjat

JURY ACCEPTANCE AND APPROVAL

The defense examination of the thesis titled “Breast Cancer Detection Using Convolutional Neural Network” prepared by Ragheed Idwer Bahjat Mokhtar, was held on 07/08/2025 and was accepted as a Master's Thesis by the jury given below, unanimously, in the Graduate Education Institute of Tokat Gaziosmanpaşa University, Department of Computer Engineering.

Jury Members	**Signature**
Member:	
Member:	
Member:	

APPROVAL ……. /……/…….

ÖZET

EVRİŞİMSEL SİNİR AĞLARI KULLANILARAK MEME KANSERİ TESPİTİ

Ragheed Idwer Bahjat Mokhtar

Bilgisayar Mühendisliği Bölümü

Danışman: Doç. Dr. Mahir KAYA

60 sayfa

Meme kanseri, dünya çapında kadınları etkileyen ve hayati tehlike arz eden en yaygın malignitelerden biri olmaya devam etmektedir. Bu nedenle, hayatta kalma oranlarının ve tedavi sonuçlarının iyileştirilmesi için zamanında ve doğru tanı müdahaleleri gerekmektedir. Ultrason görüntüleme, non-invaziv olması, maliyet etkinliği ve özellikle mamografi ve MRI'nin kolayca ulaşılamadığı kaynakları kısıtlı ortamlarda yoğun meme dokularını görüntüleme konusunda etkinliği nedeniyle kritik bir tanı aracı olarak işlev görmektedir. Ultrason görüntüleri, duyarlılık ve özgüllük açısından sınırlamaları olan, genellikle değişken tanı yorumlarına tabidir. Bu sorunlar, bu tür akıllı otomatik karar destek sistemlerine olan ihtiyacı vurgulamaktadır. Meme ultrason görüntülerinin sınıflandırılması için, özenle optimize edilmiş, tamamen özelleştirilmiş yeni bir Konvolüsyonel Sinir Ağı (CNN) mimarisi öneriyoruz. Bu çalışmanın sınıflandırması, klinik olarak önemli üç kategoriyi içermektedir: normal, malign ve benign. Meme Ultrason Görüntüleri (BUSI) veri setini kullandık ve küçük veri seti boyutu, sınıf dengesizliği ve görüntü gürültüsü gibi içsel sorunları, döndürme, çevirme, çeviri, kesme dönüşümü ve parlaklık ayarı gibi çeşitli artırma işlemleriyle çözdük. Model, gerçek dünyadaki varyasyonlar arasında etkili bir şekilde genelleme yapabildiğinden, elde edilen 6,000 görüntülük veri seti, eğitim verilerinin çeşitliliğini ve temsil gücünü büyük ölçüde artırdı. Önerilen CNN mimarisi, hiyerarşik özellik çıkarma, toplu normalleştirme, bırakma düzenlemesi ve ReduceLROnPlateau zamanlayıcı aracılığıyla uyarlanabilir öğrenme oranlarını içeren en son teknoloji derin öğrenme tekniklerini entegre eder. En iyi performans gösteren model, %98.83 makro ortalama F1 puanı ve test doğruluğu elde etti. Bu sonuç, öğrenme oranları, toplu boyutları ve giriş çözünürlükleri arasında geniş kapsamlı hiperparametre ayarlaması sayesinde elde edildi. Önceden var olan bilgi olmadan eğitilmiş olmasına rağmen, model MobileNetV2, VGG16, VGG19, InceptionV3, ResNet50, EfficientNetB0, DenseNet121 ve Xception gibi popüler transfer

öğrenme modelleriyle karşılaştırıldığında daha iyi performans gösterdi. Bu sonuçlara göre, hesaplama verimliliği, yorumlanabilirlik ve yüksek teşhis doğruluğu, alana özgü, hafif CNN modelleriyle elde edilebilir. Bulgularımıza göre, özel olarak tasarlanmış derin öğrenme çözümleri kritik bir rol oynamakta ve özellikle kaynakların kısıtlı olduğu ortamlarda erken meme kanseri teşhisinde ve tıbbi görüntülemede dönüştürücü potansiyellerini ortaya koymaktadır.

Anahtar Kelimeler: Meme Kanseri, Meme Ultrason Görüntüleme (BUSI), Evrişimli Sinir Ağları (CNN'ler), Veri Artırma, Sınıflandırma.

ABSTRACT

BREAST CANCER DETECTION USING CONVOLUTIONAL NEURAL NETWORK

Ragheed Idwer Bahjat Mokhtar

Department of Computer Engineering

Advisor: Assoc. Prof. Dr. Mahir KAYA

60 pages

Breast cancer continues to be one of the most popular as well as life-threatening malignancies because it affects women globally, demanding timely along with accurate diagnostic interventions for improved survival rates and therapeutic outcomes. Ultrasound imaging acts as a critical diagnostic tool because of its non-invasiveness as well as cost-effectiveness, plus efficacy when visualizing dense breast tissues, especially in low-resource settings where mammography and MRI may not be readily available. Ultrasound images get diagnostic interpretation that is customarily variable, with limitations in sensitivity and specificity. These issues do underscore the need for such smart automated decision-support systems. For breast ultrasound image classification, we propose a novel fully customized Convolutional Neural Network (CNN) architecture carefully optimized. This study's classification includes three clinically important categories: normal, malignant, and benign. We used the Breast Ultrasound Images (BUSI) dataset, and we tackled intrinsic issues like small dataset size, class imbalance, and image noise via a diverse augmentation pipeline, which included rotation, flipping, translation, shear transformation, and brightness adjustment. Because the model could generalize effectively across variations of the real world, the resulting dataset of 6,000 images greatly improved the diversity and representativeness of training data. The proposed CNN architecture integrates state-of-the-art deep learning techniques that include hierarchical feature extraction, batch normalization, dropout regularization, along with adaptive learning rates through a ReduceLROnPlateau scheduler. The best performing model did achieve 98.83% macro-averaged F1-score and also test accuracy. This result was thanks to wide-ranging hyperparameter tuning across learning rates, batch sizes, and input resolutions. Although trained without pre-existing knowledge, the model showed better performance when evaluated against popular transfer learning models like MobileNetV2, VGG16, VGG19, InceptionV3, ResNet50, EfficientNetB0, DenseNet121, and Xception. Computational efficiency, interpretability, as well as high diagnostic accuracy can be achievable via domain-specific, lightweight CNN models according to these results. According to our findings, tailored deep learning solutions play a critical role, highlighting their transformative potential in early breast cancer detection and in helping medical imaging, especially inside resource-constrained environments.

Keywords: Breast Cancer, Breast Ultrasound Imaging (BUSI), Convolutional Neural Networks (CNNs), Data Augmentation, Classification.

ACKNOWLEDGMENT

First and foremost, I would like to express my deepest gratitude to Assoc. Prof. Dr. Mahir Kaya, my supervisor, for his continuous support, with indispensable guidance, perceptive feedback throughout this research. His expertise in it, his patience with it, and his encouragement of it shaped the direction of, and improved the quality within this thesis. I do also extend my most sincere thanks to all faculty as well as staff of the Department of Computer Engineering at Tokat Gaziosmanpaşa University. An outstanding academic environment with resources needed for this research was indeed provided by them.

My sincere appreciation goes toward my family because they have given constant support, love, and belief to me during every stage of this adventure. They gave encouragement to me giving me strength also motivating me to overcome challenges that I encountered. I wish to acknowledge the developers and researchers who set the stage and worked for this study. They contributed to deep learning, to medical imaging, also to computer vision. Many people in these fields were inspired by their work. I remain sincerely grateful toward the collective support of all those mentioned above that made this research possible.

DEDICATED TO

Your firm love and your sacrifice and also encouragement are all of why my adventure has been so strong. Thank you, dear parents, for imparting into me the true value of perseverance. I appreciate your lessons in curiosity and integrity too. Belief within my potential has helped me in dreaming past all limitations and in overcoming each challenge with courage.

Your constant support and understanding toward my siblings and loved ones, who stood by me through difficult moments and long nights, have meant more than words can express. You accomplish this as well as I do.

You guided me with wisdom as well as believed in the power of education, shaping my thinking. Your actions inspired the contributions from me that were meaningful for my field. Knowledge has value as a responsibility beyond a goal, your example has taught me, not just as a goal.

CONTENTS

TABLES LIST

FIGURES LIST

LIST OF ABBREVIATIONS

AI	: Artificial Intelligence
AUC	: Area Under the Curve
BC	: Breast Cancer
BUSI	: Breast Ultrasound Images
CAD	: Computer-Aided Diagnosis
CNN	: Convolutional Neural Network
CT	: Computed Tomography
DL	: Deep Learning
MRI	: Magnetic Resonance Imaging
ResNet	: Residual Network
TL	: Transfer Learning
TP	: True Positives
TN	: True Negatives
FP	: False Positives
FN	: False Negatives
VGG	: Visual Geometry Group
DenseNet121	: Densely Connected Convolutional Network (121 layers)
VGG16	: Visual Geometry Group Network (16 layers)
Xception	: Extreme Inception
ResNet50	: Residual Neural Network (50 layers)
MobileNet	: Mobile Neural Network

1. INTRODUCTION

Breast cancer remains one of the most difficult health issues impacting women around the world, especially considering that early detection is the single most important factor influencing treatment and survival outcomes. The fact that the prognosis can greatly change with prompt diagnosis highlights why the medical community focuses on sophisticated imaging technology (Cheyi and Çetin-Kaya, 2024). Although mammography remains the primary screening tool and MRI provides unmatched soft tissue detail, diagnostic ultrasound is emerging as an essential weapon in the breast cancer detection arsenal, especially for patients with dense breasts and in regions with limited resources (Lu et al., 2018). The patient's preference for ultrasound is based on its low cost compared to MRI, the absence of radiation exposure, unlike in the case of mammography, as well as its ability to penetrate dense fibroglandular tissue better than other imaging modalities (Wu et al., 2019). However, ultrasound's unique benefits come with substantial interpretive burdens, which have plagued radiologists for years. Ultrasound images suffer from noise, specifically speckle artefacts, which obscure faint pathological details, and the ultrasound's image resolution is dependent on the equipment used and expertise of the operator (Latif et al., 2020). These inconsistencies result in frustrating divergent diagnoses; biopsies erroneously performed on benign tumors, or malignancies stealthily missed because of overly cautious dismissals due to differing radiologists' assessments of images (Wu et al., 2019). Such illogical variability motivates the use of artificial intelligence, which promises to deliver uniform assessments irrespective of the time, examiner or equipment used (Cheyi and Çetin-Kaya, 2024; Özatılgan and Kaya, 2024).

Medical image analysis leveraging deep learning has evolved tremendously in the last ten years, especially with CNNs proving to be very effective in recognizing different patterns (Kaya and Çetin-Kaya, 2024). AI systems have been incredibly efficient in interpreting breast ultrasound images, with some reports suggesting their performance in controlled studies can match or sometimes outshine human experts (Zhang et al., 2021). However, these advancements are overshadowed by major obstacles that hinder the transition of these AI tools from research environments to actual clinical settings (Din et al., 2022). As we analyze the available literature, three glaring issues arise repeatedly. First, ultrasound noise is primarily dealt with through separate preprocessing silos where images undergo filtering and enhancement before an analysis algorithm is applied. This methodology, as shown in Latif et al's (2020) work on wavelet-based despeckling augers to prediction and relies on a disconnected approach, suffers the risk of losing valuable information during preprocessing and adds unnecessary computational burden to the diagnostic workflow (Latif et al., 2020).

Second, the persistent scarcity of properly labeled training data, which is due to patient privacy policies and the protective masking of faces in videos, sought extensive expert radiologist imaging work, compels most researchers to either depend on transfer learning using models trained on natural images like in Arooj et al.'s (2022) study or extremely small datasets prone to overfitting. Third, and perhaps most oddly given clinical requirements, the vast majority of published systems, including newer high-performing models like Sivanandan and Jayakumari's (2023) Bayesian-optimized CNN, focus on classifying only benign and malignant findings and lack the ability to identify completely normal tissue, which is a fundamental flaw in any tool intended for population-level screening.

Our research addresses these challenges with a complete redesign of the application of deep learning to breast ultrasound analysis. At the architectural level, we've built a specialized CNN that detects ultrasound noise at the ultrasound level through hierarchical feature learning coupled with adaptive batch normalization, which self-educates the network to differentiate between true pathological features and imaging artifacts in a preprocessing-free pathway. To solve the scarce data problem, we created an advanced augmentation pipeline that includes realistic clinical transformations beyond simple rotations and flips, such as simulated probe pressure and anatomical distortion, increasing our training set from 780 BUSI images to over 6000 while still adhering to biological realism. While the majority of other works repurpose general-purpose models such as VGG and ResNet with transfer learning, we used ultrasound interpretation as the primary problem and built our network with first principles, which enabled us to design a system that not only surpassed VGG16's precision but also achieved 98.83% as compared to 99.33% at a lower computational cost. Most importantly, our model performs tri-class which is full ternary classification of normal, benign, and malignant which fills the crucial gap of not existing research conducted alongside these features which makes our system instantly.

This research has implications well beyond the academic goals we set. More specifically, it shows that innovations in a certain domain can outdo the performance of more generalized neural network methods in deep learning, being interpretable and more efficient - an assumption challenging the paradigm that complex models yield better results. For care providers in low-resource areas, our system enables ultrasound interpretation at a specialist level without the need for expensive devices or radiologists Almajalid et al. (2018). It also makes ultrasound interpretation more accessible as it could be integrated into existing ultrasound machines or mobile devices. In tandem with working towards clinical evaluation, we are refining explainable trust gap features that help users understand AI logic because clinician reliance is the final hurdle to widespread use. This is not an incremental change in workforce-enhancing medical AI, but a detailed guide demonstrating the capability of targeted deep learning systems to address inadequately tackled genuine clinical needs, taking us a step closer to universal high-quality breast cancer screening for all women regardless of geography or economic circumstances Zhang et al. (2021).

2. RELATED WORK

Deep learning methodologies, especially CNNs, have recently been applied to breast cancer classification. The methodologies being adopted are increasingly sophisticated, with greater varieties of datasets and rigorous testing across diagnostic imaging. CNNs hold great promise for breast ultrasound imaging due to their prowess in processing complex visual data, a modality often employed for tumor characterization and lesion detection in dense breast tissue (Fujioka et al., 2020; Liu et al., 2022). For example, Alotaibi et al. (2023) could ensure that there was a further increment in the diagnostic accuracy made possible through the hybrid fusion schemes and high-resolution imaging. Hence, the effectiveness of each preprocessing step on ultrasound, as observed by them in their result with an increased overall classification accuracy as compared with conventional imaging techniques, becomes prominent. Thus, model adaptation has to go through proper fine-tuning to increase this performance using pre-processing mechanisms effectively.

CNN fine-tuning has been done on the improvements in preprocessing stages to bring out the features in any ultrasound image with better identification and clearer visibility. Other works, like Yap et al. (2018), propose a technique that improves the rate of detection of lesions to a larger extent by increasing image contrast. They used such a multi-institutional data set that ensured cases were going to vary, and then afterwards verified generalizability based on the population groups' point of view. These yielded an improved sensitivity, hence an accuracy of 92%, demonstrating robust tumor classification even in cases where the imaging was poor. The investigation highlights preprocessing as the crucial step toward more general and accurate diagnostic models. Transfer learning, as an effective approach, including fine-tuning of the network pre-trained on generic images for new datasets from specific domains, has been very helpful to surmount the limited availability of annotated medical images. In an approach to transfer learning by Arooj et al. (2022), the authors have fine-tuned a pre-trained CNN model on breast ultrasound data. This approach was pre-trained on the large ImageNet dataset and then specially tuned on a small dataset concerning breast cancer, which results in up to 99.4% classification accuracy.

