T.C.

TOKAT GAZİOSMANPAŞA UNIVERSITY

GRADUATE EDUCATION INSTITUTE

DEPARTMENT OF COMPUTER ENGINEERING

M.SC RESEARCH

BREAST CANCER DETECTION WITH CONVOLUTIONAL NEURAL NETWORK USING ULTRASOUND IMAGES

Prepared by

Jehad Jamal Othman

Department of Computer Engineering

Master Thesis

Advisor: Asst. Prof. Dr. Yasemin ÇETIN KAYA

TOKAT
October, 2024

ETHICAL AGREEMENT

I hereby state that this Master's thesis entitled "Detection of Breast Cancer by Using Convolutional Neural Networks with Ultrasound Images" 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 Asst. Prof. Dr. Yasemin Çetin 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.

…./…./…….

Jehad Jamal Othman

JURY ACCEPTANCE AND APPROVAL

The defense examination of the thesis titled “**Breast Cancer Detection With Convolutional Neural Network Using Ultrasound Images**” prepared by Jehad Jamal Othman, was held on 10/10/2024 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**

Chairman: Asst. Prof. Dr. Yasemin Çetin Kaya (Advisor)

Member: Assoc. Prof. Dr. Turgut Özseven

Member: Asst. Prof. Dr. Yavuz Ünal

APPROVAL

……./……/…….

Assoc. Prof. Dr.Yusuf TEMÜR

Director of the Graduate Education Institute

ÖZET

BREAST CANCER DETECTION WITH CONVOLUTIONAL NEURAL NETWORK USING ULTRASOUND IMAGES

Jehad Jamal Othman

Bilgisayar Mühendisliği Bölümü

Danışman: Dr. Öğr. Üyesi Yasemin ÇETIN KAYA

2024, xiv ve 54 sayfa

Bu araştırma, Meme Kanserinin (MK) erken teşhisi ve tanısının iyileştirilmesi amacıyla Meme Ultrason Görüntülerinin (BUSI) sınıflandırılması ve segmentlere ayrılması için gelişmiş Evrişimsel Sinir Ağı (ESA) modellerinin geliştirilmesini ve değerlendirilmesini amaçlamaktadır. MK, en sık görülen kötü huylu tümörler arasında yer almakta ve dünya genelinde kadınlarda kansere bağlı ölümlerin en büyük ikinci nedenidir; bu da erken ve doğru tanıyı, mümkün olan en yüksek sağkalım sonuçlarıyla etkili tedavi için öncelikli bir görev haline getirmektedir. Çalışmada, erken tespit için derin öğrenme modellerini eğitmek ve değerlendirmek amacıyla BUSI veri seti kullanılmıştır. Mevcut yaklaşımlar arasında görüntülerin ön işlenmesi, veri artırımı ve transfer öğrenmesi yer almaktadır. Görüntü ön işleme bölümünde, segmentasyon, iyileştirme ve üst üste bindirme ile bazı iyileştirmeler gerçekleştirdik. Bu ön işleme adımları, geliştirilen ESA modellerinin iyi huylu dokuları kötü huylu dokulardan yüksek bir doğruluk seviyesinde ayırt edebilmesini sağlamak için gereklidir. Araştırma çalışmasında özel bir model önerildi ve bu modelin performansı doğruluk, kesinlik, geri çağırma ve F1 puanı gibi performans ölçütleriyle VGG-Net, DenseNet, EfficientNet, MobileNet, InceptionV3, Xception ve ResNet önceden eğitilmiş modelleriyle karşılaştırıldı.. Uygulanan görüntü ön işleme teknikleriyle birlikte önerilen ESA modeli tarafından elde edilen %97'lik yüksek doğruluk, bu ön işleme olmadan eğitilen modellere göre önemli ilerlemeler göstermiştir. Ayrıca, önceden eğitilmiş bir modele, geliştirilmiş ve üst üste bindirilmiş görüntülerden oluşan veri setimizle ince ayar yaparak transfer öğrenmenin etkisini inceledik. Bu yaklaşımı kullanarak dikkate değer performans iyileştirmeleri elde ettik; en iyi durumda, ulaşılan en yüksek doğruluk VGG19 ve EfficientNetB0 ile %99 oldu. Ayrıca, tıbbi görüntü analizinde son derece iyi performans gösterdiği kanıtlanmış olan U-Net mimarisini etkili görüntü segmentasyonu için test ettik.

Ayrıca modeli yorumlanabilir hale getirmek için Grad-CAM'i kullandık, böylece ESA'ların karar verirken görüntülerin hangi kısımlarına odaklandığını görebiliriz. Bu yorumlanabilirlik, modellerin karar verme süreçlerine dair içgörü kazanmak ve tahminlerine güven oluşturmak için çok önemlidir. Çalışmamızın sonuçları, önerilen özel CNN modelinin etkili görüntü ön işleme teknikleriyle birleştirilmesiyle BUSI için daha gelişmiş sınıflandırma ve segmentasyona ulaşılabileceğini göstermektedir. Radyologlar tarafından MKnin erken tespitine yardımcı olma ve sonuçta hastaların durumunu iyileştirme potansiyeline sahiptir. Bu derin öğrenme metodolojileri, MK tanısı için gelişmiş, daha doğru ve güvenilir araçlar sağlamak için kullanılacaktır; bu, daha iyi sağ kalım ve tedaviye yanıt sağlamada önemlidir.

Anahtar Kelimeler: Derin Öğrenme, CNN, Meme Kanseri, Sınıflandırma, Görüntü Işleme

ABSTRACT

BREAST CANCER DETECTION WITH CONVOLUTIONAL NEURAL NETWORK USING ULTRASOUND IMAGES

Jehad Jamal Othman

Department of Computer Engineering

Advisor: Asst. Prof. Dr. Yasemin ÇETIN KAYA

2024, xiv and 54 pages

This research aims to develop and evaluate advanced Convolutional Neural Network (CNN) models for classifying and segmenting Breast Ultrasound Images (BUSI) for improved early detection and diagnosis of Breast Cancer (BC). BC is among the most frequent malignancies and the second largest cause of cancer-related mortality in women globally, making early and accurate diagnosis a prime task toward effective treatment with maximal possible survival outcomes. In the study, BUSI dataset was used to train and evaluate deep learning models for early detection. The available approaches include preprocessing images, data augmentation, and transfer learning. In the image preprocessing part, we performed some enhancements with segmentation, enhancement, and overlaying. These preprocessing steps are essential to ensure that the developed CNN models can distinguish benign from malignant tissues at a high level of accuracy. In the research work, a custom model is proposed and its performance is compared with the pre-trained models VGG-Net, DenseNet, EfficientNet, MobileNet, InceptionV3, Xception and ResNet on performance metrics such as accuracy, precision, recall and F1 score. The high accuracy of 97%, achieved by the proposed CNN model in combination with the image preprocessed techniques applied, showed significant advancements over those models trained without this preprocessing. Also, we studied the impact of transfer learning by fine-tuning a pre-trained model with our dataset of enhanced and overlayed images. We obtained remarkable performance improvements using this approach; in the best case, the highest accuracy reached was 99% with VGG19 and EfficientNetB0. Furthermore, we tested the U-Net architecture for effective image segmentation since it has been proven to perform exceptionally well with medical image analysis. We further used Grad-CAM to make the model interpretable so we may see which parts of the images the CNNs focus on when making decisions. This interpretability is crucial for gaining insights into the models' decision-making processes and building trust in their predictions. The results of our study suggest that more advanced classification and segmentation for BUSI can be reached by combining the custom CNN model presented with effective image preprocessing techniques. It has the potential to help in the early detection of BC by radiologists, ultimately bettering the condition of patients. These deep learning methodologies will be used to provide advanced more accurate, and reliable tools for the diagnosis of BC—essential in ensuring better survival and response to treatment.

Keywords: Deep Learning, CNN, Breast Cancer, Classification, Image Processing

ACKNOWLEDGMENT

I convey my sincerest thanks to my advisor, Dr. Yasemin Çetin Kaya, for her constant support, direction, and encouragement during my research path. Her knowledge and insightful suggestions have been crucial in shaping this work. I convey my heartfelt thanks to the faculty members and staff of the Computer Engineering Department at Tokat Gaziosmanpaşa University for providing a conducive learning atmosphere and the necessary resources to carry out this research. A particular thanks to my family and friends for their continual support, patience, and understanding during my education. Their belief in me has been a constant source of motivation. I am also grateful to my colleagues and fellow researchers who have contributed to this effort through constructive talks and cooperation. Their different perspectives have enriched this research. Lastly, I would like to recognize the help of Tokat Gaziosmanpaşa University for providing the facilities and support necessary to accomplish this research.

DEDICATED TO

To my family, who through their support and encouragement have provided a sustaining source of strength throughout this journey. To my guide and supervisor who, through their valuable guidance, wisdom, and believing in me, have been a mainstay. To the countless scholars and pioneers whose work has provided a basis for this endeavour. And finally, to those inquirers after knowledge and truth, may this work, in however humble a manner, contribute to the improvement of science and society.

I now dedicate this research with deep gratitude and respect to you all.

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
NF-Net	: Noise Filter Network
ResNet	: Residual Network
SSD	: Single Shot Multibox Detector
TL	: Transfer Learning
TP	: True Positives
TN	: True Negatives
FP	: False Positives
FN	: False Negatives
U-Net	: U-shaped Network
VGG	: Visual Geometry Group
YOLO	: You Only Look Once

1. INTRODUCTION

The most prevalent cancer among women globally is BC (T. Wu et al., 2019). It is the second biggest cause of death for women (Luo et al., 2024). BC affects roughly 2.89 million women globally each year, accounting for 24.2% of all instances of female cancer and ranking first(Kabir et al., 2021; PACAL, 2022). Early detection of the disease is essential since it aids in lowering the number of premature deaths (Cruz-Ramos et al., 2023; Jabeen et al., 2022; Pourasad et al., 2021; Wu et al., 2019). Over 30% of all cancer fatalities are caused by BC, making it the deadliest cancer for women (Vigil et al., 2022). BC is identified with a variety of imaging methods, including mammography, ultrasound, scanning with magnetic resonance imaging (MRI), and electronic pathology images. Mammography is widely utilized for early detection. However, it has limits in sensitivity and specificity, particularly in thick breasts (Badawy et al., 2021). For BUSIs, state-of-the-art CNN methods like YOLO and Single Shot Multibox Detector have shown significant success in detecting breast lesions (Fujioka et al., 2020). MRI provides detailed images of the breast and is especially helpful in assessing high-risk patients and assessing tumor extent (Peng et al., 2023). Digital pathology pictures are the "golden standard" for recognizing cancer and are critical in cancer detection. Deep learning has demonstrated encouraging outcomes in increasing BC detection and categorization precision and efficiency. Deep learning has also been utilized to identify BC via ultrasound images, where transfer learning and data augmentation approaches have been applied to increase performance(G.-G. Wu et al., 2019). Deep learning methods, such as CNNs, have been used to assess medical images, which are mammography, ultrasound, MRI, and images from pathology (Mahoro & Akhloufi, 2022). Deep learning has significantly improved diagnosis accuracy and treatment efficiency in a variety of medical areas (Çetin-Kaya & Kaya, 2024; Kaya & Çetin-Kaya, 2024). Deep learning algorithms can extract meaningful characteristics from imagery and make classifications based on them (Luo et al., 2024). For example, in mammogram-based screening and diagnosis, deep learning models have been used to classify mammograms into different categories, such as malignant, benign, or normal (Y. Zhang et al., 2021). Transfer learning is pre-training a network on a big dataset and fine-tuning it for a specific job. It has also been employed to enhance the efficiency of deep learning algorithms in BC imaging (Luo et al., 2024). There are challenges, such as the need for interdisciplinary cooperation, standardized and open databases, and addressing deep learning models' poor generalization ability and interpretability (Han et al., 2017). Figure 1-1 shows the ultrasound imaging process in deep learning. In this article, we start on a journey into the transformative potential of deep learning in BUSI, characterized by a diversity of innovative solutions that reach far beyond simple classification, with the development of a novel strategy and solutions. This study seeks to overcome these constraints by investigating complex approaches such as deep learning, notably CNNs, to improve diagnostic precision and help radiologists in ultrasound imaging. Our research seeks to enhance BC diagnosis utilizing ultrasound pictures and powerful deep-learning algorithms. Our contributions are as follows:

- Developed a custom CNN model achieving 97% classification accuracy for BC detection in ultrasound images, outperforming various state-of-the-art algorithms.
- Applied advanced image processing techniques, including U-Net for segmentation and Grad-CAM for model interpretability, to improve diagnostic accuracy and transparency.

The following describes the latter bits of this research. Section 2 addresses related works; Section 3 offers the dataset and approach of the investigation. Section 4 offers

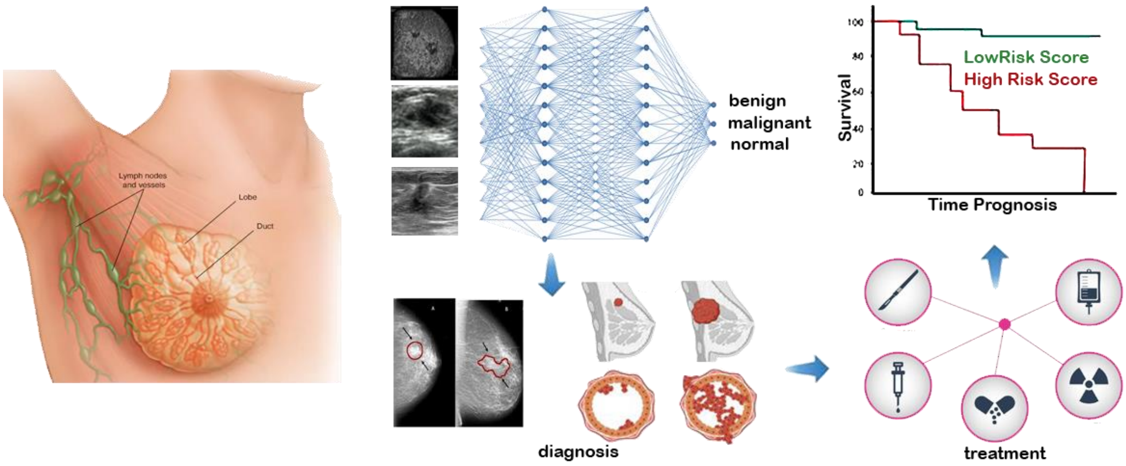


Figure 1-1 Overview of ultrasound imaging using deep learning for breast cancer. Prognosis, treatment response prediction, screening, and diagnosis are common uses of deep learning.

the findings of the research. Section 5 presents the findings of the study together with suggestions for the next investigations.

While these results are excellent for the present investigation, some limitations have to be discussed. The dataset collected is on ultrasound pictures from 600 patients, which are private and used for this study. Although it forms a good basis for developing and testing our models, it is rather small compared to the datasets referred to in other works applying deep learning. This may constitute some limit to generalizing the performance of the models. Additionally, the dataset may lack diversity in demographics, imaging equipment, and clinical settings, potentially impacting the robustness of the models across different populations and environments.

The quality of ultrasound images can vary significantly due to operator skill, equipment quality, and patient characteristics, introducing noise and artefacts that affect model performance. Although our image preprocessing steps aimed to mitigate these issues, some variability in imaging quality remains, which could influence the accuracy and consistency of the models. Our study focused exclusively on ultrasound images, while other imaging techniques such as mammography, MRI, and digital pathology images also play crucial roles in diagnosis and treatment planning. The applicability of our models to these other modalities remains unexplored, requiring further research to evaluate their performance across different imaging types.

Although using Grad-CAM, we aimed to increase the interpretability of our models, there remains a gap between model predictions and clinical decision needs. Such tools must be designed to allow accurate predictions with clear and actionable insights for radiologists and clinicians. This can be achieved if efforts are continually made to enhance transparency and usability in AI models, especially when powering diagnostic tools in clinical workflows. There are numerous challenges in deploying deep learning models within real-world clinical settings, mainly due to issues such as computational resources and infrastructure. Our CNN model, despite being custom-made, is efficient but would still require quite a lot of computational power in the training phase and for inference. It is essential to support their seamless use in clinical environments with some considerations about real-time processing and hardware constraints. AI integration in healthcare raises some regulatory and ethical issues.

Besides, the safe and ethical implementation of these technologies must be guaranteed: compliance with medical standards, protection of patient privacy, and coverage of potential biases in AI models. Our study considers these problems, but there is work to do concerning regulation frameworks and guidelines for the ethical use of AI for BC diagnosis. Future work should strive to alleviate these limitations by increasing dataset size and diversity, investigating the applicability of models in other imaging modalities, rendering models more interpretable, and developing means for clinical integration and regulatory compliance. Such work would go further in fostering the use of deep learning in BC diagnosis and in improving patient outcomes.

2. RELATED WORK

The related work analysis in breast ultrasonography covers a broad spectrum of deep learning and transfer learning applications for classification, diagnosis, detection, and segmentation. Researchers have proposed novel deep neural network architectures and inventive approaches to enhance automated breast ultrasound systems' performance, specificity, and accuracy. Their combined efforts improve the possibility of an early diagnosis and better patient outcomes by advancing the detection and evaluation of BC. Table 2-1 summarizes previous research on the use of ultrasonography to diagnose BC.

2.1 Classification

Deep-learning models have achieved incredible accuracy in the classification of breast ultrasound, often outperforming conventional methods (Jabeen et al., 2022). These models prove their capability to distinguish between normal, benign, and malignant cases, thus providing considerable support to the radiologist in improving diagnosis accuracy (Alrubaie et al., 2023). Classification-oriented studies make a considerable contribution to the improvement in BC diagnosis and risk assessment through automation. Classifiers refer to the basic task of machine learning and deep learning, which means predicting the category or label of a given input based on learned patterns from training data (Liu et al., 2022). In this regard, classification involves training the model from a dataset where each input is associated with a tentative label such that the model can accurately assign labels to new, unseen inputs. For example, the classification in medical image analysis might be to identify whether an ultrasound image contains a benign or malignant tumor (Cao et al., 2020). In most applications, classification is preferred since it simplifies decision-making tasks that are complex and reduces them into simple, discrete categories. It is particularly useful when one wants to dichotomize data into distinct classes or groups, usually based on characteristics (Momot A & Zaboluieva M, 2022). For instance, taking the case of BUSI classification, it allows the model to classify between normal and abnormal tissues; hence, this model will enable early detection and diagnosis of diseases such as cancer (Kim et al., 2021). More so, classification models are preferred since they are interpretable and their output is actionable, for example, classifying a tumor into malignant or benign (Gómez-Flores & Coelho de Albuquerque Pereira, 2020). Another reason for the preference of classification is the ease with which models like CNNs handle high-dimensional data. In medical imaging, classification helps in automating and improving the accuracy of tasks that would otherwise be labor-intensive for radiologists, example as differentiating between different types of tissue or identifying irregularities (Z. Zhang et al., 2021). These models can process enormous volumes of visual data rapidly and efficiently, leading to faster and more reliable diagnostic decisions, especially in high-stakes environments like healthcare (Masud et al., 2021).

