A Multi-Modal Deep Learning Framework for Breast Cancer Prediction and Stage Identification Using Mammography and Ultrasound Images

Mymona Akter Mim

Bachelor of Science in Software Engineering

DAFFODIL INTERNATIONAL UNIVERSITY

DECLARATION OF THESIS AND COPYRIGHT

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name	:	Mymona Akter Mim
Date of Birth	:	08/01/2002
Title	:	A Multi-Modal Deep Learning Framework for Breast Cancer Prediction and Stage Identification Using Mammography and Ultrasound Images
Academic Session	:	

I declare that this thesis is classified as:

☐	CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997)*
☐	RESTRICTED	(Contains restricted information as specified by the organization where research was done)*
☐	OPEN ACCESS	I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Daffodil International University reserves the following rights:

1. The Thesis is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the thesis for the purpose of research only.
3. The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.

Certified by:

Mymona	27-11-25
(Student's Signature)	(Supervisor's Signature)
221-35-957	Dr. S M Hasan Mahmud
Student ID Date:	Name of Supervisor Date:

NOTE : * If the thesis is CONFIDENTIAL or RESTRICTED, please attach a thesis declaration letter.

THESIS DECLARATION LETTER (OPTIONAL)

Librarian,
Daffodil International University,
Daffodil Smart City,
Ashulia.Dhaka,Bangladesh

Dear Sir,

CLASSIFICATION OF THESIS AS RESTRICTED

Please be informed that the following thesis is classified as RESTRICTED for a period of three (3) years from the date of this letter. The reasons for this classification are as listed below.

Author's Name	Mymona Akter Mim
Thesis Title	A Multi-Modal Deep Learning Framework for Breast Cancer Prediction and Stage Identification Using Mammography and Ultrasound Images
Reasons	(i)
	(ii)
	(iii)

Thank you.

Yours faithfully,

27-11-25

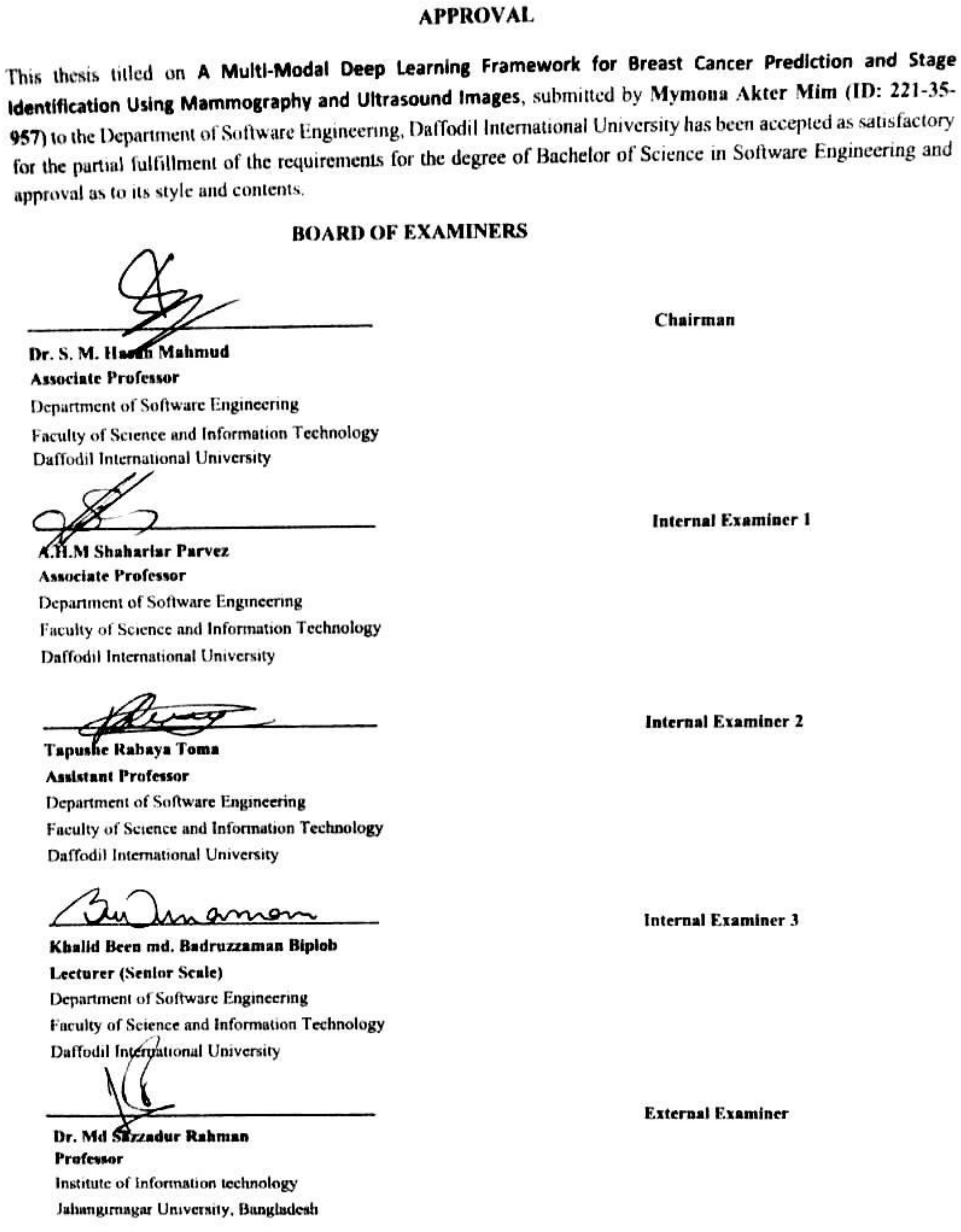
____
(Supervisor's Signature)

Date:

Stamp:

Note: This letter should be written by the supervisor and addressed to the Librarian, *Daffodil International University* with its copy attached to the thesis.

APPROVAL FORM



APPROVAL

This thesis titled on **A Multi-Modal Deep Learning Framework for Breast Cancer Prediction and Stage Identification Using Mammography and Ultrasound Images**, submitted by **Mymona Akter Mim (ID: 221-35-957)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Chairman

Dr. S. M. Hasan Mahmud
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 1

A.H.M Shahariar Parvez
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

Tapushe Rabaya Toma
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 3

Khalid Been md. Badruzzaman Biplob
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

External Examiner

Dr. Md Sazzadur Rahman
Professor
Institute of Information technology
Jahangirnagar University, Bangladesh

SUPERVISOR’s DECLARATION

I/We* hereby declare that I/We* have checked this thesis/project* and in my/our* opinion, this thesis/project* is adequate in terms of scope and quality for the award of the degree of *Bachelor of Science/ Master of Science.

27-11-25

(Supervisor’s Signature)

Full Name : Dr. S M Hasan Mahmud
Position : Associate Professor
Date :

STUDENT’S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.

Mymona

(Student’s Signature)

Full Name : Mymona Akter mim
ID Number : 221-35-957
Date :

A Multi-Modal Deep Learning Framework for Breast Cancer Prediction and Stage Identification Using Mammography and Ultrasound Images

Mymona Akter Mim

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Science/Master of Science

Department of Software Engineering (Major in Data Science)

DAFFODIL INTERNATIONAL UNIVERSITY

OCTOBER 2024

ACKNOWLEDGEMENTS

This thesis is the result of a great deal of assistance from many people. I want to express my sincere gratitude to my supervisor for all of their help with this assignment, including their critical evaluation, excellent intellectual advice, and unwavering support. I also want to express my gratitude to my committee members for their effort and professional insights, which greatly improved our study. I want to express my gratitude to my lab partners and coworkers for creating a cooperative and encouraging atmosphere. Lastly, this work is dedicated to my family, whose unflinching love and sacrifice served as the crucial basis for my academic journey.

DEDICATION

To my family, whose unwavering love, encouragement, and infinite patience served as the foundation of this work. Your faith in me fueled every late night and early morning, making this achievement possibale.

To my esteemed mentors and dedicated colleagues, whose insightful guidance shaped not only the direction of this research but also my growth as a scientist.

And finally, to the data itself, which patiently revealed its secrets through the persistent pursuit of knowledge and discovery.

©Daffodil International University

ABSTRACT

Breast cancer remains one of the most prevalent causes of mortality among women worldwide, highlighting the urgent need for accurate, accessible, and automated diagnostic systems. This study investigates the effectiveness of deep learning models in classifying breast tumors as malignant or normal using two imaging modalities—mammography and ultrasound—sourced from publicly available Kaggle datasets. A combined dataset of 7,091 images was preprocessed through resizing, normalization, and augmentation before training four convolutional neural network architectures: a custom-designed CNN and three transfer learning models, DenseNet121, MobileNetV3, and ShuffleNetV2. The objective was to evaluate model performance, examine the benefits of multimodal imaging, and explore the feasibility of breast cancer stage prediction.

Experimental results demonstrate that deep learning models can reliably classify breast tumors, with MobileNetV3 achieving the highest overall accuracy of 88.7%, followed closely by DenseNet121 at 88.2%. ShuffleNetV2 achieved the highest malignant recall (0.894), indicating stronger sensitivity in identifying cancer cases. These outcomes confirm that multimodal training enhances classification robustness by capturing complementary anatomical information from both imaging modalities. However, the study also reveals that stage prediction is not feasible using current public datasets, as they lack clinical annotations related to tumor staging.

The findings underscore the potential of deep learning for breast cancer screening and highlight the need for stage-labeled datasets and richer clinical metadata. Future research should focus on multimodal diagnostic frameworks that integrate imaging, clinical records, and advanced explainable AI techniques to support more comprehensive and clinically valuable breast cancer detection systems.

TABLE OF CONTENTS

Contents

LIST OF TABLES

©Daffodil International University

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI — Artificial Intelligence
ANN — Artificial Neural Network
AUC — Area Under the Curve
BC — Breast Cancer
BCE — Binary Cross-Entropy
BUSI — Breast Ultrasound Images Dataset
CAM — Class Activation Map
CNN — Convolutional Neural Network
CEM — Contrast-Enhanced Mammography
CT — Computed Tomography
DL — Deep Learning
DNN — Deep Neural Network
DICOM — Digital Imaging and Communications in Medicine
FN — False Negative
FNR — False Negative Rate
FP — False Positive
FPR — False Positive Rate
GPU — Graphics Processing Unit
HI — Histopathological Imaging
IoT — Internet of Things
ML — Machine Learning
MRI — Magnetic Resonance Imaging
NBCC — National Breast Cancer Centre (Lesion Scoring)
PACS — Picture Archiving and Communication System
PPV — Positive Predictive Value
RGB — Red, Green, Blue
ReLU — Rectified Linear Unit
ROC — Receiver Operating Characteristic
SE — Squeeze-and-Excitation
SGD — Stochastic Gradient Descent
TP — True Positive
TNR — True Negative Rate
TPR — True Positive Rate (Sensitivity)
TN — True Negative
US — Ultrasound

CustomCNN — Custom Convolutional Neural Network
DenseNet121 — Densely Connected Convolutional Network 121-Layer Model
MobileNetV3 — Mobile Network Version 3
ShuffleNetV2 — Shuffle Network Version 2

ROI — Region of Interest
FNAC — Fine Needle Aspiration Cytology
CAD — Computer-Aided Diagnosis
FFDM — Full-Field Digital Mammography
TNM — Tumor-Node-Metastasis (Cancer Staging System

LIST OF SYMBOLS

X — Input image or input tensor
W — Convolutional kernel / weight matrix
b — Bias term in convolution or dense layers
F — Output feature map after convolution
σ — Activation function (ReLU or Sigmoid depending on context)
$*$ — Convolution operation
$\mathbf{x}_l$ — Output of the lthl^{th}lth dense layer in DenseNet
$[\mathbf{x}_0,\mathbf{x}_1,\ldots,\mathbf{x}_{l-1}]$— Concatenation of all previous feature maps
$\mathbf{H}_l$ — Composite transformation (BatchNorm + ReLU + Convolution)
z — Logit (input to the sigmoid function)