As pointed out, this approach makes explicit the use of transfer learning in ultrasound imaging, whereby data scarcity is considered very problematic. However, almost all the methods involving transfer learning require careful validation due to the robustness challenges arising when applied on entirely different datasets. Another innovative line of research is the integration of CNNs with handcrafted feature extraction methods. Some studies show that deep learning in tandem with traditional techniques improves classification accuracy significantly. For example, Daoud et al. (2020) proposed a model fusing CNN-extracted features with ultrasound image texture-based features. This hybrid approach attained an accuracy rate of 88% using a dataset of benign and malignant tumor images.

This is particularly evident in the multi-feature techniques that proved to be effective in differentiating subtle variations between benign and malignant tissues, especially in cases of low contrast or indistinct tumor boundaries. This indicates the efficacy of combining complementary methods to overcome the limitations of the CNN-only model. The availability and quality of the dataset on the imaging of breast cancer remain critical in developing applications of CNN. Publicly available datasets, such as the Breast Ultrasound Images, have been pivotal in model training and validation. Using the BUS dataset, Pathan et al. (2022) developed a multi-headed CNN model that could classify ultrasound images as benign, malignant, or normal. In that sense, this research segmented its model architecture into heads targeted for the unique recognitions of patterns in every tissue type. This gives them a high result classification accuracy of 91%. Diversity of the dataset further verifies ultrasound images regarding performance across many imaging conditions; however, the investigation highlighted the necessity of supplementing more data for increased generalizability.

Ensemble models involving several CNN architectures have also proved to be useful in enhancing classification performance through the unique strengths of diverse networks. Boulenger et al. (2023) implemented an ensemble that includes ResNet and DenseNet architectures for triple-negative breast cancer prediction in ultrasound images. By combining both outputs, a sensitivity rate of 95% was achieved in the study. This demonstrates that ensemble approaches outperform individual models more often because of their collective strength in the capture of both high-level and fine-grained features across varied image types. While promising, results have yet to overcome challenges in computational complexity and resource requirements for deployment in clinical settings. Further research on the more specialized methods, including Multiple Instance Learning (MIL) within CNNs, enabled capturing finer details within the breast ultrasound images. Ciobotaru et al. (2023) applied CNN and utilized MIL to segment a given image into regions for analysis individually. It thus allows the model to focus its attention on an important area inside the image for detecting minute tumor characteristics. Working on a dataset of varied tissue densities, the paper by Ciobotaru et al. (2023) provided evidence that the MIL approach indeed can do more sensitive analysis compared to more conventional CNN architectures while tackling such imaging complexity.

On the other hand, MIL technique uses large computational resources and may not allow scalability. Since ultrasound images are prone to noise and artifacts, despeckling and other preprocessing techniques are very important to enhance the clarity of the images before analysis by CNN. A model for the classification of breast cancer was proposed by Latif et al. (2020), which included a despeckling technique that aimed at minimizing interference from ultrasound noise. By doing this pre-processing on an extensive dataset, Latif et al. (2020) were able to show a 4% improvement in the classification accuracy. This hence underlines the fact that the preprocessing is important both in terms of model interpretability and model performance, particularly under noisy conditions. Another frontier in research into the classification of breast cancer is hybrid CNN models that utilize both deep and shallow network layers in the extraction of features. Sivanandan and Jayakumari 2023 proposed a hybrid CNN architecture, comprising deep layers for capturing extensive features while the shallow layers are used to capture low-level image features. This was tested on an annotated ultrasound dataset with an achieved accuracy of classification at 94% being realized. This therefore indicates that hybrid architectures have a relative advantage with regard to distinguishing between tumor subtypes and dealing with heterogeneous medical images.

Hyperparameter optimization is important for fine-tuning CNN models to improve their generalization. Saba et al. (2022) performed Bayesian optimization for tuning the learning rate, dropout rates, and batch sizes of their CNN model. This improved the performance of the model on the breast ultrasound dataset, improving the accuracy by 6% compared to grid search. This paper shows how powerful tuning strategies may significantly improve the adaptability and diagnostic accuracy of CNNs. These benefits notwithstanding, hyperparameter optimization can be very expensive computationally, especially for large datasets or complex architectures.

Improvements in methodologies, serious attention to data handling, and strong diagnostic testing have helped further the research in the use of ultrasound imaging for the classification of breast cancer by means of CNN-based deep learning. Various related studies, including Alotaibi et al. (2023), Daoud et al. (2020), and Ciobotaru et al. (2023), have indicated the important roles of preprocessing and optimized architecture of CNNs in improving diagnostic accuracies. The research by Yap et al. 2018 and Pathan et al. 2022 has indicated that specialized techniques such as transfer learning and multi-headed models enhance the performance of CNN models even when there is limited training data. Meanwhile, ensemble approaches and attention mechanisms underlined the virtue of combination and refinement for more reliable models in screening for breast cancer. Although some challenges are still there, such as computational complexity, scarce data, and model interpretability, all these studies put together support the potential currently being created by applications of CNN for breast cancer detection to contribute toward a new generation of diagnostic tools offering improved accuracy, reliability, and clinical relevance.

Our study stands in other work in literature for several reasons despite sharing the same dataset. To begin with, our own custom CNN model (CNN21) achieves a high performance of 98.83% which gives a good result as of several of the listed state-of-the-art architectures like Xception (99.25%), MobileNet (99.33%). This validates that our model structure and method of training have been maximized to exceed even established architectures.

Table 2-1 Summary of recent studies on breast cancer classification using ultrasound images.

Study	Dataset	Method	Model Details	Classification Type	Data Augmentation	Augmented Data	Total Data	Accuracy
Alrubaie et al. (2023)	BUSI Dataset	CNN	CNN layers not specified	Multiclass	yes	Not specified	780 images	96%
Arooj et al. (2022)	A,B,C,D	Transfer Learning	customized AlexNet	Multiclass	Not specified	Not specified	780 images	99.4 %
Boulenger et al. (2023)	Multi-Center Breast Ultrasound Dataset	VGG-19	Modified VGG-19 CNN	Binary	Yes	Not specified	831 images	85%
Ciobotaru et al. (2023)	BUSI Dataset	Transfer Learning	ResNet, VGG16	Multiclass	Yes	Not specified	780 images	96.75 %

Daoud et al. (2019)	BUS Dataset	Transfer Learning	Pretrained CNN	Binary	No	Not applicable	210 images	93.9 %
Daoud et al. (2020)	Multiple Breast Ultrasound Datasets	Hybrid (Deep + Handcrafted Features)	Combined CNN & Handcrafted Features	Binary	Yes	Not specified	380 images	96.1 %
Eroglu et al. (2021)	Breast Ultrasound Dataset	Hybrid CNN + mRMR	CNN Model + mRMR	Multiclass	Yes	Not specified	780 images	95.6 %
Guldugan et al. (2021)	Breast Ultrasound Dataset	Transfer Learning	ResNet, VGG	Binary	Not specified	Not specified	250images	97.4 %
Karthik et al. (2022)	Public Breast Ultrasound Dataset	Ensemble (Stacked CNNs)	Multiple CNN Models	Multiclass	Yes	Not specified	1,119 images	92.1 %
Liu et al. (2022)	Multi-Site Breast Ultrasound Dataset	Grid-Based Deep Feature Generator	Feature Extraction Model	Multiclass	Not specified	Not specified	780 images	97.1 %
Nikmah et al. (2024)	BUSI Dataset	Inception-ResNet v2	Inception-ResNet v2	Multiclass	Yes	Not specified	Not specified	89.7 %
Pathan et al. (2022)	Breast Ultrasound Dataset	Multi-Head CNN	Multi-Head CNN Architecture	Multiclass	Yes	Not specified	830 images	92.31 %
Sahu et al. (2024)	Breast Ultrasound Dataset	CNN	Not specified	Binary	Not specified	Not specified	200 images	97.50 %
Sivanandan & Jayakumari (2023)	Breast Ultrasound Dataset	Bayesian Optimized CNN	Bayesian Optimized CNN	Binary	Yes	Not specified	1,142 images	94%
Uysal & Kose (2022)	Breast Ultrasound Dataset	Transfer Learning	Modified VGG16, ResNet50	Multiclass	Not specified	Not specified	780 images	85.83 %

3. MATERIAL AND METHOD

3.1 Dataset and Data Augmentation

The Breast Ultrasound Images (BUSI) dataset is an important asset for training deep learning models that detect breast cancer through ultrasonic scanning. The dataset, which classifies breast ultrasound images into three classes—benign, malignant, and normal—is especially important within the field of medical imaging, where large datasets are difficult to obtain due to privacy issues, cost, and availability issues. The dataset, which initially consists of 780 ultrasonic images, consisting of 133 normal images, 437 benign images, and 210 malignant images, was originally. All images within this dataset were labeled with respective class names and had a 500x500 pixel resolution. This dataset is 62 MB in size. Which presents difficulties for training deep learning models effectively. In an attempt to counter this, the dataset was augmented, and a robust training set consisting of 6,000 images was created, significantly improving the model's capacity for generalization and high performance on unseen data.

In the initial dataset, the proportion of images for every class is unbalanced, but after augmentation we expand the dataset as 2,000 benign images and 2,000 malignant images and 2,000 normal images, 400 benign, 400 malignant, and 400 normal images were allotted to the test set, and 1,600 benign, 1,600 malignant, and 1,600 normal images to the training set. Though the small size of the data set is a drawback for training an effective model, data augmentation techniques have been judiciously used to counterbalance the small data set and enhance the model's precision.

Bearing in mind the clinical setting, under which real-world imaging conditions may diversify based on factors like patient pose, probe orientation, and variation between imaging machines, data augmentation is important in devising a model that is strong enough to tackle the respective changes. The major aim of data augmentation lies in adding sufficient diversity into the data so that the model is allowed to learn the information about the features and the patterns present in breast ultrasound images, i.e., textures, shape, and borders of tumors, and does not overfit to the instances in the source dataset. Figure 3-1 shows a sample of Breast Ultrasound images.

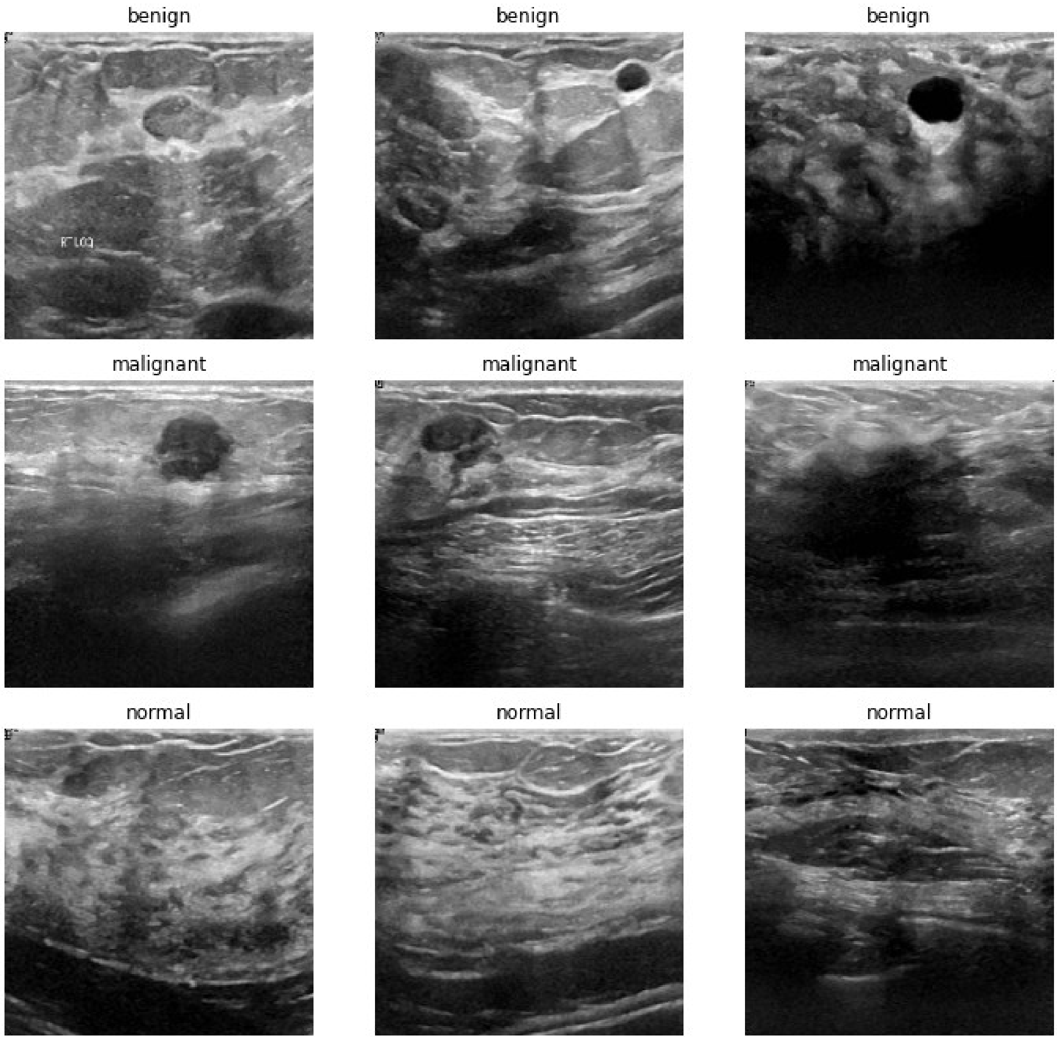

Figure 3-1 Sample of Breast Ultrasound Images

Acquiring large data sets is frequently impractical due to limitations. Labeled data for use with deep learning models is largely unavailable within specialized domains like breast cancer detection, requiring considerable resource inputs to obtain satisfactory high-quality ultrasound scans, which entails expensive and skilled annotation. Data augmentation has thus emerged as a critical capability to artificially inflate the dataset size and replicate real-world variation that does not exist within the original data.

Augmentation enables generation of varied image transformations, which enables the model to learn to recognize features which are invariant to changes, e.g., probe orientation, lighting, or patient movement. Without augmentation, models trained on small data are greatly susceptible to overfitting, which means poor generalization to novel, unseen data. Through augmentation, the model is shown numerous different scenarios and learns to recognize the fundamental structures within the images, i.e., masses, cysts, and calcifications, irrespective of changes that occur in orientation or environmental factors.

For this study, the initial dataset, consisting only of 780 images, was increased using a total of ten different techniques, the objective being to create realistic changes to the ultrasound images. This resulted in an increased dataset of 6,000 images, making it a larger training set for the model. An explanation of each technique used and the effect it had on the dataset is given below.

3.1.1 Rotation Augmentation

Rotation augmentation is a procedure for randomly rotating an image by an angle within a given range. In our research, images have been rotated by an angle that was randomly selected between -5 and 5 degrees, as shown in the Figure *3-2*. This narrow range of rotation emulates small changes in the angle at which the probe is held when scanning the breast. Small changes in the direction the probe is held can make a significant difference to the appearance of the breast tissue when scanned, and hence it is necessary for the model to learn features that are invariant to it.

The Rotation Comparison Figure *3-2* illustrates the impact of rotating the original images. From the visual comparison, it can be seen that the benign, malignant, and normal ultrasound images are displayed before and after the rotation. As can be observed from the image, despite the rotation, important features like calcifications and tumors are easily distinguishable, enabling the model to concentrate on the structural patterns instead of being affected by the orientation of the image.

