CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Breast Cancer Detection via Multi-Modal Imaging with Independent Evaluation and Clinically Admissible Fusion

by

Nashrah Khan

A thesis submitted in partial fulfilment for the degree of Master of Science

in the

Faculty of Computer Science
Department of Computer Science

2026

CERTIFICATE OF APPROVAL

Breast Cancer Detection via Multi-Modal Imaging with Independent Evaluation and Clinically Admissible Fusion

by

Nashrah Khan

(MCS241002)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. M. Nouman Noor	FAST, Islamabad
(b)	Internal Examiner	Dr. M. Masroor Ahmed	CUST, Islamabad

Dr. Sabeen Masood
Thesis Supervisor
June, 2026

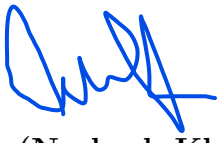
Dr. M. Masroor Ahmed
Head
Dept. of Computer Science
June, 2026

Dr. M. Abdul Qadir
Dean
Faculty of Computing
June, 2026

Author's Declaration

I, **Nashrah Khan** hereby state that my MS thesis titled "**Breast Cancer Detection via Multi-Modal Imaging with Independent Evaluation and Clinically Admissible Fusion**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time, if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.



(Nashrah Khan)

Registration No: MCS241002

Plagiarism Undertaking

I solemnly declare that research ,work presented in this thesis titled **"Breast Cancer Detection via Multi-Modal Imaging with Independent Evaluation and Clinically Admissible Fusion"** is solely my research work with no significant contribution from any other person. Small contributions/help wherever taken have been duly acknowledged, and the complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I, as an author of the above-titled thesis, declare that no portion of my thesis has been plagiarized, and any material used as a reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of the MS Degree, the University reserves the right to withdraw/revoke my MS degree, and that HEC and the University have the right to publish my name on the HEC/University website, on which names of students are placed who submitted plagiarized work.

(Nashrah Khan)

Registration No: MCS241002

Acknowledgement

"And whoever puts all his trust in God, He will be enough for him." The Holy Book [65:1]. I would like to say Thank you for everything God has blessed me with that enabled me to reach here. I want to express my gratitude to my supervisor, Dr. Sabeen Masood, who guided me and helped me with her valuable suggestions throughout this work. And gave her valuable time, May God keep her in his blessings.

I am Thankful to my parents for their love, prayers, and everything that I needed, and they keep on pushing me for higher education.

Lastly, I would also like to thank everyone who has helped me along the way. Special thanks to Dr. M. Abdul Qadir, Dr. Nadeem Anjum and Dr. M. Masroor Ahmad for increasing my knowledge and helping me with technical aspects, and for giving motivational support throughout my research journey.

(Nashrah khan)

Abstract

Breast cancer is one of the leading cause of cancer-related mortalities across the globe. Timely detection is important to enhance survival outcomes. Although there has been improvement in this domain, but the majority of the CAD systems focused on either single or dual modalities and few on tri modalities but having limitations. Three modalities i.e mammography, ultrasound and histopathology have specific features; first one are projectional x-rays that have overlapping white or gray tissue structures on dark background. Second are real time echogenicity with fat lobules and speckle noise, while third are microscopic cellular architecture without macroscopic breast anatomy. Collectively they cover macro, meso and micro levels of breast details respectively. In this study, a tri-modal framework for breast cancer detection is proposed that independently evaluates mammography, ultrasound, and histopathology and then fusion is employed to give combined result to overcome the prior work limitation. To improve quality and contrast of the downloaded experimental data, preprocessing is employed on each modality. To extract features, DenseNet-121 is used on mammography and ultrasound, while histopathology uses Domain Adversarial Neural Network with it to learn domain-invariant features that occur due to staining protocols and acquisition locations. Consistent vector space is maintained in classification head across modalities for predictability. Later, feature level and decision level fusion is also implied for combined results. At decision level, predictions from modalities are fused to achieved clinical admissibility that prior work on forced fusion failed to provide. The model achieved accuracies of 96% for mammography, 90.8% for ultrasound, 96.4% for histopathology and decision level fusion achieved improved results of 98.90% which proves that model in each modality support each other detection. Data variability that occurs due to acquisition or domain change also accommodated through techniques like z- scoring, minimalist preprocessing, class balancing and deep learning feature extraction. In order to improve clinical transparency and explainability, Grad-CAM is employed for localization of influential regions at image modality level to support decision with standard diagnostic metrics for evaluation of the model performance.