2.1.1 Deep Learning

Several studies on BC have made significant contributions to diagnosis and classification by implementing deep learning techniques. Jabeen et al. (2022) introduced a probability-based optimal deep learning feature fusion approach, achieving an impressive accuracy of 99.1% in BC classification. Their method utilized the BUSI dataset, including 780 images categorized as normal, malignant, or benign. Zhuang et al. (2021) took a different approach, using image decomposition and fusion, including adaptive multi-model spatial feature fusion. This method yielded a remarkable accuracy of 95.48% and high precision on the BUSI dataset. Momot A & Zaboluieva M (2022) utilized deep neural networks, the EfficientNet B0, pre-trained on the ImageNet dataset, to automate ultrasound BC image classification. They reached an accuracy of 81.26%, highlighting the promise of deep learning in ultrasound diagnostics. Alrubaie et al. (2023) implemented CNN and transfer learning to achieve a high classification accuracy, with a 96% accuracy rate for one dataset and a perfect 100% accuracy for another. This study involved datasets divided into Group A, comprising 780 images with three classes (benign, malignant, normal), and Group B, consisting of 9,016 images with benign and malignant BCs. A noise filter network (NF-Net) was used by Cao et al. (2020) to learn from noisy tagged ultrasound images to classify breast tumors. This approach attained a classification accuracy of 73%. Furthermore, Kim et al. (2021) presented a work on Deep learning with little supervision for detecting ultrasounds of BC. Moon et al. (2020) focus on the diagnosis of BC. The researchers achieved high accuracy and AUC values using CNNs for ensemble learning. They used a private dataset containing 1687 tumors, with 953 benign and 734 malignant cases. Along with an open BUSI dataset with 697 tumors, comprising 437 benign, 210 malignant, and 133 normal cases. The study presents a computer-aided diagnosis (CAD) system that utilizes CNN architectures for tumor diagnosis. performed BC diagnosis using artificial intelligence and deep learning techniques. The suggested grid-based deep feature generator model classified breast ultrasonic images with 97.18% accuracy across cancerous, benign, and normal classes. The performance metrics, including accuracy, recall, precision, F1-score, and geometric mean, exceeded 96 for all classes. The study used a BUSI dataset comprising images from 600 female patients. The method involves grid-based deep feature generation, pre-developed CNN models, incremental feature selection, and a deep classifier to improve BC diagnosis accuracy.

2.1.2 Transfer Learning for Breast Cancer Diagnosis.

Transfer learning is a potent approach in machine learning where a model created for a certain task is adapted to work on another similar but different task. In this way, by reusing an already pre-trained model, which has learned from a huge and heterogeneous dataset, transfer learning allows the application of previously acquired knowledge to a new problem, besides avoiding training on a usually much smaller dataset that frequently lacks comprehensiveness (Cruz-Ramos et al., 2023). This is the preferable approach, as it significantly reduces training time and computational cost, hence yielding high performance with limited data. That also enhances the accuracy of the model by transferring learned features from the original task to improve its generalization capability (Masud et al., 2021). Transfer learning will also ease the development process since a researcher or a developer will build upon a robustly pre-trained model rather than starting from scratch. This further avoids overfitting, as realized in limited datasets, since it capitalizes on the diverse knowledge of the pre-trained model. Generally, transfer learning provides a practical and efficient solution for many problems in machine learning. It thereby turns it not only into an integral tool but one of great value in a number of applications. Wu et al. (2019) the use of machine learning, especially transfer learning in the diagnosis of triple-negative BC, using ultrasonography. Grayscale and color doppler features were used for classification. In the dataset, there were 140 confirmed surgically, BC patients' ultrasound and clinical data were integrated. This study showed that machine learning with quantified ultrasonic image characteristics, including color Doppler information, efficiently separated the triple-negative BC cases. Moustafa et al. (2020) used color doppler ultrasound introduced to enhance BC detection using machine learning. The dataset contains 159 solid masses, with 95 benign and 64 malignant cases. The study illustrates that incorporating color doppler and grayscale features in the training dataset improves the receiver operating characteristic area, thus enhancing the diagnosis of BC. Qi et al.

(2019) a competitive and automated diagnosis approach was proposed for processing breast ultrasound pictures using deep neural networks. Diagnosis of the BreAst Ultrasound image is performed based on a cascade of two neural networks: Mt-Net for malignant tumor classification and Sn-Net for solid nodule classification. News Dataset: The large-scale ultrasound image collection, annotated by experts, is used in their work and has three pieces: a training set, a validation set, and a test set. The proposed approach has achieved performance equal to that of human sonographers, with high accuracy and specificity obtained on BC diagnosis. Various studies have used transfer learning to enhance the diagnosis and classification in BC from ultrasound imaging. Zhang et al. (2021) the work presented here exploits transfer learning to tune a model to predict BC subtypes better. It was concluded that transfer learning remarkably improved the accuracy of the classification. Similarly, Pang et al. (2021) recognition of result railway and bus in augmented breast ultrasound mass classification using a semi-supervised GAN-based radiomics model was done with an accuracy of 90.41% and sensitivity at a high value, etc. Coronado-Gutiérrez et al. (2019) studied quantification ultrasound evaluation of images to diagnose Metastatic BC invasion. Their proposed method achieved an accuracy of 86.4% and a sensitivity of 84.9% using a dataset of 118 lymph node ultrasound images from 105 patients selected from two hospitals. Cruz-Ramos et al. (2023) applied transfer learning on a pre-trained architecture to classify benign and malignant BCs in ultrasound and mammography pictures, resulting in 97.6% accuracy. These studies' results show that transfer learning methods improve the efficiency of BC diagnosing and classification techniques using ultrasound imaging.

2.2 Detection using Deep Learning

Several research has investigated the utilization of Deep learning algorithms for detecting BC in radiography and ultrasound imaging. Mahoro & Akhloufi (2022) evaluated therapy response using reference images from the Evaluating Therapy Response Breast MRI dataset (1,500 DICOM format images) and the BUSI dataset (250 BMP type BC images). Masud et al. (2021) uses CNNs that have been previously trained to identify BC using ultrasound imaging. Their study demonstrated impressive results, with models like DenseNet201, Xception, and ResNet18 achieving 100% accuracy using various optimizers. This study used publically accessible breast ultrasound scans from Rodrigues; there are 250 ultrasound scans, including 100 normal and 150 cancerous instances. The study used a transfer learning technique using pre-trained models and a K-fold cross-validation procedure for assessment. Li et al. (2022) presented the "BUSnet" deep learning model for breast tumor lesion detection in ultrasound images. Their dataset included 780 samples, with 133 normal, 487 benign, and 210 malignant cases. The testing dataset was collected from a prior source. The method included unsupervised region proposal, bounding-box regression methods, and a post-processing strategy to improve detection accuracy. Zhang et al. (2021) focused on employing a Deep learning algorithm used in automated breast ultrasonography to detect cancers. They achieved a sensitivity of 0.88 with 0.19 FP/S. The study used an ABUS imaging dataset uniquely provided by Peking University People's Hospital, which contains 170 ABUS tumor volumes from 124 female patients. Their method first comprises the Bayesian YOLOv4 network combined with Monte Carlo dropout and some new YOLOv4 modifications to better detect ABUS tumors.

2.3 Segmentation

Segmentation is a fundamental approach within computer vision that refers to dividing an image into discrete segments or regions with the intent of simplifying its analysis and enhancing its interpretability. This process will enable the exact location and identification of certain objects or structures within an image, including tumors within medical imaging and various components in autonomous driving contexts (Zhang et al., 2021). Segmentation, apart from making an image more meaningful by segmenting it into meaningful parts, increases the precision and reliability of further analyses while enabling more detailed and focused studies of regions of interest (Gómez-Flores & Coelho de Albuquerque Pereira, 2020). Such enhanced precision in the analysis serves to facilitate tasks such as object recognition and classification, along with the provision of valuable contextual understanding in complex scenes (Ilesanmi et al., 2021). Moreover, the segmentation methodology can be tuned for a wide range of image types and applications-from classical thresholding and clustering methods to state-of-the-art deep-learning methods like the U-Net (Vakanski et al., 2020). Therefore, segmentation is preferred because it turns complex images into manageable components; their analysis and processing are thus easier and more effective for further insights in many fields (Xu et al., 2019).

2.3.1 Deep Learning-Based Segmentation Studies

Gómez-Flores & Pereira (2020) elaborate on the performance evaluation of pre-trained CNNs in segmentation tasks on BCs in ultrasound pictures. Finally, they concluded that SegNet and DeepLabV3+ had performed the best; ResNet18 showed great potential to be applied in the CAD system. The study included more than 3000 BUSI from seven different ultrasound machine models. Ilesanmi et al. (2021) suggested a segmentation strategy producing high dice measures for malignant and normal breast ultrasound images. The study used two datasets: one with 264 photos (100 malignant, 164 benign) and another with 830 images (487 malignant, 210 benign, and 133 normal). Their approach includes a preprocessing stage and a deep learning segmentation stage. Vakanski et al. (2020) incorporating attention blocks into deep-learning models for breast tumor segmentation resulted in models that outperformed the basic U-Net model. The dataset included 510 ultrasound images converted to grayscale 8-bit data and resampled into floating points using normalization. Xu et al. (2019) primarily focused on applying machine learning to segment breast ultrasound pictures, with quantitative measurements achieving more than 80% accuracy. The dataset utilized is not specified. However, their method categorizes BUSI into four tissue types: skin, fibroglandular, mass, and fatty tissue. Zhang et al. (2019) introduced a novel method, Boundary-aware Semi-Supervised Deep Learning (BASDL), for breast ultrasound lesion CAD. BASDL achieved a classification accuracy of around 92.00±2.38% and was evaluated on two breast ultrasound datasets. Chiang et al. (2019) specifically targeted tumor detection in automated ultrasound image with a dataset that includes 230 pathology-proven lesions from 187 individuals, 90 of which are benign and 140 of which are malignant. The proposed method involved preprocessing, segmentation, and feature extraction using a sliding window detector for localized analysis. Lei et al. (2018) obtained the highest segmentation performance by developing a technique for segmenting breast anatomy in full BUSI. The ConvEDNet is a boundary-regularized deep convolutional encoder-decoder network. The limited dataset used manual annotation due to the associated cost. Gong et al. (2020) explore a bi-modal approach to BC diagnosis using a multi-view deep neural network support vector machine. The classification accuracy attained was 86.36%, with an AUC of 0.9079. The dataset employed in this study contained 264 pairs of breast ultrasound and ultrasound elastography pictures from 129 patients with benign tumors and 135 patients with malignant malignancy. The recommended model is separated into two parts: Multi-Depth Neural Network and Fusion Deep Neural Network. The multi-view approach enhances diagnosis by incorporating information from both sorts of ultrasound pictures.

2.3.2 Transfer Learning-Based Segmentation Studies

Negi et al. (2020) proposed the WGAN-RDA-UNET technique, which uses Wasserstein GANs. It obtained an overall accuracy of 0.98 and a PR-AUC of 0.95. The research used the PASCALVOC2012 dataset for training and the Berkeley Segmentation Database (BSDS 300 and BSDS500) for assessment.

Table 2-1 Overview of deep learning-based ultrasound screening and diagnosis. BUSI: breast ultrasound image, DL: deep learning, TL: transfer learning, C=Classification, SE=Segmentation, DE=Detection.

Author	Task	Methods Used	Dataset	Accuracy
Jabeen et al. (2022)	DL, C	Modified DarkNet-53 deep model	BUSI 780 images	99.1%
Liu et al. (2022)	DL, C	grid-based deep feature generation-based	BUS 600 images	97.18%
Zhuang et al. (2021)	DL, C	Fuzzy enhancement	BUS 1328 images	95.48%
Momot A & Zaboluieva M (2022)	DL, C	Efficient Net B1	ImageNet dataset	81.26%
Alrubaie et al. (2023)	DL, C	CNN Inception-v3	A: BUSI 780 images B: BUS 9016	96% 100%
Cao et al. (2020)	DL, C	Noise filter network (NF-Net)	BUS	73%
Kim et al. (2021)	DL, C	Weakly-supervised deep learning algorithms	BUS 1000 images	-
Moon et al. (2020)	DL, C	Dense Net VGG-16 and VGG-Like ResNet	Private 1687 images BUSI: 697 images	91.10% 94.62%
Moustafa et al. (2020)	DL, C	Color Doppler AdaBoost ensemble classifier	159 solid masses	-
Qi et al. (2019)	DL, C	Automated BC diagnosis model	BUS 8145	87.79%
Y. Zhang et al. (2021)	TL, C	(CLSTM)	N/A	91%
Cruz-Ramos et al. (2023)	TL, C	fusion and handcrafted features	BUSI 780 images	97.6%
Pang et al. (2021)	TL, C	Semi-supervised GAN model –	BUS 1447	90.41%

		TGAN model		
Coronado-Gutiérrez et al. (2019)	TL, C	QUS image analysis techniques	BUS 217 images	86.4%
Mahoro & Akhloufi (2022)	DL, DE	different screening methods for breast cancer	(RIDER) 1500 BUS 250 images	-
Masud et al. (2021)	DL, DE	DenseNet201 - ResNet50	BUS 250 images	100%
Li et al. (2022)	DL, DE	BUSnet	BUSI 780 images	100 %
Z. Zhang et al. (2021)	DL, DE	Bayesian YOLOv4	BUSI 21,624 images	-
Negi et al. (2020)	TL, DE	(WGAN) (RDAU-NET)	PASCALVOC2012	0.98
Gómez-Flores & Coelho de Albuquerque Pereira (2020)	DL, SE	Multi models	BUS 3000 images	-
Ilesanmi et al. (2021)	DL, SE	End-to-end deep learning segmentation stage	Dataset 1: 264 images Dataset 2: 830 images	89.73
Vakanski et al. (2020)	DL, SE	U-Net - U-Net-SA	BUS 510 images	-
Xu et al. (2019)	DL, SE	CNNs	Private dataset	80%
E. Zhang et al. (2019)	DL, SE	BASDL	UDIAT, UTWS	92.00% 83.9%
Chiang et al. (2019)	DL, SE	3-D CNN - 2-D CNN	Automated whole BUSIs	-
Lei et al. (2018)	DL, SE	ConvEDNet - Adaptive Domain Transfer (ADT)	Automated Whole Breast Ultrasound images	86.8%
Gong et al. (2020)	DL, SE	Multi-view deep neural network	BUS 264 images	86.36%

3. MATERIAL AND METHOD

3.1 Deep Learning and Convolutional Neural Networks (CNNs)

Deep learning is, in general, one of the advanced sub-areas of machine learning, which engages with mostly multilayered artificial neural networks to model and learn complicated patterns in data. This approach has transformed many fields- especially image recognition, speech processing, and natural language understanding to enable machines to extract meaningful features from raw data naturally (Jabeen et al., 2022). Among deep learning models, CNNs are one of the most basic forms. They are originally designed to work on grid-structured data, such as images and videos (Alrubaie et al., 2023). The unique arrangement of CNNs instantly captures the spatial relationships of the data while information flows through the network by being able to spot patterns like edges, textures, and more complicated structures (Kim et al., 2021). The architecture of CNNs paces the combination of various types of layers that act in learning features of incoming data: Convolutional layers form the core-a small set of filters that acts to locate the presence of any pattern, whether spatial relationships enable the extraction of features across a picture (Masud et al., 2021). These comprise a few more with the incorporation of activation functions like the Rectified Linear Unit, which introduces non-linearity into the model so it can learn higher-order representations. Further pooling layers are added to reduce the computational load and loss of overfitting (Alrubaie et al., 2023). These comprise a down-sampling operation that reduces the size of the dimensions in each feature map while maintaining very important information, such as max-pooling. The architecture of this network, after multiple convolutional and pooling layers, makes use of fully connected layers where each neuron in the fully connected layer, conventionally, is connected to every neuron in the preceding layer. These layers take up information from the extracted features of the early layers to provide a model that has more insight in making decisions about the data it gets presented with. To improve generalization, dropout layers can be included that randomly stop a percentage of neurons at any given time while training to avoid overfitting. Finally, the network is usually terminated with an output layer that typically includes a softmax activation function to build a probability distribution across the possible output classes, yielding the final classification or prediction (Kim et al., 2021). The structured combination of layers especially empowers CNNs to easily learn from raw images both low-level and high-level features adaptively, which is so crucial in accomplishing complex image analysis tasks (Alrubaie et al., 2023). This thesis adopts CNNs mainly for the classification and segmentation of pictures in breast ultrasound. It enlightens how such models may be used to catch intricate patterns in medical images that could help in locating and diagnosing abnormalities (Jabeen et al., 2022).

3.1.1 Evaluation Metrics

We have used several key metrics to evaluate model performance, including Accuracy, Precision, Recall, and F1-Score. These metrics are critical for assessing classification models, particularly in the context of medical imaging where false negatives and false positives can have significant consequences. Performance metrics are the quantification of machine learning models and algorithms in terms of their usefulness and accuracy (Momot A & Zaboluieva M, 2022). These metrics make known how a model performs well on a particular task at hand, be it classification, regression, or clustering. They form the basis for judging the qualities of the model predictions and for knowing its strengths and limits for making informed choices in model selection and optimization. Several key performance indicators were considered for this work to evaluate the different models applied and some of them gave insights into important performance features concerning these models. These include but are not limited to accuracy, precision, recall, and the F1 number. Each of these measures is considered a critical attribute in assessing the quality of the predictions and interpreting model behavior on these tasks, mostly classification and segmentation (Cao et al., 2020).

One of the simplest and most frequently used performance metrics is accuracy. It is a measure of the overall correctness of the model, which forecasts the ratio of correctly classified cases, positives and negatives, in between the total number of examples (Ilesanmi et al., 2021): The formula for accuracy looks as follows:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Number of Instances}} \qquad \text{3-1)}$$

Here, True Positives means examples that were correctly predicted as positive. True Negatives are examples that were correctly predicted as negative. The overall Number of Examples encompasses the contents in TP, which are FP and FN. Accuracy gives a general view of the model's performance but may not really pin its effectiveness in situations where class imbalance is present, one class is considerably more common than another (Ilesanmi et al., 2021). Precision or Positive Predictive Value is the fraction of the model's identifications that were correct (Alrubaie et al., 2023). More formally, precision measures what percent of identifications truly were correct. Precision is calculated using the formula:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \qquad 3\text{-}2)$$

The True Positives are the correct identifications as positive in this light, and the False Positives are the incorrect ones reported as positive. High precision suggests that when the model is predicting a good result, it tends to be right. Precision is especially crucial when the cost associated with a false positive is high. For example, in medical tests, a false positive may result in unnecessary stress or further testing. Recall-sometimes referred to as Sensitivity or True Positive Rate-is the measure that tells how well the model can detect all the important positives existing in the data (Alrubaie et al., 2023). It is defined as:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \qquad 3\text{-}3)$$

True Positives are examples correctly identified as positive, and False Negatives (FN) are examples that were not identified as being positive but indeed should have been (Kim et al., 2021). Recall is important in applications where it is important to have all possible positive cases present, even at the cost of some false alarms. For example, in cancer detection, with high recall, most cases of the disease are immediately highlighted, hence minimizing the chances of overlooking a potential diagnosis (Cao et al., 2020). The F1-score can be used to calculate the harmonic mean to get a proper insight into the balance between accuracy and memory (Masud et al., 2021). It is quite useful when dealing with imbalanced information-the reason being that instances of one class can greatly dominate another class. It is calculated based on:

$$\text{F1} - \text{Score } 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad 3\text{-}4)$$

The F1-score combines both precision and recall in a single metric in such a way that the trade-offs between the two are taken into consideration. High scores in F1 mean that the model has a decent balance between precision and recall. Therefore, F1 is useful when models have to be compared in instances when false positives and false negatives come at costly repercussions. These performance measures were used comprehensively to evaluate the success of each of the models in this study (Masud et al., 2021). They do indeed provide rich insights into the strengths and weaknesses of the models to make better decisions regarding model choice and optimization for BUSI classification and segmentation.