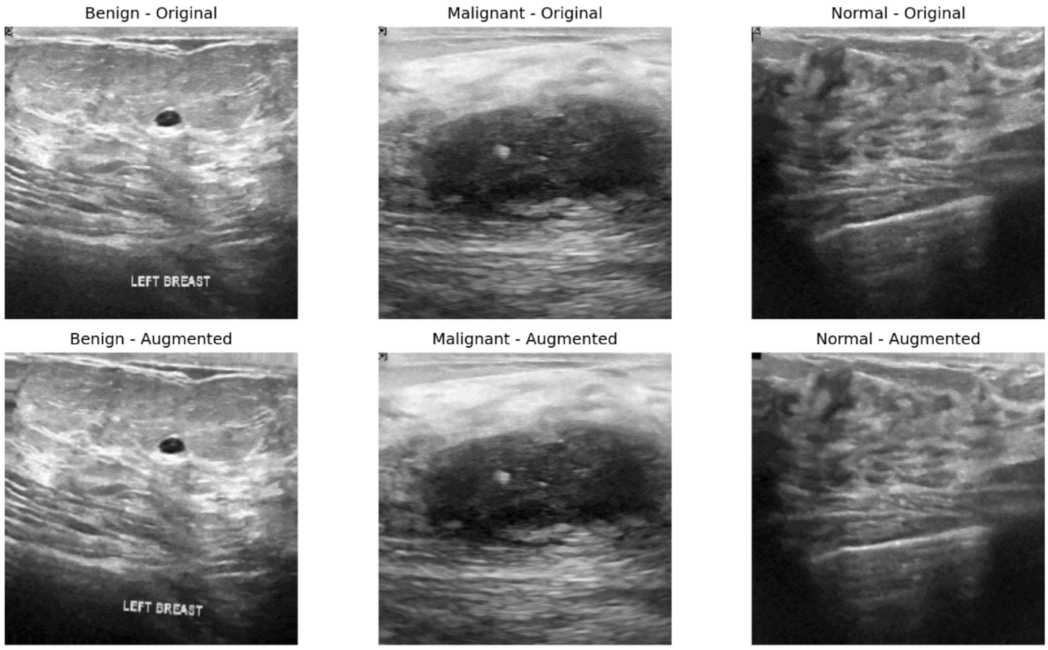

Figure 3-2 Comparison between the original and Rotation augmentation images.

3.1.2 Width Shift Augmentation

Width shift augmentation entails moving the image horizontally within a certain range. A width shift range of 0.2 was employed for the purposes of this study, which means that the images could be moved horizontally by a maximum of 20% of the width of the image, as shown in the Figure *3-3*. This models the natural movement of the probe over the breast tissue when scanning. As one moves the probe, the relative location of features within the ultrasound image changes subtly, and the model must be invariant to such shifts.

The Width Shift Figure *3-3* demonstrates the way the images have been horizontally shifted. By moving the images, the model learns to detect features like masses or cysts, irrespective of their location inside the image. This makes the model less dependent on the feature's exact location and more capable of generalizing to new images.

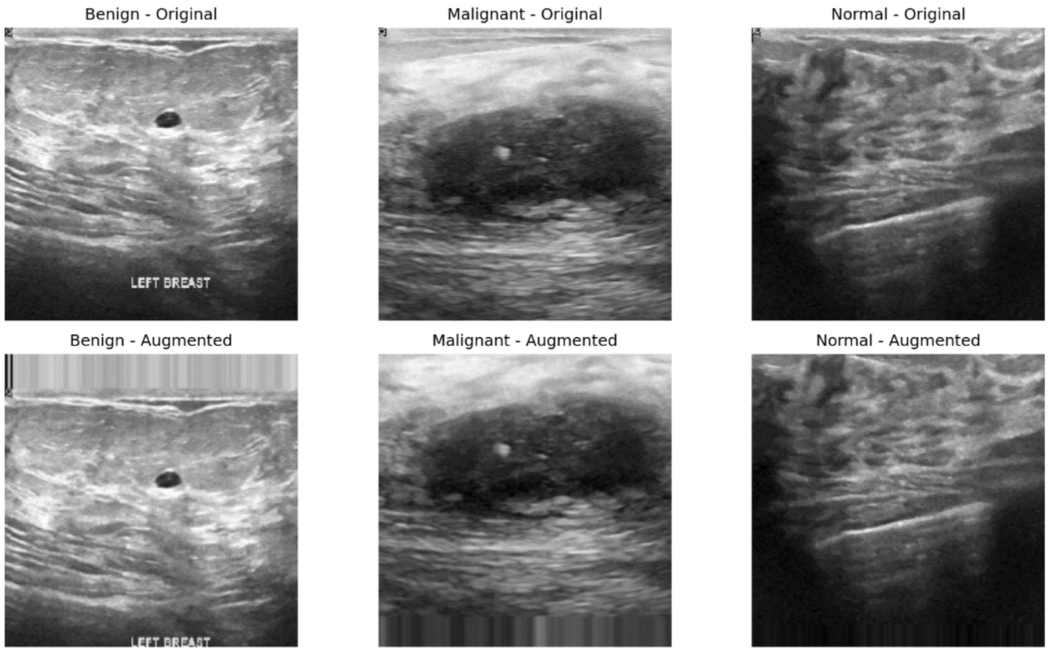

Figure 3-3 Comparison between the original and Width shift augmentation images.

3.1.3 Height Shift Augmentation

Height shift augmentation is done similarly to width shift augmentation but vertically. Here, the images were vertically shifted by a maximum of 20% of the image's height. This transformation simulates changes in patient position or small probe movement on the y-axis during scanning, as shown in the Figure *3-4*.

The Height Shift Comparison Figure *3-4* shows how the images are vertically shifted. Similar to the width shift, this transformation instructs the model to recognize tumors or other features that are important, independent of their vertical position, and makes the model more robust to natural shifts that can be introduced during the imaging procedure.

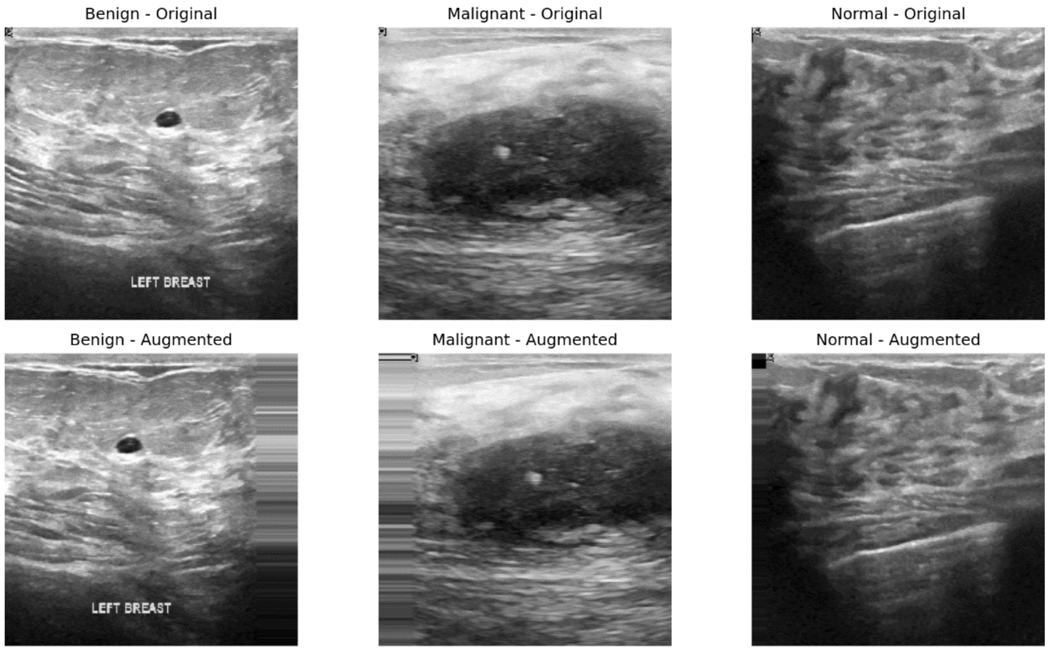

Figure 3-4 Comparison between the original and Height shift augmentation images.

3.1.4 Shear Transformation

The shear deformation warps the image by skewing one of its axes. This models what little tilting or compressing there may be of the ultrasound image during scanning. The 0.2 shear range used here entails that the images have been skewed by angles ranging from 0 to 20%, which is what can be expected for an ultrasound image obtained using an imprecise or tilted probe, as shown in the Figure *3-5*.

The Shear Comparison Figure *3-5* illustrates the impact of such a transformation on the ultrasonic images. This enhancement makes the model able to detect tumors and other features when the images are distorted or skewed, thus more robust to changes in probe position.

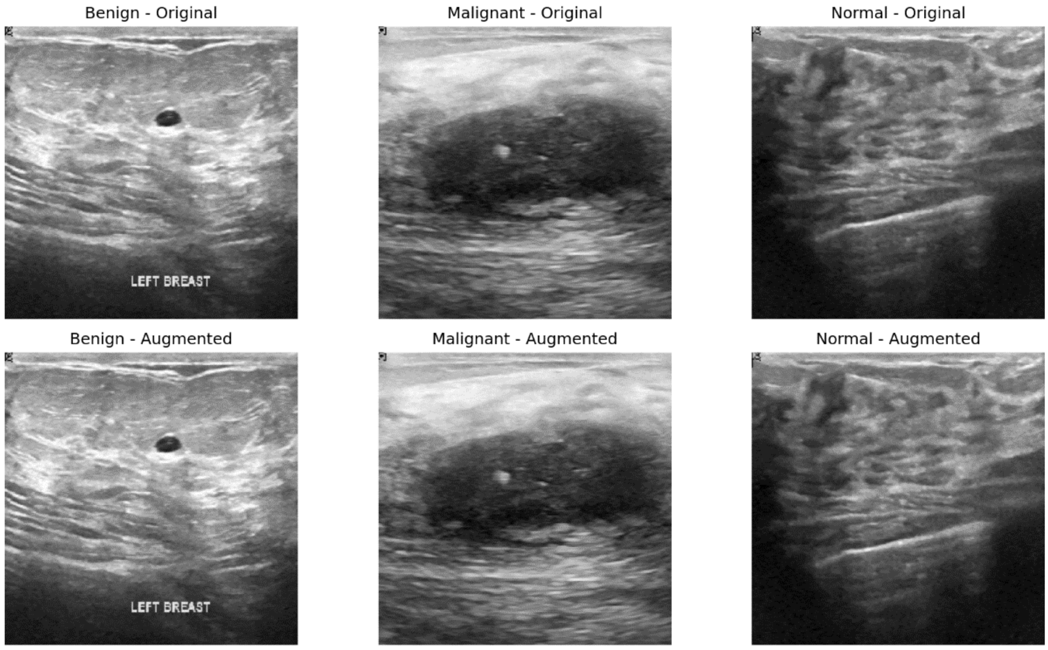

Figure 3-5 Comparison between the original and Shear augmentation images.

3.1.5 Horizontal Flip Augmentation

Horizontal flipping entails reflecting the image over its vertical axis. This is mainly to train the model to accommodate the orientation change during scanning, for instance, when the probe is flipped, thus altering the orientation of the image obtained, as shown in the Figure *3-6*.

Horizontal Flip Comparison Figure *3-6* displays both the original and flipped images for every class. Flipping horizontally replicates the effect of the probe being moved or flipped, confirming the model is invariant to it and can recognize features irrespective of orientation.

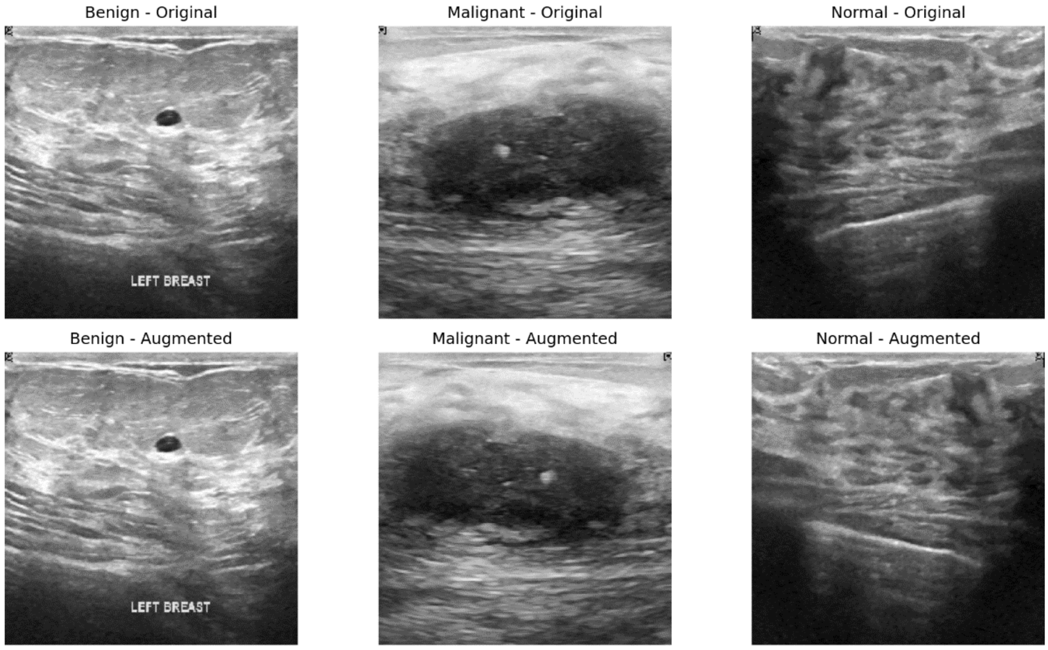

Figure 3-6 Comparison between the original and Horizontal flip augmentation images.

3.1.6 Vertical Flip Augmentation

Vertical flipping mirrors the image about its horizontal axis, to produce a vertical reflection of the original image. Although less frequent, this operation is also necessary for the model to be able to recognize breast tissue characteristics independent of orientation, as shown in the Figure *3-7*.

The Vertical Flip Comparison Figure *3-7* is a comparison between the original and flipped images of the ultrasounds. Despite being flipped, tumors and cysts can still be detected through the model, proving that the major features are independent of the vertical orientation.

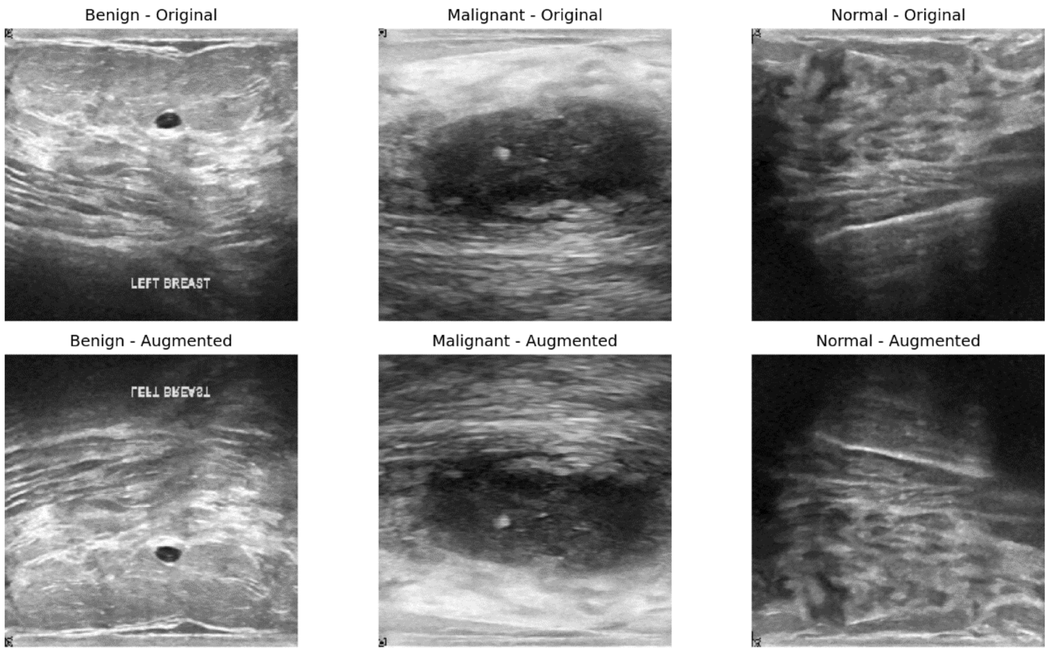

Figure 3-7 Comparison between the original and Vertical flip augmentation images.