Contents

List of Figures

List of Tables

Abbreviations

ACS	American Cancer Society
AI	Artificial Intelligence
AUC	Area Under the Curve
AUROC	Area under the Receiver Operating Characteristic Curve
BC	Breast Cancer
BI-RADS	Breast Imaging Reporting and Data System
CAD	Computer Aided Diagnostic
CBIS-DDSM	Curated Breast Imaging Subset of DDSM
CC	Cranio Caudal
CLAHE	Contrast Limited Adaptive Histogram Equalisation
DANN	Domain Adversial Neural Network
DDSM	Digital Database for Screening Mammography
DL	Deep Learning
DNBCD	Deep Neural BC Detection
FN	False Negative
FP	False Positive
GHE	Global Histogram Equalization
GRL	Gradient Reversal Layer
GRUs	Gated Recurrent Unit
HDI	Human Development Index
HP	Histopathology
IoT	Internet of Things
KNN	K-Nearest Neighbour
LBP	Local Binary Patterns

LGBM	Light Gradient Boosting Machine
LoG	Laplacian of Gaussian
LSTMs	Long Short-Term Memory
MG	Mammograms
MIAS	Mammography Image Analysis Society
ML	Machine Learning
MLO	Medio Lateral Oblique
NFMs	Neural Factorisation Machines
PNG	Portable Network Graphics
RF	Random Forest
RNNs	Recurrent Neural Networks
ROI	Region of Interest
SMOTE	Synthetic Minority Over-sampling Technique
SRAD	Speckle Reducing Anisotropic Diffusion
SVM	Support Vector Machine
TCNs	Temporal Convolutional Networks
TN	True Negative
TP	True Positive
US	Ultrasound
ViT	Vision Transformer
XAI	Explainability AI
XGBoost	Extreme Gradient Boosting
YOLOv8	You Only Look Once Version 8

Symbols

μ	Global Mean
σ	Standard deviation
p	Training process that is linearly increased from 0 to 1
L_y	Cross-entropy loss for diagnostic information
L_d	Cross-entropy loss for domain labels
λ	Dynamic adaption factor
z	Z-scoring
x	Data instance
N	Samples
C	Different classes
n_i	Current number
θ_f	Parameters of the feature extractor network
∂L_y	Gradient of the diagnostic (label) loss w.r.t. θ_f
∂L_d	Gradient of the domain (adversarial)

Chapter 1

Introduction

1.1 Background

In recent years, the diagnosis of breast cancer(BC) has attracted the attention of many bioinformatics communities. It is still a global health issue with millions of women being affected and a significant proportion of cancer-related death has occurred. The World Health Organisation (WHO) [1] and Global Breast Cancer Statistics [1] reported a global increase in BC. In 2022, around 2.3 millions were diagnosed, and 670,000 deaths were recorded [1], [2]. American Cancer Society (ACS) disclosed the statistics of United States of America(USA). In 2023 and 2024, new cases of BC reported were 300,590 and 310,720 respectively. In 2025, an estimated 316,950 women in the U.S were diagnosed from BC, with additional 59,080 new cases of dutal carcinoma. Furthermore, in 2026, an estimated 321,910 new cases of BC are expected to be diagnosed in U.S along with 60,730 cases of ductal carcinoma [3]. Human Development Index (HDI) estimated that approximately 3.5 million BC cases will be reported till 2050 [4]. BC is the abnormal growth of tissues in the breast region that further spreads to the body. Therefore, timely detection of it is necessary to make early decision. BC tissues categorized into benign and malignant. Former, is localized and is not usually dangerous, but latter develops actively and spreads to other areas, posing a significant threat to health. For early detection, different BC diagnostic tools are used such as digital mammography, x-ray, ultrasound, histopathology, magnetic resonance imaging

(MRI), 3D Mammography, breast thermograms and positron emission tomography [5]. Current screening modalities although helps in early detection but they posses a limitation of sensitivity, specificity and cost-effectiveness. Standard treatments such as chemotherapy, targeted therapy, hormonal therapy and surgical radiations has also improved the survival rates but they often associated with side effects which is a major clinical challenge. It is accurate and gives definitive diagnosis. Manually, high level of expertise are required to diagnose the BC from Mammography(MG), Ultrasound (US) and Histopathology (HP) images. Traditional diagnostic processes relying on manual analysis of medical images having certain limitations of variability, subjectivity, expert knowledge, complexity and ambiguity [6] which highlights the urgent need of automated BC detection systems [7] on these imaging modalities that cover macro, meso and micro level insights of cells and tissues . The rationale of automation in the BC diagnosis is to increase sensitivity, specificity, accuracy with faster and more efficient diagnosis . Human interpretations of medical images i.e MG, US, Magnetic Resonance Imaging (MRI) are prone to inter reader variability. Different radiologists have different perceptions on the images. There must be some system/third party that gives confidence on decisions given by radiologists. Misinterpretations often lead to both false positives (FP) and false negatives (FN). FP cause unnecessary biopsies that cause patient anxiety, discomfort and increased health care costs. However, FN cause missed cancers. It is often difficult to identify BC in dense tissues with the human eye, and radiologists often make wrong decisions. Pathologists rely on limited factors i.e. age, breast density, family history etc and they don't capture the complexity of individual BC tissues. Moreover, experienced radiologists are not accessible to all areas, especially in rural areas where BC health risk is high. It is further added that some images like MG is prone to high sensitivity in women making it difficult to detect tissues. Human are prone to errors due to their natural limitations of fatigue, distractions and cognitive bias-ness. This leads to errors in interpretations. To empower it, superior tools like Deep Learning(DL) is used that extract the intricate and complex patterns efficiently. Machine Learning (ML) techniques are also incorporated for early prediction of BC i.e. free of subjectivity and variability in less time constraint and adds an additional layer of confidence to pathologists [8].