3.1.2 Experimental Setup

The server used during this work was a Lenovo laptop with an 11th Generation Intel® Core™ i7-1165G7 processor, associated with 8 GB DDR4-3200 system memory and 1 TB HDD at 5400rpm. The system was equipped with Wi-Fi 6, up to 2x2 AX, as well as Bluetooth® 5.1 or higher for wireless communication. It features a total of two USB 3.2 Gen 2 Type-C ports with functionality supporting USB, DisplayPort, and Power Delivery through dual-direction functionality; two USB 3.2 Gen 2 Type-A ports, of which one supports always-on charging, alongside an RJ-45 Ethernet port, an HDMI port, a headset and microphone combo jack, and finally, a 4-in-1 media card reader supporting SD, MMC, SDHC, and SDXC formats. The laptop also featured a 720p HD camera with a privacy cover and dual mics. The bottom had fitted an NVIDIA GeForce MX450 GPU with 2GB of dedicated memory. The sCREEN went up to 15.6 inches horizontally, availed FHD resolution, and captured 45% of color coverage. The type of service made available-maintenance-was Courier or Carry-in.

3.2 Dataset

BUSI is used as part of the validation procedure in this study. The BUSI Dataset includes 780 breast ultrasound pictures categorized as benign, malignant, or normal. Images were taken from 600 female patients ranging in age from 25 to 75. The images are 500 by 500 pixels on average and saved in PNG format. Images were classified into three categories: normal (133 images), benign (487 images), and malignant (210 images). The normal images reveal healthy breasts, whereas the benign images show benign masses and the malignant images indicate cancerous masses (Al-Dhabyani et al., 2020). The original images were preprocessed to reduce noise and improve image quality. This involved cropping the images to remove irrelevant information, scaling the images to a uniform size, and applying contrast enhancement. The data was collected in 2018 using LOGIQ E9 ultrasound systems. The images were captured using standard ultrasound protocols for breast imaging. The images were then manually labeled by expert radiologists (Oluwasanmi et al., 2021). In addition, ground truth the appropriate B-mode images are supplied with binary mask images. Figure 3-1 shows the ground truth images.

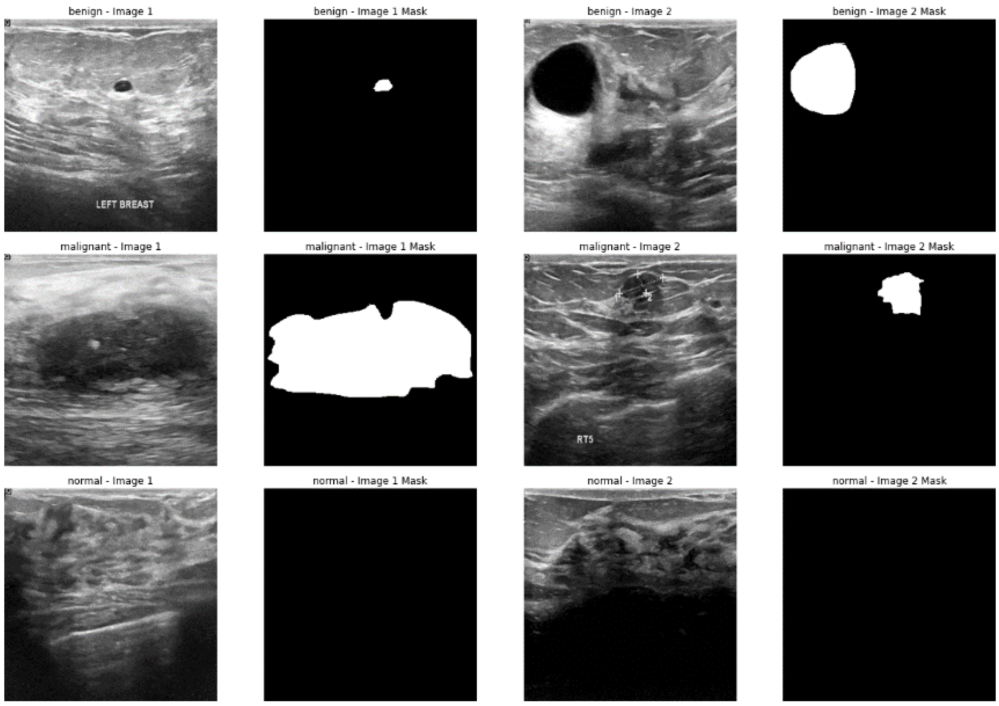


Figure 3-1 Samples of ultrasound images with his ground truth.

3.3 Methodology

3.3.1 Image Preprocessing

Preprocessing images is an essential step in BUSI analysis for improving quality and interpretability. The pipeline in preprocessing is made up of three main stages: segmentation, enhancement, and overlaying. Every stage is a very vital part of the ensuing analysis and diagnosis processes that assure accuracy and effectiveness.

3.3.2 Image Segmentation

Image segmentation is partitioning an image into multiple segments to simplify its representation and make it more meaningful for analysis. In breast ultrasound imaging, segmentation is critical for identifying and isolating regions of interest from background tissues. We used a U-Net architecture for the segmentation process, similar to the studies reviewed in Section 2.3 (Y. Zhang et al., 2021; Gómez-Flores & Coelho de Albuquerque Pereira, 2020). As discussed in the literature, U-Net has proven effective for medical image segmentation, particularly in breast cancer detection. The U-Net architecture is one of the best structures suited for medical image segmentation because it captures both spatial and contextual information. The U-Net model can be segmented into two main parts: the contracting path and the expanding path. This part, the encoder, will derive contextual information regarding the image. It consists of a series of convolutional layers followed by several max pooling operations, progressively reducing the input image. The feature is further refined, coming to the deepest level of the network in the bottleneck layer. The decoder or expanding path gradually restores the image's spatial resolution by upsampling. It has transposed convolutions for resolution up-sizing of the feature maps. Additionally, the skip connections concatenate the high-level features from the contracting path and the layers in the corresponding expanding path. These skip connections help recover spatial information lost during downsampling, thereby ensuring a detailed and accurate segmentation map.

The segmentation process starts with careful preparation of the dataset. This involves resizing the images and their corresponding masks to a uniform size of 224x224 pixels, such that the model can be trained on a dataset with normalization in the image. Normalizing the photos is another most critical step because it standardizes the pixel values, hence giving the dataset so we can train the model. Next, after building and compiling the model with the Adam optimizer and binary cross-entropy loss using the U-Net model, the dataset is subdivided into 80:20 training and validation datasets. It generalizes to any data unknown during training: the model can learn enough from patterns in the training data yet is not so dependent on a specific dataset that the task becomes overfitting. The model, through multiple iterations, called epochs, learns to segment these images accurately. Figure 3-2 shows an overlayed segmentation image on sample data. The validation set is used to calculate model performance, and the model re-adjusts to update its final Accuracy. Consequently, the trained model will be saved and applied to new images; thus, segmenting new ultrasound images will be done efficiently and accurately. This automated process of segmentation advances the possibility of early and accurate diagnosis of BC, which consequently leads to treatment effectiveness and improved outcomes for patients.

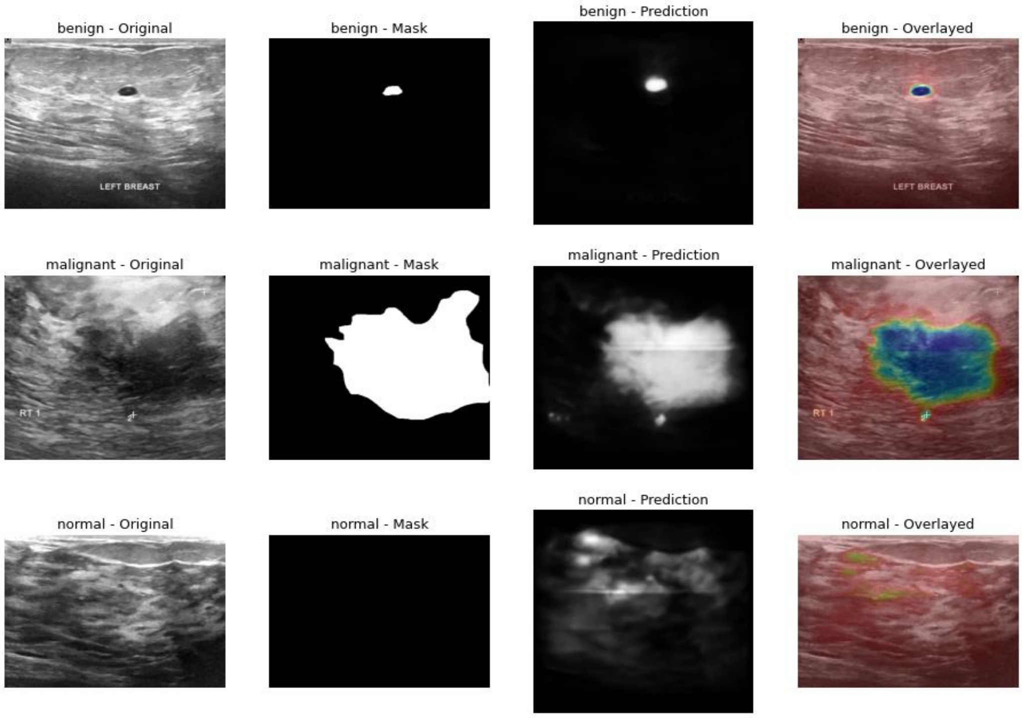


Figure 3-2 Visual Results of Breast Ultrasound Image Classification and Segmentation

3.3.3 Image Enhancement

Enhancement techniques are essential in medical imaging because they make vital structures and features clear and, thus, allow the images to be informative and helpful in making a diagnosis. The BUSI dataset enhanced by these techniques targets different kinds of information reflected in the quality of images (Zhuang et al., 2021). The enhancement of contrast by the fuzzy process helps to adjust the image to emphasize differences in density and compositions of body tissues. In that light, the contrast enhancement of medical images becomes indispensable for the finer details that distinguish healthy tissue from pathological tissue. By using fuzzy enhancement, the images are more detailed and precise to diagnose the conditions—for instance, tumors or lesions (Badawy et al., 2021). Figure 3-3 presents a fuzzy enhancement of ultrasound images.

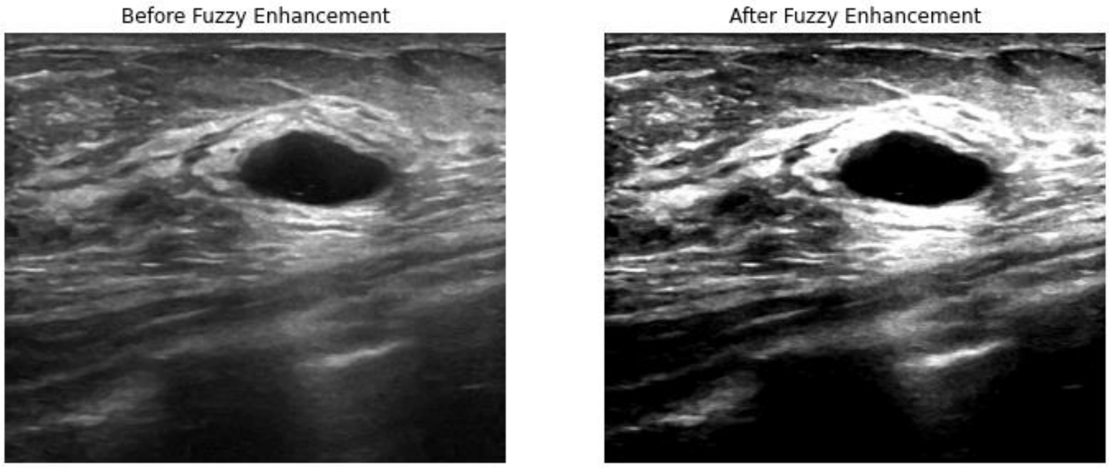


Figure 3-3 Fuzzy enhancement on the image

Sharpness Enhancement, Enhancement of an image is aimed at enhancing the clarity of the edges within an image. Edge definition in lesions or tumors is required for effective and accurate analysis of images in ultrasound imaging. Sharpness enhancement techniques apply filters that increase the contrast at the boundary between different tissues to have their edges stand out and be easily separated (Luo et al., 2024). Figure 3-4 shows the sharpness enhancement on the image.

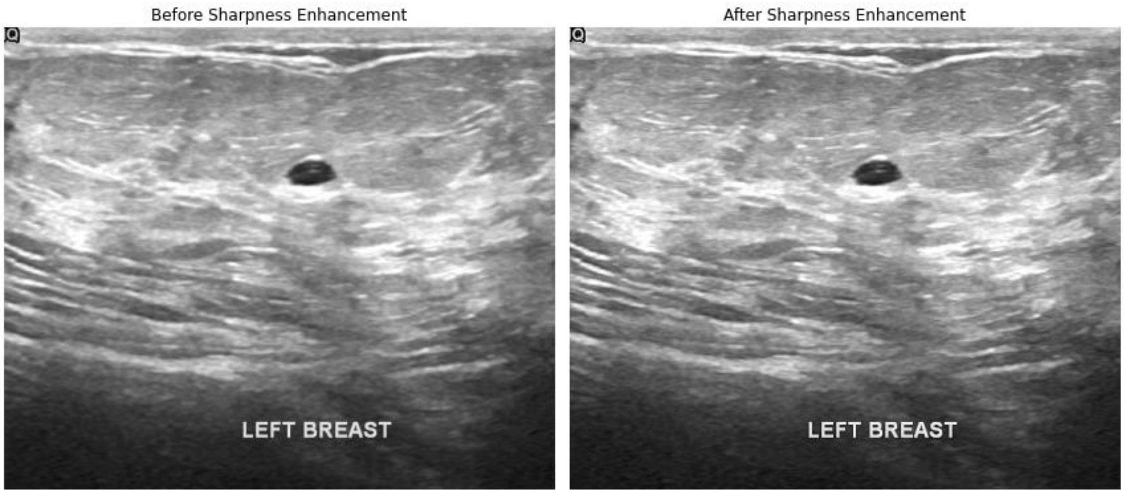


Figure 3-4 Sharpness enhancement on the image.

Brightness Adjustment: Balancing the brightness of the image so that important regions in the image do not go unnoticed due to the lighting or intensity difference with the background. This is the process of changing pixel intensity values to have a uniform light distribution around all the pixels. This is particularly advantageous in enhancing overexposed or underexposed areas of the original image so that all the exciting places are visible (Moustafa et al., 2020). Figure 3-5 illustrates the enhancement of brightness to the image.

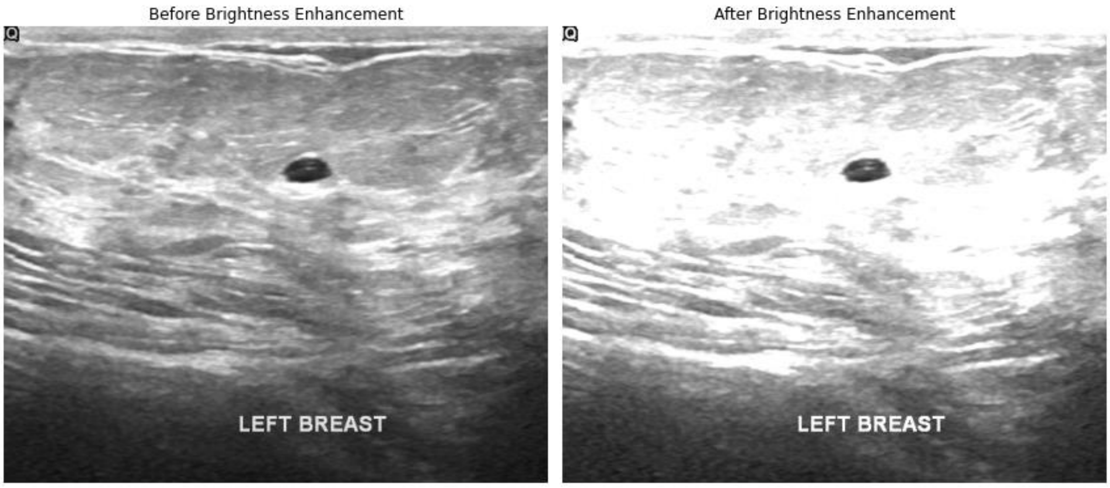


Figure 3-5 Brightness enhancement on the image

In rescaling for visualization and fair comparison, one needs to make sure the images are of equal size. Rescaling refers to adjusting image sizes to a fixed dimension. This is so important in the treatment of comprehensive datasets or training model pipelines. Resizing the image will enable it to be fed into algorithms with input of a specific size, keeping uniformity and reliability in the analysis. In Figure 3-6 , one can see the resized image. These techniques work together to enhance the quality and informativeness of ultrasound images through the maximization of contrast, sharpness, brightness, and size to a point where these images are easy to view, interpret, and analyze by radiologists and automated systems (PACAL, 2022). Enhanced images are more accessible to explore; they will be much better in accuracy and reliability for making diagnoses that are aimed to benefit the patient in the long run.

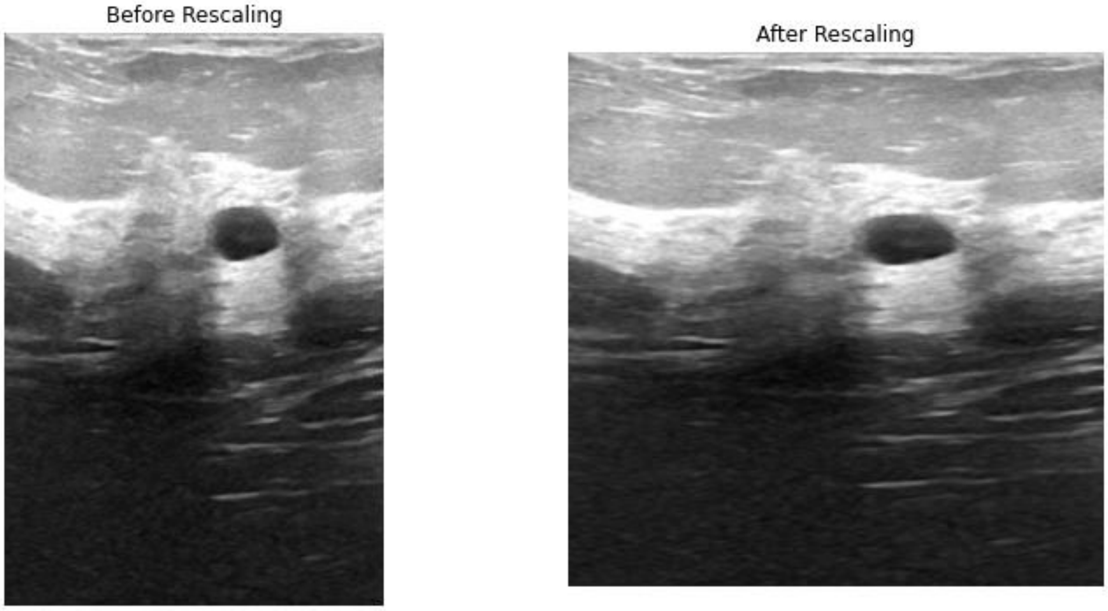


Figure 3-6 Rescale image size.