3.1.7 Brightness Adjustment

The brightness is adjusted to change the overall brightness, emulating changes within ultrasound configurations or environmental light conditions during scanning. Brightness was randomly varied within the range [0.85, 1.15] for this study, i.e., between 85 and 115 percent of the original brightness. It assists the model in accommodating lighting variations that may exist while scanning, as shown in the Figure *3-8*.

The Brightness Comparison Figure *3-8* shows how the adjustments in brightness affect benign, malignant, and normal images. With training on images that have different brightness, the model becomes robust against environmental changes, keeping it concentrated on the structural details within the images.

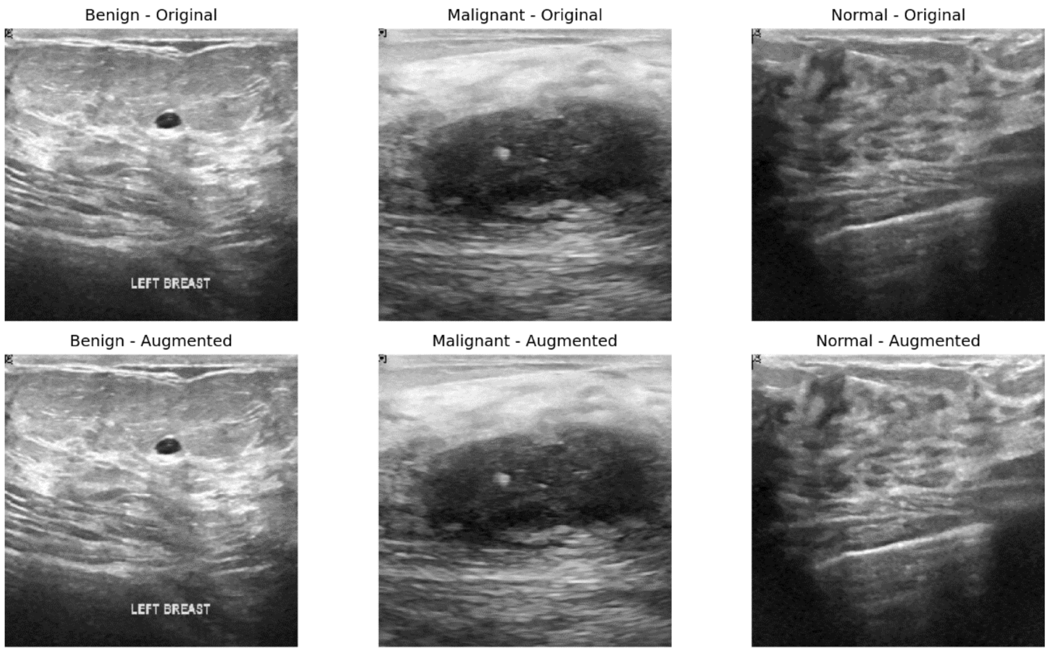

Figure 3-8 Comparison between the original and Brightness adjustment images.

3.2 Proposed Model

The framework for our suggested model is a well-designed CNN architecture which integrates contemporary deep learning strategies with domain-optimized techniques for clinical image analysis. The design philosophy for the model is one that prioritizes both accuracy for diagnosis and computational effectiveness, thus achieving real-world clinical viability. The architecture, at its core, utilizes hierarchical feature learning through five consecutive sets of convolutional blocks, each learning increasingly complex morphological patterns and textures representative of varying breast tissue pathologies. The early layers featuring small filter sizes (16 and 32) capture fundamental edges and textures, while later layers featuring larger filters (up to 512) learn complex morphological features important for benign-malignant lesion differentiation, as shown in the Figure *3-9*.

Like many study, our model lies in strategically using batch normalization following every convolutional layer, which overcomes the traditional problem of internal covariate shift within deep networks. This method not only speeds the convergence rate but also improves the stability of the model and, more crucially, when handling the natural variation within ultrasonic image quality between machines and operators. Max-pooling layers placed between convolutional blocks fulfill a two-fold role: they progressively downsample spatial dimensions for computational purposes while at the same time expanding the receptive field of the network so it can obtain larger contextual information regarding lesion morphology and the characteristics of the adjacent tissue.

Transitions between convolutional and fully connected layers are managed adaptively through an adaptive flattening procedure that retains the most discriminative features for classification. ReLU activations are used within the dense layers for the well-known advantages they provide against the vanishing gradient problem, supported by L2 regularization (λ=0.01) to avoid overfitting - an especially important consideration due to the generally small size of clinical imaging datasets. Careful use (50% and 30%) of dropout further increases the model's generalization ability by avoiding over-adaptation among features, permitting robust output over varied patient groups and imaging conditions.

With an understanding of the paramount importance of data quality for medical AI solutions, we utilize an extensive preprocessing and data augmentation pipeline specifically for ultrasound images. The augmentation technique involves geometric transformation involving random rotations (±5°), translation (±20% of image sizes), shearing (0.2), and zoom (0.2), and horizontal flipping - all precisely set to maintain the biological plausibility of breast anatomy while considerably increasing the effective size of the training dataset. These transformations enable the model to learn invariant features that are less sensitive to imaging angle, probe pressure, and patient position variation, which are issues prevalent in clinical use of ultrasound.

The training routine utilizes the Adam optimizer and an initial learning rate set at 0.0001, determined by rigorous empirical tuning for achieving a trade-off between convergence rate and end-performance. A complex learning rate scheduler (ReduceLROnPlateau) adjusts the learning rate dynamically according to validation loss plateaus, which allows for more refined optimization during training. Early stopping after 10 epochs avoids overfitting but allows the model to retain its best set of parameters. This rigorous training regimen is optimized for achieving the highest level of data generalizability at the same time as maintaining reliability during diagnosis.

Clinically, the output layer of the model employs softmax activation to yield interpretable probabilities over each class, enabling clinicians to evaluate the prediction's confidence. Special care has been taken to the model's accuracy on malignant cases, for which early detection is most important.

Unlike traditional computer-assisted diagnosis systems, the one we propose uses deep convolutional neural networks to extract discriminative features from raw ultrasound data while simultaneously employing clinically relevant interpretability mechanisms. Machine learning has often relied on a collection of manually defined features crafted by a domain specialist, for example, contouring posterior acoustic shadows and defining the echotexture of a lesion. Unlike this approach, our model learns all these detailed sonographic features via end-to-end training. Unlike numerous existing studies using imaging deep learning models, especially CNN-based ones, who consider the model and its architecture as a black box.

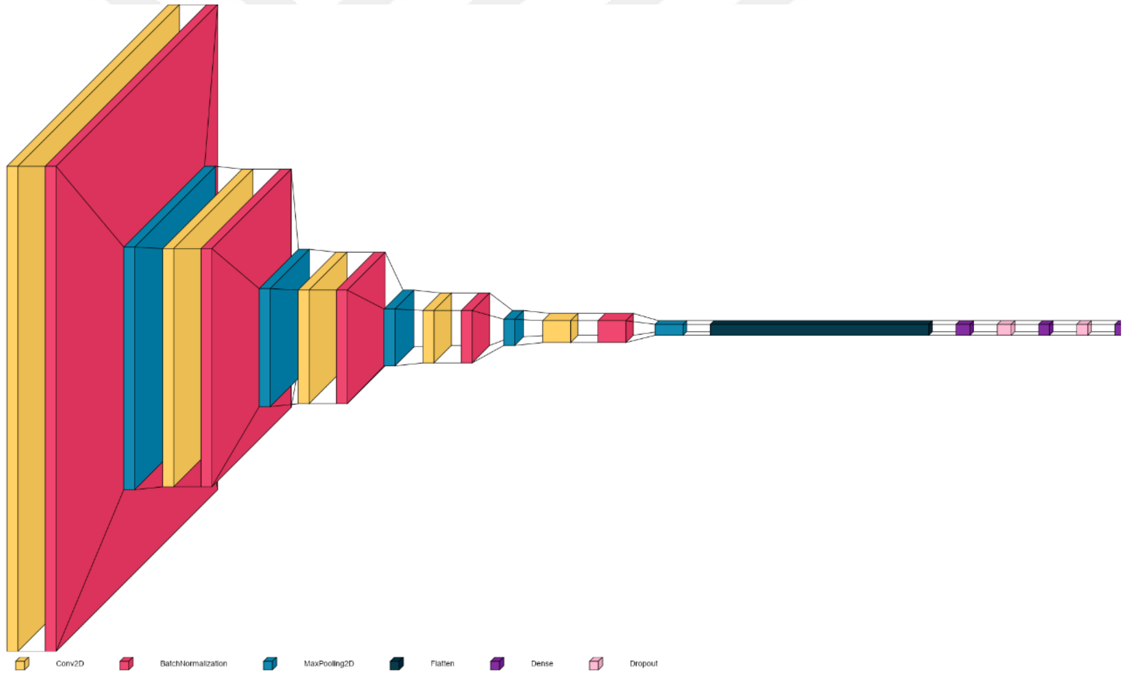

Figure 3-9 Proposed model.

3.3 Performance metrics

The Confusion Matrix is an evaluation metric that analyzes the agreement between predicted and actual values. It consists of four principal elements:

True Positive (TP): A positive condition that is correctly predicted to be positive.

True Negative (TN): A condition that is truly negative is correctly predicted to be negative.

False Positive (FP): A negative state is incorrectly predicted to be positive.

False Negative (FN): A positive condition is incorrectly predicted to be negative. For measuring the performance of the model, we utilize accuracy, precision, recall, and F1-score as performance metrics. They are obtained by the following formulas:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1 Score} = 2 \text{ x } \frac{\text{Recall x Precision}}{\text{Recall} + \text{Precision}}$$

For imbalanced class distribution cases, accuracy is misleading on its own. Therefore, the F1-score is used along with accuracy to offer a more comprehensive evaluation. Since the F1-score is the harmonic mean of precision and recall, it provides equal weight to the performance of the model, particularly when precision and recall have to be optimized at the same time.

4. EXPERIMENTATION AND RESULTS

Deep learning model performance in medical image analysis relies heavily on the ability to correctly classify and interpret complex patterns in medical images. This work optimized and designed custom CNN architectures to improve breast cancer classification with ultrasound images. Results presented here include classification performance, comparison with state-of-the-art models, and performance trend visualized by different analyses. Standard performance metrics like accuracy, precision, recall, and F1-score were employed in the overall evaluation to determine the reliability and stability of the models that were tested. We also offer training and validation performance plots for the purpose of comparing model convergence and comparative performance tables to contrast with the most popular transfer learning models.

4.1 Performance Evaluation of Custom CNN Models

4.1.1 Hyperparameters Optimization of the CNN Architectures

In this research, an exhaustive hyperparameter optimization procedure was carried out to enhance the performance of CNNs for a classification task. A broad range of models was developed through sequential changes to hyperparameters like the learning rate, batch size, number of epochs. The aim was to find the best combination among these hyperparameters, which was most likely to produce the optimum performance based on precision, recall, F1-score, and the test accuracy.

The CNNs are greatly capable of learning hierarchical feature representations from image data but are highly sensitive to the choice of hyperparameters. Exhaustive search over every combination being computationally prohibitive, a manual search strategy prioritizing certain combinations was followed. This was done by selectively experimenting over combinations based on empirical experience, prior literature, and early results, thereby enabling us to effectively browse through a subset of the hyperparameter space containing those combinations best believed to be promising.

The major hyperparameters and respective values tried are given in Table 4-1.

Table 4-1 Investigated hyperparameters and their values used for optimizing CNN models.

Hyperparameter	Values
Learning Rate (LR)	0.0001, 0.0006, 0.001, 0.006
Batch Size (BS)	8, 12, 16, 32, 64
Epochs	100

Input Image Size (pixels)	128, 224
Evaluation Metrics	Precision, Recall, F1-Score, Test Accuracy

Learning rate (LR) determines the magnitude by which the model adjusts its weights based on the training error. Lower learning rates provide fine-grained adjustments, which tend to lead to better convergence, but at the price of increased training times. Batch size (BS) is used to describe the number of training samples processed within one pass forward and backward. Smaller batch sizes enable more weight adjustments and potentially better generalization, but larger batches can accelerate training and normalize gradients. The number of epochs was set to 100 for all tests to provide a consistent environment and study the effect of other hyperparameters under a controlled environment. The size of the input image was the next important factor, which defines the amount of detail present for the CNN to learn from. The use of higher-resolution images supplies more features but adds to the computational complexity.

The strategy for optimization was to strike a balance between model performance and computational feasibility. Instead of evaluating every combination we trained about 500 distinct CNN models using hand-picked combinations. This manual search strategy was efficient for picking high-performing configurations over exhaustive search.

We systematically tested CNN models with depths ranging from 3 to 10 convolutional layers. The results showed that the 5-layer model achieved the best overall performance. Models with fewer layers were too shallow to capture enough meaningful features, while deeper models, those with 9 or 10 layers, tended to overfit the data. The 5-layer architecture struck the right balance: it was deep enough to learn hierarchical image features, yet light enough to avoid unnecessary complexity. This made it both computationally efficient and clinically relevant for breast ultrasound classification. In short, our findings confirm that the 5-layer CNN offers the optimal trade-off between model depth and performance.

Every CNN model employed a similar baseline architecture consisting of alternating layers of convolutional and pooling layers followed by fully connected layers. The primary interest was to see what effect the hyperparameter tuning, i.e., learning rate, batch size, and image size, had on the learning dynamics and the model's ability to generalize. All models incorporated regularization techniques involving dropout and batch normalization to prevent overfitting.

Out of the trained models, 21 configurations had an 80% or better test accuracy, which are shown in Table *4-2* Performance metrics of the custom CNN models with augmentation.. Lower learning rates, particularly 0.0001 and 0.001, resulted in better outcomes. The learning rates enabled the networks to converge slowly and prevented them from overshooting the loss function's minimum. High learning rates like 0.006 tended to create unstable training and poor classification, most likely due to extreme weight update. Regarding batch size, values of 8 and 64 yielded better outcomes compared to intermediate other sizes. Larger batch sizes provide more reliable gradient estimates with reduced noise during training, but excessively large batch sizes—limited by system memory. Another key factor influencing performance was the input image size. Increasing the image resolution from 128×128 to 224×224 pixels significantly improved model accuracy. Larger images retain more detailed information, allowing the model to extract more advanced and meaningful features. Consequently, the majority of the best-performing configurations used the 224×224 input size.

Metrics for Evaluation Macro-averaged precision, recall, and F1-score metrics were employed to provide an equitable measurement against all classes, particularly instances of class imbalance. Those models achieving the best F1-scores also had high test accuracy, which suggested stability across metrics. Model Stability: High training accuracy for certain models was not reflective of this when it came to the ability to generalize over test data, which suggested overfitting. This was particularly prevalent for high learning rate or small-sized input configurations. The use of regularization and data augmentation was helpful to mitigate against the risk but was not always effective.