$\hat{y}$— Predicted probability of the malignant class
$\mathbf{Conv}_{\mathbf{DW}}\mathbf{(X)}$ - Depthwise convolution in MobileNet
$\mathbf{Conv}_{\mathbf{PW}}\mathbf{(X)}$— Pointwise (1×1) convolution in MobileNet
$\mathbf{X}_{\mathbf{shuffle}}$ — Channel-shuffled output in ShuffleNetV2
t — Expansion factor for MobileNetV3 inverted residual block
s — Channel attention scaling factor in squeeze-and-excitation block

CHAPTER 1

INTRODUCTION

1.1 Background

Breast cancer happens to be one among them posing a threat to the lives of all women in the world. Early detection is the most essential part of improving the survival, easing treatment, and causing the inability to be healthy in the long term according to the world statistics of health. Diagnosis of breast cancer Medical imaging forms the core of the detection of the disease and the most popular screening and diagnostic methods are the mammography and ultra sound imaging. Screen of a population is conventionally conducted through the assistance of mammography and additional modality is ultrasound that is especially used with younger women and dense breast tissues. With the creation of machine learning and deep learning, image-based diagnostic systems are increasingly gaining importance through automated systems. These systems can assist the radiologists who are enhancing accuracy of detecting, diagnostic load, and human error. In recent years, numerous countries investigated the idea of the possibility of applying the computer vision algorithms and convolutional neural networks (CNNs) to recognize the tumors in the breast as the malignant ones, or benign. However, the problem of low quality of datasets, as well as low inclusion and clinically annotated metadata, is still to be problematic to model the performance and clinical value. In this regard, the utilization of publicly available data about the mammography and ultrasound images will be an excellent opportunity to comprehend the potential of the AI-driven breast cancer detection, especially in the resource-deprived regions where professional radiologists are not common.

1.2 Problem Statement

Despite the above, major gains have been realized in the development of machine-learning-based systems of identification of breast cancer, though various issues are yet to be addressed. Published datasets are also typically not detailed with clinical annotations and thus the bearing of more advanced diagnostic models are constrained. Majority of studies have only dealt with binary classification- usually distinguishing between normal and malignant cases- and not carried out deeper diagnostic information. Also, most of the currently used models are based on a single imaging source, despite the fact that it is a combination of multiple imaging sources that could be possibly used to improve the accuracy of the diagnoses. Considering these restrictions, it is necessary to approach the analysis of the possibility of machine learning and deep learning models to reliably label the images of the breast cancer as either malignant or normal on the basis of mammography and ultrasound data. More so, given that the stage-labeled information is absent, it can be questioned whether it is possible to predict cancer stage using the existing datasets, or whether further information sources are needed.

1.3 Motivation

The reason as to why this research is to be done is in the fact that there is a growling need to ensure that effective diagnostic support systems whether manual or automated are readily available to identify breast cancer. In most of the developing world such as Bangladesh, diagnostic facilities are minimal and no accessibility to specialized radiologists to interpret results of the radiology tests. The use of machine learning-supported automated systems is able to decrease the differences in the diagnoses and improve the timely diagnosis. It also has the possibility of applying multi-modal diagnostic modelling where mammography image and ultrasound images will be involved because of the open-source data provided in breast cancer image. Further increase in classification accuracy could help to benefit clinical decision-making, especially in cases where resources available to effect a diagnosis are scarce. Prediction of cancer-stages is another factor that motivated the researchers to generalize the study to predicting the stage of cancer, and this has restricted due to the limitation of the datasets, and this resultant ability of research has created a significant gap in research.

1.4 Significance of the Study

It is an important study in terms of academia and practical applications. Scholarly, it adds to the existing literature on the use of artificial intelligence in medical imaging, especially breast cancer detection in multi-modal data. The work conducts a performance analysis of various machine learning and deep learning models that give information on how the models interact with various imaging techniques. In practice, the results can be used to design low-cost and automated screening devices to help healthcare professionals in their early diagnosis. The study will investigate the potential to increase the accuracy of checking tumor through multi-modal learning by addressing the question of the feasibility in using mammography and ultrasound images. Moreover, recording the constraints pertaining to stage prediction assists in the future research orientation where the scholars are advised by the data that they require and where to undertake more extensive research technique to direct more extensive diagnostic systems.

1.5 Research Questions

1. Do machine learning or deep learning models with mammography and ultrasound data treat tumour images of the breast with high accuracy in malignant or normal images?
2. Is the combination of mammography and ultrasound imaging better in improving the classification accuracy to using one modality?
3. Can stage prediction of breast cancer be done based on publicly available imaging datasets? Otherwise, what are the constraints to stage prediction?

1.6 Research Objectives

1. To preprocess and analyze the data of mammography and ultrasound systems on Kaggle available on breast cancer classification.
2. To engineer machine learning and deep learning models that will be able to distinguish between malignant and normal breast tumors
3. To explore the possibilities of predicting stage of cancer and determine the limitations present in the dataset that limits prediction of stage
4. To assess the performance between models that are trained on single imaging modalities and mixed modalities
5. To make suggestions as to further studies that will include stage-labeled datasets and superior diagnostic modelling.

1.7 Limitations and Scope of the Research

paper is limited to classification of breast cancer using mammography and ultrasound datasets that are freely available to the public. Data preprocessing, model construction, performance analysis, and modality analysis are all in scope. Nevertheless, the results of the study are affected by a number of constraints: Absence of stage-labeled data: The data are missing information about the clinical staging (Stage I-IV), and it is impossible to predict stage using the existing structure of the dataset. Image quality and resolution difference: For different sources of data, variability may exist as a result of different data sources. Small scale data: Kaggle public data sets are typically smaller than a clinical dataset which can affect generalizability. No patient metadata: Tumor size, patient age and lymph node involvement i.e. staging relevant information are unavailable. Modality imbalance: The modalities of some modalities (e.g. ultrasound) might not have as many malignancies compared to mammography, and the imbalance of the classes of the modality should be consciously managed. The proposed study concentrates on binary classification alone, and stage prediction is presented as an unmet objective because of the limitations with data.

1.8 Thesis Organization

This thesis is organized into five main chapters. Chapter 1 presents the research background and outlines the problem statement, followed by the objectives, significance, scope, and limitations of the study.

Chapter 2 Reviews existing studies on breast cancer detection using machine learning, deep learning, mammography, and ultrasound imaging. Identifies key research gaps and positions this study within current scientific work.

Chapter 3: Methodology Describes the datasets used, the preprocessing techniques applied, feature extraction pipelines, and the machine learning and deep learning models implemented. Also explains the training, validation, and evaluation strategies.

Chapter 4: Results and Discussion Presents the experimental results, compares model performance across modalities, and interprets findings in relation to the research questions.

Chapter 5:Conclusion and Recommendations Summarizes key findings, outlines study limitations, and provides recommendations for future work, particularly regarding cancer-stage prediction and multi-modal medical imaging research.

©Daffodil International University

CHAPTER 2

LITERATURE REVIEW

2.1 Related works

By presenting a multimodal deep learning system to integrate mammography with ultrasound images, Chen et al. (2025) studied the drawbacks of single-modality breast cancer prediction models. They took 790 samples of patients, 2,235 mammography-based and 1,348 ultrasound images, to test six models in deep learning to determine which architecture would yield the best results on fusion. The AUC, sensitivity, specificity, precision, and accuracy were used to evaluate the model performance, and the results indicated that the use of multimodal model was significantly more important than single-modality models, with 96.41% specificity, 93.78% accuracy, 83.66% precision, and an AUC of 0.968, as compared to the single-modality models. The model interpretability was also confirmed by the visualization of heatmap. On the whole, their study indicated that a combination of two types of imaging increases the diagnostic reliability and facilitates a better clinical decision-making (Chen et al., 2025).

Zhang et al. (2025) provided a deep learning ensemble classifier to detect breast cancer which was developed to work well even with small data. They combined AlexNet, ResNet, and MobileNetV2 and applied residual learning, depthwise separable convolutions, and inverted bottleneck architectures to achieve better performance in terms of speed and optimisation. Laplacian of Gaussian filtering and modified high-boosting further improved the image quality and augmented the feature-extraction power, and the ensemble model was highly effective on a variety of datasets: 99.17% abnormality and 97.75% malignancy accuracy on mini-DDSM; 96.92% on BUSI; and 97.50% on the BUS2 ultrasound dataset. These findings indicate the versatility and high generalization capacity of the model on the multimodal breast cancer imaging (Zhang et al., 2025).

Jabeen et al. (2022) introduced a deep learning design classification system that aims to enhance the quality of detection of breast cancer via ultrasound bruise of the skin. They use a procedure that has 5 primary steps, which include augmenting the dataset to magnify CNN education, transforming a pre-trained DarkNet-53 backbone to the augmented category structure, transfer learning to extract features using the global average pooling layer, choosing the most discriminatory features by reformed differential evolution (RDE) and reformed gray wolf (RGW) optimization algorithms, and fusing these most effective features through a probability-based serial fusion strategy and then classification with machine learning algorithms. With augmented data of BUSI, the framework obtained the highest accuracy of 99.1% and outperforms some recent methods. The article emphasizes the efficiency of using transfer learning, optimisation-based feature selection and feature fusion to achieve strong ultrasound-based breast cancer classification (Jabeen et al., 2022).

Alom et al. (2025) proposed Deep Neural Breast Cancer Detection (DNBCD) model that is an explainable deep learning architecture aimed to enhance the accuracy and interpretability of breast cancer using histopathological and ultrasound images. Their model is based on a Densenet121 backbone, which is provided with extra CNN layers, GlobalAveragePooling2D, Dense and Dropout, and a learning of transfer is applied to enhance the feature representation. The authors used the following steps of preprocessing such as normalization, resizing, and data augmentation and managed the issue of class imbalance with class weighting. One of the contributions that can be noted about the study is the incorporation of Grad-CAM that allows one to see model decisions, and enhance the interpretability of the model to facilitate its beneficial use in clinical settings. On two datasets, BreakHis-400x and BUSI, the model that is being tested was tested (DNBCD model) with an accuracy of 93.97 and 89.87 respectively. Such findings allow concluding that the model is more accurate than some of the currently available approaches and provides a higher level of transparency, which suggests its practical usefulness in the context of medical diagnostic (Alom et al., 2025).

To address the need to overcome the resource constraint in rural and resource-limited areas that tend to have a difficult time with mammography interpretation, Dwivedi et al. (2025) suggested a CNN-based breast cancer detection system, which utilizes ultrasound images to offer dependable diagnostic support. In their paper, they were interested in creating an automated classification tool that can differentiate normal and malignant cancer and the system should become a part of Healthcare-IoT networks to provide quick and convenient diagnosis. The authors intended to boost the early detection and seek patient outcomes by leveraging the safety and low cost of ultrasound imaging with the use of deep learning. Accuracy, precision, recall, and F1 score were used to determine model performance, and it showed promising performance according to clinical and real-world healthcare services. This strategy shows the possibility of CNN-based ultrasound analysis as a convenient diagnostic instrument to underserved areas (Dwivedi et al., 2025).