3.3.4 Overlaying

Overlaying is a process where the original ultrasound image is taken, and a mask with features like the tumor or lesion is added. The overlaying technique combined with the methods mentioned above makes the enhanced image a more effective and informative visualization. Implementing the overlaying process commences by applying the segmentation masks from the U-Net model on the original ultrasound images. These masks define areas of interest that separate tumors or lesions from the surrounding tissue. Superimposing these masks on the original images generates a composite image with transparent regions of interest, which, on further enhancement using the above techniques, can be enhanced. The salient regions of the overlayed images are enhanced for higher quality, using features like fuzzy, sharpness, brightness, and rescaling. This way, a more complete view is created where the critical regions are demarcated and emphasized. The images of Figure 3-7 are below, before and after segmentation overlaying. This can aid radiologists in diagnosing with increased accuracy and confidence as they view these regions of interest in more detail and contrast compared to the original ultrasound images.

The preprocessing pipeline thus allows for a robust workflow in an effective way to create segmentation, enhancement, and overlaying of the image in the analysis of breast ultrasound. All these elements work together, enhancing the accuracy of automated diagnostic systems and giving radiologists the best means to detect early stages of BC and plan treatment. An increase in enhancement and segmentation of images helps identify pathological regions accurately, which is helpful for making a well-informed decision toward better patient care.

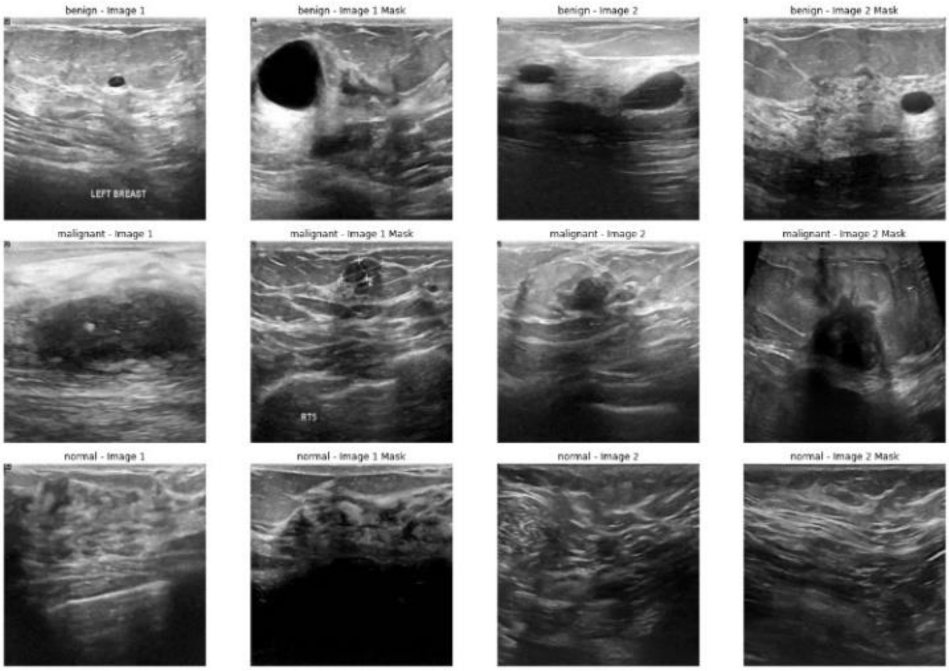


Figure 3-7 Before image overlaying.

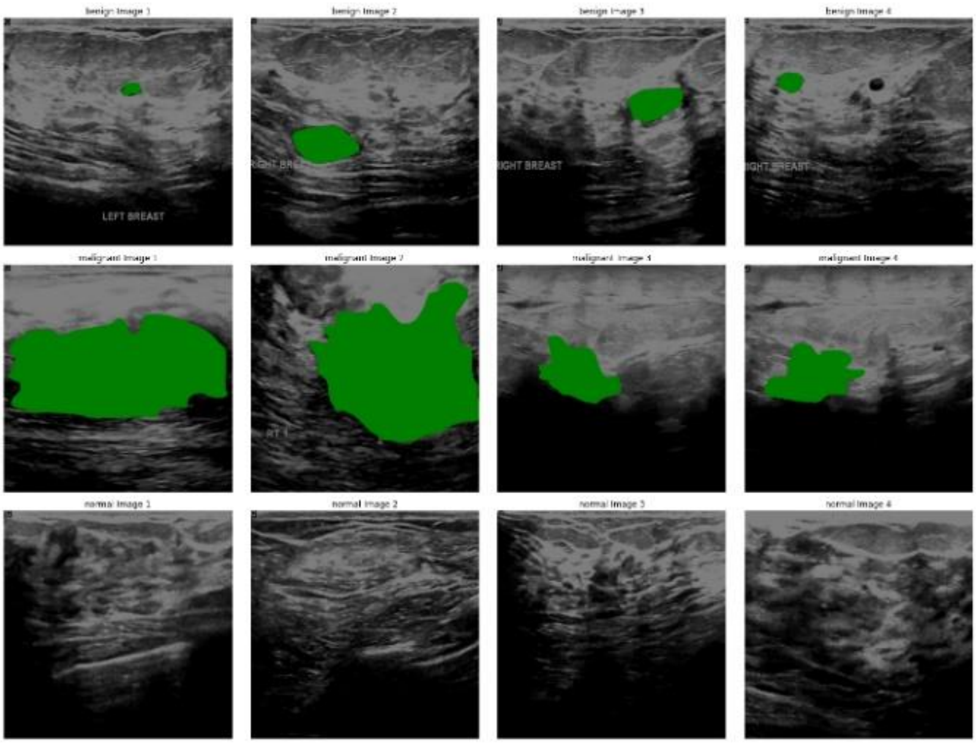


Figure 3-8 after overlaying segmentation image.

3.3.5 Proposed Model

This section involved extensive experiments using different convolutional neural network architectures that supported the proposed model. We performed a systematic variation of parameters related to the number of layers, sizes of filters, and activation functions. We first evaluated models of three, five, and seven layers. The general trend was that deeper networks performed better, but beyond a certain depth, performance gains diminished. We also tried different filter sizes and observed that smaller filters, for example 3x3, captured finer details well, thus performing better than larger filters, such as 5x5 and 7x7. With a best performing proposed model accuracy of 97%, those of the benchmark models obtained very low accuracies. It has been reiterated as an iterative process, showing that hyperparameter tuning and model selection form important steps in the development of an effective ultrasound image-based breast cancer detection system.

We developed a custom CNN for breast image classification. As shown in Figure 3-9, which have shown significant success in breast ultrasound classification. Our model differentiates itself by incorporating additional preprocessing and fine-tuning, as outlined in Section 3.3.1. our model captures key properties from images using a sequence of convolutional and pooling layers that preserve the spatial hierarchy. The model starts with a convolutional layer that has 32 filters, each measuring 3x3 pixels. This layer uses a stride of (1, 1) and the same padding to extract key characteristics from the input pictures. It is followed by a max-pooling layer with a kernel size of (2, 2) and a stride of (2, 2), which reduces spatial dimensions while enhancing translation invariance and model efficiency. Our network has additional convolutional layers with larger filter sizes (64, 128, 256, 256, and 512), followed by max-pooling layers. Except for the last convolutional layer, the output is decreased to 1*1 size to further reduce the sample size of the feature maps while improving abstraction. We use ReLU (Rectified Linear Unit) activation functions in these layers to help the network understand complex patterns in the input. Following the convolutional layers, we used the GlobalAveragePooling to fully connected (dense) layer. The model has three dense layers: the first has 512 neurons and employs ReLU activation in conjunction with L2 regularization (with a value of 0.01) to prevent overfitting. We also use a 0.5-rate dropout layer to deactivate neurons during training, randomly increasing generalization. The second layer has 1024 neurons and employs ReLU activation in conjunction with L2 regularization (with a value of 0.01). The last dense layer contains three neurons and employs a softmax activation function to generate probabilistic outputs for the three classes (benign, malignant, and normal). Our results show that the proposed CNN model is highly effective. When we trained it on enriched overlayed datasets, the model had an F1-score of 0.97, an accuracy of 0.97, a precision of 0.98, and a recall of 0.96. These metrics were much higher than those obtained while training with all layers without the enhanced overlayed datasets, which yielded an accuracy of 0.83, precision of 0.80, recall of 0.83, and F1-score of 0.81. These findings demonstrate the robustness of our algorithm in detecting subtle patterns in medical images, resulting in extremely accurate categorization. We experimented with various models, architectures, and hyperparameters to achieve an accuracy higher than 97%, only this

specific model managed to exceed that benchmark on our dataset.

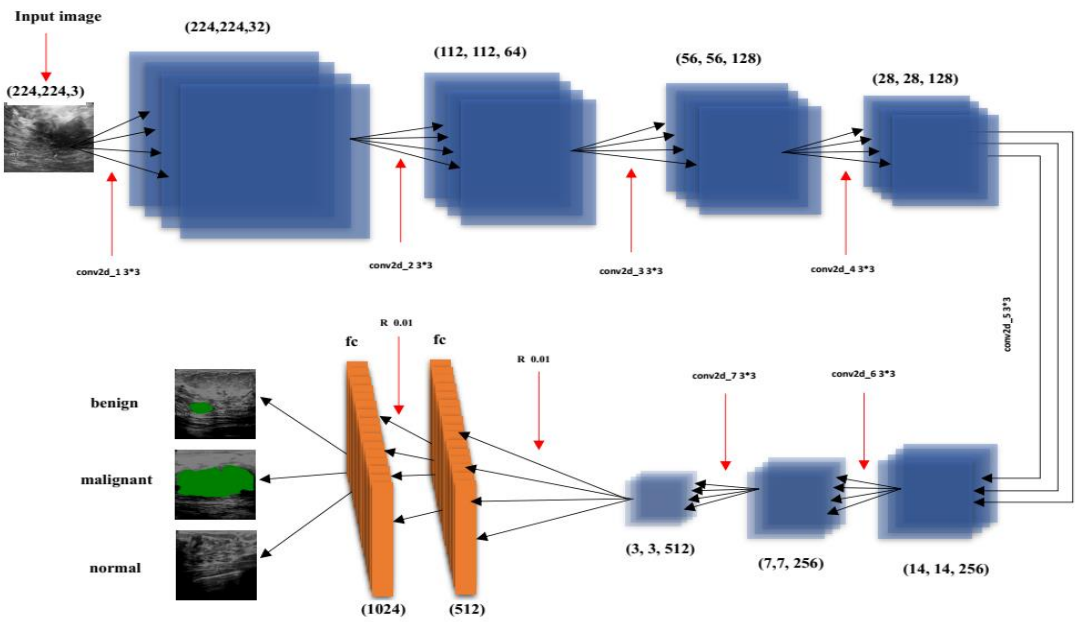


Figure 3-9 proposed model.

3.3.6 State-of-the-art Models

The state-of-the-art models like VGG-Net and DenseNet were fine-tuned using transfer learning techniques, as highlighted in Section 2.1.2 (Cruz-Ramos et al., 2023; G.-G. Wu et al., 2019). Transfer learning enables models to leverage pre-existing knowledge, improving classification performance, especially when dealing with limited datasets.In this section, different state-of-the-art models implemented for the classification of images are presented. Detailed descriptions of each model's architecture, salient features, and benefits in our study are discussed. The VGG16 and VGG19 models, developed by the Visual Geometry Group at the University of Oxford, are famous for being profound and superficial. VGG16 has 16 layers, while VGG19 has 19, both with convolutional and max-pooling layers. These models have a small convolution filter size of 3x3; that was a new concept at the time of the model proposal. The primary contributions made by these models are simplicity and depths that make detailed feature detection possible in images (Ferguson et al., 2017; Çetin-Kaya & Kaya, 2024). The architecture for VGG16 and VGG19 starts with a few convolutional layers and after that proceeds with one activation function, rectified linear unit (ReLU), following each convolutional layer. These are stacked so that complex features of the input image are gradually extracted. At the end of every two convolution layers, a max-pooling layer is applied to reduce the feature maps' spatial size and, therefore overall computation complexity by emphasizing the most critical features. The final layers are fully connected layers that produce class probabilities, given by a softmax layer. The architecture of VGG-Net is displayed in Figure 3-10 . VGG models possess much depth and a uniform structure; therefore, they are much easier to implement and fine-tune. Profound in architecture but quite simple in models, hence very reliable for different tasks in image classification (Çetin-Kaya & Kaya, 2024).

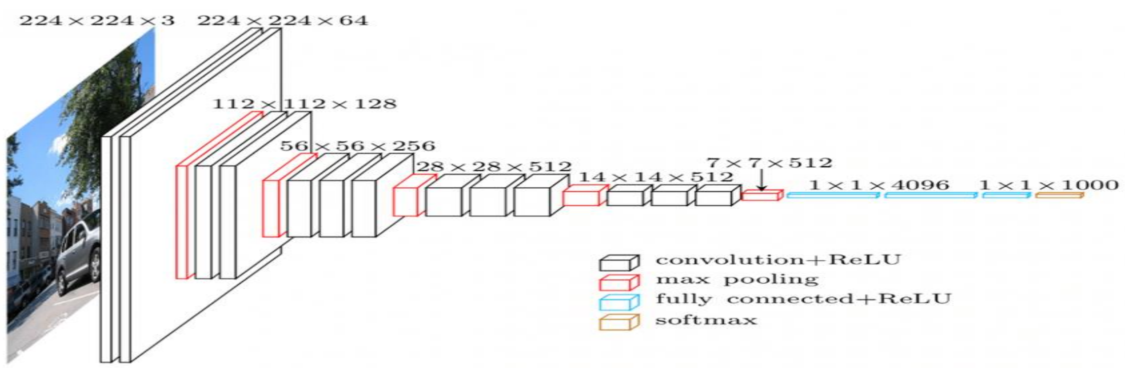


Figure 3-10 VGG-Net architecture (Ferguson et al., 2017).

The DenseNet121 architecture is made up of dense blocks that are separated by transition layers responsible for down-sampling. Each dense block contains multiple layers, with each layer receiving inputs from all previous layers and sending its outputs to all subsequent layers. This pattern of direct connections ensures efficient gradient flow and feature reuse, which helps prevent the vanishing gradient problem and results in a more compact and efficient model (He et al., 2015). Figure 3-11 presents an overview of the DenseNet121 architecture. Arguably because of its dense connectivity, DenseNet121 is very parameter-efficient and less prone to overfitting. It's this efficiency in parameters that allows DenseNet 121 to be highly accurate for the number of parameters with respect to conventional architectures, hence making it quite a good option for this image classification task, especially considering the somewhat meager computational resources.

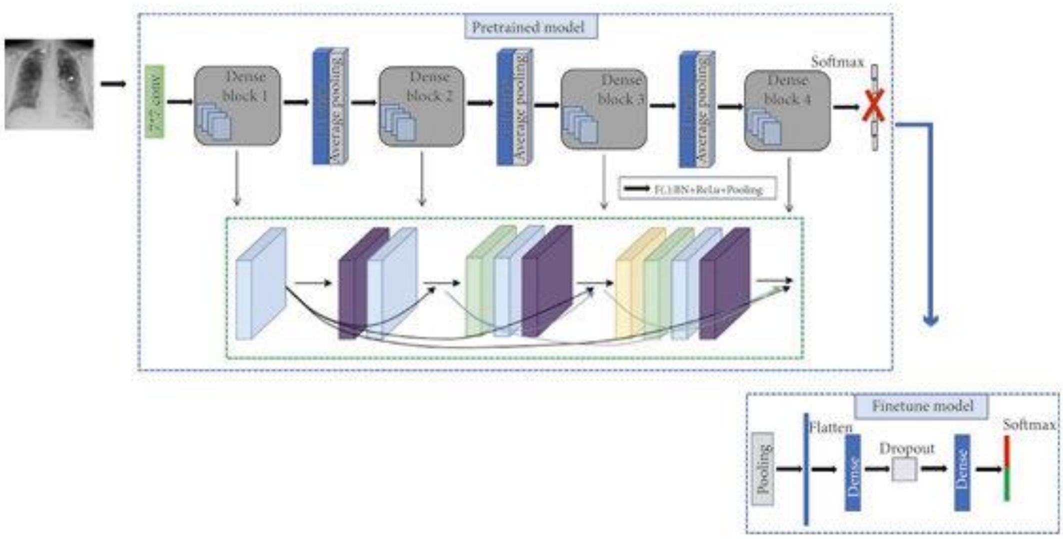


Figure 3-11 DensNet121 Architecture(He et al., 2015) .

EfficientNetB0 uses a compound scaling method by which it scales up the network uniformly in depth, width, and resolution. This approach leads to outstanding performance with the same order of computational efficiency. From the base model, EfficientNetB0 scaled up the dimensions of a network through compound scaling that means increasing the depth and width along with the resolution in due proportion to achieve the best performance (Magán et al., 2022).

EfficiencyNetB0 contains a type of mobile inverted bottleneck convolution called MBConv blocks, which increases performance by first doing an expansion of channel number, then applying depthwise convolutions, and reducing the channels to a smaller number again (Saini & Devi, 2023). Figure 3-12 introduces the EfficientNetB0 architecture. EfficientNetB0 uses compound scaling to develop a model that provides a perfect balance between accuracy and efficiency and, hence, is effective in cases of low computational resources and memory (Mahoro & Akhloufi, 2022). Since the model requires lower resources with high performance, EfficientNetB0 becomes one of the most promising models for a variety of image classification tasks (Al-Dhabyani et al., 2020).

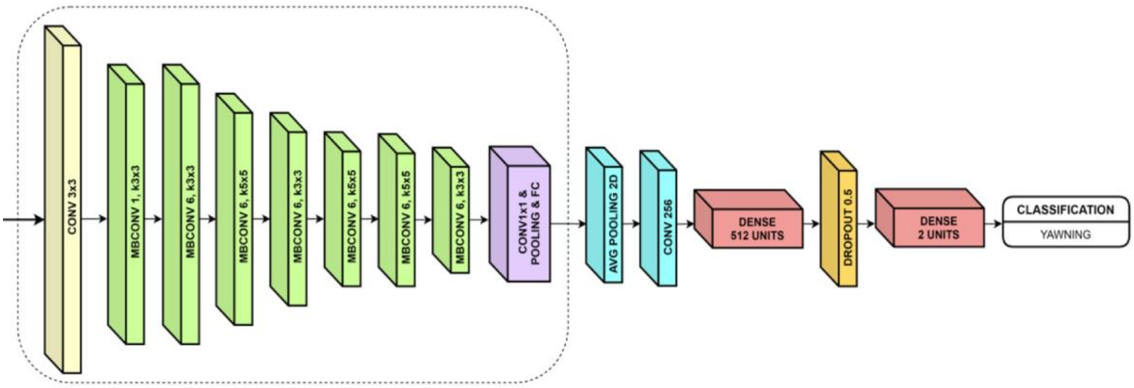


Figure 3-12 EfficientNetB0 architecture (Magán et al., 2022).

In particular, MobileNet was proposed for mobile and embedded vision applications by adopting depthwise separable convolutions instead of standard convolutions to reduce the number of parameters and computational costs significantly (Saini & Devi, 2023). In the MobileNet architecture, a depthwise separable convolution is essentially divided into two steps: a depthwise convolution and a pointwise convolution. While depthwise convolution applies one filter to every input channel, pointwise convolution joins the results from the depthwise convolution using a 1x1 convolution (Ilesanmi et al., 2021). This greatly lowers the complexity of parameters and computational load compared with standard convolutional layers—Figure 3-13: MobileNet Architecture. MobileNet is lightweight and efficient; this design is therefore very suitable for real-time applications on mobile devices (Mahoro & Akhloufi, 2022) (Al-Dhabyani et al., 2020). Even with less parameters, MobileNet is very strong in accuracy, making it very ideal for resource-constrained settings like mobile and embedded systems (Saini & Devi, 2023).