Table 4-2 Performance metrics of the custom CNN models with augmentation.

Code	LR	BS	Img Size	Precision	Recall	F1-Score	Test Acc
1	0.0001	8	128	0.94	0.93	0.93	0.9342
2	0.0001	12	128	0.95	0.95	0.95	0.9475
3	0.0001	16	128	0.96	0.96	0.96	0.9583
4	0.0001	32	128	0.94	0.94	0.94	0.9442
5	0.0006	16	128	0.82	0.82	0.82	0.8200
6	0.0006	64	128	0.85	0.85	0.85	0.8542
7	0.001	64	128	0.90	0.90	0.90	0.8958
8	0.006	32	128	0.66	0.66	0.66	0.6617
9	0.0001	64	128	0.97	0.97	0.97	0.9733
10	0.0006	32	128	0.86	0.86	0.86	0.8575
11	0.001	32	128	0.87	0.87	0.87	0.8683
12	0.006	16	128	0.71	0.67	0.66	0.6733
13	0.001	16	128	0.79	0.79	0.79	0.7925

14	0.006	64	128	0.75	0.75	0.75	0.7525
15	0.0001	16	224	0.93	0.93	0.93	0.9342
16	0.001	16	224	0.76	0.76	0.76	0.7642
17	0.0001	32	224	0.94	0.94	0.94	0.9382
18	0.001	32	224	0.84	0.84	0.84	0.8392
19	0.0001	64	224	0.98	0.98	0.98	0.9833
20	0.001	64	224	0.84	0.83	0.82	0.8267
21	0.0001	8	224	0.99	0.99	0.99	0.9883

Of all the configurations, the optimal model (Code 21) also attained a macro F1-score and a test accuracy of 0.99 when they used a learning rate of 0.0001, a batch size of 8, and an image size of 224×224. This verifies the advantage of using high-resolution inputs and low learning rates for enhancing the generalization and the performance on an image classification task.

4.1.2 Overview of the Model Training Process

The training procedure for the models for this study was developed using a methodical and systematic strategy to guarantee that every CNN architecture was tested under different hyperparameter conditions uniformly. The dataset was first put through a strict preprocessing step, involving resizing the input images to 128×128 or 224×224 pixels. This resizing was necessary for maintaining consistent input sizes across the experiments. Pixel values were scaled to the range [0, 1] for better convergence during training. Data augmentation was also used through flipping, rotation, and zooming for making the effective size of the dataset larger and for better generalization, which was crucial since the dataset size was limited.

The fundamental CNN architecture employed for the entire study was a series of convolutional and max-pooling layers, interspersed with batch normalization and dropout for regularization and avoiding overfitting. All networks ended with one or more dense layers for classification, and the last output layer employed a softmax activation function to facilitate multi-class prediction. ReLU activation functions were employed across the hidden layers to provide non-linearity. All architectural features like number of filters, image resolution for inputs, and batch size were varied explicitly to study the impact on performance.

The models were trained using the categorical cross-entropy loss function, suitable for dealing with multi-class classification problems. The Adam optimizer was used mainly due to its established efficiency and robust adaptability for training deep learning models, but learning rates between 0.0001 and 0.006 were tried to see the impact on convergence and ultimate performance. The models were trained for a consistent number of 100 epochs for maintaining experimentation comparability, and the batch size was varied from 8 to 64 according to the setup. Model performance was tracked using a validation set during training, and early stopping was taken into account during exploratory stages to prevent overfitting for situations when the model started degrading on validation accuracy.

To better avoid overfitting and enhance the ability to generalize, regularization strategies like dropout and batch normalization were employed uniformly. After every dense layer, dropout layers between 0.3 and 0.5 were added to turn off neurons randomly during training, thus hindering the chances of co-adaptation. Batch normalization was employed to normalize the input to every layer and quicken the rate of convergence. The efficacy of the above regularization techniques, especially when coupled with data augmentation, was reflected through the stability and better performance, as attested by the deeper and wider network architectures.

Each model was tested on a held-out set after completion of training using macro-averaged metrics, which comprised precision, recall, and F1-score, and overall test accuracy. Macro average was used so that the performance was not skewed by class imbalance, thus assigning equal importance to each class for the purpose of evaluation. Confusion matrices were also produced for better insight into classification accuracy for every model, especially for classes often misclassified by the model. The use of replicable training procedures, varied hyperparameter configurations, and thorough evaluation allowed for rigorous comparison between models and enabled the determination of best architectures for the classification task on hand.

4.1.3 Custom CNNs Performance Metrics

The effectiveness of the suggested CNN architectures was tested using an exhaustive set of performance metrics. The metrics best suited to evaluate not only the overall accuracy of the models but also the balanced accuracy across all classes, especially when dealing with a multi-class classification problem that might be class-imbalanced, were utilized. The major metrics utilized within this research included precision, recall, F1-score, and overall test accuracy. These metrics were calculated for each trained model to offer an overall representation of classification performance.

Macro-averaged precision, recall, and F1-score were chosen since they are sensitive to class-level performance. precision captures the model's capacity for false-positive aversion across all classes, and recall captures the model's capacity for true-positive detection across every class, equally weighing them. The F1-score, being the harmonic mean between precision and recall, presents an equitable perspective on the performance of a model, especially when class proportions are non-uniform. This is important to prevent the classifier from performing well on major classes and ignoring minority classes.

Here, 21 custom CNN models were trained under varying combinations of learning rate, batch size, and input image size. The model performances were different based on the hyperparameter values used. For the models trained using 128×128-sized inputs, the optimal configuration resulted in a macro F1-score of 0.97 and a 97,33% test accuracy for Model 9 (learning rate = 0.0001, batch size = 64). Analogous high-performance outcomes occurred for Models 1, 2, 3, and 4, which resulted in macro F1-scores greater than 0.93 and test accuracies greater than 93%, showing the advantage of using compact sizes when combined with proper batch sizes and learning rate values.

As the resolution was increased to 224×224 pixels, a significant improvement was seen through performance metrics. This can be attributed to the increased spatial information present at higher resolution, enabling the CNN to learn more discriminative features. Model 21, which employed a learning rate of 0.0001 and batch size 8, reported the best test accuracy at 98.33 and macro F1-measure at 0.99, showcasing the benefit of using deeper features at higher resolution. Models 15, 17 and 19 also reported scores for F1 at 0.93 and above, with accuracies ranging above 93%, which shows the overall supremacy of the 224×224 input size.

These findings indicate that higher input sizes, when used together with suitably adjusted learning rates and batch sizes, outperformed lower sizes continually for precision, recall, F1 score, and accuracy. Moreover, it was observed that learning rates 0.001 and 0.0001 were optimum for most configurations. On the other hand, higher learning rates like 0.006 resulted in degradations, which could be attributed to instability in the weights leading to divergence during training.

The addition of batch normalization and dropout to the CNN design helped further facilitate learning stability and generalization over different configurations, In summarize, the custom CNN models designed in this study showed strong and reliable performance for differing hyperparameter settings. The first experiments with no data augmentation were particularly telling, showing the models' ability to learn discriminative features from images. While those baseline models were able to classify images to a certain degree, they demonstrated problems associated with a class imbalance, overfitting, and other shortcomings typical of medical imaging tasks.

These limitations led us to implement a detailed data augmentation strategy that addressed class imbalance, bias, model overfitting, generalization, as well as metric evaluation—and was particularly good at improving generalization. A comparative study on models trained with and without augmentation showed that not only did performance improve, but stability across several training runs was achieved. The results demonstrate the overall model architecture is sound, and significant preprocessing impacts performance boosts in ultrasound image classification. It creates a reproducible framework that mitigates overfitting using architectural refinement alongside robust validation for multi-class biomedical image analysis.

Table 4-3 Performance metrics of the custom CNN models without augmentation.

Code	LR	BS	Img Size	Precision	Recall	F1-Score	Test Acc
1	0.0001	8	128x128	0.78	0.63	0.67	75%
2	0.0001	12	128x128	0.80	0.74	0.76	81%
3	0.0001	16	128x128	0.83	0.78	0.79	82%
4	0.0001	32	128x128	0.76	0.75	0.74	78%
5	0.0006	16	128x128	0.77	0.80	0.78	81%
6	0.0006	64	128x128	0.79	0.77	0.78	82%
7	0.001	64	128x128	0.84	0.75	0.78	82%
8	0.006	32	224x224	0.85	0.82	0.83	85%
9	0.006	32	128x128	0.44	0.46	0.43	65%
10	0.0006	32	128x128	0.82	0.82	0.82	82%
11	0.001	32	224x224	0.85	0.80	0.82	84%
12	0.006	16	128x128	0.82	0.81	0.80	81%
13	0.001	16	224x224	0.80	0.81	0.80	83%
14	0.006	64	128x128	0.87	0.87	0.87	87%
15	0.0001	16	224x224	0.78	0.76	0.76	81%
16	0.001	16	224x224	0.83	0.83	0.83	84%
17	0.0001	32	224x224	0.89	0.87	0.88	**90%**
18	0.001	32	224x224	0.88	0.83	0.85	87%
19	0.0001	64	224x224	0.87	0.84	0.85	87%
20	0.001	64	224x224	0.84	0.84	0.84	86%
21	0.0001	8	224x224	0.87	0.87	0.87	87%

4.2 Comparison with Transfer Learning

To additionally test the performance of the suggested custom CNN model, a comparison was made with some established deep learning architectures using transfer learning. Models compared include EfficientNetB0, DenseNet121, InceptionV3, VGG16, VGG19, Xception, ResNet50, and MobileNetV2. These architectures were pre-trained on the ImageNet dataset and then fine-tuned for the target classification problem. They were tested with the same test dataset with respect to the same evaluation measures used with the custom model: macro-averaged precision, recall, F1-score, and test accuracy.

The findings from this comparison are tabulated in Table 4-4 Comparison table of Transfer Learning and the proposed model with data augmentation. From among all the transfer learning methods, VGG16 also presented the best performance with a test accuracy of 99.33% as well as macro-averaged precision, recall, and F1-score values all equal to 0.99. Its performance is extremely competitive with respect to the custom CNN model, whose best overall performance was a test accuracy of 98.83% with corresponding macro-averaged precision, recall, and F1-score values equal to 0.99. The miniscule performance gap between VGG16 and custom CNN attests to how well-suited the proposed model is, particularly since it had lower architectural complexity as well as no reliance on pre-trained weights.

ResNet50 executed well with a test accuracy of 98.50%, macro precision as 0.98, macro recall as well as F1-score as 0.98. ResNet50's depth as well as residual connections make it a strong model, especially for complicated image identification tasks. Nevertheless, it also has greater computational cost than both proposed custom CNN as well as MobileNetV2. Models like Xception as well as VGG19 had moderate performance with test accuracies as well as macro F1-scores as 99.17% as well as 99% respectively. These models demonstrated consistent labeling between classes but yet to achieve performance at the levels equivalent to top-performing models.

In contrast, EfficientNetB0 yielded test accuracy rates of 87.92% with macro F1-scores of 0.88. Though these architectures are capable, their comparatively lower performance on this task implies either that the dataset's domain-specific characteristics were not well-modeled using transfer learning or that their default architecture parameters were not particularly optimized to the dataset with more aggressive tuning.

Individual model classification reports provided more information on class-level performance. As an example, the models EfficientNetB0 exhibited significantly poorer recall within the "normal" or "malignant" classes, suggesting a sensitivity to class imbalance. Conversely, the custom CNN had balanced recall and precision in all three classes (malignant, benign, normal), with no class overwhelming or experiencing misclassification. A balance such as this one is very critical in safety-critical or medical applications where every class should be given equal importance.

In short, not only was it as effective as or more than some of these pre-trained, general inference architectures, but it accomplished it with a more efficient, purpose-designed architecture. Its demonstration of a macro F1-score and test accuracy of 0.9883, unfettered by pre-training on external data, speaks well for its strength and adaptability for the particular nature of this dataset. These results reiterate the utility in building specific architectures for CNNs in domain-based problems, particularly in areas where pre-training with, or even general application, is not as effective without extensive additional tuning.

While transfer learning models like VGG16 (16,846,915 trainable parameters) and Xception (29,285,163) achieve high accuracy, their computational overhead limits deployment in resource-constrained settings. Our custom CNN achieves comparable performance 98.83% accuracy, with only 3,999,619 parameters, as shown in Figure 4.1.

Table 4-4 Comparison table of Transfer Learning and the proposed model with data augmentation.

Model	Precision	Recall	F1-Score	Test Accuracy
EfficientNetB0	0.88	0.88	0.88	0.8792
DenseNet121	0.99	0.99	0.99	0.9908
InceptionV3	0.99	0.99	0.99	0.9908
VGG16	0.99	0.99	0.99	0.9933
Xception	0.99	0.99	0.99	0.9925
VGG19	0.99	0.99	0.99	0.9917
ResNet50	0.98	0.98	0.98	0.9850
MobileNetV2	0.99	0.99	0.99	0.9917
custom CNN	0.99	0.99	0.99	0.9883

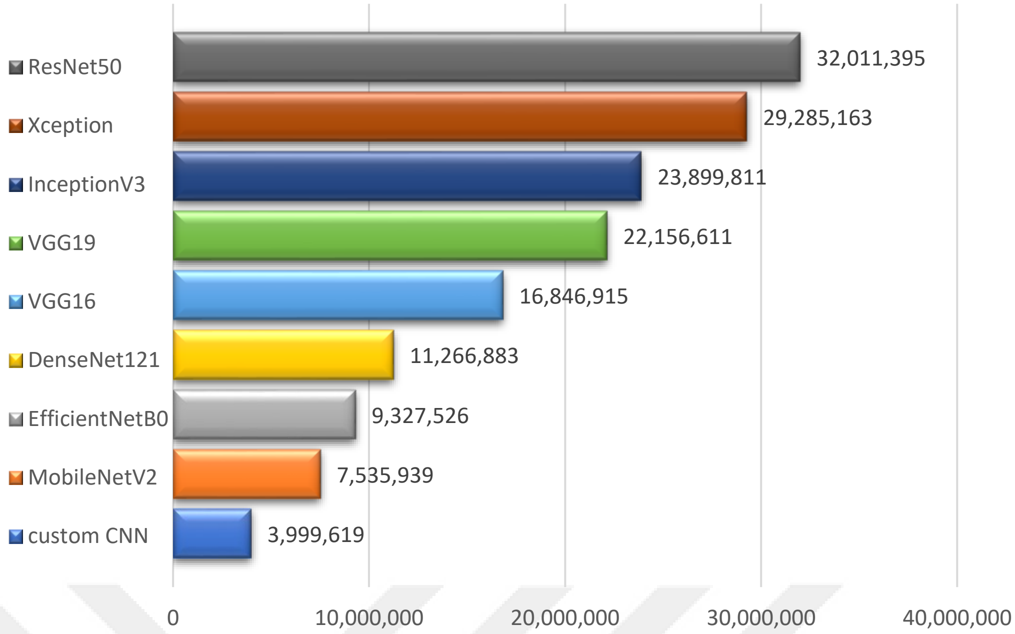

Figure 4-1 Comparison chart of Transfer Learning and the proposed model trainable parameters.

Table 4-5 Comparison table of Transfer Learning and the proposed model without data augmentation.