Wei et al. (2025) proposed a multimodal deep learning framework that was intended to discern benign and malignant breast masses based on the radial and anti-radial ultrasound image and radiological reports. They have special image and text encoders in the architecture to identify the modality specific features and these are then converted to the same by a transformation layer after which they are jointly classified. The model reported an AUC of 85% and it showed definite improvements in comparison to unimodal baselines as it presented an increase of 6 percent and 8 percent in Youden index. It also had superior performance on zero-shot predictions on popular foundation models like CLIP and MedCLIP. The multimodal system was diagnosed with more accuracy compared to physician assessed rating, which highlights its practical applicability. The value of presenting image and textual data to create more efficient and understandable AI-assisted diagnostics tools comes out in this work (Wei et al., 2025).

Taylor et al. (2023) utilized a comparative study that compared contrast-enhanced mammography (CEM) to MRI in terms of identifying known and additional malignant lesions during preoperative mammography staging of breast cancer. The authors examined the results of using both imaging modalities on 59 women with invasive breast cancer, and the level of assessment of independent radiologists by incorporating the National Breast Cancer Centre (NBCC) scoring. Out of 68 malignant lesions that had been initially identified by mammography and ultrasound and later confirmed by a biopsy, 66 lesions (97%) of these were identified with MRI and 67 lesions (99%), with CEM. Nevertheless, discrepancies arose with regard to locating other occult lesions as 41 new sites were found in 29 patients. Only 15 percent of these extra lesions were found by CEM, 56 percent by MRI and 29 percent by both. Notably, CEM was sensitive to 1 of 6 malignant additional lesions, whilst MRI was sensitive to all 6 as shown by positive predictive values of 8% and 23% respectively. The authors have found that overall sensitivity of both modalities was high in the detection of known lesions, but MRI was more effective as an implicator of additional malignant foci, not detected by conventional imaging, thus its greater usefulness in the accurate staging of breast cancer (Taylor et al., 2023).

Chelloug et al. (2025) have tested the transfer learning in detecting breast cancer based on the ultrasound image, which efficiently resolves the risks of manual diagnosis and high cost in training the deep learning models without transfer learning. The authors have tested a number of ready-made architectures, such as modified InceptionV3, GoogLeNet, ShuffleNet, AlexNet, VGG-16, and SqueezeNet, and additionally, they came up with a deep neural network, which incorporates the improved features of the modified InceptionV3 architecture. Their experiments revealed that the modified InceptionV3 performs the best, achieving an accuracy of 99.10, recall of 98.90, precision of 99.00, and F1-score of 98.80, which is better than the rest of the models that were tested. The article pinpoints the potential of the transfer learning to reach high diagnostic accuracy and validates the adapted InceptionV3 design as a more effective solution to the classification of breast cancer through ultrasound (Chelloug et al., 2025).

Din et al. (2022) carried out an exhaustive review of the research on detecting breast cancer, reviewing the application of machine learning and deep learning in various imaging modalities, such as mammography, histopathology, ultrasound, PET/CT, MRI, and thermography. The authors pointed out the weaknesses of conventional diagnostic processes that require the involvement of radiologists and pathologists and are typified by high cost, time-intensive and low human error and error rates. Their review revealed that deep learning models have proven themselves highly successful in their automatic prediction and detection of complex features in high-dimensional medical data, and in certain cases, they even beat traditional machine learning models. Nevertheless, the Review also focused on the fact that even though most AI systems may be offering promising retrospective results, the majority of them do not have enough outside validation to be used in the real clinical setting. Din et al. described widely utilized datasets, reviewed the significant conclusions of previous research, and reported that the three critical issues most of the challenges in this context such as scarcity of data, generalizability, and low interpretability had to be resolved to enable deep learning tools to be used in clinical environments. Their efforts contribute to upcoming studies with a valuable premise through the synthesis of the advancements in the various imaging modalities and the elucidation of the current restrictions in the radiation of an AI-based system of breast cancer detection (Din et al., 2022).

The study performed by Sood et al. (2019) was a systematic review and a meta-analysis that aimed to determine the efficacy of ultrasound as a primary diagnostic instrument used in the detection of breast cancer at an early stage, mostly in the context of environments where mammography is either not available or not feasible. The authors filtered 526 studies that were published in 2000-2018, and finally identified 26 eligible articles that performed handheld, portable ultrasound systems as independent screening modalities. Their combined analysis showed that the sensitivity was 80.1 and the specificity was 88.4 with much better performance in low and middle-income countries with sensitivity rate of 89.2 and specificity of 99.1. Subgroup analysis revealed no significant differences in diagnostic accuracy by age, risk factor, signs and symptoms, type of study, or environment although heterogeneity was observed. The research made a conclusion that ultrasound is a highly promising diagnosis tool and that it is a major diagnostic tool in limited resources settings and when mammography is either not available or when it is not feasible (Sood et al., 2019) due to its potentials.

2.2 Research gaps and proposed solutions

After completing the literature review of related work, their findings, results, identifiable gaps and Proposed solutions are shown at Table 1

Table 1: Gaps and Proposed solution of related works

Research Works	Models used	Model Evaulation	Identifiable research gaps	Proposed Solution
Chen et al. (2025)	Six DL Classifier multimodal fusion	AU C 0.968, Acc 93.78% Spec 96.41% Prec 83.66%	Sensitivity worse with multimodal model; multimodal needs large data; less readable	Combine feature-level fusion + Grad-CAM to interpret better; compare multimodal to single-modality in my data

Zhang et al. (2025)	Ensemble AlexNet, ResNet, MobileNetV2; LoG; high-boost filtering	99.17% abnormality, 97.75% malignancy (mini-DDSM); 96.92% abnormality (BUSI)	Ensemble can be computationally expensive; cannot necessarily generalize to all modalities	Use lightweight TL models; performance at both mammography and ultrasound
Jabeen et al. (2022)	DarkNet-53 + RDE + RGW + ML classifiers	99.1% accuracy (BUSI)	Computationally expensive optimization algorithms; ultrasound-only	TL and feature fusion followed by feature fusion simplification; has multimodal comparison.
Alom et al. (2025)	DenseNet121 + custom CNN + Grad-CAM	93.97% (BreakHis), 89.87% (BUSI)	Worse performance on ultrasound;	separated modality evaluation
Dwivedi et al. (2025)	Custom CNN for ultrasound	Measured on Accuracy, Precision, Recall, F1 Measures of Multimodality	Measures of Multimodality measures not presented; no multimodal evaluation done.	Use robust multimodal dataset; perform precise metric-based evaluation
Wei et al. (2025)	Multimodal image encoder + text encoder + fusion model	AUC 85%; improved Youden index	Needs radiology reports; worse AUC than image-only SOTA	Have image-only multimodal fusion (mammography + ultrasound) to avoid text dependency.

Taylor et al. (2023)	Radiologist-based CEM vs MRI comparison	MRI 97% sensitivity; CEM 99% sensitivity; MRI detects more additional lesions	Not ML-based; focuses only on lesion staging	Use insights on modality strengths to improve DL multimodal classification
Chelloug et al. (2025)	Modified InceptionV3, GoogLeNet, ShuffleNet, AlexNet, VGG16, SqueezeNet	Best model: Accuracy 99.10%, Recall 98.90%, Precision 99.00%	Uses only ultrasound; possible dataset overfitting	Evaluate multiple TL models across both imaging modalities
Din et al. (2022)	Review of ML, DL, DRL methods	Synthesized cross-modal performance	None of them were experimented on; datasets are scarce, not well-generalized, etc.	Use clean dataset, multimodal inputs, XAI, and balanced training methods
Sood et al. (2019)	No ML model; systematic review of ultrasound diagnostic accuracy	Sensitivity 80.1%, Specificity 88.4%; LMIC: Sensitivity 89.2%	Ultrasound alone insufficient for high-confidence diagnosis	Combine ultrasound and mammography using DL for higher diagnostic power

This table provides a comparative overview of existing breast cancer detection studies, highlighting the models used, their performance, limitations, and research gaps

CHAPTER 3

METHODOLOGY

This chapter provides the methodology of identifying breast cancer and exploring the future diagnosis of the disease based on multi-modal medical pictures employed in this work. The technique entails dataset acquisition, preprocessing, model design and selection, training methods, and evaluation measures and implementation specifics. The process is summarized in figure 3(a).

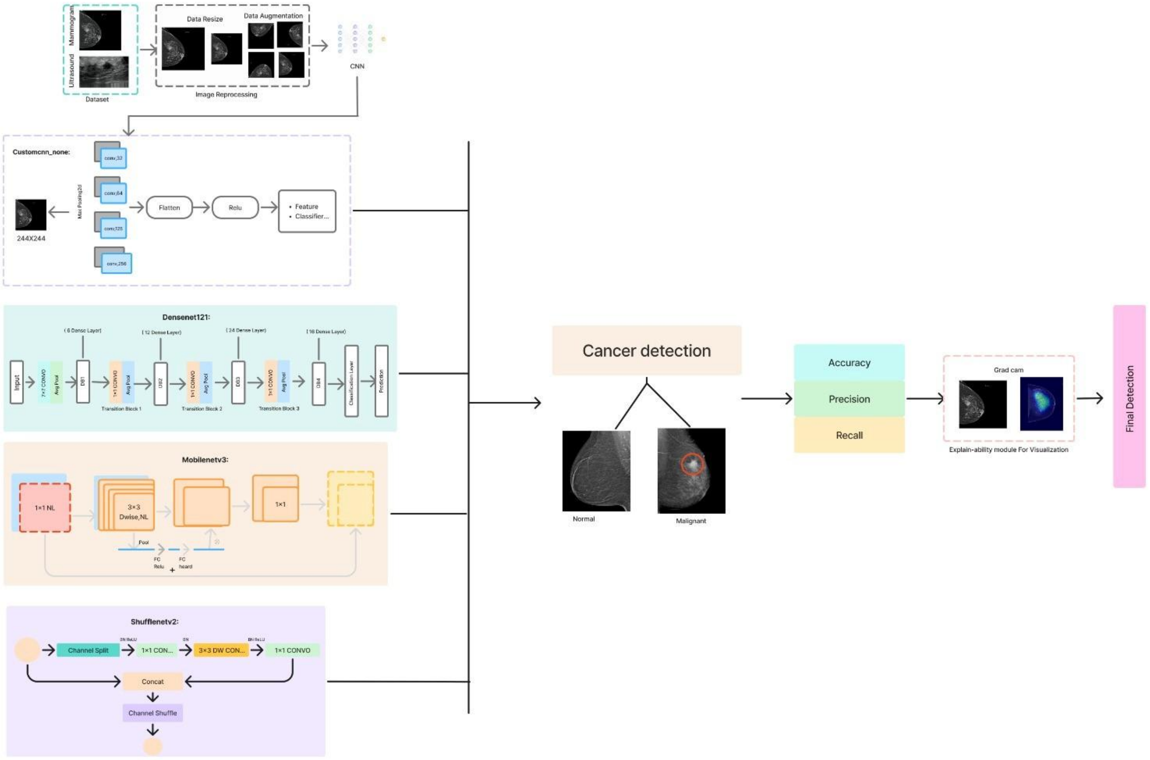


Fig 3(a): Methodology Architecture

3.1 Data Collection

Publicly available repositories of breast-cancer images on Kaggle were used as datasets in this study. Two different imaging modalities were incorporated, i.e., mammography and ultrasound. The mammography data were set on 3,008 images (1,549 normal and 1,469 malignant). The ultrasound data were set on 4,083 images (1,991 normal and 1,892 malignant). As both datasets consisted of images only and did not have any tabular or clinical metadata, the classification task only depended on the visual patterns in the images. The datasets were left together as a single multimodal set to achieve a unified experimental setting and comparative learning across modalities and modulate the models. The images had binary class labels (normal or malignant) allowing the models to develop cross-modal visual information applied to detecting breast tumors. Combining both mammography and ultrasound images also allowed the deep learning models to extrapolate to images with various features, augmenting the strength of the ultimate classification system.