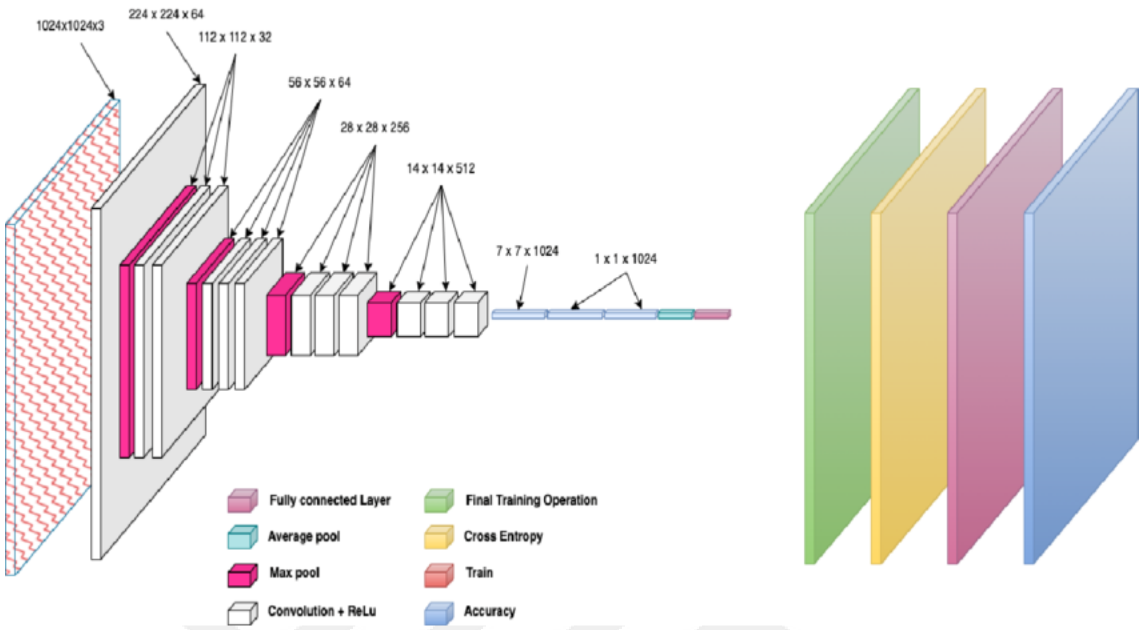


Figure 3-13 Mobile Net Architecture (Saini & Devi, 2023).

InceptionV3 proposed the network architecture of inception modules, which consisted of applying convolutions based on different sizes in the same module to capture multiscale features (Szegedy et al., 2015; Magán et al., 2022). This provided the solution to process multiscale features simultaneously within the network. The architecture of InceptionV3 was such that it had several inception modules, with each one consisting of multiple convolutional filters of varying sizes followed by max-pooling layers (Ilesanmi et al., 2021; Liu et al., 2022). These modules enable the network to capture an extensive range of spatial features by processing the input through filters of different sizes in parallel. The output from all these filters is then concatenated to be fed into the next layer, ensuring that fine and coarse features are captured. What makes InceptionV3 particularly good on complex images in which features are significantly different in size because of its multi-scale feature extraction capability (Szegedy et al., 2015). Figure 3-14: InceptionV3 Architecture. Because of the above versatility, InceptionV3 is very well applied to many tasks with regards to image classification and robust across dataset(Saini & Devi, 2023) .

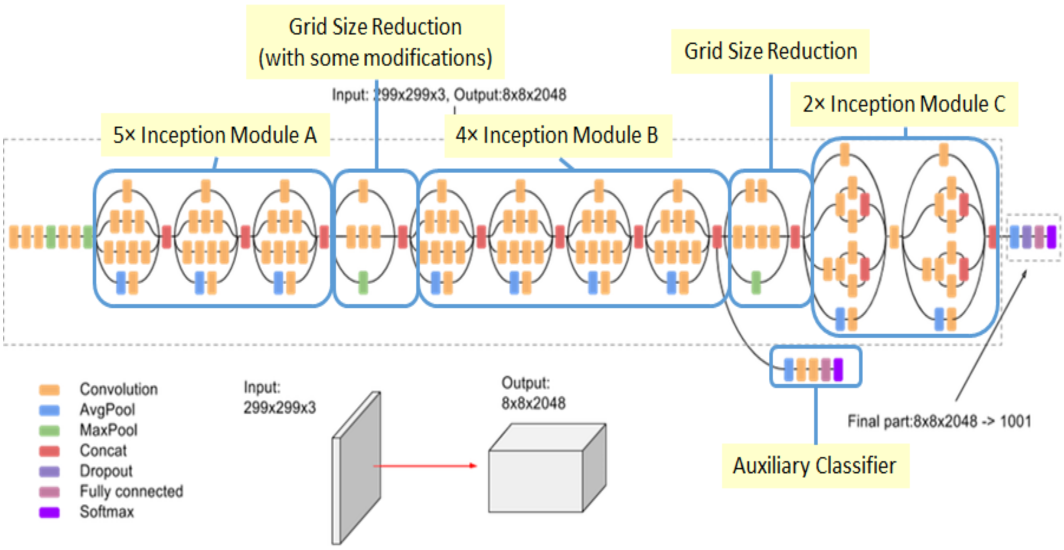


Figure 3-14 InceptionV3 Architecture (Szegedy et al., 2015;.

Xception stands for "Extreme Inception." It is a better inception model, which, in turn, leads to the substitution of depthwise separable convolutions with standard convolutions. This change in architecture enhances the model's efficiency and performance dramatically. For an Xception architecture, depthwise separable convolutions are incorporated throughout the network (Srinivasan et al., 2021). Every depthwise separable convolution is composed of a depthwise convolution followed by a pointwise one. Similar to the MobileNet architecture, this significantly reduces parameters and computational cost while retaining high Accuracy (Mahoro & Akhloufi, 2022).

It has higher efficiency and accuracy than traditional inception models because Xception leverages depthwise separable convolutions. Thus, Xception is a efficient and robust model for tasks of image classification and in scenarios where the computation ability is limited—Figure 3-15: An Overview of the Xception Architecture.

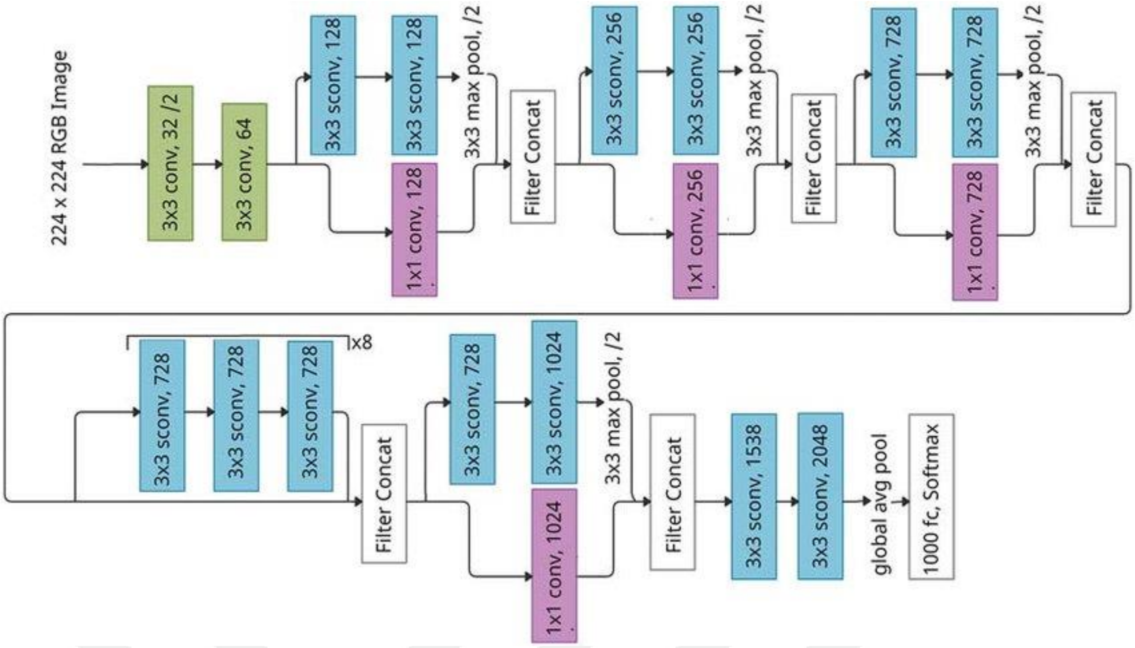


Figure 3-15 Xception Architecture (Srinivasan et al., 2021).

ResNet50 introduced residual learning with skip connections to jump over some layers. It avoids the vanishing gradient problem and allows for intense network training. Now, ResNet50 has 50 layers with identity skip connections bypassing one or more layers. These skip links maintain the flows of gradients during backpropagation and thus allow a deeper model to be learned without suffering from the degradation problem (Magán et al., 2022). Every residual block of ResNet50 includes convolutional layers followed by group normalization with ReLU activation. Also, the skip link adds the input to the output immediately. Figure 3-16: ResNet50 Architecture The residual connections in ResNet50 enable the training of intense networks without much complexity; hence, it would be expected to perform well on complex image classification tasks. Due to the ability to train deep networks without degradation in performance that ResNet50 has been one of the most influential models in deep learning and has provided a foundation for many subsequent architectures (Srinivasan et al., 2021).

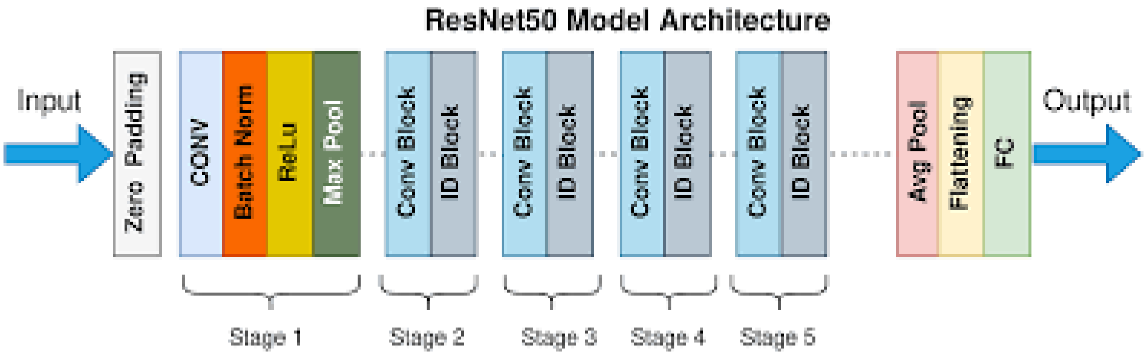


Figure 3-16 ResNet50 Architecture (Srinivasan et al., 2021).

In this regard, each of the pre-trained models was fine-tuned using the enlarged dataset for training and assessment concerning the last four layers. Since this technique employs feature extraction capabilities from pre-trained models, it adjusts them to our classification goal (Mahoro & Akhloufi, 2022). The models were trained using the Adam optimizer with a learning rate of 0.0001 alongside Sparse Categorical Cross-entropy Loss. This combination makes the training efficient and steady, therefore perfect for models to perform correct classification tasks. The multiple data augmentation techniques utilized to the training dataset include flipping, rotating, shifting, shearing, and zooming to make the model more versatile and resilient. Training was done throughout 100 epochs with a batch size of 32. Early stopping was utilized with patience of 5 epochs to prevent overfitting guaranteeing the models did not continue training after their performance on the validation set had reached its best point (Mahoro & Akhloufi, 2022; Srinivasan et al., 2021).

In these experiments, a custom top-layer architecture was designed that allows pre-trained models to be tuned for 3-class classification. It includes a few significant components: First, it has a base model that works as a feature extractor and extracts high-level features from the input image. After that, GlobalAveragePooling2D is applied to reduce the spatial dimensions of the feature maps and transform them into a single vector. It also has dense layers with a dropout that ensures the model does not overfit; this is done by randomly shutting down neurons during training. The last layer is the output layer with three neurons, activated by softmax, enabling multi-class classification. Figure 3-17 displays a Custom layer added to the models. Such a top-layer design will maximize the performance of this model in accurately classifying images under the three distinct categories, taking full advantage of pre-trained models while adapting to your needs.

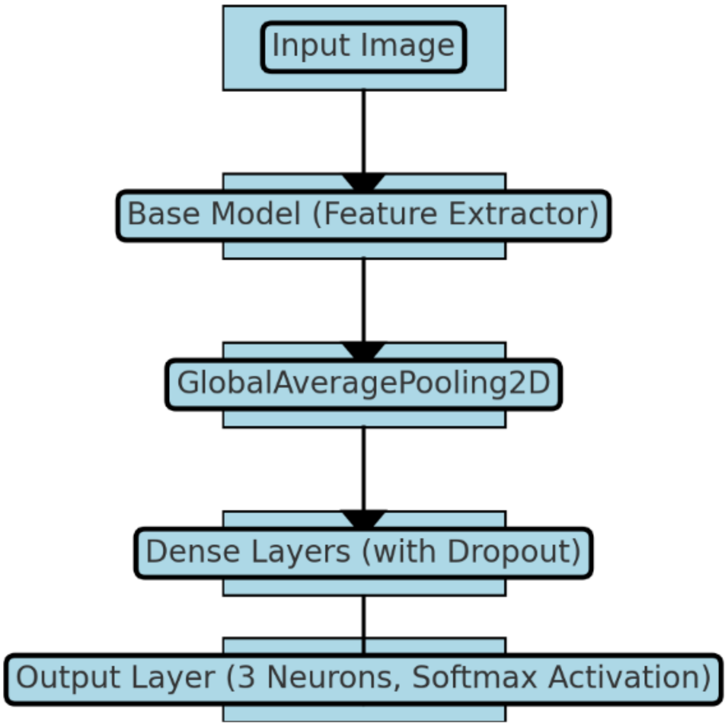


Figure 3-17 Custom layer.

4. RESULTS AND DISCUSSION

4.1 Results of the Proposed CNN Model

We found fascinating information about their performance in custom CNN models. More interestingly, when we used image preprocessing methods, the Accuracy went as high as 0.97. Our custom CNN performed amazingly with these preprocessing methods in terms of other essential metrics as the model's precision and recall were high, indicating good overall performance and the ability to minimize false positives and false negatives: F1-Score was 0.97, precision was 0.98, and recall was 0.96. As shown in Figure 4-1 the accuracy of the CNN model improved significantly after data preprocessing. When this model was trained without preprocessing, its performance dropped significantly; it still has respectable metrics. More concretely, the Accuracy dropped to 70%, meaning that the classification will be less accurate than if preprocessing were applied.

The confusion matrix elaborated on the classification capabilities of this model, showing it to be proficient at distinguishing classes. Train validation accuracy and loss graphs represented its learning journey during training and provided valuable insights into model convergence and generalization. Such findings strongly emphasize the quests of image preprocessing toward improving custom CNN performance but also point to further ways of optimization and research in this field. The designed CNN model performed outstandingly in terms of precision, recall, and F1-score for all

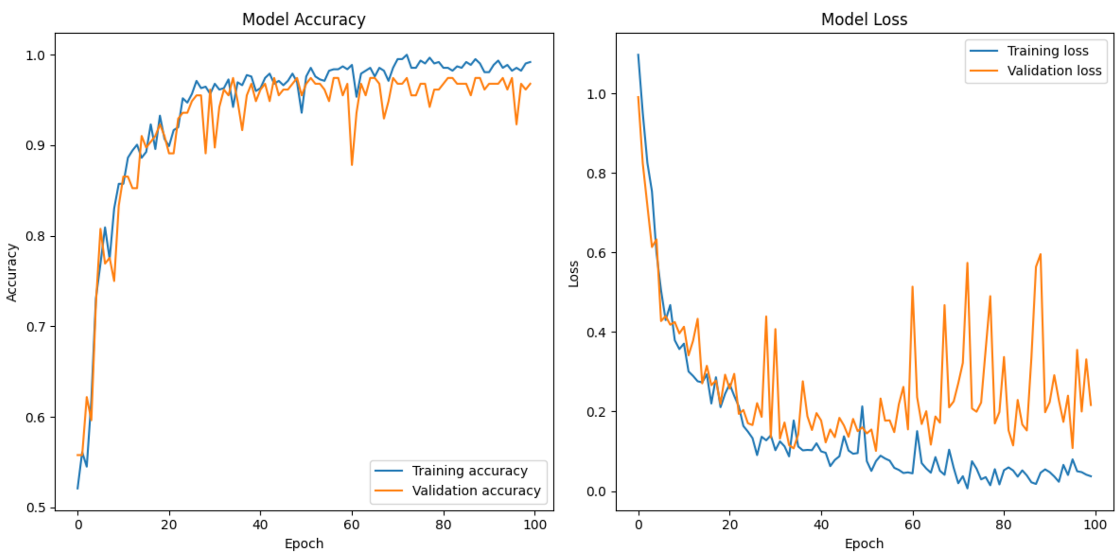


Figure 4-1 The Accuracy and the Loss of the CNN model after data preprocessing.

classes. For class 0, the model recorded a precision of 72%, a recall of 82%, and an F1-score of 77%.

For class 1, this gives a precision of 0.64, a recall of 0.79, and an F1-score of 0.71. Class 2 had a precision of 0.83, a recall of 0.19, with an F1 score of 0.30. The confusion matrix also presented the model accuracy, showing minimal misclassification, as shown in Figure 4-3.Generally, these results serve to underscore the effectiveness of the custom CNN in correctly classifying instances across multiple classes while at the same time showing how data preprocessing may have a tremendous impact on the model's performance.

Without data enhancement, there was a slight decline in the classification results. Figure 4-2 represents the accuracy and loss graphs of the custom CNN model before data preprocessing. For class 0, the model achieved a precision of 72%, a recall of 82%, and an F1-score of 77%. For class 1, precision stood at 64% and recall at 79%, resulting in an F1 score of 71%. Class 2 showed a precision of 83%, a recall of 19%, and an F1 score of 30%.