Globally, BC is one of the most common cancer disease in women and remains one of the primary causes of cancer-related death [1]. Its early diagnosis is necessary for prediction, implementation of relevant treatment and improvement in survival rates. Medical imaging used for screening to diagnose BC provides clinicians to detect pathological anomalies. The use of mammography has been widely adopted due to its capability of identifying structural changes and micro calcifications [9] in the breast tissue. US imaging is a complementary procedure that provides further details about the morphology of lesions, margins, and tissue structure especially when using dense breasts [10]. However, US is very operator-dependent and can show inconsistency in interpreter conclusions [11]. Another technique of HP is used that act as a gold standard [6] which gives detailed information regarding cell morphology and tissues. Convolutional Neural Network(CNN) is renowned model for feature extractors in medical images as compared to traditional hand-crafted methods. Most of the existing Computer Aided Designs (CAD) systems that are based on DL, are based upon single or dual modalities [12] - [13] limiting their ability to effectively reflect the heterogeneity of BC modalities. Research on tri-modal systems is limited [14] - [15].

1.2 Motivation

Although studies are published based upon single/dual/tri modalities for classification of BC tumours MG, US and HP. Most single or dual modality researches limiting multi-scale exploration of the deep learning models. Few tri-modal researches prioritize accuracy alone, with limited focus on explainability, which are essential for clinical adoption. There are certain loopholes in the prior work for example, enforced fusion on data by fusing the features of different patients across modalities. To achieve accuracy, researchers also used cleaned data despite the availability of the heterogenous data. This brings methodological constraints in previous work. Some models failed to perform in dense breast tissues and others use XAI tools i.e. LIME ans SHAP that are not user friendly. Clinicians find them hard to interpret results from them. The motivation behind this paper is to develop a framework that is clinically admissible and enforced fusion of data

and features of different patients should be avoided. Due to use of DL model, dense breast tissues can be also classified. Further, for localization of the regions, GradCam is employed.

1.3 Problem Statement

Although significant improvements in the BC CAD systems are made through DL models, but several key challenges remain unsolved. The current dominant methodologies rely on single or dual imaging modalities [12] - [13] which limits their ability to capture the heterogeneous, multi-scale properties of BC that span the structural, morphological and cellular dimensions. Even though MG, US, and HP are regularly used in clinics, but work on these three modalities is limited [14] - [15]. Recent study, [14] used forced fusion by aggregating the data of different patients across modalities which is clinically inadmissible approach. Few studies failed to diagnose dense breast tissues and others despite the presence of noisy and variable data, clean data is used to achieve high accuracy. Moreover, some modern DL systems failed to provide transparency of black box systems[16] which undermines the clinical trust and practical applicability. To cater it, tri-modal DL framework for BC detection is proposed that evaluates individual modality with explainability and also incorporated clinical admissible fusion to achieve combined accuracy.

1.4 Research Questions

This thesis investigates the following key questions:

RQ1: How can explainable artificial intelligence provide transparency and build clinician trust for each modality?

RQ2:How can a deep learning architecture independently diagnose breast lesions across US, MG and HP?

RQ3: What quantitative performance gains are achieved when moving from independent modality evaluation to clinically admissible fusion framework?

1.5 Research Objectives

The objectives of this thesis are summarized as follows:

RO1: To implement XAI framework that builds clinician trust and provide black box transparency by localising of the influenced region within each modality.
RO2: To develop and evaluate an independent deep learning framework that extracts diagnostic features and classifies breast lesions across modalities
RO3: To evaluate changes in diagnostic performance metrics i.e. sensitivity, specificity, accuracy, F1-Score, AUROC when moving from individual to fusion.

1.6 Technological Foundations

This study is based upon recent technological innovations in the field of DL, medical image analysis and XAI. Minimalist approach for preprocessing is used along with multi-level splitting. DenseNet-121 is used for feature extraction. Due to its dense connectivity of layers, feature reuse, and ability to learn low-level textures and high-level semantic features, it demonstrated excellent performance on medical imaging tasks [17]. Transfer learning based on large-scale databases like ImageNet also encourages model convergence and generalization in the situation where medical data are less marked.