3.2 Data Preprocessing

One of the most important steps in the processing of the mammography and the ultrasound data in deep learning was preprocessing. Given that both datasets were only image files with no tabular data attached to them, there was need to standardize the input, such that the four models would learn uniformly with such an amalgamation. The methodology diagram indicates that the preprocessing pipeline consisted of several steps, which are cleaning of images, resizing, normalization, augmentation, and merging of datasets. Such measures helped to make sure that these imaging modalities did not play any lesser part in the learning process, and that the models got high quality images of uniform processing quality.

3.2.1 Image Cleaning and Formattin

The preprocessing stage was initiated by loading all the images in the two datasets and changing them to a uniform format. The initial files were in mixed formats, i.e., the JPEG and PNG formats, and color channels were different according to the source. To gain consistency, all images were placed into the format needed by the neural networks, usually in the form of RGB images, with some grayscale images being kept depending on compatibility with the model. In this step, images which were illegible, corrupted or incomplete were filtered out to avoid deficits in training. This cleaning made sure that only clear and valid images were passed through in the pipeline. Since mammography and ultrasound display different images, mammography is grayscale with a high contrast, and ultrasound is characterized by noise speckles with low contrast, this procedure did not violate the integrity of the visual information and prepares the images to be ingested by the model.

3.2.2 Image Resizing

After cleaning and formatting these images the second thing was to resize them to a standard resolution. All images were downsampled to 224 x 224 which fits each of the chosen architectures, such as DenseNet121, MobileNetV3, and ShuffleNetV2, which are usually ImageNet-pretrained and require standardized input. Resizing enabled the same size of tensors throughout the dataset, which enabled effective batch processing and prevented errors of mismatched dimensions when training a model. Growth was done to avoid compromising the visual definition of tumor boundaries, masses or micro calcifications that were the crucial features in the anatomy. The consistent image resolution also enabled the models to acquire patterns at equivalent spatial resolution, which enhanced the performance of the convolutional operations in extracting the features.

3.2.3 Normalization

Normalization was done to regulate the numerical measure of pixel values. Medical images are normally in the range of 0 to 255, whereas deep learning models can make use of inputs scaled in a less unstable range. The values of all images were normalized (divided by 255) to bring the values to 0-1. This operation enhanced the stability of numbers in the gradient descent process since it ensured that the gradients were not overly large to destabilize learning. In the case of transfer learning architectures, like DenseNet121 and MobileNetV3, normalization was again chosen to support pretrained weights, which use normalized input. Normalization made training more efficient and gave the models smoother loss curves, allowing models to converge more reliably.

3.2.4 Data Augmentation

Data augmentation was used to increase generalization and decrease overfitting, and because the classes of normal and malignant cases were uneven, the method was used to a large extent. The process increased the size of the data artificially through the generation of transformed versions of the original images. Many changes were implemented and these included rotations, horizontal, vertical flips, zoom effects and positional changes that mimic changes in the real world that transpire in the course of imaging. Further brightness and contrast adjustments were introduced to overcome changes in the conditions of ultrasound and mammography acquisition. Ultrasound images are usually affected by speckle noise, light inconsistency and tissue density variation, and the mammography images may vary in terms of exposure and breast compression. The data augmentation process made sure that the models would acquire invariant features that do not change in response to such alterations. Augmentation provided the classifiers with a large variety of potential distortions, as well as discouraged their ability to learn the training set by heart.

3.2.5 Dataset Merging

Once the cleaning, resizing, normalization and augmentation of every dataset was done, the mammography and ultrasound datasets were combined into one united dataset. This measure allowed the model to learn both sets of imaging modalities at the same time, as opposed to considering them as independent datasets. The combination of the mammography and ultrasound enabled the models to take in data regarding the dense breast tissue patterns, as well as the soft-tissue contrast differences, extending their traces of detecting malignancies. The combined data also contributed to the possibility of reducing the bias of modality due to exposing the models to a wide range of visual characteristics. After combining, the dataset was segmented into training, validation, and test sets and balanced the classes carefully to have a fair evaluation.

3.3 Model Description

Four different deep learning models, including Customnet, DenseNet121, MobileNetV3, and ShuffleNetV2, were used in this research to classify merged mammography and ultrasound images into normal and malignant. All these designs were carefully chosen to span a wide range of design philosophies, such as basic bare-bones network, down to the very optimized state-of-the-art design lightweight architecture. It became possible to assess the effect of network depth, connectivity, architecture complexity and efficiency of its parameters on breast cancer detection performance by evaluating these models jointly.

3.3.1 Custom_net CNN

Customnet is a simple convolutional neural network that was created in this study to put up a baseline against more advanced architectures. It consists of a series of convolutional and the ReLU activation and max-pooling layers in that order. These layers turn the input image into a series of feature maps of local edges, intensity gradients, contours and simple textural structure (Grant, 2015). Customnet uses binary representation of breast images by flattening them and feeding them into dense layers. This simplicity in model introduces it as one of the best references because its performance can be connected to the basic concepts of learning of standard CNNs irrespective of using pretrained weights or secondary architectural blocks in the model. Applying Customnet to detect breast cancer allows it to be stated that the degree to which predictive power is attained as a direct result of the visual patterns in the images and the level of improvement that can be achieved by more advanced or specialized architectures. Its simple architecture also guarantees the interpretability it is easy to determine the influence of higher architectural sophistication on accuracy, generalization and feature extraction.

Model equation:

$F = \sigma(W * X + b)$.................. (1)

where:

- W = learnable kernels
- X = input feature map
- b = bias
- σ = ReLU activation

The final classifier uses a sigmoid unit:

$$\hat{y} = \frac{1}{1 + e^{-z}}$$

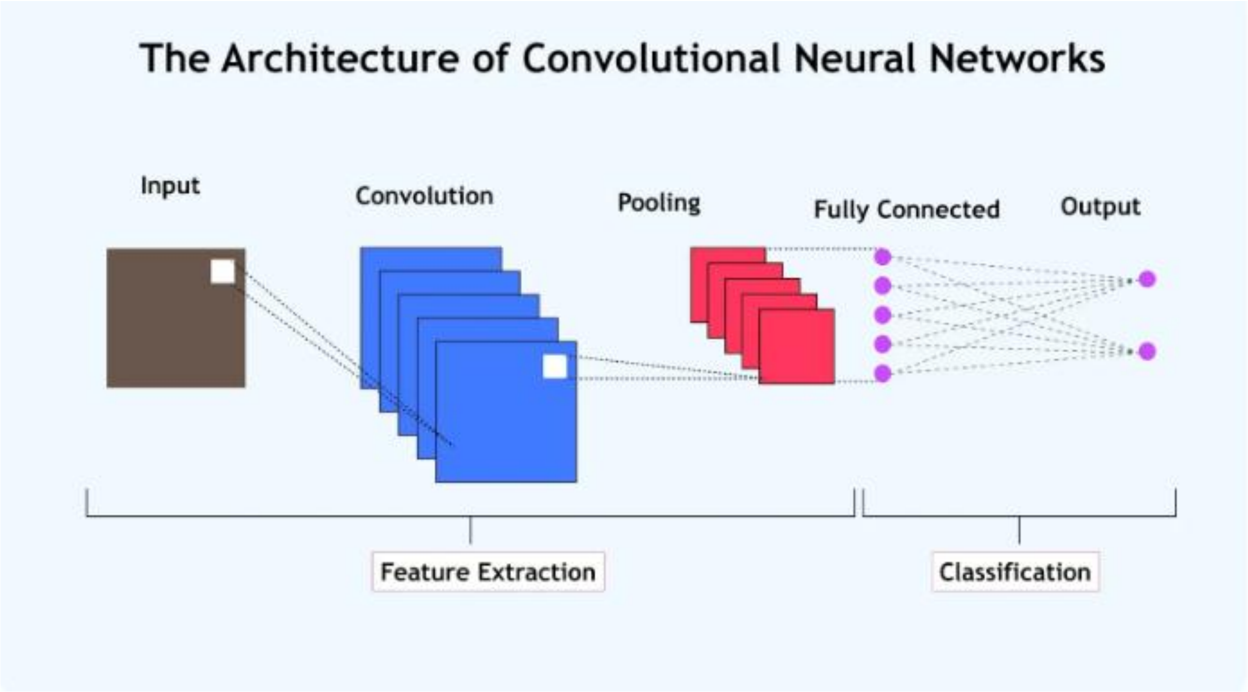


Fig 3.3.1: CNN Architecture

3.3.2 DenseNet121

DenseNet121 is a convolutional network that is characterized by the dense connectivity property, that is, each layer is connected to all the layers before it. This design facilitates gradient flow greatly, enables features to flow more efficiently through the network, and eliminates redundancy by promoting re-use of features. DenseNet121 consists of several dense blocks with transition layers between them that summarise and normalise the features learned. This architecture has proven to be very effective in medical imaging processes like breast cancer detection as it records fine-detailed information and intricate tissue structures that could have been missed under basic models. Images in mammography usually have small abnormalities such as micro-calcium, small masses and architectural distortions and ultrasound images include soft-tissue textures caused by the speckle noise. The capability of denseNet121 to combine low-level and high-level features is what makes it particularly appropriate in detecting these variations. It can also be further improved through transfer learning where the model is allowed to begin training with the pretrained knowledge acquired using large-scale datasets and then fine-tune on the particular breast cancer dataset in this study. Consequently, DenseNet121 is one of the best and most dependable architectures in this comparison. The dense connectivity mechanism used by DenseNet121 is given by:

$x_l = H_l([x_0, x_1, x_2, \ldots, x_{l-1}])$ ……… (2)

where the concatenation of all previous feature maps serves as input to the current layer

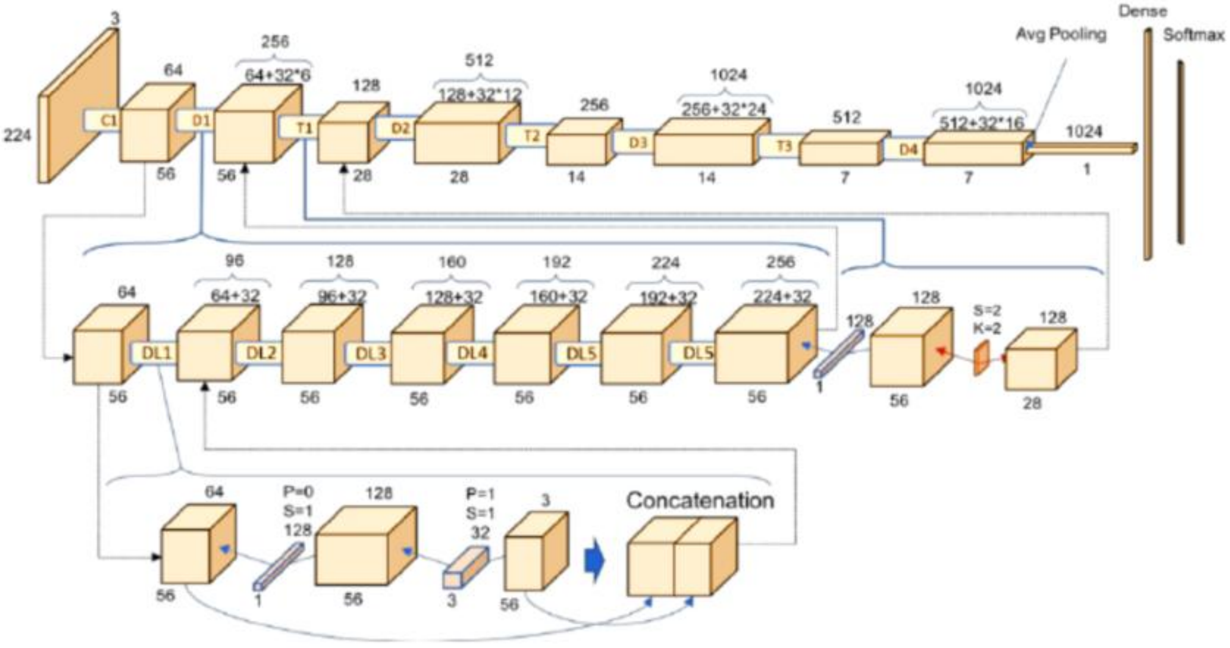


Fig 3.3.2: DenseNet121 Architecture

3.3.3 MobileNetV3

MobileNetV3 is a deep learning model that is efficient-oriented and aimed at delivering high classification accuracy at a minimum cost in terms of computation. It balances this by the use of inverted residual blocks, depthwise separable convolutions, nonlinear activation functions like h-swish along with squeeze-and-excitation attention modules. These elements are combined to extract significant features and the number of parameters and floating-point operations is extremely small. The inverted residual block expands feature space, implements depth wise convolution to learn the spatial pattern of a feature and directly reduces the representation into a light format. Squeezing and excitation method assigns weights of importance on channels channel by channel, which allow MobileNetV3 to prioritize channel diagnostically significant features, including mass borders and localized fluctuations in texture typically found in breast cancer images. This architecture has been specifically useful to ultrasound images in which one can only see patterns that are weak and occasionally covered by noise. MobileNetV3 was found to be the most computationally efficient, which enabled it to be extremely suitable in training on Kaggle GPUs, since it provided fast training and averaging convergence with competitive classification. The convolution in MobileNetV3, which consist of depth-wise separable convolution, is designed as follows:

$\mathrm{Conv}_{DW}(X) + \mathrm{Conv}_{PW}(X)$

SE module attention score:

$s = \sigma\left(w_2 \cdot \delta(w_1 \cdot z)\right)$...........(3)

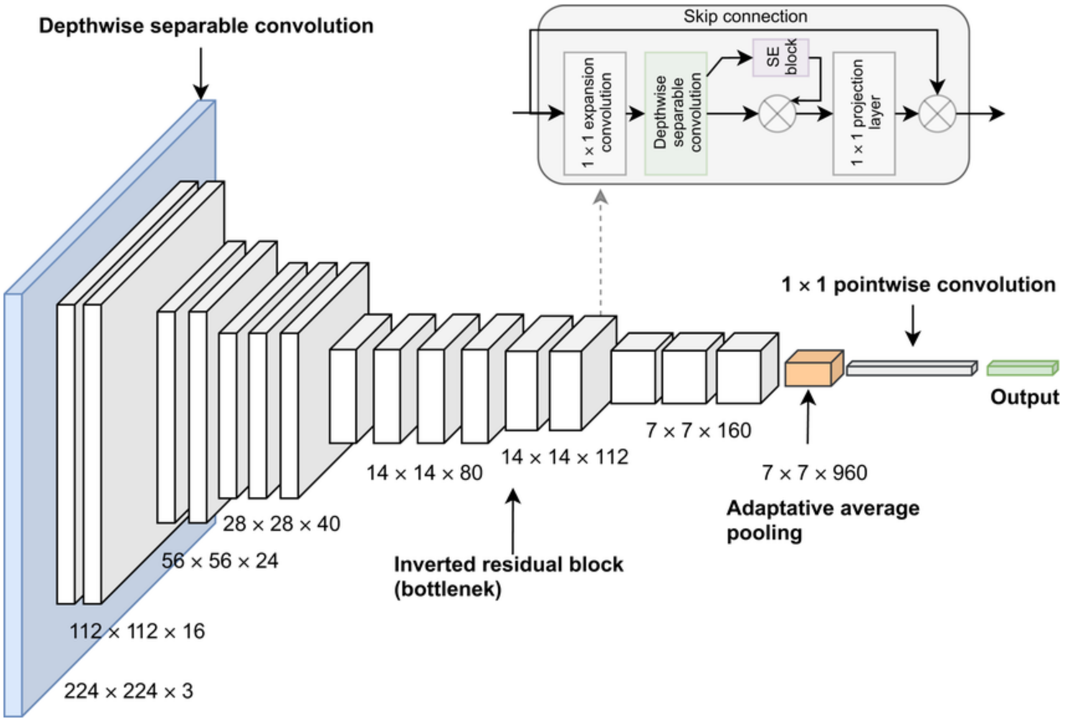


Fig 3.3.3: MobileNetV3 Architecture

3.3.4 ShuffleNetV2

ShuffleNet V2 is very lightweight and convolutional network that has been optimized in terms of both speed and efficiency and hence it is suitable in applications where computing power is a constraint. It is based on the idea of channel splitting, channel shuffling and depthwise convolution to improve information flow throughout the network without additional computational cost. The ShuffleNetV2 architecture splits input channels into disconnected branches which jointly experience various transformations following which they are combined together, it provides a rich feature representation of information. Channel shuffle operation enables the mixing of the same information between spatial and across feature channel, which enhance the expression capacity of the network. Through this architecture ShuffleNetV2 will be able to maintain competitive accuracy despite having a low number of parameters. ShuffleNetV2 could provide the opportunity of real time classification in a fast manner in the imaging processes of breast cancer, particularly in ultrasound, where it is necessary to provide a diagnostic result within a very brief period of time. Its portability further indicates that it can be used in the future in point-of-care or portable diagnostic devices particularly in low resource clinics. It then switches channels in order to communicate: $X_{\text{shuffle}}=\text{reshape}(\text{transpose}(\text{reshape}(X)))$ ………… (4)

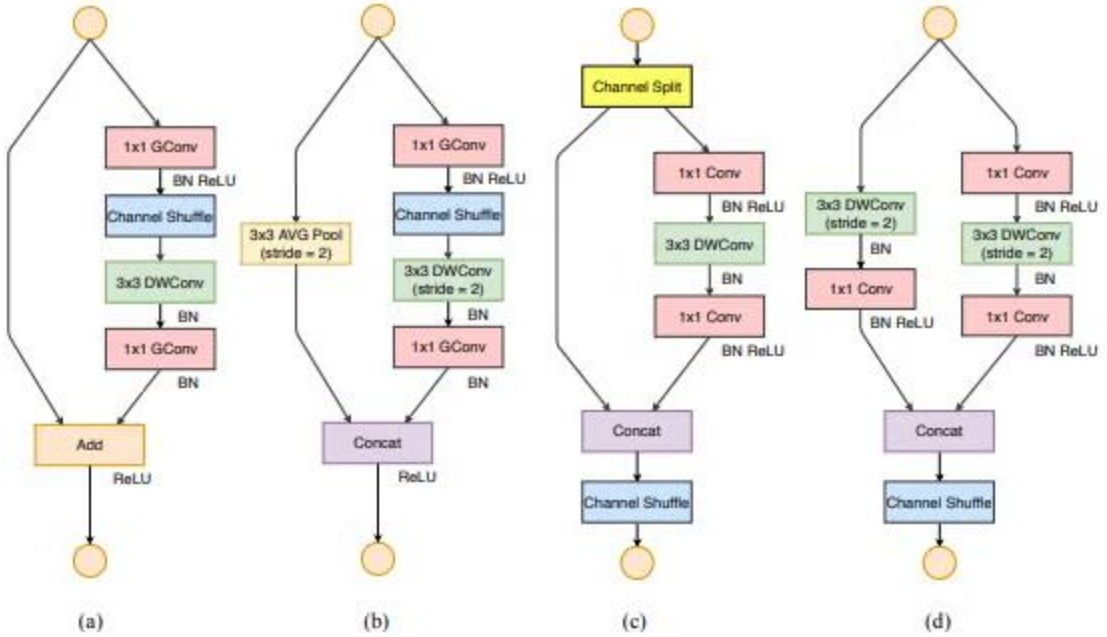


Fig 3.3.4: ShuffleNetV2 Architecture

3.4 Experimental Setup

Experiments were all done in Kaggle Notebook environment that is free to use NVIDIA Tesla GPUs that are applicable in the training of deep learning models. To provide the view of the unbiased model evaluation, the dataset was divided into training (70%), validation (15%), and testing (15%) layers.

Hardware Setup

- Kaggle GPU (NVIDIA Tesla P100 or T4 depending on session)
- 16 GB RAM
- 2-CPU virtual machine

Software Setup

- Python 3.20
- TensorFlow / Keras

- OpenCV for image preprocessing
- Scikit-learn for metrics
- Matplotlib for visualization
- Grad-CAM for explainability outputs

After training, grad-CAM was used to visualize the discriminative regions that had effects on model decisions, which presents interpretability as per the methodology diagram.

3.5 Model Metrics Evaluations

The quality of the four deep learning models to differentiate an image of a breast cancer as either normal or the presence of cancerous cells was evaluated through various standard evaluation measures. These measures offer statistical quantification of the various elements of classification behavior and are useful in determining the effectiveness of the models in discriminating between the two categories.

Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$ ……….. (5)

TPR (True Positive Rate) = $\frac{TP}{TP+FN}$ ………. (6)

TNR (True Negative Rate) =$\frac{TN}{TN+FP}$ …………… (7)

FPR (False Positive Rate) =$\frac{FP}{FP+TN}$ ………… (8)

FNR (False Negative Rate) =$\frac{FN}{FN+TP}$ ………. (9)

Precision = $\frac{TP}{TP+FP}$ …………….. (10)

F1-score = $\frac{2 * Precision * Recall}{Precision+Recall}$ …… (11)

Accuracy gives a general estimate on the number of times the model correctly classifies the images. In medical diagnosis, however, correctness is not enough in that it can conceal an imbalance in the detection of malignancy cases. Precision provides a hint on the reliability of the model to predict malignant images since it is seen to measure the percentage of images that are classified as positive and which are really malignant. The high precision means that the model has fewer false alarms production which is significant in minimizing unjustified clinical tests. True positive rate repeats the notion of sensitivity which concentrates on capturing malignant instances. The true negative rate or specificity on the other hand measures how well the model detects normal cases so that healthy patients are not falsely identified as having.

CHAPTER 4

RESULTS

The chapter contains experimental findings of training and evaluation of four variants of the Convolutional Neural Network (CNN) architecture based on solving the two-class problem in the field of malignant vs. normal samples. The models include three pre-trained models that were trained under transfer learning: DenseNet-121, MobileNetV3, and ShuffleNetV2 and a single specific architecture (CustomCNN). Each model was evaluated using the training dynamics, overall accuracy, and class-specific measures that get derived through the confusion matrix on the test set.

4.1. Training Convergence and Dynamics.

The learning process of each of the four models was assessed by testing the Training Loss, Validation Loss, Training Accuracy, and Validation Accuracy throughout the process of their respective training periods

4.1.1 CustomCNN Training History

The CustomCNN model was trained on more than 40 epochs

Loss History (Figure 4.1a): The Training Loss was declining in a continuous and regression manner and ranged around \$0.36\$. However, the Validation Loss was highly random, and the fluctuations were quite significant across the epochs, which implies that it was sensitive to the composition of the validation set. The validation loss stabilized near the training loss indicating controlled bias and the general behavior was down even though it was not stable.