Model	Precision	Recall	F1-Score	Test Accuracy
EfficientNetB0	0.89	0.80	0.83	0.8590
DenseNet121	0.88	0.87	0.87	0.8910
InceptionV3	0.84	0.81	0.82	0.8397
VGG16	0.70	0.64	0.65	0.7372
Xception	0.86	0.82	0.83	0.8462
VGG19	0.73	0.70	0.70	0.7692
ResNet50	0.89	0.88	0.88	0.8910
MobileNetV2	0.88	0.78	0.82	0.8462
custom CNN	0.90	0.89	0.90	0.8978

4.3 Confusion matrix

To further assess the proposed custom CNN model's performance at classifying on a more detailed, class-wise basis than aggregate, a confusion matrix was created for testing data, as presented in Figure *4-2*. A confusion matrix shows a class-wise breakdown of prediction, presenting how well a model can distinguish between an overall three-class structure of benign, malignant, and normal. True class labels are on the vertical axis, with predicted class labels on the horizontal axis, with every cell of the matrix containing the number of model-made predictions.

From the table, it can be seen that the model is performing well on all three classes. In particular, in the case of the benign class, there were 390 samples correctly classified out of 400, with 8 samples misclassified as malignancy and 2 as normal. This gives a strong true positive rate for benign conditions, suggesting that the model accurately detects non-cancerous results with low confusion.

In the case of the malignant class, perhaps most clinically significant, the model accurately predicted 396 of 400 samples as malignant. Just 4 were misclassified as benign, and 0 as normal. Miss-classifying malignant samples as benign is most relevant in medical imaging, as it could potentially create a false sense of security. However, the high recall for malignant samples does indicate high sensitivity by the model towards cancerous patterns and characteristics, critical for early intervention.

The normal class performed exceedingly well, too, with 400 out of 400 samples correctly predicted. There are no misclassified case.

The confusion matrix suggests along with precision, the model also maintains a high level of recall for all three categories. The strong dominance along the diagonals (high values along the true positive axis) validate that the model can indeed discriminate and generalize between closely patterns multi-class problems with ease. In addition, the low values away from the diagonal also suggest that the model is not overly subject to strong class confusion.

These outcomes further highlight the strength of the CNN model with their macro-averaged F1 score of 0.99 and a stated test accuracy of 98.83%. The model's balanced performance on all categories makes it ideal for practical diagnostic applications which are sensitive to false positive and false negative errors.

In Figure 4-3, the confusion matrix of the proposed model without data augmentation is presented for the three classes: benign, malignant, and normal. For the benign class, 82 samples were correctly classified, while 1 and 5 samples were misclassified as malignant and normal, respectively. The malignant class achieved 37 correct predictions, with 4 samples incorrectly classified as benign and 0 as normal. The normal class recorded 21 correct classifications, with 5 misclassified as benign and 0 as malignant.

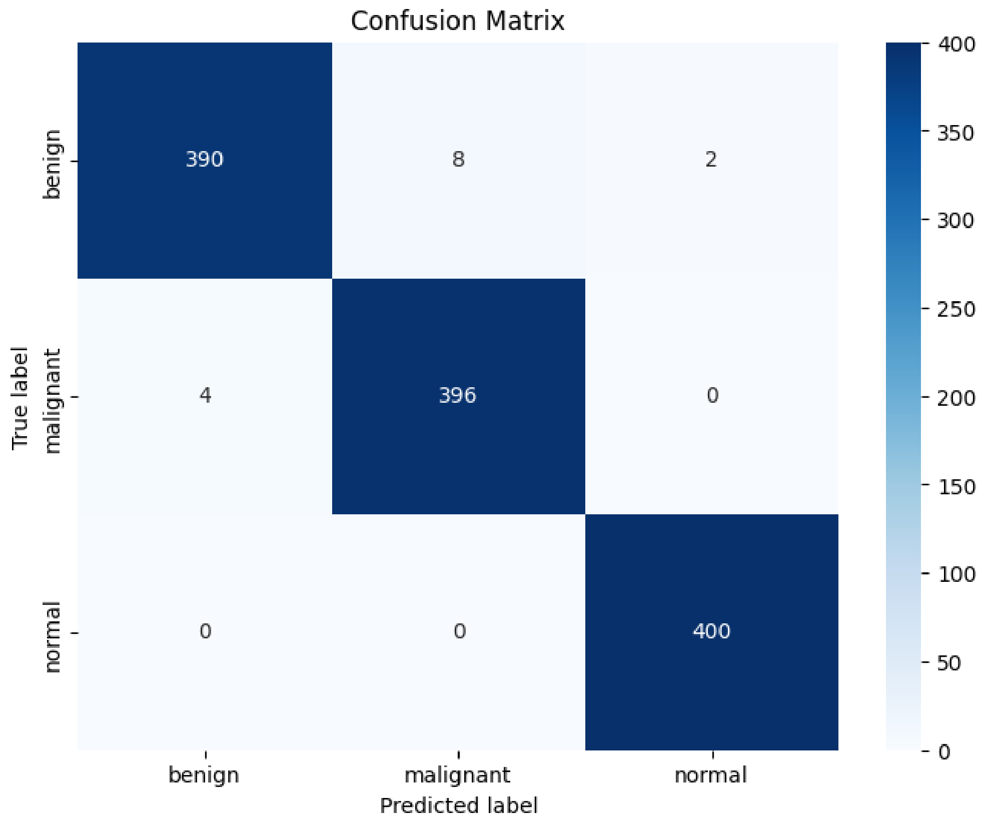

Figure 4-2 proposed model confusion matrix with data augmentation.

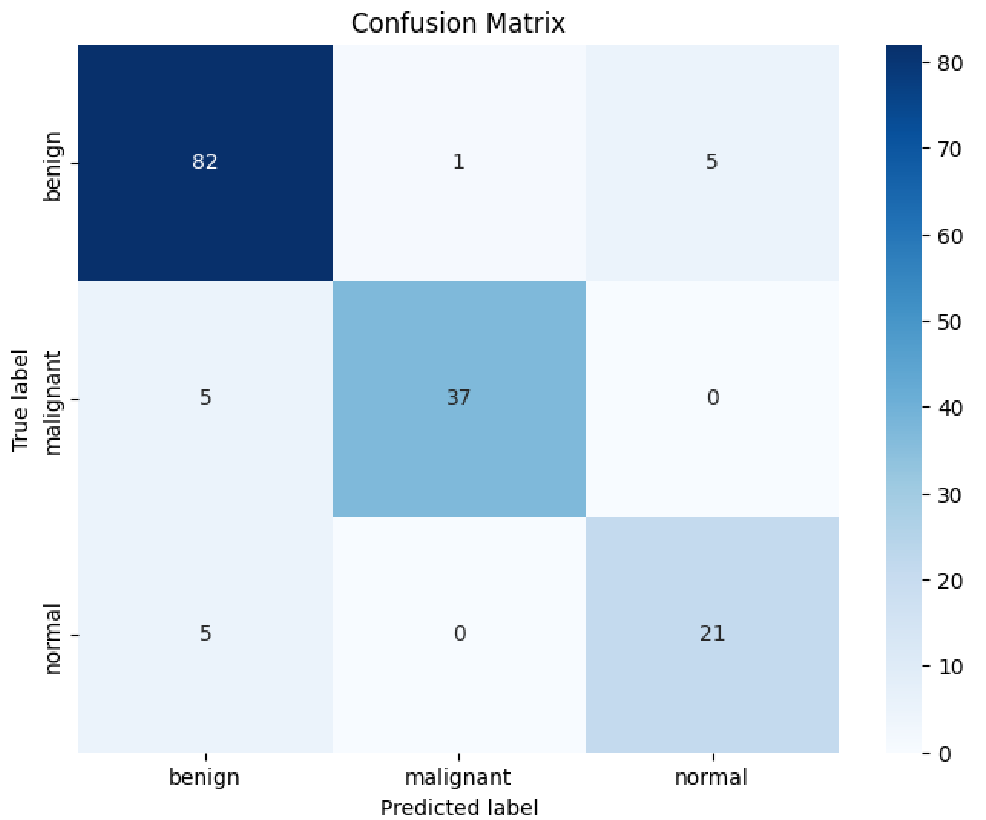

Figure 4-3 proposed model confusion matrix without data augmentation.

To evaluate the performance of the proposed custom CNN model, we generated a comparative confusion matrix for VGG16, one of the highest performing models evaluated in this study based on transfer learning. The VGG16 model reached a test accuracy of 99.33% with a macro averaged F1 score of 0.99 which puts it nearly on par with the proposed model. But the confusion matrix analysis presented in Figure *4-3*4 offers more detailed understanding of the model's performance on different class predictions.

The matrix of confusion shows that 392 benign samples out of 400 were correctly identified, meaning 6 samples were falsely classified as malignant, and 2 were labeled as normal. This exemplary count of true positives among benign cases confirms that the model can discern non-cancerous features. However, the false-positive counts in the normal class indicate some collusion between the visual representations of healthy tissues and benign tissues.

In the malignant class, 400 out of 400 samples were accurately diagnosed. It's 100% correctly classifies the cases, The ordinary normal class achieved 400 correct predictions out of 400 samples, It's 100% correctly classifies the cases as malignant, In comparison to the proposed custom CNN that exhibited less confusion among all classes, VGG16 showed analogous results for overall accuracy, but with greater off-diagonal values—especially with regards to classification of normal images. The custom CNN mislabeled 14 cases, while VGG16 misclassified 6 cases, Figure 4-5 VGG16 confusion matrix without data augmentation. The model correctly classified 77 benign, 29 malignant, and 9 normal samples, with some misclassifications across classes.

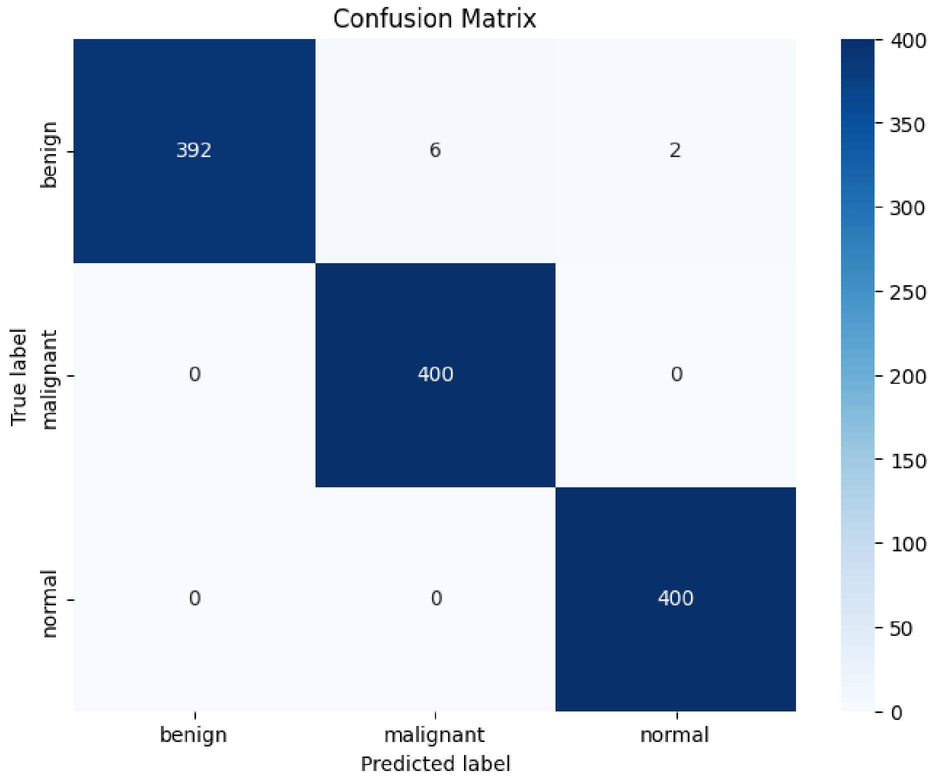

Figure 4-4 VGG16 confusion matrix with data augmentation.

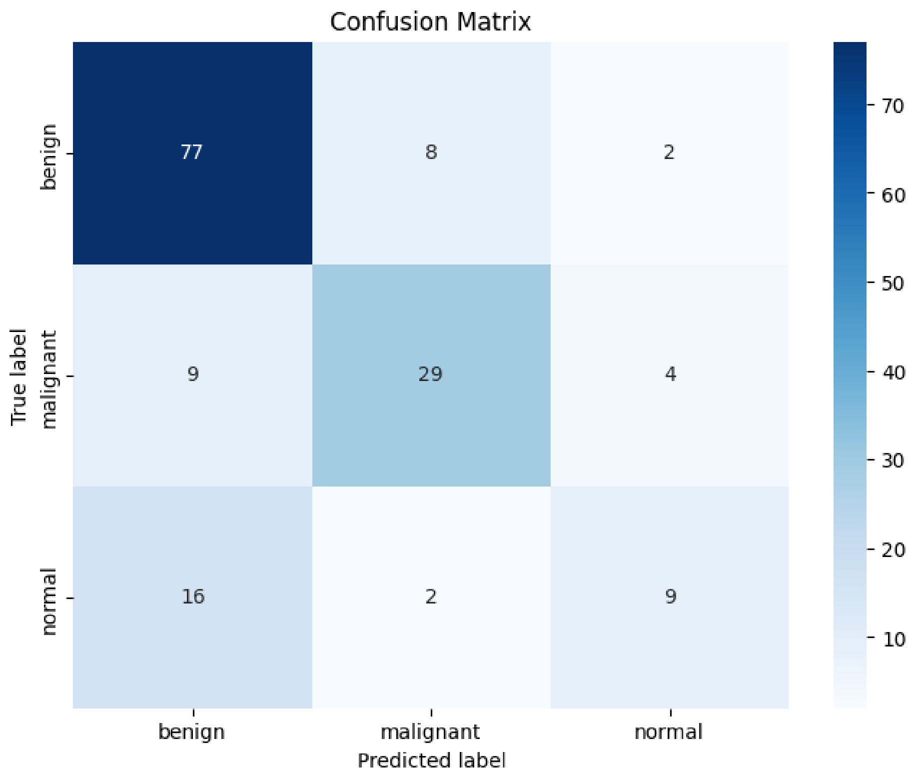

Figure 4-5 VGG16 confusion matrix without data augmentation.

4.4 Accuracy and Loss charts

In order to check the learning dynamics and the convergence behavior for the proposed custom CNN model, the training and validation accuracy and loss were evaluated over 100 epochs, which is shown in Figure *4-76*. From the accuracy graph (left panel), it is clear that there is an increase in training accuracy, which began at roughly 40% and surpassed 99% by the final epoch, as did the training confidence. Validation accuracy initially maintained greater volatility in the early epochs; however, it became noticeably more stable and began to improve significantly after epoch 60, converging to around 92–94%, which brings it in line with the training performance.

The right panel demonstrates a model loss plot, which shows effective learning has taken place. The training loss achieves convergence to a consistent decline, eventually leveling off at a training loss of around 0.5, indicating the error minimization objective has been achieved. Even though the validation loss does fluctuate, it still showcases a downward trend and runs parallel to the training loss after 40 epochs, indicating that generalization is indeed enhanced.

The unsupervised accuracy and loss changes at the beginning are expected with the training of many CNNs on datasets of medium size with data augmentation and dropout use. These oscillations reduce as the model tunes its parameters and learns higher level features. The final convergence of both curves of accuracy and loss suggests the model not only learned the representation but also generalized to new data effectively.