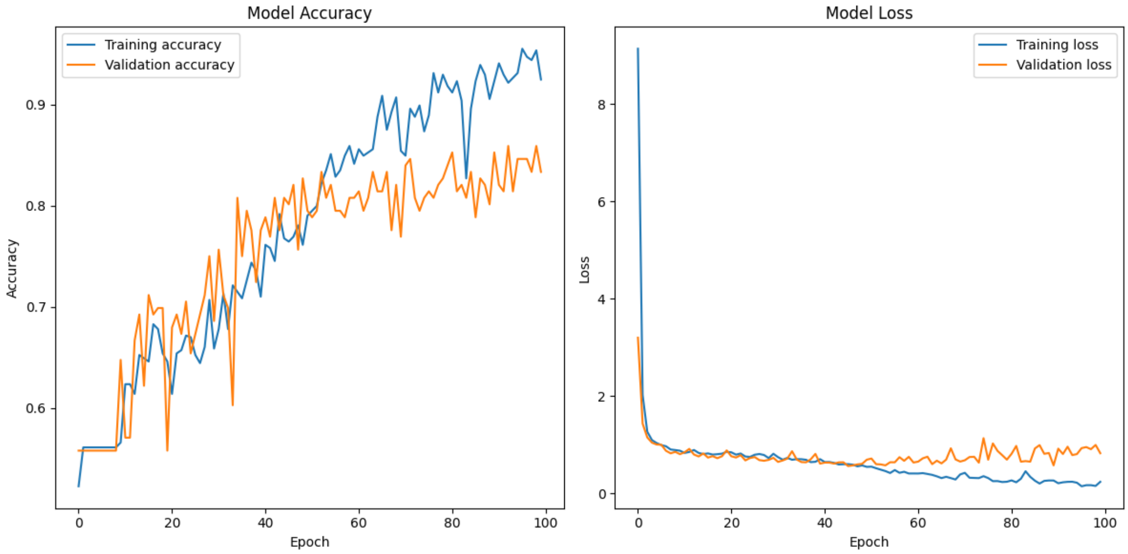
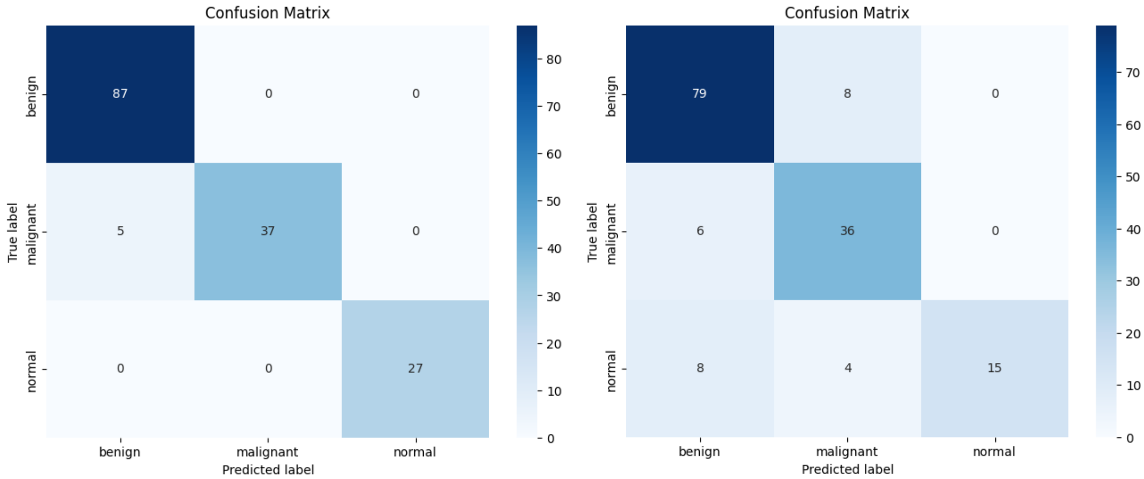


Figure 4-2 The Accuracy and the Loss of the CNN model before data preprocessing.

Despite these variations, the confusion matrix showed the model's Accuracy, with minimal misclassifications. These findings highlight the custom CNN's effectiveness in accurately classifying instances across multiple classes while demonstrating the significant impact of data preprocessing on model performance.



a) After image processing b) Before image processing.

Figure 4-3 Confusion Matrix of the CNN model after and before data preprocessing.

4.2 Results of Transfer Learning Models

In this study, we investigated the performance of fine-tuned models on a dataset augmented with enhanced overlayed images, considering various training scenarios. The performance metrics for each scenario are summarized in Tables Table 4-1Table 4-2Table 4-3 Table 4-4. Table 4-1 presents the performance of models fine-tuned with the last four layers using the enhanced overlayed dataset. EfficientNetB0 demonstrated the highest accuracies of 0.99, with VGG19 achieving a precision of 0.98, a recall of 0.98, and an F1-score of 0.98. VGG16 and ResNet50 also performed exceptionally well, with accuracies of 0.98 and 0.96, respectively. This demonstrates the effectiveness of the enhanced dataset in improving model performance, achieving balanced and high scores across all metrics.

Table 4-1 Performance metrics of the models trained with the last four layers with enhancing and overlaying.

Model	Accuracy	Precision	Recall	F1-Score
VGG16	0.98	0.99	0.98	0.98
VGG19	0.99	0.98	0.98	0.98
DenseNet121	0.88	0.91	0.87	0.89
EfficientNetB0	0.97	0.99	0.97	0.98
MobileNet	0.82	0.82	0.81	0.80
Inceptionv3	0.63	0.59	0.55	0.56
Xception	0.69	0.67	0.61	0.63
ResNet50	0.96	0.96	0.96	0.96

In contrast, Table 4-2 shows the performance of models fine-tuned with the last four layers without enhancements. There was a noticeable drop in performance across all models. VGG16's Accuracy decreased to 0.82, and Inceptionv3's Accuracy fell to 0.60. EfficientNetB0 and VGG19, while performing relatively higher with accuracies of 0.86 and 0.84, respectively, also exhibited declines in their precision, recall, and F1-scores. This indicates the negative impact of the absence of enhancements on model generalization and performance.

Table 4-2 Performance metrics of the models trained with the last four layers without enhancing and overlaying.

Model	Accuracy	Precision	Recall	F1-Score
VGG16	0.82	0.79	0.82	0.80
VGG19	0.84	0.81	0.85	0.82
DenseNet121	0.76	0.75	0.71	0.72
EfficientNetB0	0.86	0.85	0.84	0.84
MobileNet	0.74	0.69	0.72	0.70
Inceptionv3	0.60	0.57	0.50	0.51
Xception	0.66	0.61	0.58	0.59
ResNet50	0.81	0.83	0.75	0.78

Table 4-3 showcases the results of models trained with all layers using the enhanced overlayed dataset. These models exhibited outstanding performance, with several achieving near-perfect metrics. EfficientNetB0 reached an accuracy of 0.99, with precision, recall, and an F1-score of around 0.99. VGG19 and DenseNet121 both achieved an accuracy of 0.98. The Custom CNN also showed strong performance with an accuracy of 0.97. These results underscore the benefits of using an enhanced dataset for training models across all layers.

Table 4-3 Performance metrics of the models trained with all layers with enhancing and overlaying.

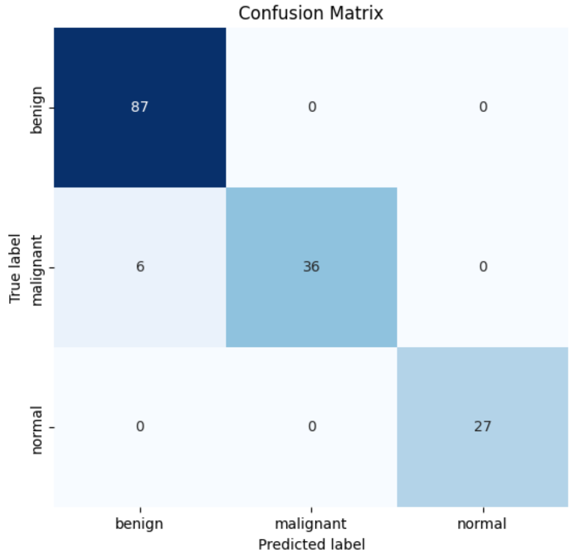
Model	Accuracy	Precision	Recall	F1-Score
VGG16	0.96	0.98	0.95	0.96
VGG19	0.98	0.99	0.98	0.98
DenseNet121	0.98	0.99	0.98	0.98
EfficientNetB0	0.99	0.99	0.98	0.99
MobileNet	0.98	0.98	0.98	0.98
Inceptionv3	0.98	0.99	0.98	0.98
Xception	0.98	0.99	0.98	0.98
ResNet50	0.91	0.91	0.93	0.92
Custom CNN	0.97	0.98	0.96	0.97

Table 4-4 details the performance of models trained with all layers without enhancements. Although there was a general decrease in performance compared to their enhanced counterparts, models like DenseNet121 and EfficientNetB0 still maintained relatively high accuracies of 0.87 and 0.86, respectively. However, models such as VGG16 and MobileNet exhibited more substantial drops, emphasizing the importance of dataset enhancements. The Custom CNN, without enhancements, showed a moderate decrease in performance with an accuracy of 0.83.

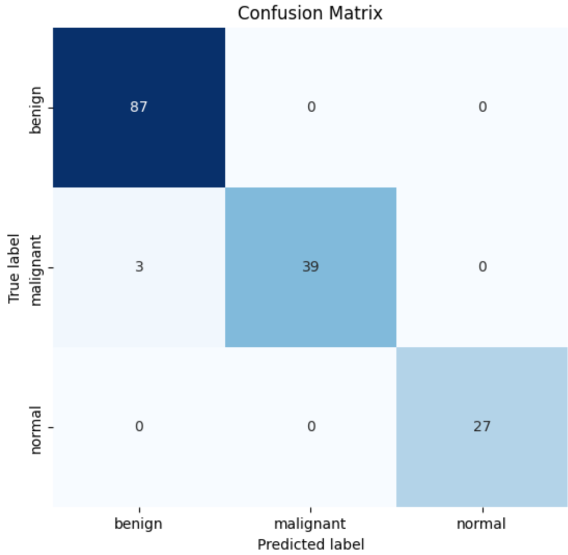
Our study demonstrates the critical role of dataset quality and preprocessing techniques in enhancing deep learning models' robustness and generalization capabilities. The enhanced overlayed dataset consistently improved model accuracy, precision, recall, and F1-score across various architectures. These findings reaffirm the value of data enhancement in machine learning tasks, highlighting its importance for achieving superior model performance. Figure 4-4 and Figure 4-5 represent the confusion matrices of the transfer learning models after and before data preprocessing.

Table 4-4 Performance metrics of the models trained with all layers without enhancing and overlaying.

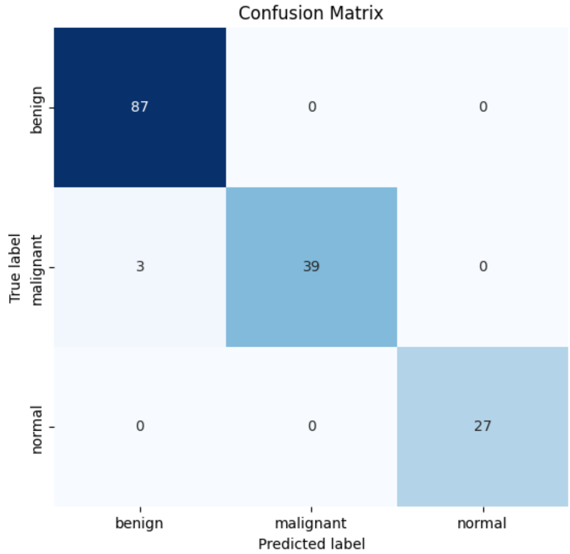
Model	Accuracy	Precision	Recall	F1-Score
VGG16	0.76	0.75	0.73	0.71
VGG19	0.77	0.74	0.77	0.75
DenseNet121	0.87	0.89	0.83	0.85
EfficientNetB0	0.86	0.85	0.84	0.85
MobileNet	0.86	0.85	0.84	0.83
Inceptionv3	0.82	0.80	0.80	0.80
Xception	0.85	0.84	0.81	0.82
ResNet50	0.85	0.84	0.79	0.81
Custom CNN	0.83	0.80	0.83	0.81



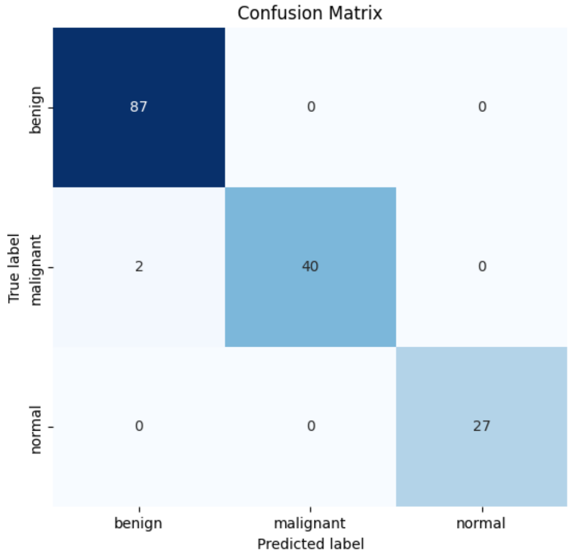
a) VGG16



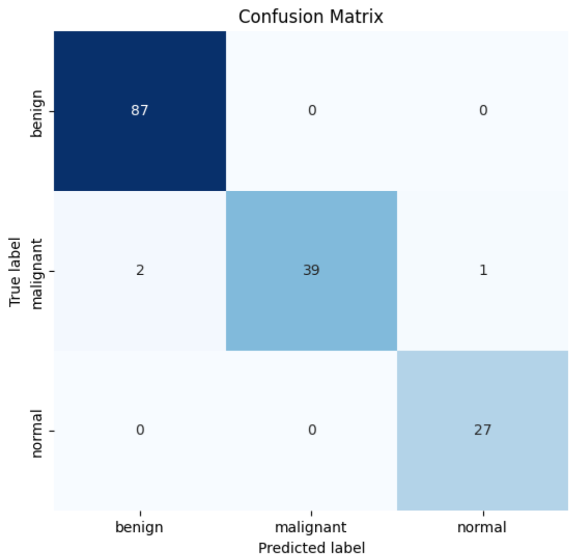
b) VGG19



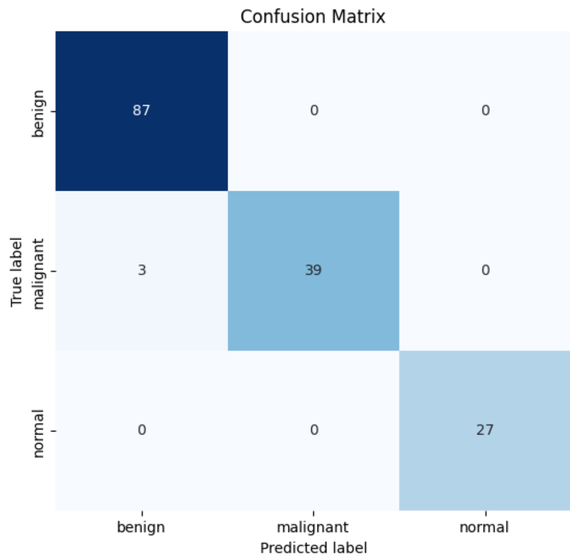
C) DensNet121



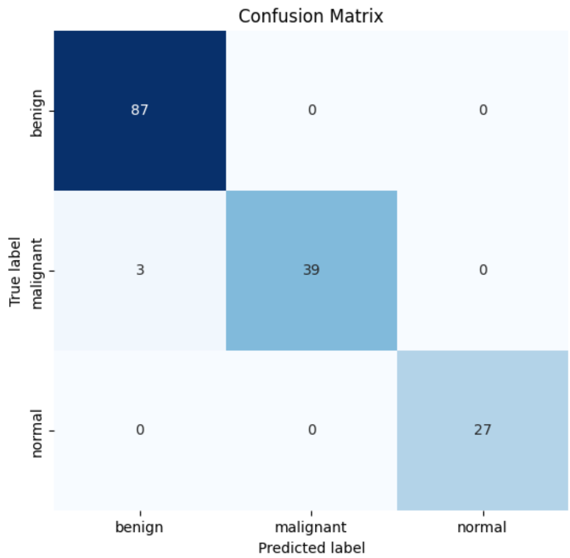
d) EfficientNetB0



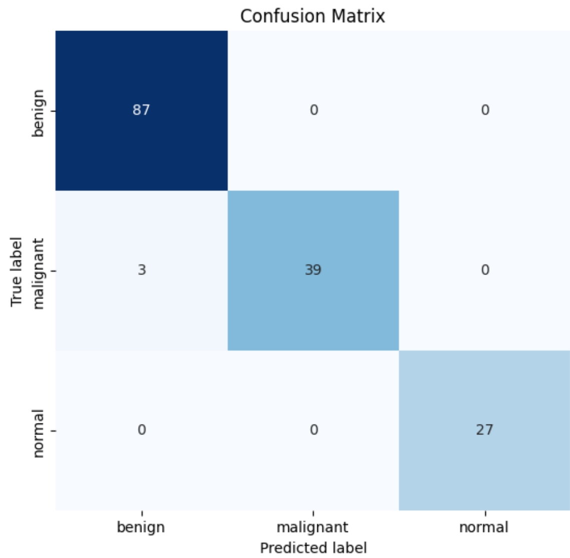
e) MobileNet



f) InceptionV3

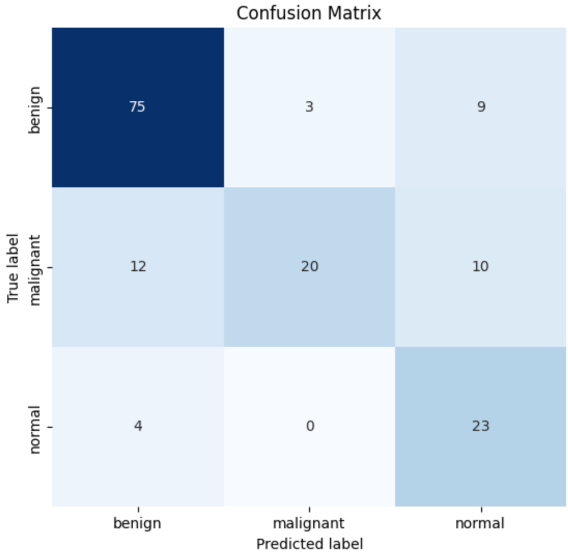


g) Xception

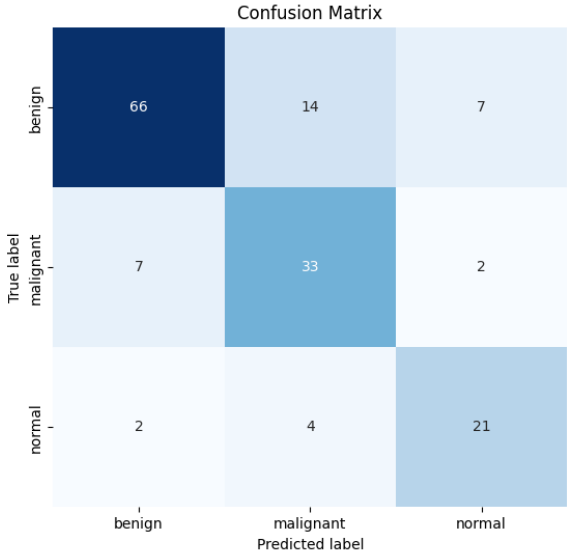


h) ResNet50

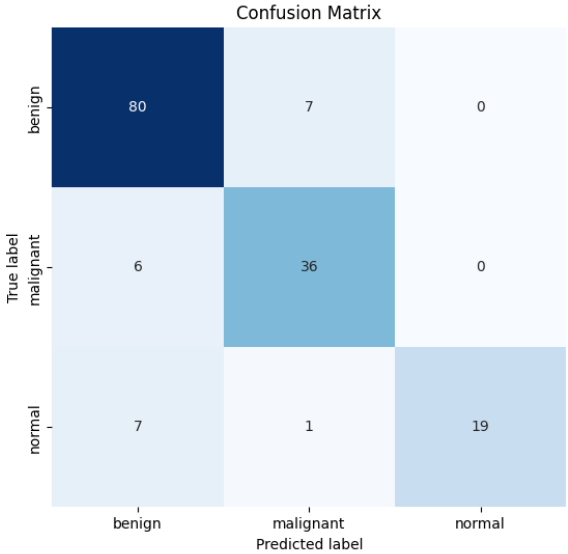
Figure 4-4 Confusion matrix of transfer learning models after data preprocessing.