Accuracy History (Figure 4.1b): The training accuracy was approximately 81.0. The final state of the model did not exhibit serious overfitting as was exemplified by the validation accuracy which was comparable to the volatility of the loss curve yet terminated at similar accuracies as in training

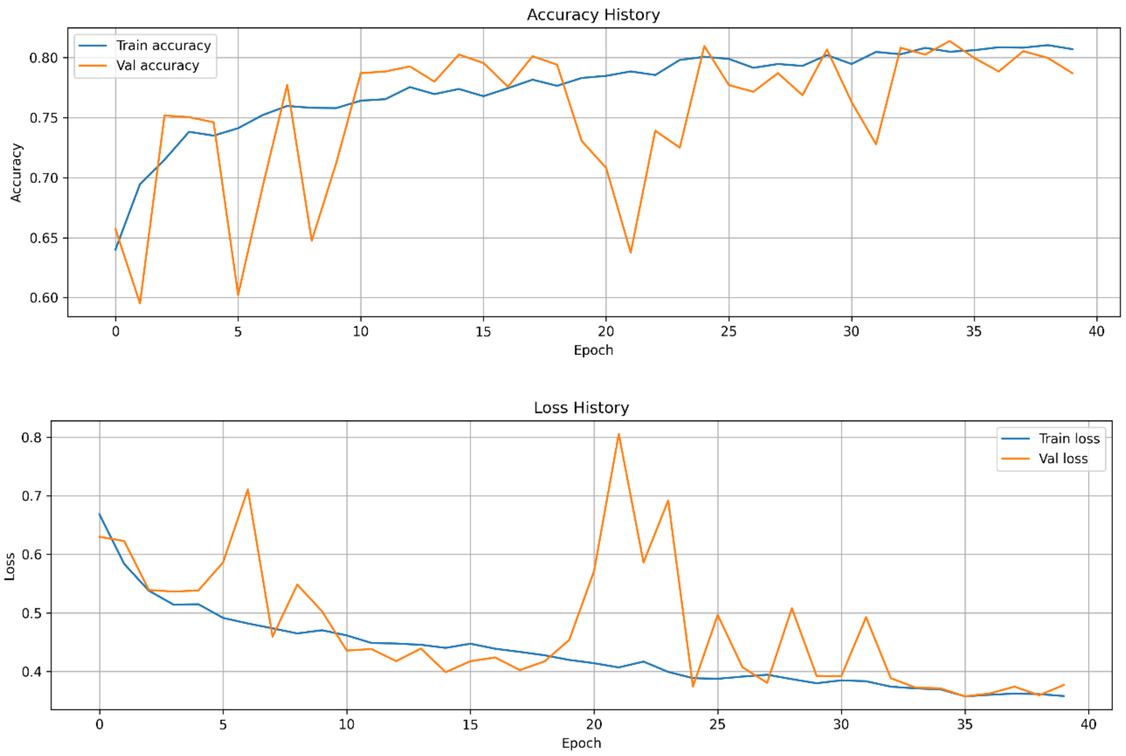


Figure 4.1.1: CNN Training Metrics Visualization of Accuracy and Loss

4.1.2 DenseNet121 Training History

The DenseNet-121 model was trained on fifteen epochs. Loss and Accuracy History (Figures 4.2a, 4.2b): This model provided a typical case of major overfitting. Training Accuracy jumped over 97 percent whereas Training Loss went down within seconds to under 0.1. Conversely, the Validation Accuracy began to decrease since the highest point of $86.5\%% at approximately Epoch 7 and the Validation Loss began to increase at a high rate after Epoch 4. This deviation indicates that the model began to store training specific information after the Epoch 4 and that premature termination was required in order to achieve maximum generalization.

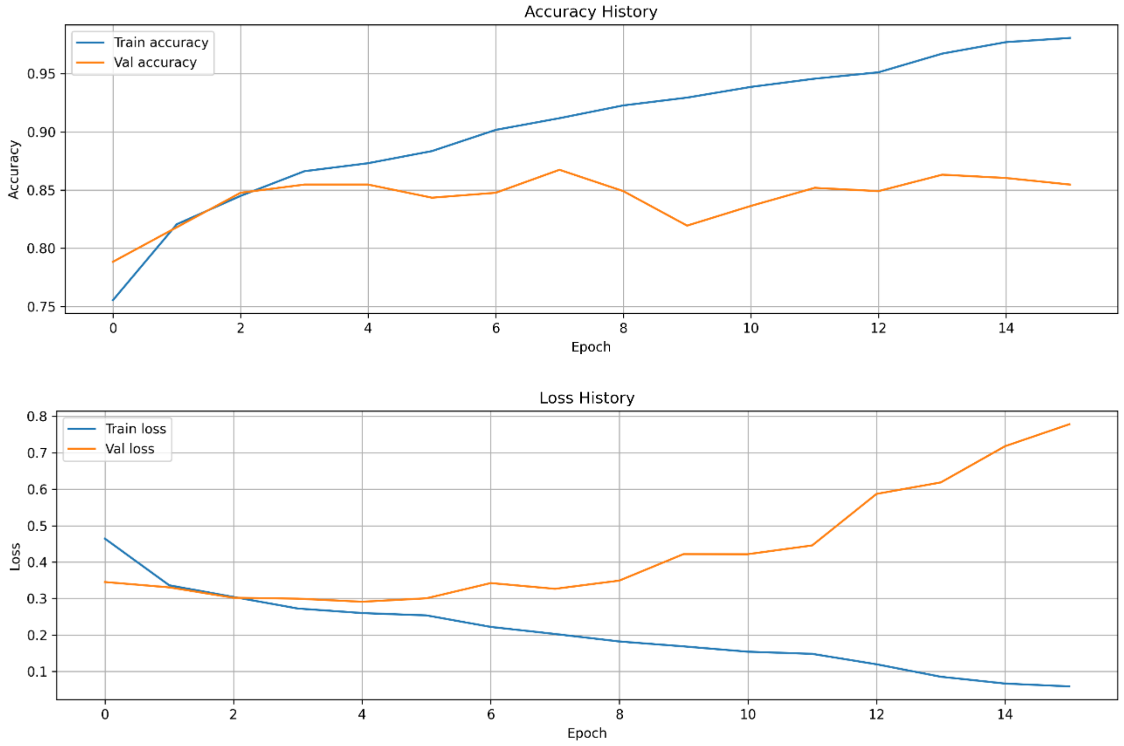


Figure 4.1.2: DenseNet121 Training Metrics Visualization of Accuracy and Loss

4.1.3 MobileNetV3 Training History

The training history of MobileNetV3 was trained using the leaner model, rejecting any learning errors and aiming to compare the model results with known ones.The training history of MobileNetV3 was trained on the leaner model, and it rejected the possibility of learning errors and sought to compare the model result with known ones. The MobileNetV3 was trained using 16 epochs. Loss and Accuracy History (Figures 4.3a, 4.3b): This model also used overfitting, as did DenseNet-121. Then in Epoch 4, and then in Validation Loss, which began to steadily increase (between ~0.30 and $\sim0.46), the Training Loss continued to drop to approximately 0.12. The Validation Accuracy level was reduced modestly after reaching value of $87.5 percent near

Epoch 10, which means additional training did not help decrease the level of negatively influence the ability of the model to generalize.

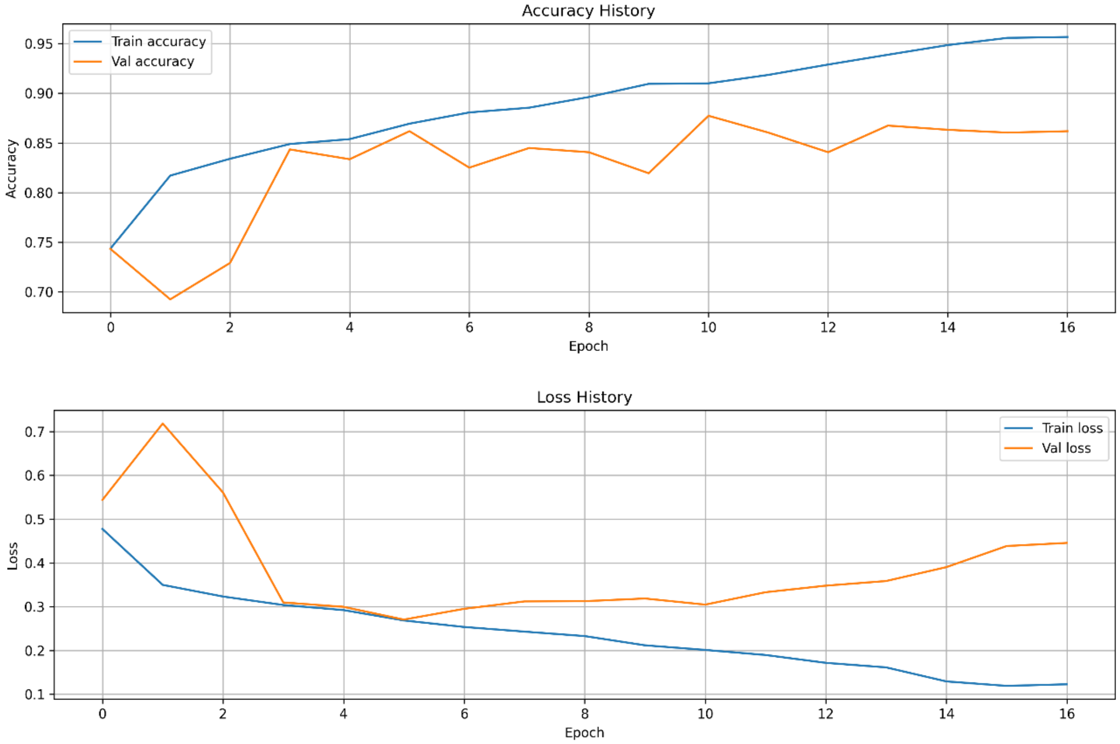


Figure 4.1.3 MobileNetV3 Training Metrics Visualization of Accuracy and Loss

4.1.4 ShuffleNetV2 Training History

ShuffleNetV2 The training history of ShuffleNetV2 is extensive and can be summarized as follows: ShuffleNetV2 Training History The training history of ShuffleNetV2 is long and can be described in the following way: The ShuffleNetV2 model was trained with 23 epochs. Figure 4.4a/Figure 4.4b: The overfitting tendency of this model was observed to be relatively less pronounced than that of the other pre-trained models. Validation Accuracy was the highest at Epoch 11 (Epoch 11) was 84.7 percent), and Validation Loss was the highest at Epoch 11 (Epoch 11) was 0.28. The ShuffleNetV2 architecture is designed with an efficient design, which by default could offer more regularization as shown by the subsequent reduction gap between training and validation.

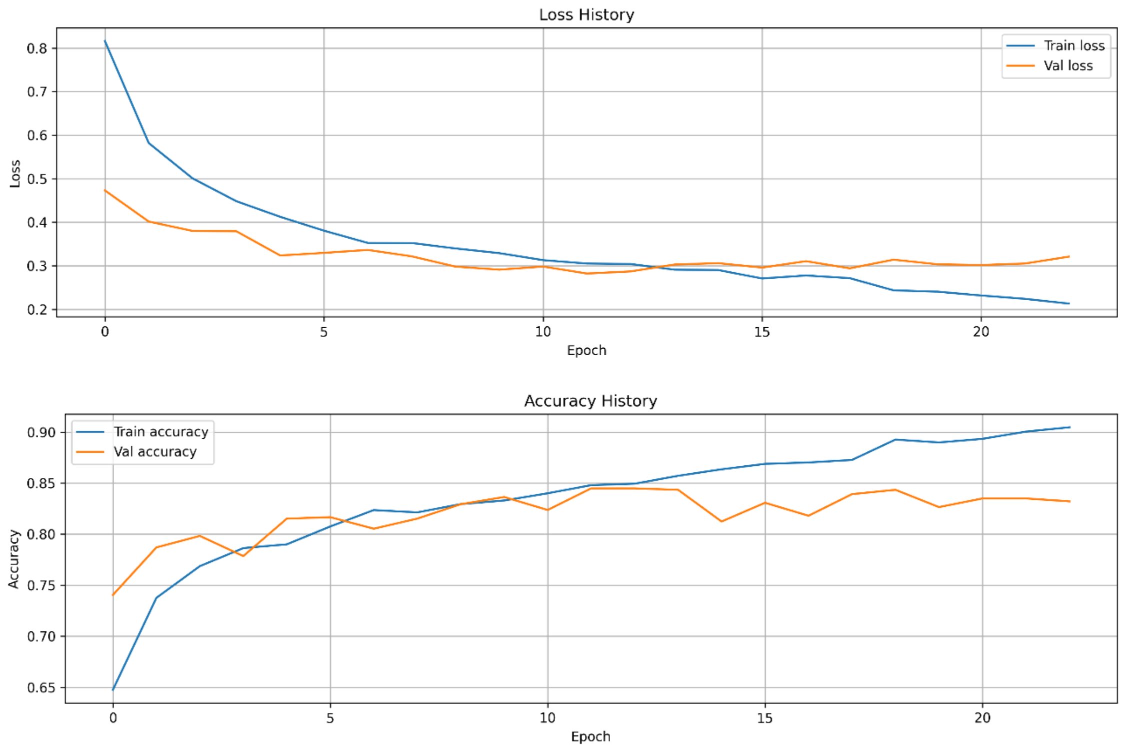


Fig 4.1.4: ShuffleNetV2 Training Metrics Visualization of Accuracy and Loss

4.2 Test Set Performance Comparison

Test set performance comparison will be done in the context of 4.2. Each model was quantitatively tested on a held-out test set. The primary measures are in the form of Overall Accuracy and the class-specific measures of Precision, Recall (Sensitivity), and F1-Score.