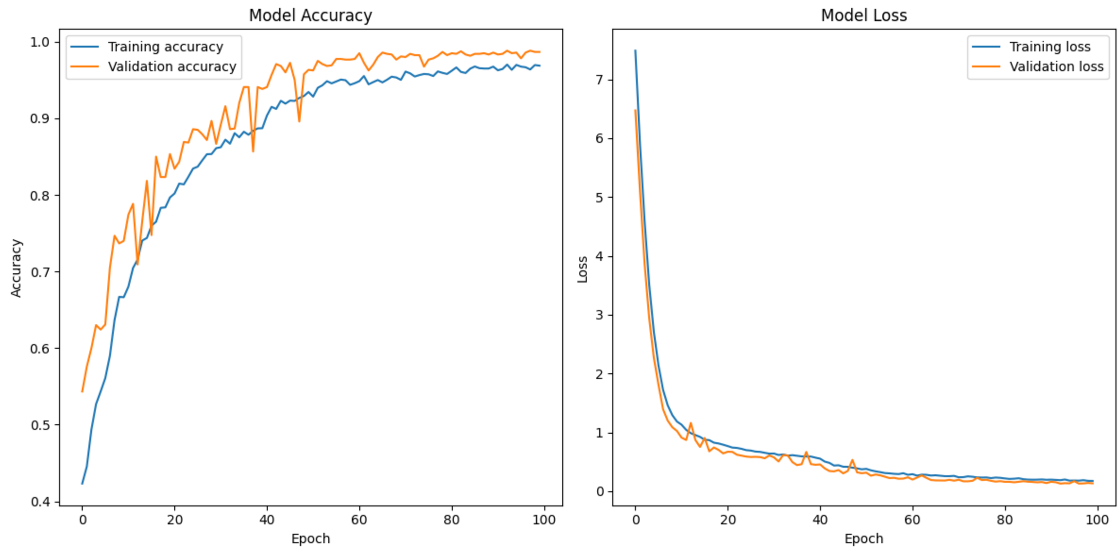

Figure 4-6 Proposed model loss and accuracy with data augmentation.

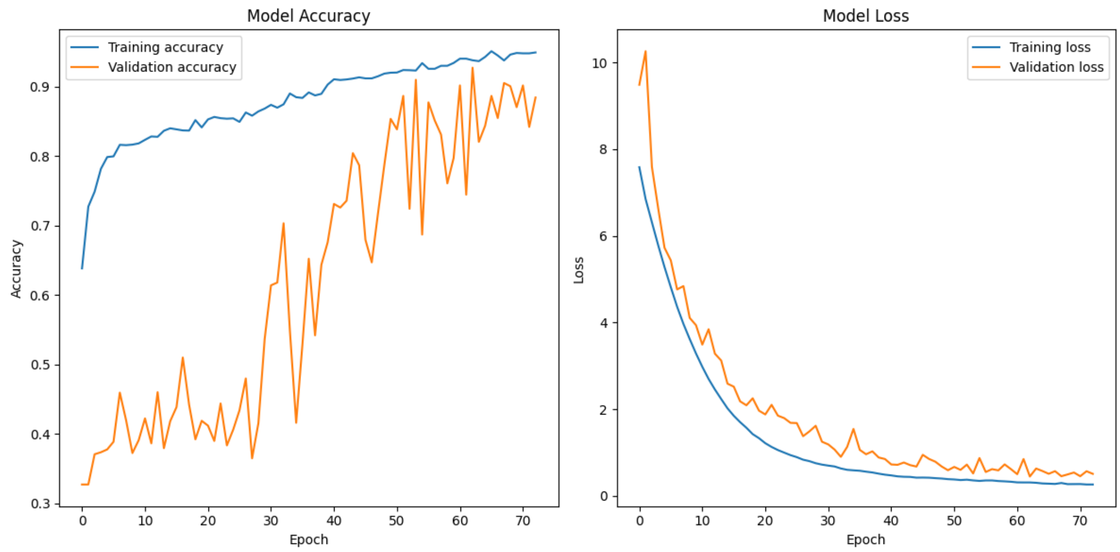

Figure 4-7 Proposed model loss and accuracy without data augmentation.

So that the performance of the proposed model can be assessed in more detail, it was compared to some recent works in the area of classification from images using deep learning techniques. A number of works between the years 2018 to 2024 were gathered, and their methodologies and corresponding accuracies were summarized and are showcased in Table *4-5* Comparison table of Transfer Learning and the proposed model without data augmentation.

In recent years, many researchers have developed CNN-based methods for classifying breast ultrasound images, as shown in

Table 4-6 Comparison of the proposed model with related studies. These methods vary in complexity, ranging from simple CNNs to more advanced models that use transfer learning, ensemble techniques, and optimization strategies.

For example, Daoud et al. (2020) combined CNNs with handcrafted features and achieved 96.1% accuracy, while Arooj et al. (2022) reached the highest reported accuracy of 99.4% using transfer learning with ResNet50 and VGG16. Similarly, Saba et al. (2022) applied transfer learning along with optimization techniques and achieved 92.8%, and Nikmah et al. (2024) used the Inception-ResNet v2 architecture to reach 89.7%.

Some studies explored alternative methods. Pathan et al. (2022) designed a multi-headed CNN and achieved 92.3%, while Karthik et al. (2022) used an ensemble of stacked CNNs to reach 92.1%. A Bayesian-optimized CNN by Sivanandan and Jayakumari (2023) reached 94%. Meanwhile, Alrubaie et al. (2023) achieved 96% using a basic CNN, and Alotaibi et al. (2023) used image fusion with CNNs, resulting in a lower accuracy of 87.8%.

In comparison, the model proposed in this study achieved a test accuracy of 98.83%, without relying on pre-trained models, ensemble methods, or external optimizers. While slightly below Arooj et al.'s top score, this result is still higher than most other studies, including several that used more complex or resource-intensive approaches.

This shows that with careful design and tuning, a custom CNN trained from scratch can achieve excellent results. The simplicity of the model also makes it more practical for real-world use, especially in situations where training time, model size, and computational resources are limited. Unlike many recent studies that depend on heavyweight methods to boost performance, our approach strikes a solid balance between accuracy, efficiency, and ease of implementation.

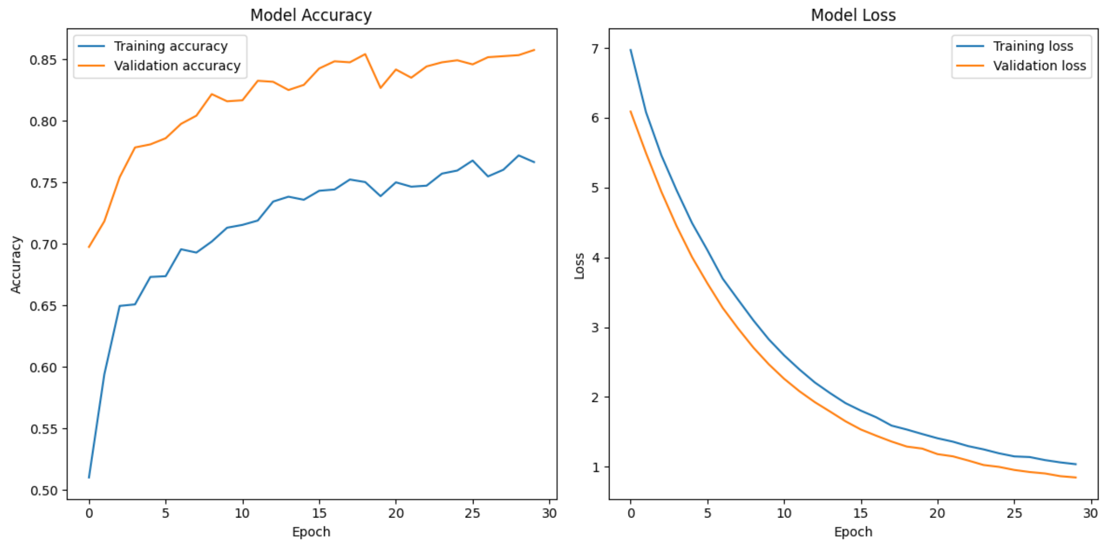

Figure 4-8 VGG16 loss and accuracy with data augmentation.

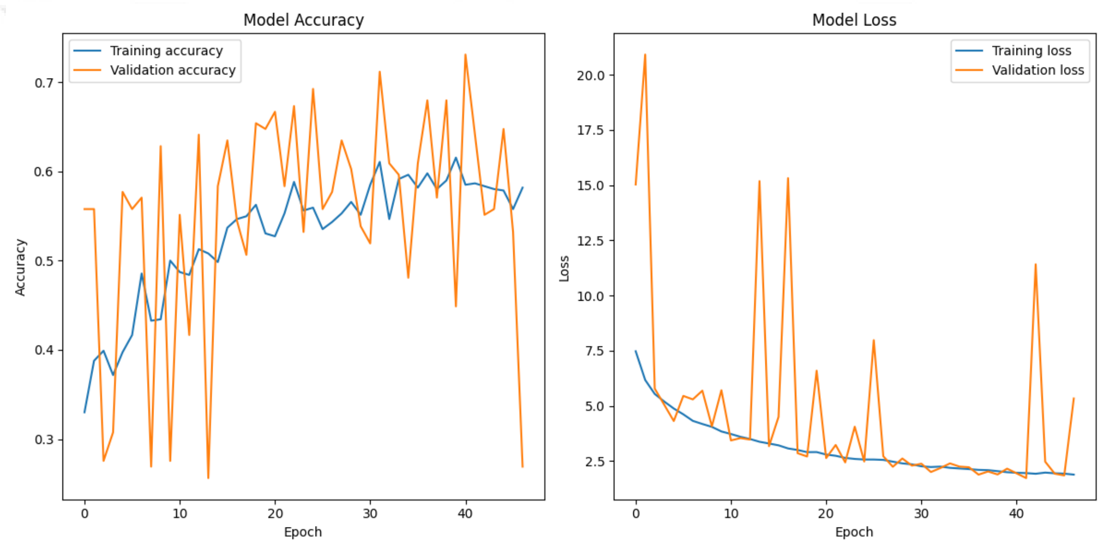

Figure 4-9 VGG16 loss and accuracy without data augmentation.

Table 4-6 Comparison of the proposed model with related studies.

Model	Year	Dataset	Method	Accuracy (%)
Daoud et al.	2020	Ultrasound images	Hybrid (CNN + Handcrafted Features)	96.1
Arooj et al.	2022	Ultrasound images	Transfer Learning (ResNet50, VGG16)	99.4
Pathan et al.	2022	BUSI	Multi-Headed CNN	92.31
Saba et al.	2022	Ultrasound images	Transfer Learning + Optimization	92.8
Karthik et al.	2022	Ultrasound images	Ensemble CNNs (Stacked)	92.1

Sivanandan &Jayakumari	2023	Ultrasound images	Bayesian Optimized CNN	94
Alotaibi et al.	2023	BUSI	CNN + Image Fusion	87.8
Alrubaie et al.	2023	BUSI	CNN	96
Nikmah et al.	2024	BUSI	Inception-ResNet v2	89.7
Proposed Model	2025	BUSI	Custom Optimised CNN	98.83

4.5 Gradient-weighted Class Activation Mapping

To increase the clinical interpretability of the proposed CNN model, we applied Gradient-weighted Class Activation Mapping (Grad-CAM) to highlight the regions of breast ultrasound images which were most influential to the model's predictions. With Grad-CAM, we are able to explain the model's predictions by generating heatmaps which show the most relevant spatial regions of the final convolutional layer that are activated for a given class. This approach allows us to analyze whether the model pays attention to clinically relevant features. The heatmaps produced from Grad-CAM show that for malignant cases, the model focus hypoechoic regions with irregular margins and posterior acoustic shadowing which are associated with malignancy. In benign and normal cases, the focus shifts to more homogeneous or cystic structures which matches established patterns radiologists use. These forms of explanations not only support that the model proves to be useful in learning various meaningful diagnosis distinguishing features, but these also serve as a basis for accepting trust in automated decisions. Hence, the model's interpretability is enhanced, emphasizing its application in clinical decision systems, especially where the experience of a radiologist can be a constraint.

The custom CNN produces heatmaps that are relatively fragmented, with multiple scattered activation regions across the input, particularly for malignant and normal cases as showen in Figure *4-10* Custom CNN Grad-CAM of breast ultrasound images.. This indicates that the model is attending to fine-grained textural variations, but in a less consistent and more noisy manner, which may reduce interpretability. In contrast, the VGG16-based model generates smoother and more centralized activation maps that focus strongly on the lesion regions. For benign and malignant samples, the VGG16 Grad-CAM highlights broader and more structured areas that correspond more closely to the underlying pathology, while for normal tissue it shows a more localized but still coherent region of attention. Overall, the VGG16 transfer learning model demonstrates clearer and clinically more interpretable attention patterns, whereas the custom CNN captures subtle details but with less spatial consistency.

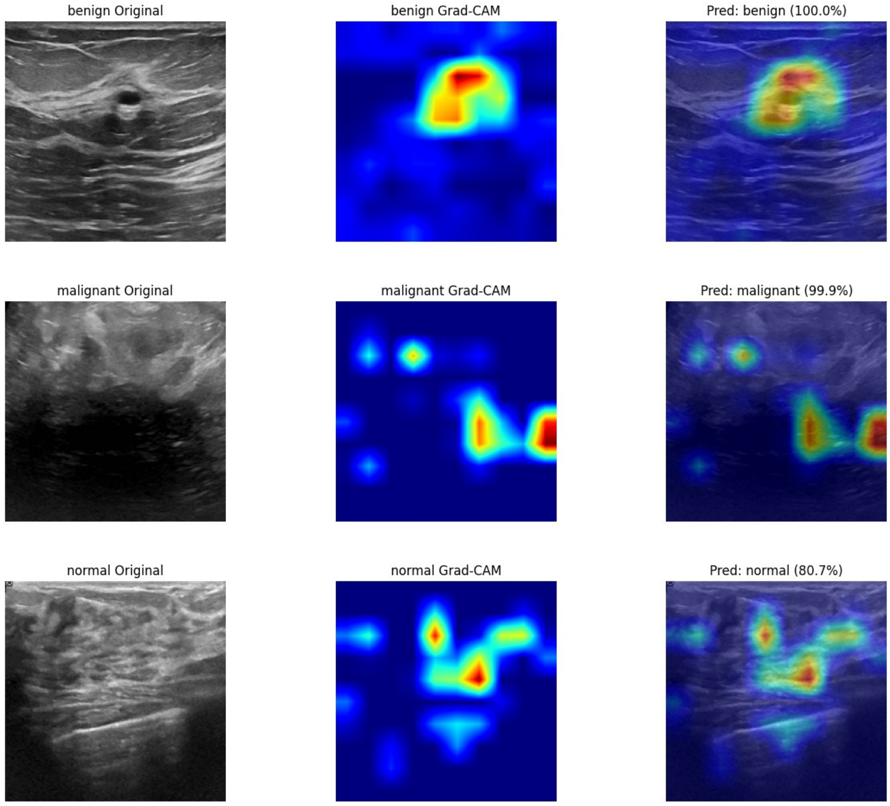

Figure 4-10 Custom CNN Grad-CAM of breast ultrasound images.