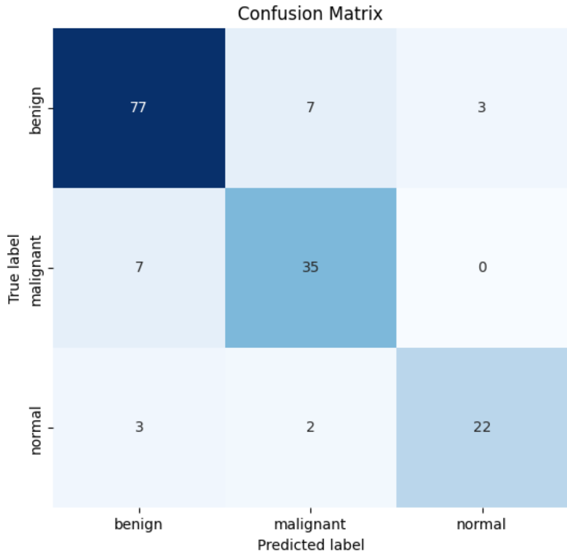
a) VGG16



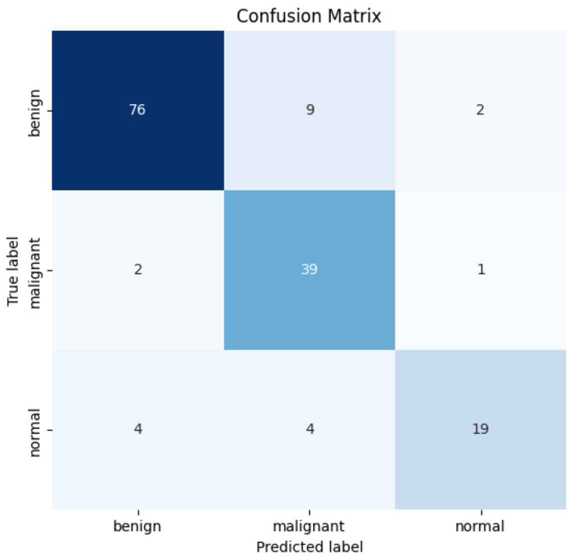
b) VGG19



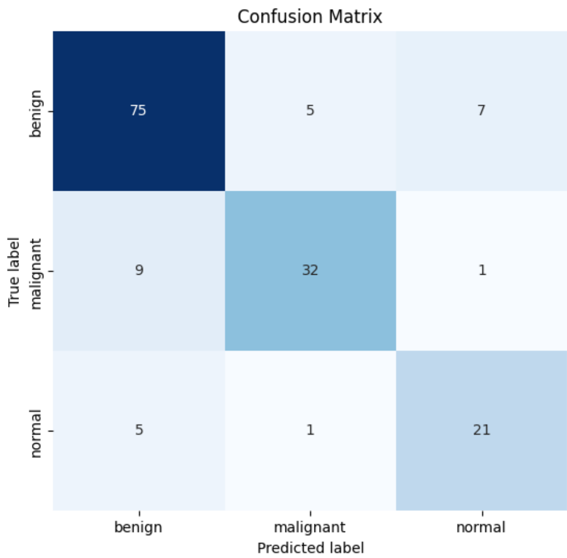
C) DensNet121



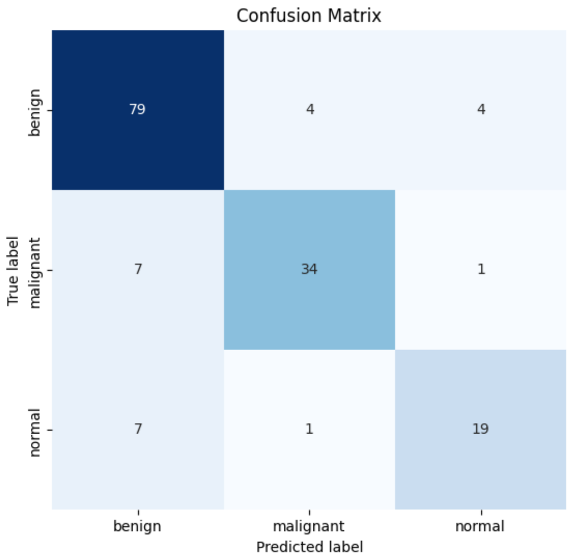
d) EfficientNetB0



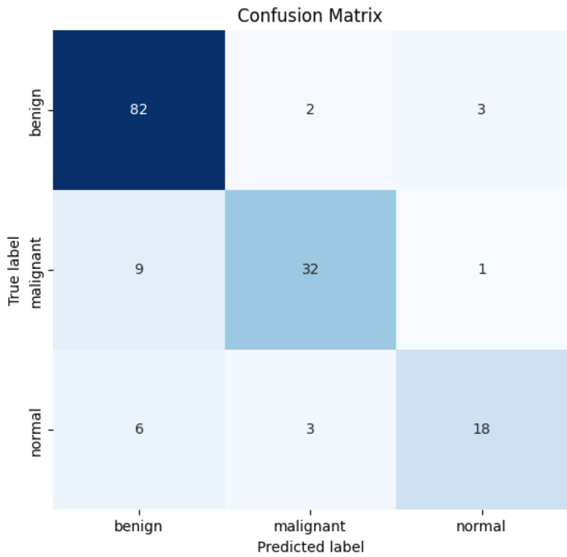
e) MobileNet



f) InceptionV3



g) Xception



h) ResNet50

Figure 4-5 Confusion matrix of transfer learning models before data preprocessing.

4.3 Performance Comparison of Transfer Learning and Custom CNN Models

When comparing our study's findings with other transfer learning models, EfficientNetB0 performed exceptionally well, achieving an accuracy score of 0.99. This indicates its high confidence and precision in categorizing images across various datasets. VGG19, MobileNet, DenseNet121, InceptionV3, and Xception all performed well with an accuracy of 0.98. VGG16 achieved an accuracy of 0.96. ResNet50, although slightly lower, still maintained strong performance with an accuracy of 0.91. The custom CNN also demonstrated excellent performance, achieving an accuracy of 0.97. The custom CNN achieved high precision, recall, and F1 scores with significantly fewer parameters than other architectures except EfficientNetB0 and Mobilenet. VGG16 with 14879299, VGG19 with 20188995, Densnet121 with 7333187, EfficientNetB0 with 4410790, Mobilenet with 3524547, InceptionV3 with 22360611, Xception with 21419307, ResNet50 with 24145539. With only 3097283 parameters, our model efficiently extracts relevant features and makes precise classifications without needing an excessive number of parameters. Figure 4-6 show the comparison of transfer learning and our custom CNN. This emphasizes not only its technical excellence but also its efficiency in using computational resources. Therefore, our custom CNN emerges as a standout performer, offering a strong case for its adoption in image classification tasks where both efficiency and accuracy are crucial.

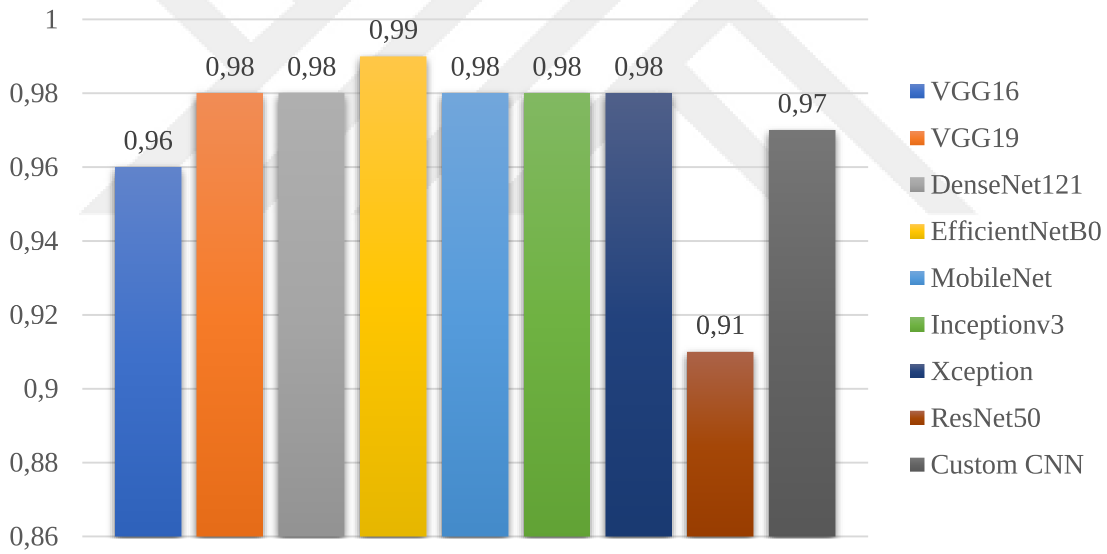


Figure 4-6 Comparison of accuracy between transfer learning models and the custom CNN model on breast ultrasound images.

4.4 Comparison with Results from Similar Studies

Several important insights emerge when comparing our study's outcomes with similar studies utilizing deep learning or transfer learning methodologies and comparable datasets for BC analysis.(Liu et al., 2022) followed suit with a grid-based deep feature generation approach on the BUS dataset, yielding a commendable accuracy of 97.18%, showcasing the efficacy of this method in achieving high precision. (Zhuang et al., 2021) employed fuzzy enhancement techniques on the BUS dataset, resulting in a notable accuracy of 95.48%, demonstrating the potential of such preprocessing methods in improving classification performance. (Alrubaie et al., 2023) utilized the Inception-v3 CNN architecture on both BUSI and BUS datasets, achieving remarkable accuracies of 96% and 100%, respectively, underscoring the robustness of their approach across different datasets. (Cruz-Ramos et al., 2023) integrated fusion and handcrafted features on the BUSI dataset, attaining an accuracy of 97.6% and a precision of 98%, showing the effectiveness of feature engineering in enhancing classification outcomes. (Pang et al., 2021) This work utilized a semi-supervised GAN model on the BUS dataset, demonstrating an accuracy as high as 90.41%, hence proving the potential of generative models in augmenting the performance in classification tasks. (Z. Zhang et al., 2021) The large-scale BUSI dataset was used, where the Bayesian YOLOv4 application resulted in a sensitivity of 0.88 with low false positives, indicating that indeed their approach attained an accurate tumor detection. All these aforementioned works are in contrast to ours, where a custom CNN architecture is tailored for the analysis of breast ultrasound images, providing competitive accuracy. We achieved 97% accuracy, demonstrating the effectiveness of our methodology for accurately classifying breast ultrasound. Through this comparative analysis, we gain valuable insights into the diverse methods and their respective successes in BC analysis tasks. Table 4-5 shows the comparison of our model with related studies.

Table 4-5 Comparison table of Custom CNN with related studies.

Study	Methods Used	Dataset	Accuracy
(Zhuang et al., 2021)	Fuzzy enhancement	BUS (1328 images)	95.48%
(Cruz-Ramos et al., 2023)	Fusion and handcrafted features	BUSI (780 images)	97.6%
(Pang et al., 2021)	Semi-supervised GAN model	BUS (1447 images)	90.41%
(Z. Zhang et al., 2021)	Bayesian YOLOv4	BUSI (21,624 images)	Sens: 0.88 FPs/S: 0.19
(Alrubaie et al., 2023)	Inception-v3 CNN	A: BUSI (780 images) B: BUS (9016 images)	96%
Proposed Model	CNN	BUSI (780 images)	97%

4.5 Grad-CAM Visualizations and Interpretability of the Model

Grad-CAM was used to understand the internal mechanisms of a CNN model when making predictions using images from the test dataset. Grad-CAM visualizes the activations of a trained convolutional neural network, due to any input image, at the last convolutional layer; it uses modulus, max across channels, after passing through ReLU activations. In these analyses, specific layers of the CNN architecture were chosen and used to visualize different class activation patterns. Figure 4-7Figure 4-8Figure 4-9 show images in the Grad-CAM heatmap process. There are four different visualizations in each figure, combined to give an all-rounded understanding of how the model is making its decisions.

First, there is the original input image to provide context. Second, there is an overlay of Grad-CAM's heat map on top of the input image to visualize those areas in the image that strongly influenced the model's prediction. To further increase interpretability, two other kinds of visualizations were included: one with the Grad-CAM heatmap overlaid using transparency to clarify the dissemination of the gradient in the heatmap and another kind that used a transparent jet colourmap, which substantially enhances the content intuition within the heatmap. Besides being helpful in understanding which image regions most contributed to the model's decision, such visualizations can be very useful when studying CNN's feature extraction and classification processes. Grad-CAM, most importantly, is enabling researchers and practitioners to validate models for attention mechanisms in the elucidation of a prediction, hence improving transparency and interpretability in deep learning models for image analysis tasks.

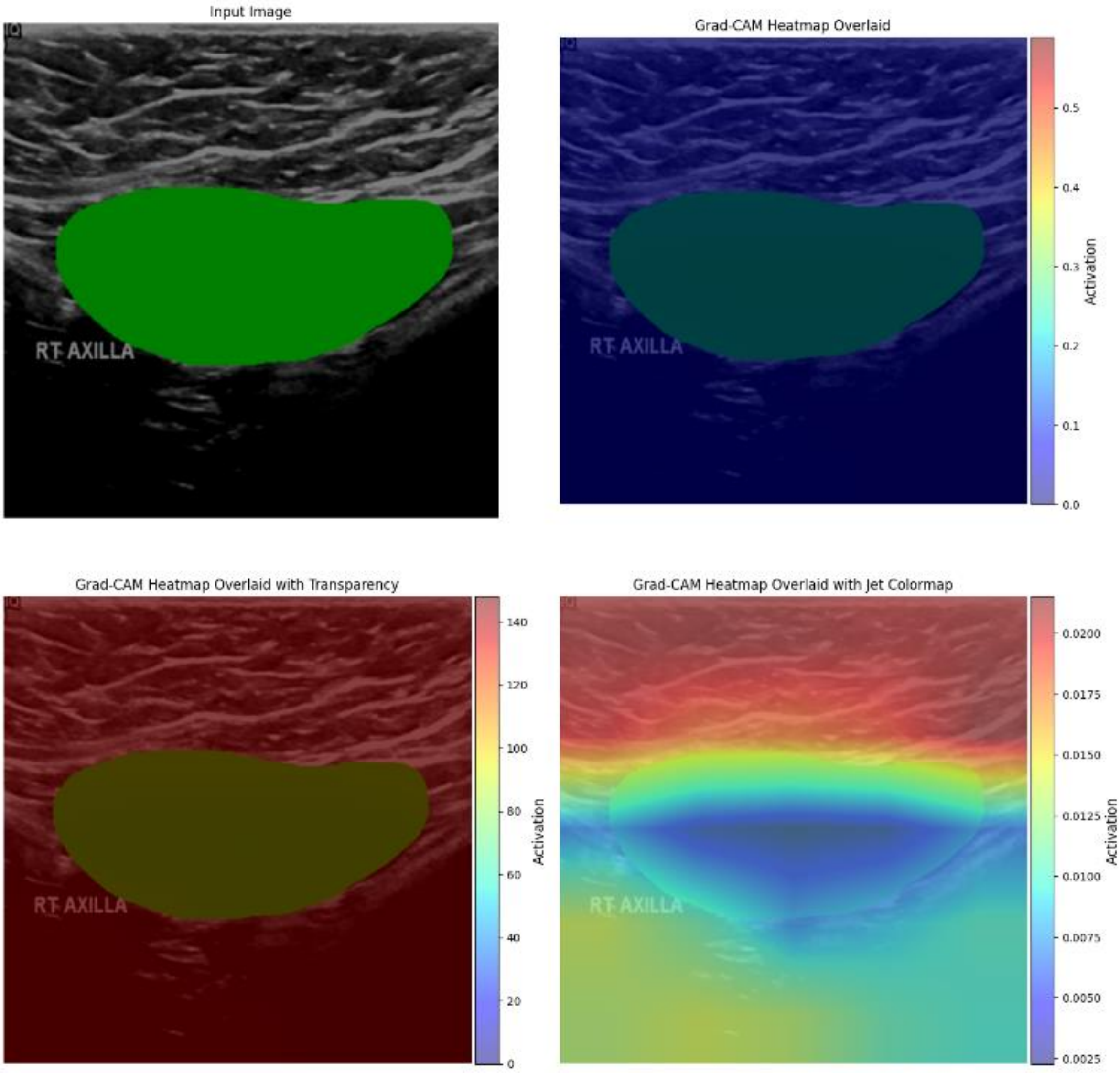


Figure 4-7 The Grad-CAM heatmap for a malignant case.

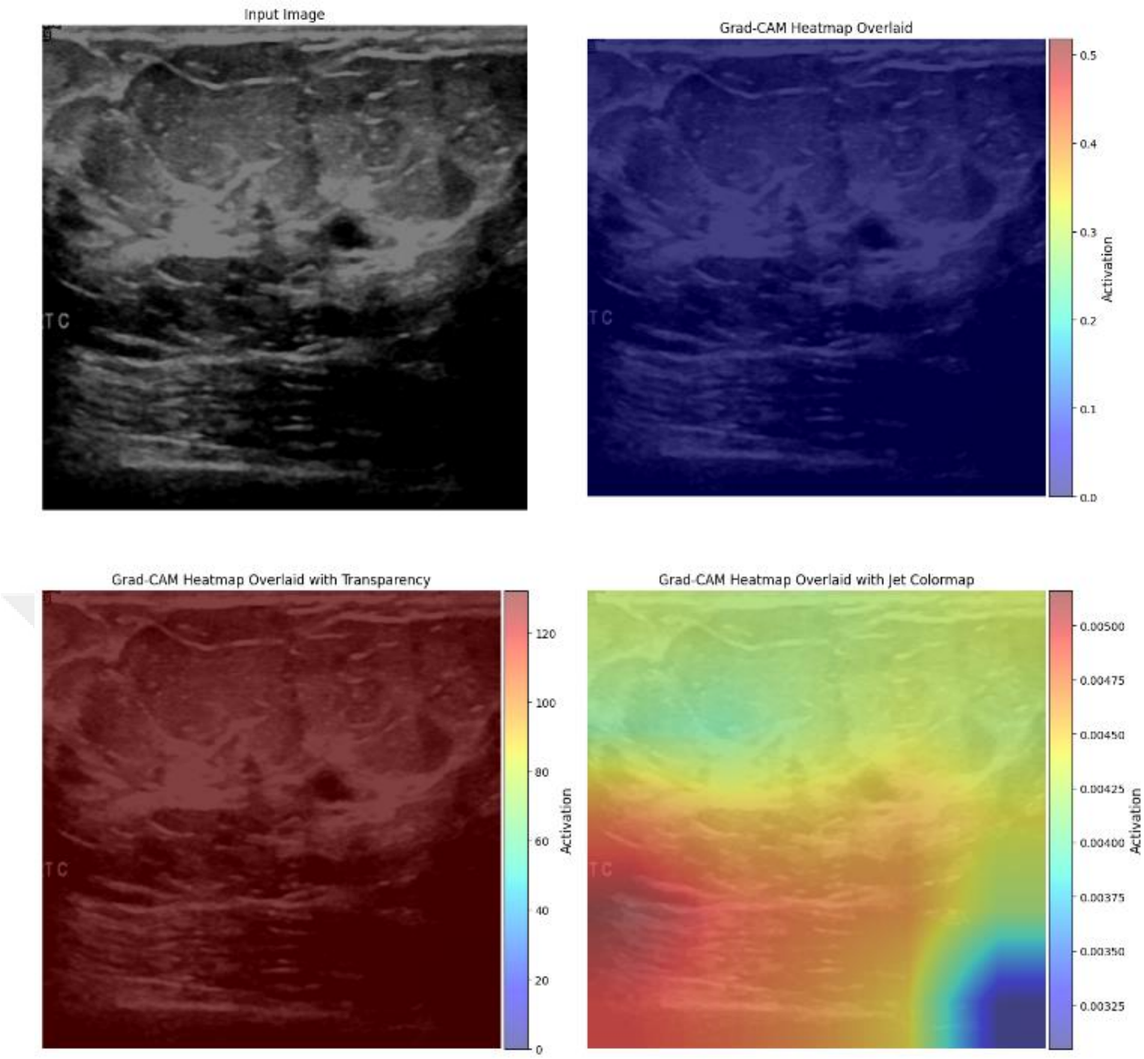


Figure 4-8 The Grad-CAM heatmap for a normal case.