4.2.1 Confusion Matrix Analysis

The comprehensive confusion matrices of each of the four models that are presented below can be analyzed by considering true positives (TP), false negatives (FN), false positives (FP) and true negatives (TN). The total number of test instances was 710 and across all models.

These are All model confusion matric:

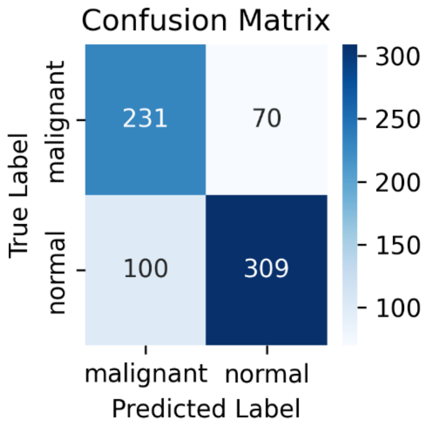


Figure 4.2.1: CustomCNN Confusion Matrix

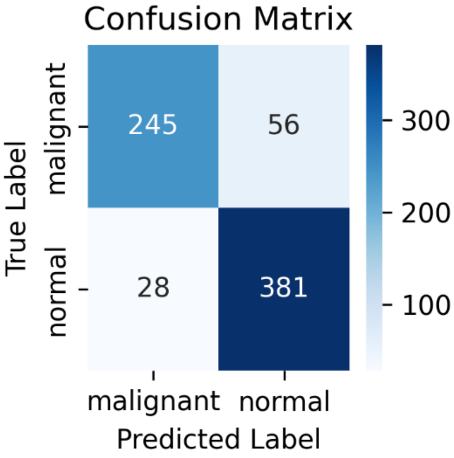


Figure 4.2.2: DenseNet121 Confusion Matrix

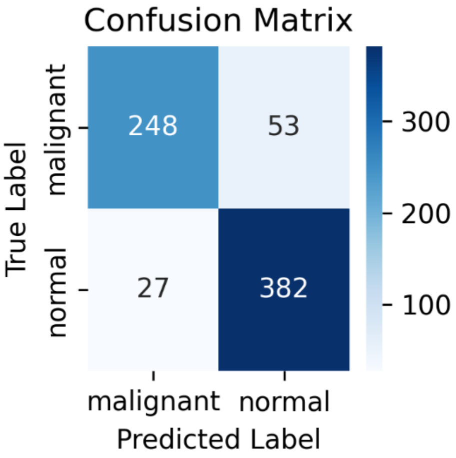


Fig 4.2.3 MobileNetV3 Confusion Matrix

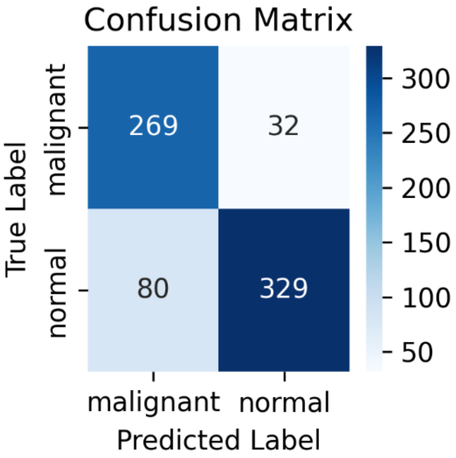


Fig 4.2.4: ShufflenetV2 Confusion Matrix

Table 2: All model confusion matric comparison.

Model	True Label	Prediction: Malignant(TP/FP)	Prediction: Normal(FN/TN)	Total Malignant	Total Normal
customCNN	Malignant	231(TP)	70(FN)	301	N/A
	Normal	100(FP)	309(TN)	N/A	409
DenseNet_121	Malignant	245(TP)	56(FN)	301	301
	Normal	28(FP)	318(TN)	N/A	N/A
Mobilnet3	Malignant	248(TP)	53(FN)	301	301
	Normal	27(FP)	382(TN)	N/A	N/A
ShuffleNetV2	Malignant	269(TP)	32(FN)	301	301
	Normal	80(FP)	329(TN)	N/A	N/A

4.2.2 Quantitative Metrics Comparsion

The calculated performance metrics for all four models are summarized in Table 4.1.

Table 3: All model Accuracy comparison.

Model	Overall Accuracy	Malignant Precision	Malignant Recall (Sensitivity)	Malignant F1-Score	Normal F1-Score

MobilnetV3	0.882	0.882	0.882	0.882	0.882
DeseNet_121	0.882	0.882	0.882	0.882	0.882
ShuffleNetV2	0.842	0.842	0.842	0.842	0.842
CustomCNN	0.761	0.698	0.768	0.732	0.785

4.3 Comparative Discussion

Best Overall Accuracy: The MobileNetV3 model is the most accurate with 88.7 accuracy and at a close, DenseNet-121 has the equal highest accuracy at 88.2. This suggests that huge, pre-trained architectures that are powerful feature extractors (that was developed purposefully to deploy on mobile and on the edge) performed rather well in this area. Custom Model Performance: The smallest and smallest capacity of the CustomCNN compared to the depthier, pre-trained models explains why its minimum accuracy of $76.1percent is also very minimal.

Malignant Case Detection (Recall)

Sensitivity or the ability to correctly identify the cases of true positives, which is also known as Malignant Recall, might be the most significant aspect of a medical diagnostic system since when it is low, the chance of a False Negative (a disease is missed) increases. highest Recall: The ShuffleNetV2 model with the highest Malignant Recall was at 0.894 (with 32 cases missing). Trade-off: This is best characterized by a very active, risk-averse, categorization process which had the highest False Positives (80 instances), however at the sacrifice of the worst Malignant Precision value (0.771).

Malignant Prediction Reliability (Precision)

Maximum Precision: The best Malignant Precision was achieved in the MobileNetV3 at 0.902. This indicates that MobileNetV3 is correct in over 90 percent cases when it states that a case is malignant. This results in a decrease in patient anxiety and wasting of resources due to unnecessary follow-up activities.

Final Selection Criteria

The hierarchical nature of clinical objectives arrives at the best model by the quantitative outcomes: Reduction of False Negatives (Putting Safety First): ShuffleNetV2 is superior due to its great Recall (\$0.894\$). First Placing Efficiency and Overall Accuracy (General Performance): MobileNetV3 is the most appropriate choice to have the most reliable predictions (Precision: \$0.902\$), and the overall-accuracy is the largest (0.887).

CHAPTER 5

CONCLUSION

5.1 Findings and contribution

This study examined the efficiency of breast cancer detection through the deep learning method and utilized two commonly used imaging techniques: mammography and ultrasound, and the finding that varied CNNs are effective in image classification (malignant vs. normal). The aim of the work was not only to compare the performance of the models in single and mixed modalities, but also to examine the possibility of the prediction of the stage of breast cancer by use of publicly available datasets. This chapter gives the most important findings of the work, evaluates how far the research questions were fulfilled, discusses the limitations of the study and also gives suggestions of what could be done in future studies. The initial research question was whether the deep learning models which integrate mammography and ultrasound images can be able to accurately classify the normal and cancerous breast tumors. The outcomes of the experiment showed that the deep learning approaches especially transfer learning frameworks performed well in the binary classification problem. MobileNetV3 was the most accurate at 88.7 the next was DenseNet121 at 88.2 and ShuffleNetV2 was the most accurate in identifying malignant cases with the best recall of 0.894. These findings validate the assumption that contemporary deep learning architectures are capable of classifying tumours in the breast with great precision when the techniques of preprocessing, augmentation, and model optimization are employed. This makes the CustomCNN with its simplicity still able to reach a decent accuracy in 76.1% which shows that simple models can still learn the key patterns, but more profound pretrained models are better. Therefore the initial research question has been answered positively as it demonstrated that machine learning and deep learning platforms are useful solutions to the problem of breast tumor classification in case of using image-based data. The second research question was to determine whether the combination of mammography and ultrasound imaging would result in better classification results better than the use of a single modality. This research did not divide the training of models according to modality, but the integrated dataset provided a more vivid and representative training of breast cancer images. This multimodal combination enabled the models to acquire complementary structure and textural information by the two imaging systems. The good results of the transfer learning models imply that the combination of the two modalities increases the feature space and helps in the enhancement of the generalization. In addition, paradigms that were reported to be inhibited by single-modality constraints, like noisy ultrasound or low contrast in mammography, were less inconsistent when trained on the combination set. Hence, this finding can be used to infer that multimodal imaging increases the strength and accuracy of classification that can be used to answer the second research question in the affirmative. The third research question focused on the question of the feasibility of the stage prediction of breast cancer based on publicly available image datasets. After conducting a comprehensive analysis of the datasets we found that stage data or data related to the size of the tumor, the presence of the lymph nodes and their metastasis are not included in the mammography and ultrasound dataset utilized in this study. These data sets merely have nominal values of the normal and the malignant condition and stage cannot be predicted. The staging of breast cancer needs detailed pathological annotations and clinical information, which can not be represented in simple imaging data sets. This weakness is also one of the limitations that can reveal the limitation of the problem; it is not possible to carry out automated stage classification on publicly available breast cancer images without stage-labeled datasets or multimodal clinical information. As a result, the third research question is answered knowing that in the state of the existing open-source datasets, stage prediction is not possible. With regards to the research objectives, the research was able to achieve its objectives. Mammography and ultrasound dataset were preprocessed and analyzed using standard methods so that high-quality inputs are obtained to train the models. Four deep learning models, such as Custom_net, DenseNet121, MobileNetV3, and ShuffleNetV2 were designed and tested showing their ability to classify breast tumors that are malignant and those that are normal. The publication has also explicitly investigated the drawbacks of open-ended data sets with the absence of stage labels being the major barrier to carrying out cancer stage forecasting. The relative and comparative performance analysis of the models further unraveled the effects of multimodal training on the classification performance, and this entailed the goal of comparing single or mixed modality training performance. Lastly, suggestions regarding the further research were made, such as the necessity to have stage-labeled datasets and more proficient models of diagnosis.