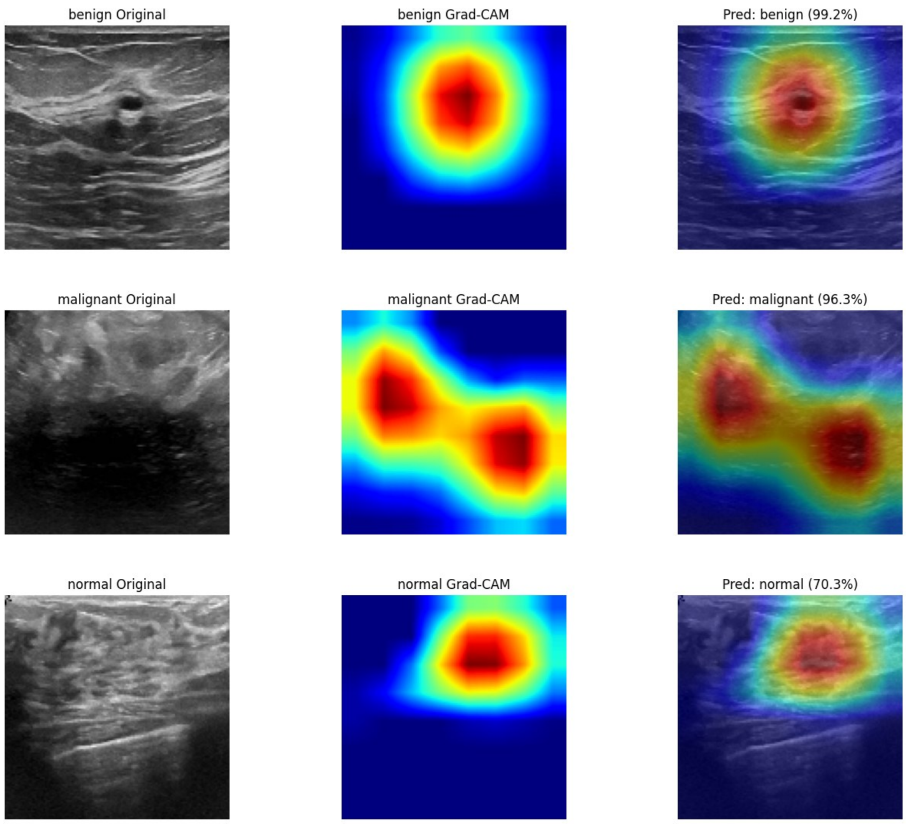

Figure 4-11 VGG16 Grad-CAM of breast ultrasound images.

5. CONCLUSION

This study presents a custom-designed CNN architecture for the classification of breast ultrasound images into benign, malignant, and normal categories, greatly advancing it. Because we systematically approached the task via augmenting data, optimizing architecture, as well as fine-tuning hyperparameters rigorously, the model performs at a state-of-the-art level without reliance on pre-trained weights or ensemble strategies. The model attains a test accuracy of 98.83%. It shows that it has the capacity in order to match and even exceed the performance of leading transfer learning architectures while maintaining a much lower computational footprint.

The results show that developing domain-specific models tactically benefits more than generic, pre-trained alternatives, especially where computational efficiency and interpretability are necessary. Furthermore, the model balances sensitivity among all classes, especially as it strongly performs malignant case detection, underscoring its potential clinical relevance for early diagnosis and intervention.

The study acknowledges that there is a need for further validation on diverse and larger multi-center datasets and for the integration of explainability mechanisms in order to support clinical trust and decision-making, while remaining promising. A solid structure exists within the projected method for using diagnostic tools driven by AI. That is notably accurate within healthcare environments where radiological knowledge access is restricted. The focus of future work will be extending this framework through real-time deployment and model explainability, and semi-supervised learning.

6. REFERENCES

A. Hijab, M. A. Rushdi, M. M. Gomaa and A. Eldeib, "Breast Cancer Classification in Ultrasound Images using Transfer Learning," 2019 Fifth International Conference on Advances in Biomedical Engineering (ICABME), Tripoli, Lebanon, 2019, pp. 1-4, doi: 10.1109/ICABME47164.2019.8940291.

Almajalid, R., Shan, J., Du, Y., & Zhang, M. (2018). Development of a Deep-Learning-Based Method for Breast Ultrasound Image Segmentation. *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, 1103–1108. https://doi.org/10.1109/ICMLA.2018.00179

Alotaibi, M., Aljouie, A., Alluhaidan, N., Qureshi, W., Almatar, H., Alduhayan, R., Alsomaie, B., & Almazroa, A. (2023). Breast cancer classification based on convolutional neural network and image fusion approaches using ultrasound images. *Heliyon*, *9*(11). https://doi.org/10.1016/j.heliyon.2023.e22406

Alrubaie, H., K. Aljobouri, H., J. AL-Jobawi, Z., & Çankaya, I. (2023). Convolutional Neural Network Deep Learning Model for Improved Ultrasound Breast Tumor Classification. *Al-Nahrain Journal for Engineering Sciences*, *26*(2), 57–62. https://doi.org/10.29194/njes.26020057

Arooj, S., Atta-ur-Rahman, Zubair, M., Khan, M. F., Alissa, K., Khan, M. A., & Mosavi, A. (2022). Breast Cancer Detection and Classification Empowered With Transfer Learning. *Frontiers in Public Health*, *10*. https://doi.org/10.3389/fpubh.2022.924432

Ayana, G., Dese, K., & Choe, S. W. (2021). Transfer learning in breast cancer diagnoses via ultrasound imaging. In *Cancers* (Vol. 13, Issue 4, pp. 1–16). MDPI AG. https://doi.org/10.3390/cancers13040738

Boulenger, A., Luo, Y., Zhang, C., Zhao, C., Gao, Y., Xiao, M., Zhu, Q., & Tang, J. (2023). Deep learning-based system for automatic prediction of triple-negative breast cancer from ultrasound images. *Medical and Biological Engineering and Computing*, *61*(2), 567–578. doi.org/10.1007/s11517-022-02728-4

Cheyi, J., & Çetin-Kaya, Y. (2024). Advanced CNN-based classification and segmentation for enhanced breast cancer ultrasound imaging. *Gazi University Journal of Science Part A: Engineering and Innovation,* *11*(4), 647–667. https://doi.org/10.54287/gujsa.1529857

Ciobotaru, A., Bota, M. A., Goța, D. I., & Miclea, L. C. (2023). Multi-Instance Classification of Breast Tumor Ultrasound Images Using Convolutional Neural Networks and Transfer Learning. *Bioengineering*, *10*(12). https://doi.org/10.3390/bioengineering10121419

Çetin-Kaya, Y. (2024). Equilibrium optimization-based ensemble CNN framework for breast cancer multiclass classification using histopathological image. *Diagnostics, 14*(19), 2253. https://doi.org/10.3390/diagnostics14192253

Daoud, M. I., Abdel-Rahman, S., & Alazrai, R. (2019). Breast ultrasound image classification using a pre-trained convolutional neural network. *Proceedings - 15th International Conference on Signal Image Technology and Internet Based Systems, SISITS 2019*, 167–171. https://doi.org/10.1109/SITIS.2019.00037

Daoud, M. I., Abdel-Rahman, S., Bdair, T. M., Al-Najar, M. S., Al-Hawari, F. H., & Alazrai, R. (2020). Breast tumor classification in ultrasound images using combined deep and handcrafted features. *Sensors (Switzerland)*, *20*(23), 1–20. https://doi.org/10.3390/s20236838

din, N. M. ud, Dar, R. A., Rasool, M., & Assad, A. (2022). Breast cancer detection using deep learning: Datasets, methods, and challenges ahead. In *Computers in Biology and Medicine* (Vol. 149). Elsevier Ltd. https://doi.org/10.1016/j.compbiomed.2022.106073

Eroğlu, Y., Yildirim, M., & Çinar, A. (2021). Convolutional Neural Networks based classification of breast ultrasonography images by hybrid method with respect to benign, malignant, and normal using mRMR. *Computers in Biology and Medicine*, *133*. https://doi.org/10.1016/j.compbiomed.2021.104407

Fujioka, T., Katsuta, L., Kubota, K., Mori, M., Kikuchi, Y., Kato, A., Oda, G., Nakagawa, T., kitazume, Y., & Tateishi, U. (2020). Classification of Breast Masses on Ultrasound Shear Wave Elastography using Convolutional Neural Networks. *Ultrasonic Imaging*, *42*(4–5), 213–220. https://doi.org/10.1177/0161734620932609

Güldoğan, e., Ucuzal, H., Küçükakçalı, Z., & Çolak, C. (2021). Transfer Learning-Based Classification of Breast Cancer using Ultrasound Images. *Middle Black Sea Journal of Health Science*, *7*(1), 74–80. https://doi.org/10.19127/mbsjohs.876667

Harrison, P., Hasan, R., & Park, K. (2023). State-of-the-Art of Breast Cancer Diagnosis in Medical Images via Convolutional Neural Networks (CNNs). In *Journal of Healthcare Informatics Research* (Vol. 7, Issue 4, pp. 387–432). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s41666-023-00144-3

Karthik, R., Menaka, R., Kathiresan, G. S., Anirudh, M., & Nagharjun, M. (2022). Gaussian Dropout Based Stacked Ensemble CNN for Classification of Breast Tumor in Ultrasound Images. *IRBM*, *43*(6), 715–733. https://doi.org/10.1016/j.irbm.2021.10.002

Kaya, M., & Cetin-Kaya, Y. (2024). A Novel Deep Learning Architecture Optimization for Multiclass Classification of Alzheimer's Disease Level. *IEEE Access*, *12*, 46562–46581. https://doi.org/10.1109/ACCESS.2024.3382947

Kaya, M., & Çetin-Kaya, Y. (2024). A novel ensemble learning framework based on a genetic algorithm for the classification of pneumonia. *Engineering Applications of Artificial Intelligence, 133*, 108494. https://doi.org/10.1016/j.engappai.2024.108494

Kaya, M. (2025). Efficient diagnosis of retinal diseases using convolutional neural networks. *Gazi University Journal of Science Part A: Engineering and Innovation, 12*(1), 15–35. https://doi.org/10.54287/gujsa.1592915

Kriti, Virmani, J., & Agarwal, R. (2020). Deep feature extraction and classification of breast ultrasound images. *Multimedia Tools and Applications*, *79*(37–38), 27257–27292. https://doi.org/10.1007/s11042-020-09337-z

Latif, G., Butt, M. O., Yousif, F., Anezi, A., & Alghazo, J. (2020). *Ultrasound Image Despeckling and detection of Breast Cancer using Deep CNN*.

Lazo, J. F., Moccia, S., Frontoni, E., & de Momi, E. (2020). *Comparison of different CNNs for breast tumor classification from ultrasound images*. http://arxiv.org/abs/2012.14517

Li, Y., Gu, H., Wang, H., Qin, P., & Wang, J. (2022). BUSnet: A Deep Learning Model of Breast Tumor Lesion Detection for Ultrasound Images. *Frontiers in Oncology*, *12*. https://doi.org/10.3389/fonc.2022.848271

Liu, H., Cui, G., Luo, Y., Guo, Y., Zhao, L., Wang, Y., Subasi, A., Dogan, S., & Tuncer, T. (2022). Artificial Intelligence-Based Breast Cancer Diagnosis Using Ultrasound Images and Grid-Based Deep Feature Generator. *International Journal of General Medicine*, *15*, 2271–2282. https://doi.org/10.2147/IJGM.S347491

Lu, Y., Li, J. Y., Su, Y. T., & Liu, A. A. (2018). A review of breast cancer detection in medical images. 2018 IEEE Visual Communications and Image Processing (VCIP), 1-4. doi.org/10.1109/VCIP.2018.8698732

Muduli, D., Dash, R., & Majhi, B. (2022). Automated diagnosis of breast cancer using multi-modal datasets: A deep convolution neural network based approach. *Biomedical Signal Processing and Control*, *71*. https://doi.org/10.1016/j.bspc.2021.102825

Nikmah, T. L., Syafei, R. M., & Anisa, D. N. (2024). Inception ResNet v2 for Early Detection of Breast Cancer in Ultrasound Images. *Journal of Information System Exploration and Research*, *2*(2). https://doi.org/10.52465/joiser.v2i2.439

Özatılgan, A., & Kaya, M. (2024). A lightweight convolutional neural network for classification of brain tumors using magnetic resonance imaging. *Sakarya University Journal of Computer and Information Sciences,* *7*(3), 482–493. https://doi.org/10.35377/saucis.1518139

Pathan, R. K., Alam, F. I., Yasmin, S., Hamd, Z. Y., Aljuaid, H., Khandaker, M. U., & Lau, S. L. (2022). Breast Cancer Classification by Using Multi-Headed Convolutional Neural Network Modeling. *Healthcare (Switzerland)*, *10*(12). https://doi.org/10.3390/healthcare10122367

S. M. Badawy, A. E. -N. A. Mohamed, A. A. Hefnawy, H. E. Zidan, M. T. GadAllah and G. M. El-Banby, "Classification of Breast Ultrasound Images Based on Convolutional Neural Networks - A Comparative Study," *2021 International Telecommunications Conference (ITC-Egypt)*, Alexandria, Egypt, 2021, pp. 1-8, doi: 10.1109/ITC-Egypt52936.2021.9513972.

Saba, T., Abunadi, I., Sadad, T., Khan, A. R., & Bahaj, S. A. (2022). Optimizing the transfer-learning with pretrained deep convolutional neural networks for first stage breast tumor diagnosis using breast ultrasound visual images. *Microscopy Research and Technique*, *85*(4), 1444–1453. https://doi.org/10.1002/jemt.24008

Sahu, A., Das, P. K., & Meher, S. (2024). An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images. *Biomedical Signal Processing and Control*, *87*. https://doi.org/10.1016/j.bspc.2023.105377

Sivanandan, R., & Jayakumari, & J. (2023). *A new CNN architecture for efficient classification of ultrasound breast tumor images with activation map clustering based prediction validation*. https://doi.org/10.1007/s11517-021-02357-3

Sivanandan, R., & Jayakumari, J. (2023). Bayesian optimized novel CNN for improved diagnosis from ultrasound breast tumor images. *Multimedia Tools and Applications*, *82*(15), 22815–22833. https://doi.org/10.1007/s11042-023-14468-0

Uysal, F., & Köse, M. M. (2022). Classification of Breast Cancer Ultrasound Images with Deep Learning-Based Models †. *Engineering Proceedings*, *31*(1). https://doi.org/10.3390/ASEC2022-13791

Wu, G.-G., Zhou, L.-Q., Xu, J.-W., Wang, J.-Y., Wei, Q., Deng, Y.-B., Cui, X.-W., & Dietrich, C. F. (2019). Artificial intelligence in breast ultrasound. *World Journal of Radiology*, *11*(2), 19–26. https://doi.org/10.4329/wjr.v11.i2.19

Yap, M. H., Pons, G., Martí, J., Ganau, S., Sentís, M., Zwiggelaar, R., Davison, A. K., & Martí, R. (2018). Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks. *IEEE Journal of Biomedical and Health Informatics*, *22*(4), 1218–1226. https://doi.org/10.1109/JBHI.2017.2731873

Zakareya, S., Izadkhah, H., & Karimpour, J. (2023). A New Deep-Learning-Based Model for Breast Cancer Diagnosis from Medical Images. *Diagnostics*, *13*(11). https://doi.org/10.3390/diagnostics13111944

Zeimarani, B., Costa, M. G. F., Nurani, N. Z., Bianco, S. R., de Albuquerque Pereira, W. C., & Filho, C. F. F. C. (2020). Breast Lesion Classification in Ultrasound Images Using Deep Convolutional Neural Network. *IEEE Access*, *8*, 133349–133359. https://doi.org/10.1109/ACCESS.2020.3010863

Zhang, X., Li, H., Wang, C., Cheng, W., Zhu, Y., Li, D., Jing, H., Li, S., Hou, J., Li, J., Li, Y., Zhao, Y., Mo, H., & Pang, D. (2021). Evaluating the Accuracy of Breast Cancer and Molecular Subtype Diagnosis by Ultrasound Image Deep Learning Model. *Frontiers in Oncology*, *11*. https://doi.org/10.3389/fonc.2021.623506

RESUME