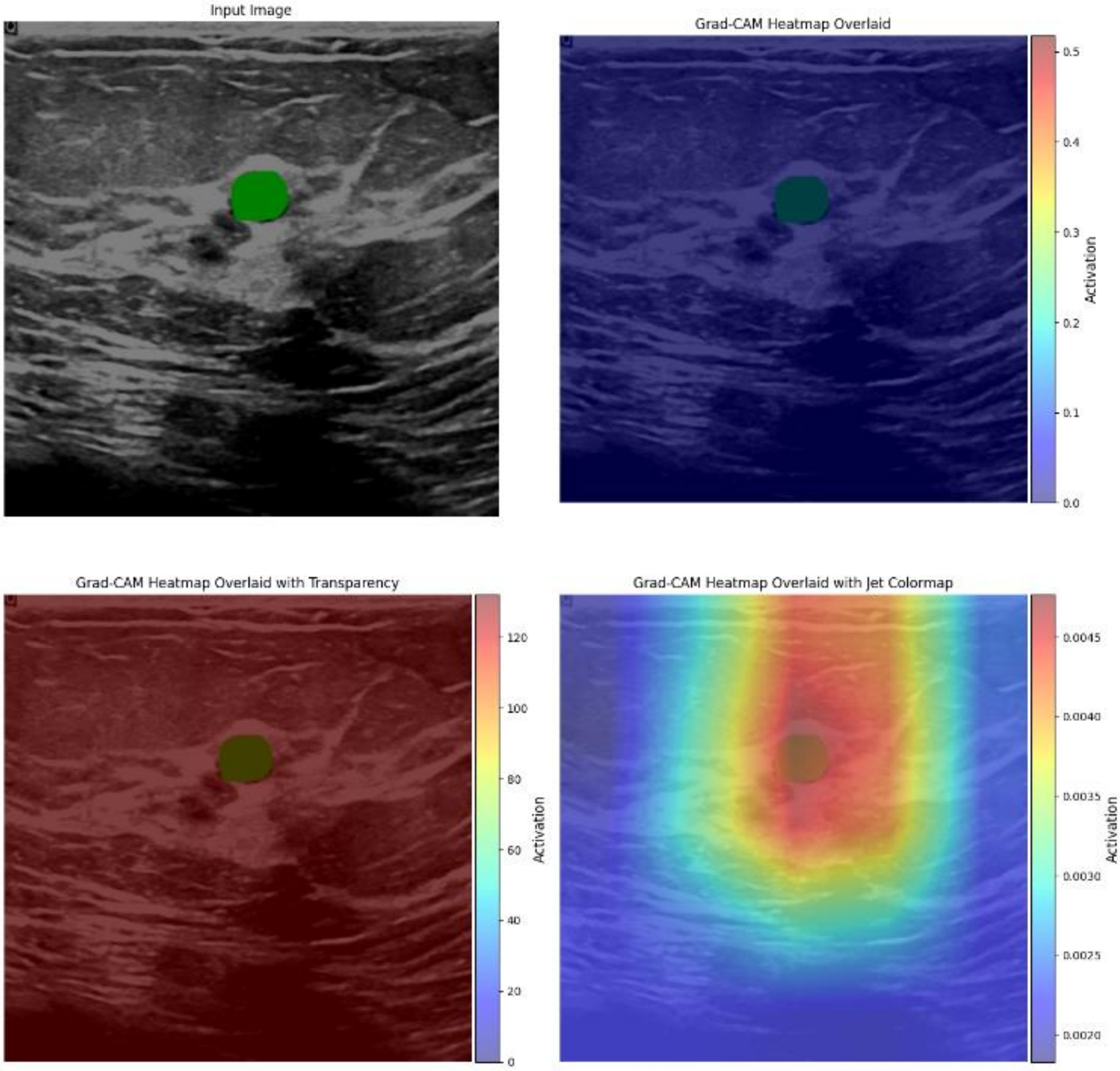


Figure 4-9 The Grad-CAM heatmap for a benign case.

5. CONCLUSION

In conclusion, this research demonstrated the enormous potential of deep learning models, especially CNNs, for the efficient classification of ultrasound images in the diagnosis of BC. The subject models were easily interpretable; besides, they showed better diagnostic performance when advanced image processing techniques-segmentation and enhancement-were applied to the images. Our proposed CNN model achieved an accuracy of 97%, compared with other state-of-the-art models such as EfficientNetB0, MobileNet, and InceptionV3. This emphasizes the need for models specially tailored with the ability to handle challenges that come with the uniqueness of medical imaging tasks. More importantly, the interpretability and transparency of our models were taken further with Grad-CAM use, making the decision-making process more comprehensible. This is critical in the clinical field. AI diagnostic tools should be not only precise but also interpretable for healthcare professionals and patients to build confidence in them. These results also suggest the importance of strong techniques for image preprocessing. Techniques such as image enhancement and overlay greatly improved model performance. This work should be extended in further studies to include more data modalities, such as 3D ultrasound or MRI images, and more complex CNN architectures that may result in even better classification outcomes. Moreover, such models' ability to generalize on larger datasets with more diverse data would be reassuring against their clinical use. The current study provided a sufficient backbone regarding the application of CNNs in the diagnosis of BC by using ultrasound images and gave important indications on how to design more efficient and trustworthy diagnostic tools that could potentially improve patient outcomes.

6. REFERENCES

Al-Dhabyani, W., Gomaa, M., Khaled, H., & Fahmy, A. (2020). Dataset of breast ultrasound images. *Data in Brief*, *28*, 104863. https://doi.org/10.1016/j.dib.2019.104863

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

Badawy, S. M., Mohamed, A. E.-N. A., Hefnawy, A. A., Zidan, H. E., GadAllah, M. T., & El-Banby, G. M. (2021). Automatic semantic segmentation of breast tumors in ultrasound images based on combining fuzzy logic and deep learning—A feasibility study. *PLOS ONE*, *16*(5), e0251899. https://doi.org/10.1371/journal.pone.0251899

Cao, Z., Yang, G., Chen, Q., Chen, X., & Lv, F. (2020). Breast tumor classification through learning from noisy labeled ultrasound images. *Medical Physics*, *47*(3), 1048–1057. https://doi.org/10.1002/mp.13966

Çetin-Kaya, Y., & Kaya, M. (2024). A Novel Ensemble Framework for Multi-Classification of Brain Tumors Using Magnetic Resonance Imaging. *Diagnostics*, *14*(4), 383. https://doi.org/10.3390/diagnostics14040383

Chiang, T.-C., Huang, Y.-S., Chen, R.-T., Huang, C.-S., & Chang, R.-F. (2019). Tumor Detection in Automated Breast Ultrasound Using 3-D CNN and Prioritized Candidate Aggregation. *IEEE Transactions on Medical Imaging*, *38*(1), 240–249. https://doi.org/10.1109/TMI.2018.2860257

Coronado-Gutiérrez, D., Santamaría, G., Ganau, S., Bargalló, X., Orlando, S., Oliva-Brañas, M. E., Perez-Moreno, A., & Burgos-Artizzu, X. P. (2019). Quantitative Ultrasound Image Analysis of Axillary Lymph Nodes to Diagnose Metastatic Involvement in Breast Cancer. *Ultrasound in Medicine & Biology*, *45*(11), 2932–2941. https://doi.org/10.1016/j.ultrasmedbio.2019.07.413

Cruz-Ramos, C., García-Avila, O., Almaraz-Damian, J.-A., Ponomaryov, V., Reyes-Reyes, R., & Sadovnychiy, S. (2023). Benign and Malignant Breast Tumor Classification in Ultrasound and Mammography Images via Fusion of Deep Learning and Handcraft Features. *Entropy*, *25*(7), 991. https://doi.org/10.3390/e25070991

Ferguson, M., Ak, R., Lee, Y.-T. T., & Law, K. H. (2017). Automatic localization of casting defects with convolutional neural networks. *2017 IEEE International Conference on Big Data (Big Data)*, 1726–1735. https://doi.org/10.1109/BigData.2017.8258115

Fujioka, T., Mori, M., Kubota, K., Oyama, J., Yamaga, E., Yashima, Y., Katsuta, L., Nomura, K., Nara, M., Oda, G., Nakagawa, T., Kitazume, Y., & Tateishi, U. (2020). The Utility of Deep Learning in Breast Ultrasonic Imaging: A Review. *Diagnostics*, *10*(12), 1055. https://doi.org/10.3390/diagnostics10121055

Gómez-Flores, W., & Coelho de Albuquerque Pereira, W. (2020). A comparative study of pre-trained convolutional neural networks for semantic segmentation of breast tumors in ultrasound. *Computers in Biology and Medicine*, *126*, 104036. https://doi.org/10.1016/j.compbiomed.2020.104036

Gong, B., Shen, L., Chang, C., Zhou, S., Zhou, W., Li, S., & Shi, J. (2020). BI-Modal Ultrasound Breast Cancer Diagnosis Via Multi-View Deep Neural Network SVM. *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 1106–1110. https://doi.org/10.1109/ISBI45749.2020.9098438

Han, S., Kang, H.-K., Jeong, J.-Y., Park, M.-H., Kim, W., Bang, W.-C., & Seong, Y.-K. (2017). A deep learning framework for supporting the classification of breast lesions in ultrasound images. *Physics in Medicine & Biology*, *62*(19), 7714–7728. https://doi.org/10.1088/1361-6560/aa82ec

Ilesanmi, A. E., Chaumrattanakul, U., & Makhanov, S. S. (2021). A method for segmentation of tumors in breast ultrasound images using the variant enhanced deep learning. *Biocybernetics and Biomedical Engineering*, *41*(2), 802–818. https://doi.org/10.1016/j.bbe.2021.05.007

Jabeen, K., Khan, M. A., Alhaisoni, M., Tariq, U., Zhang, Y.-D., Hamza, A., Mickus, A., & Damaševičius, R. (2022). Breast Cancer Classification from Ultrasound Images Using Probability-Based Optimal Deep Learning Feature Fusion. *Sensors*, *22*(3), 807. https://doi.org/10.3390/s22030807

Kabir, S. M., Bhuiyan, M. I. H., Tanveer, M. S., & Shihavuddin, A. (2021). RiIG Modeled WCP Image-Based CNN Architecture and Feature-Based Approach in Breast Tumor Classification from B-Mode Ultrasound. *Applied Sciences*, *11*(24), 12138. https://doi.org/10.3390/app112412138

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

Kim, J., Kim, H. J., Kim, C., Lee, J. H., Kim, K. W., Park, Y. M., Kim, H. W., Ki, S. Y., Kim, Y. M., & Kim, W. H. (2021). Weakly-supervised deep learning for ultrasound diagnosis of breast cancer. *Scientific Reports*, *11*(1), 24382. https://doi.org/10.1038/s41598-021-03806-7

Lei, B., Huang, S., Li, R., Bian, C., Li, H., Chou, Y.-H., & Cheng, J.-Z. (2018). Segmentation of breast anatomy for automated whole breast ultrasound images with boundary regularized convolutional encoder–decoder network. *Neurocomputing*, *321*, 178–186. https://doi.org/10.1016/j.neucom.2018.09.043

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*, *Volume 15*, 2271–2282. https://doi.org/10.2147/IJGM.S347491

Luo, L., Wang, X., Lin, Y., Ma, X., Tan, A., Chan, R., Vardhanabhuti, V., Chu, W. C., Cheng, K.-T., & Chen, H. (2024). Deep Learning in Breast Cancer Imaging: A Decade of Progress and Future Directions. *IEEE Reviews in Biomedical Engineering*, 1–20. https://doi.org/10.1109/RBME.2024.3357877

Magán, E., Sesmero, M. P., Alonso-Weber, J. M., & Sanchis, A. (2022). Driver Drowsiness Detection by Applying Deep Learning Techniques to Sequences of Images. *Applied Sciences*, *12*(3), 1145. https://doi.org/10.3390/app12031145

Mahoro, E., & Akhloufi, M. A. (2022). Applying Deep Learning for Breast Cancer Detection in Radiology. *Current Oncology*, *29*(11), 8767–8793. https://doi.org/10.3390/curroncol29110690

Masud, M., Hossain, M. S., Alhumyani, H., Alshamrani, S. S., Cheikhrouhou, O., Ibrahim, S., Muhammad, G., Rashed, A. E. E., & Gupta, B. B. (2021). Pre-Trained Convolutional Neural Networks for Breast Cancer Detection Using Ultrasound Images. *ACM Transactions on Internet Technology*, *21*(4), 1–17. https://doi.org/10.1145/3418355

Momot A, & Zaboluieva M. (2022). AUTOMATION OF ULTRASOUND BREAST CANCER IMAGES CLASSIFICATION USING DEEP NEURAL NETWORKS. *Sciences of Europe #*, *96*. https://doi.org/10.5281/zenodo.6809758

Moon, W. K., Lee, Y.-W., Ke, H.-H., Lee, S. H., Huang, C.-S., & Chang, R.-F. (2020). Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Computer Methods and Programs in Biomedicine*, *190*, 105361. https://doi.org/10.1016/j.cmpb.2020.105361

Moustafa, A. F., Cary, T. W., Sultan, L. R., Schultz, S. M., Conant, E. F., Venkatesh, S. S., & Sehgal, C. M. (2020). Color Doppler Ultrasound Improves Machine Learning Diagnosis of Breast Cancer. *Diagnostics*, *10*(9), 631. https://doi.org/10.3390/diagnostics10090631

Negi, A., Raj, A. N. J., Nersisson, R., Zhuang, Z., & Murugappan, M. (2020). RDA-UNET-WGAN: An Accurate Breast Ultrasound Lesion Segmentation Using Wasserstein Generative Adversarial Networks. *Arabian Journal for Science and Engineering*, *45*(8), 6399–6410. https://doi.org/10.1007/s13369-020-04480-z

Oluwasanmi, A., Aftab, M. U., Qin, Z., Ngo, S. T., Doan, T. Van, Nguyen, S. B., & Nguyen, S. H. (2021). Transfer Learning and Semisupervised Adversarial Detection and Classification of COVID-19 in CT Images. *Complexity*, *2021*(1). https://doi.org/10.1155/2021/6680455

PACAL, İ. (2022). Deep Learning Approaches for Classification of Breast Cancer in Ultrasound (US) Images. *Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, *12*(4), 1917–1927. https://doi.org/10.21597/jist.1183679

Pang, T., Wong, J. H. D., Ng, W. L., & Chan, C. S. (2021). Semi-supervised GAN-based Radiomics Model for Data Augmentation in Breast Ultrasound Mass Classification. *Computer Methods and Programs in Biomedicine*, *203*, 106018. https://doi.org/10.1016/j.cmpb.2021.106018

Peng, Y., Tang, W., & Peng, X. (2023). The study of ultrasonography based on deep learning in breast cancer. *Journal of Radiation Research and Applied Sciences*, *16*(4), 100679. https://doi.org/10.1016/j.jrras.2023.100679

Pourasad, Y., Zarouri, E., Salemizadeh Parizi, M., & Salih Mohammed, A. (2021). Presentation of Novel Architecture for Diagnosis and Identifying Breast Cancer Location Based on Ultrasound Images Using Machine Learning. *Diagnostics*, *11*(10), 1870. https://doi.org/10.3390/diagnostics11101870

Qi, X., Zhang, L., Chen, Y., Pi, Y., Chen, Y., Lv, Q., & Yi, Z. (2019). Automated diagnosis of breast ultrasonography images using deep neural networks. *Medical Image Analysis*, *52*, 185–198. https://doi.org/10.1016/j.media.2018.12.006

Saini, K., & Devi, R. (2023). A systematic scoping review of the analysis of COVID-19 disease using chest X-ray images with deep learning models. *Journal of Autonomous Intelligence, 7*(2). https://doi.org/10.32629/jai.v7i2.928

Srinivasan, K., Garg, L., Datta, D., A. Alaboudi, A., Z. Jhanjhi, N., Agarwal, R., & George Thomas, A. (2021). Performance Comparison of Deep CNN Models for Detecting Driver's Distraction. *Computers, Materials & Continua*, *68*(3), 4109–4124. https://doi.org/10.32604/cmc.2021.016736

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). *Rethinking the Inception Architecture for Computer Vision*. http://arxiv.org/abs/1512.00567

Vakanski, A., Xian, M., & Freer, P. E. (2020). Attention-Enriched Deep Learning Model for Breast Tumor Segmentation in Ultrasound Images. *Ultrasound in Medicine & Biology*, *46*(10), 2819–2833. https://doi.org/10.1016/j.ultrasmedbio.2020.06.015

Vigil, N., Barry, M., Amini, A., Akhloufi, M., Maldague, X. P. V., Ma, L., Ren, L., & Yousefi, B. (2022). Dual-Intended Deep Learning Model for Breast Cancer Diagnosis in Ultrasound Imaging. *Cancers*, *14*(11), 2663. https://doi.org/10.3390/cancers14112663

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

Wu, T., Sultan, L. R., Tian, J., Cary, T. W., & Sehgal, C. M. (2019). Machine learning for diagnostic ultrasound of triple-negative breast cancer. *Breast Cancer Research and Treatment*, *173*(2), 365–373. https://doi.org/10.1007/s10549-018-4984-7

Xu, Y., Wang, Y., Yuan, J., Cheng, Q., Wang, X., & Carson, P. L. (2019). Medical breast ultrasound image segmentation by machine learning. *Ultrasonics*, *91*, 1–9. https://doi.org/10.1016/j.ultras.2018.07.006

Zhang, E., Seiler, S., Chen, M., Lu, W., & Gu, X. (2019). Boundary-aware Semi-supervised Deep Learning for Breast Ultrasound Computer-Aided Diagnosis. *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 947–950. https://doi.org/10.1109/EMBC.2019.8856539

Zhang, Y., Chen, J.-H., Lin, Y., Chan, S., Zhou, J., Chow, D., Chang, P., Kwong, T., Yeh, D.-C., Wang, X., Parajuli, R., Mehta, R. S., Wang, M., & Su, M.-Y. (2021). Prediction of breast cancer molecular subtypes on DCE-MRI using convolutional neural network with transfer learning between two centers. *European Radiology*, *31*(4), 2559–2567. https://doi.org/10.1007/s00330-020-07274-x

Zhang, Z., Li, Y., Wu, W., Chen, H., Cheng, L., & Wang, S. (2021). Tumor detection using deep learning method in automated breast ultrasound. *Biomedical Signal Processing and Control*, *68*, 102677. https://doi.org/10.1016/j.bspc.2021.102677

Zhuang, Z., Yang, Z., Raj, A. N. J., Wei, C., Jin, P., & Zhuang, S. (2021). Breast ultrasound tumor image classification using image decomposition and fusion based on adaptive multi-model spatial feature fusion. *Computer Methods and Programs in Biomedicine*, *208*, 106221. https://doi.org/10.1016/j.cmpb.2021.106221