5.2 Limitations

In as much as this study has its strengths, a number of limitations have to be mentioned. The used datasets did not include much clinical information, as they did not include metadata like the tumor grade, the presence or absence of the hormone receptor status, and the staging indicators. The merged data set did not permit the comparison of the modality-specific performance, which might have provided information regarding the effects of each type of imaging used separately on the accuracy of the classification. Moreover, the dataset used was adequate to make binary classification, but a bigger dataset could also enhance the generalization of deep learning models. The other weakness is caused by the overfitting (in DenseNet121 and MobileNetV3) after a long period of training, which stresses the importance of more serious regularization measures. The limitations imply the need of a broader dataset and better training regimes.

5.3 Future Improvements :

In prospect, the effort of attaining or building up breast cancer dataset that incorporates label of the stage, clinical history as well as biopsy-proven diagnostic data ought to be incorporated in future work. These data sets would allow more complex diagnostic models, such as multi-class and survival prediction. Innovations that would include state-of-the-art multimodal architectures combining imaging and radiology reporting with genomic factors or patient demographics may greatly improve performance in diagnosis. In addition, attention-based networks or vision transformers can also be examined in future research, as they have demonstrated good results in medical image interpretation. Lastly, the future of breast cancer detection (as suggested by the next generation of the system) should be informed by real-world deployment requirements, including a model with explainable characteristics, performance, and clinical workflow connectivity. To conclude, this research paper confirms the possibilities of deep learning as an effective method of detecting breast cancer based on mammography and ultrasound images. The results of the experiment prove that the model is well trained, multimodal imaging is more effective, and existing datasets are limited in terms of stage prediction. Even though difficulties are evident, the findings present a good platform on which further studies can be carried out in order to achieve more meaningful and quantifiable breast cancer diagnostic models.

REFERENCES

1. Sahu, A., Das, P., & Meher, S. (2024). An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images. *Biomed. Signal Process. Control.*, 87, 105377. https://doi.org/10.1016/j.bspc.2023.105377.
2. Chen, J., Pan, T., Zhu, Z., Liu, L., Zhao, N., Feng, X., Zhang, W., Wu, Y., Cai, C., Luo, X., Lin, B., Wang, X., Ye, Q., Gao, R., Zhou, Z., Beatson, R., Tang, J., Ming, R., Wang, D., Deng, J., & Zhou, G. (2025). A deep learning-based multimodal medical imaging model for breast cancer screening. *Scientific Reports*, 15. https://doi.org/10.1038/s41598-025-99535-2.
3. Rahman, H., Bukht, T., Ahmad, R., Almadhor, A., & Javed, A. (2023). Efficient Breast Cancer Diagnosis from Complex Mammographic Images Using Deep Convolutional Neural Network. *Computational Intelligence and Neuroscience*, 2023. https://doi.org/10.1155/2023/7717712.
4. Wu, J., Li, Y., Gong, W., Li, Q., Han, X., & Zhang, T. (2025). Multi-modality radiomics diagnosis of breast cancer based on MRI, ultrasound and mammography. *BMC Medical Imaging*, 25. https://doi.org/10.1186/s12880-025-01767-1.
5. Rahman, H., Bukht, T., Ahmad, R., Almadhor, A., & Javed, A. (2023). Efficient Breast Cancer Diagnosis from Complex Mammographic Images Using Deep Convolutional Neural Network. *Computational Intelligence and Neuroscience*, 2023. https://doi.org/10.1155/2023/7717712.
6. Liu, B., Liu, S., Cao, Z., Zhang, J., Pu, X., & Yu, J. (2025). Accurate classification of benign and malignant breast tumors in ultrasound imaging with an enhanced deep learning model. *Frontiers in Bioengineering and Biotechnology*, 13. https://doi.org/10.3389/fbioe.2025.1526260.
7. Zhang, T., Tan, T., Han, L., Appelman, L., Veltman, J., Wessels, R., Duvivier, K., Loo, C., Gao, Y., Wang, X., Horlings, H., Beets-Tan, R., & Mann, R. (2023). Predicting breast cancer types on and beyond molecular level in a multi-modal fashion. *NPJ Breast Cancer*, 9. https://doi.org/10.1038/s41523-023-00517-2.

8. Jabeen, K., Khan, M., Alhaisoni, M., Tariq, U., Zhang, Y., Hamza, A., Mickus, A., & Damaševičius, R. (2022). Breast Cancer Classification from Ultrasound Images Using Probability-Based Optimal Deep Learning Feature Fusion. *Sensors (Basel, Switzerland)*, 22. https://doi.org/10.3390/s22030807.
9. Chen, L., Su, Y., Wang, Y., Yu, Y., Huang, D., Zhang, M., Chen, X., Ye, X., He, Y., Xue, E., Lin, L., & Chen, Z. (2025). Ultrasound Radiomics-Based Machine Learning and SHapley Additive exPlanations Method Predicting Pathological Prognostic Stage in Breast Cancer: A Bicentric and Validation Study.. *Clinical breast cancer*. https://doi.org/10.1016/j.clbc.2025.05.001.
10. Alom, M., Farid, F., Rahaman, M., Rahman, A., Debnath, T., Miah, A., & Mansor, S. (2025). An explainable AI-driven deep neural network for accurate breast cancer detection from histopathological and ultrasound images. *Scientific Reports*, 15. https://doi.org/10.1038/s41598-025-97718-5.
11. Dwivedi, A., Yadav, A., & Gamad, R. (2025). Breast Cancer Disease Prediction using Convolutional Neural Networks with Ultrasound Image Dataset for Healthcare-IOT Application. *2025 4th OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 5.0*, 1-6. https://doi.org/10.1109/otcon65728.2025.11070903.
12. Wei, T., Chang, A., Kang, Y., Patel, M., Fang, Y., & Yan, Y. (2025). Multimodal deep learning for enhanced breast cancer diagnosis on sonography. *Computers in biology and medicine*, 194, 110466 . https://doi.org/10.1016/j.compbiomed.2025.110466.
13. Taylor, D., Burrows, S., Saunders, C., Parizel, P., & Ives, A. (2023). Contrast-enhanced mammography (CEM) versus MRI for breast cancer staging: detection of additional malignant lesions not seen on conventional imaging. *European Radiology Experimental*, 7. https://doi.org/10.1186/s41747-022-00318-5.
14. Chelloug, S., Mahel, A., Alnashwan, R., Rafiq, A., Muthanna, M., & Aziz, A. (2025). Enhanced breast cancer diagnosis using modified InceptionNet-V3: a deep learning approach for ultrasound image classification. *Frontiers in Physiology*, 16. https://doi.org/10.3389/fphys.2025.1558001.

15. Din, N., Dar, R., Rasool, M., & Assad, A. (2022). Breast cancer detection using deep learning: Datasets, methods, and challenges ahead. *Computers in biology and medicine*, 149, 106073 . https://doi.org/10.1016/j.compbiomed.2022.106073.
16. Sood, R., Rositch, A., Shakoor, D., Ambinder, E., Pool, K., Pollack, E., Mollura, D., Mullen, L., & Harvey, S. (2019). Ultrasound for Breast Cancer Detection Globally: A Systematic Review and Meta-Analysis. *Journal of Global Oncology*, 5. https://doi.org/10.1200/jgo.19.00127.
17. Zhang, D., Dihge, L., Arvidsson, I., Bendahl, P., Dustler, M., Ellbrant, J., Gulis, K., Hjärtström, M., Ohlsson, M., Rejmer, C., Schmidt, D., Zachrisson, S., Edén, P., & Rydén, L. (2025). Abstract P2-09-11: Improved Prediction of Axillary Lymph Node Metastasis in Early-Stage Breast Cancer Using Deep Learning on Routine Mammography. *Clinical Cancer Research*. https://doi.org/10.1158/1557-3265.sabcs24-p2-09-11.
18. Sahu, A., Das, P., & Meher, S. (2024). An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images. *Biomed. Signal Process. Control.*, 87, 105377. https://doi.org/10.1016/j.bspc.2023.105377.
19. Elboghdady, A., Aziz, A., Bayoumi, D., & Taha, E. (2025). The effectiveness of sono-mammographic findings in predicting the molecular subtypes of breast cancer. Using a sono-mammography score in distinguishing between non-triple-negative and triple-negative breast cancer. *Egyptian Journal of Radiology and Nuclear Medicine*. https://doi.org/10.1186/s43055-025-01444-x.
20. Chen, J., Pan, T., Zhu, Z., Liu, L., Zhao, N., Feng, X., Zhang, W., Wu, Y., Cai, C., Luo, X., Lin, B., Wang, X., Ye, Q., Gao, R., Zhou, Z., Beatson, R., Tang, J., Ming, R., Wang, D., Deng, J., & Zhou, G. (2025). A deep learning-based multimodal medical imaging model for breast cancer screening. *Scientific Reports*, 15. https://doi.org/10.1038/s41598-025-99535-2.
21. Li, H., Zhang, C., Shao, H., Pan, L., Li, Z., Wang, M., & Xu, S. (2025). Prediction models of breast cancer molecular subtypes based on multimodal ultrasound and clinical features. *BMC Cancer*, 25. https://doi.org/10.1186/s12885-025-14233-6.
22. Hu, M., Zhang, L., Wang, X., & Xiao, X. (2025). Enhanced HER-2 prediction in breast cancer through synergistic integration of deep learning, ultrasound radiomics, and clinical data. *Scientific Reports*, 15. https://doi.org/10.1038/s41598-025-12825-7.

23. Liu, B., Liu, S., Cao, Z., Zhang, J., Pu, X., & Yu, J. (2025). Accurate classification of benign and malignant breast tumors in ultrasound imaging with an enhanced deep learning model. *Frontiers in Bioengineering and Biotechnology*, 13. https://doi.org/10.3389/fbioe.2025.1526260.
24. Zhang, T., Tan, T., Han, L., Appelman, L., Veltman, J., Wessels, R., Duvivier, K., Loo, C., Gao, Y., Wang, X., Horlings, H., Beets-Tan, R., & Mann, R. (2023). Predicting breast cancer types on and beyond molecular level in a multi-modal fashion. *NPJ Breast Cancer*, 9. https://doi.org/10.1038/s41523-023-00517-2.
25. Ragab, M., Albukhari, A., Alyami, J., & Mansour, R. (2022). Ensemble Deep-Learning-Enabled Clinical Decision Support System for Breast Cancer Diagnosis and Classification on Ultrasound Images. *Biology*, 11. https://doi.org/10.3390/biology11030439.
26. Shen, L., Margolies, L., Rothstein, J., Fluder, E., McBride, R., & Sieh, W. (2017). Deep Learning to Improve Breast Cancer Detection on Screening Mammography. *Scientific Reports*, 9. https://doi.org/10.1038/s41598-019-48995-4.
27. Rabiei, R., Ayyoubzadeh, S., Sohrabei, S., Esmaeili, M., & Atashi, A. (2022). Prediction of Breast Cancer using Machine Learning Approaches. *Journal of Biomedical Physics & Engineering*, 12, 297 - 308. https://doi.org/10.31661/jbpe.v0i0.2109-1403.
28. Radomyski, P., Trojanowski, M., Kijewska, K., Lis, A., Pietkiewicz, J., Matuszewski, K., Mocydlarz-Adamcewicz, M., Taraszkiewicz, Ł., & Kycler, W. (2025). Advancing clinical staging in invasive breast carcinoma: the role of contrast-enhanced mammography. *Polish Journal of Radiology*, 90, e207 - e214. https://doi.org/10.5114/pjr/202229.
29. Sahitya, J., Kilaparti, K., & Batchu, M. (2025). Diagnostic utility of ultrasound features in identifying triple-negative breast cancer. *Current Medicine Research and Practice*. https://doi.org/10.4103/cmrp.cmrp_187_24.

LIBRARY CLEARANCE

©Daffodil International University

PLAGARISM REPORT

ACCOUNT CLEARANCE

©Daffodil International University