To address domain variability of histopathology images, which is caused by variations in staining and imaging equipment, Domain Adversial Neural Network (DANN) present a conceptualization of how to learn domain-invariant feature representations [18]. DenseNet used for feature extraction that is followed by multi layer classifier to get diagnostic of each modality. Lastly, fusion is also integrated for combine accuracy. The last dense-block encourages architectural consistency, protects the integrity of the learned 1024-dimensional feature embeddings, and makes it simple to directly incorporate Grad-CAM to improve image-level localization [19].

XAI algorithm, such as Grad-CAM used for localization of the region that allows visualizing salient parts of the image, explaining model decisions and promoting transparency and clinical trust in automated systems of breast cancer diagnosis.

1.7 Scope of Study

This research study aims at the creation of an end to end DL system used to identify BC individual medical modalities features, i.e. MG, US, and HP. Fusion strategy is also implied and comparison drawn in between individual modalities and fusion results. It is limited to binary classification of breast lesions as benign and malignant based on preprocessed image data and patient-level stratified splitting to address data leakage. DenseNet-121 is used as the main feature extractor on all modalities and an additional DANN is applied to histopathology only as a means of ensuring both stain and domain-invariance.

The study is also based on the feature-level fusion, therefore, facilitating independent modality prediction in the case of using only one source of imaging and prediction from fusion strategy. The explainability is done through Grad-CAM, thus giving image-level visual explanations, to facilitate clinical trust. Performance evaluation is performed with traditional image-level measures such as accuracy, sensitivity, specificity, and the Area under the Receiver Operating Characteristic Curve (AUROC). The thesis focused on creating a tri-modal diagnostic model that could be clinically trusted.

1.8 Significance of the Study

The current thesis proposes tri-modal model of BC classification which encompasses the use of MG, US and HP as imaging modalities thus overcoming the shortcomings of the currently existing single and dual modalities which better reflects the heterogeneous and multi-scale characteristics across modalities independently and also through integration.

The use of weighted DL methods and learning domain-irrelevant features makes it more robust to class imbalance and inter-domain variations, particularly in HP images. The use of end-to-end DenseNet121 model from feature extraction to classification brings stability in the system.

In addition to this, XAI technique Grad cam is used for localization.The framework promotes modality-based prediction that is independent, thus increasing the

practicality of the framework to real-world clinical settings where not all modalities are present.

On the whole, this thesis contributes to the diagnosis of breast cancer that classify modalities independently and through fusion. Explainable AI highlights the important features on which the DL model do the prediction.

1.9 Proposed Solution

The framework of this study is proposed as follows:

i. Tri-modal data collection through MG, US and HP images to cover all diagnostic aspects

ii. Stratified splitting at patient level is used to eliminate the data leakage and to ensure that the classes are represented in all three parts of process; training, validation and test.

iii. Preprocessing of modality-specific is done which includes noise removal, contrast, resizing and intensity normalization which is done through Z-scoring.

iv. Feature extraction is carried out through Densenet-121 and for HP, DANN is employed with DenseNet-121 to facilitate domain-inspired feature acquisition and stain resistance.

v. Same DenseNet-121 classification head is used to conduct independent modality specific classification to maintain end-to-end consistency and deployment flexibility..

vi. XAI via Grad-CAM++ is integrated for the localisation of the images for interpretation.

vii. Fusion is employed to integrate the modalities at feature level and prediction level to get maximum influence of modalities' features

viii. The performance is evaluated by standard diagnostic measures, such as accuracy, sensitivity, specificity, and the AUROC. It is supported by evaluation and confusion metrics.

1.10 Contributions

i. The research study presents a tri-modal DL diagnosis framework for BC through MG, US, and HP images capturing macro, meso, and micro-level diagnostic information in a clinically realistic manner.

ii. Multi level stratification employed for splitting of data based on patient id, modality and label. It reduces the risk of data leakage and to guarantee reliability of the evaluation.

iii. Through DANN with DenseNet-121 integration, the model allows strong, domain-independent feature extraction methods of histopathological images are able to deal with variability due to staining and acquisition procedures.

iv. Smoothing functions of weighted losses and augmentation are used to reduce the imbalance of different classes, which reduces predictive bias and improves diagnostic reliability between benign and malignant groups.

v. End to end DenseNet-121 architecture is implemented on all the modalities individually for feature extraction

vi. To address the lack of transparency of DL architectures, Grad-CAM is added to provide image-level and modal visualization.

vii. For individual level classification, multi layer classifier are used for each specified modality at individual level. To corporate integration, fusion strategy is employed.

viii. The proposed framework was evaluated with the help of the known diagnostic measures, i.e., accuracy, sensitivity, specificity, and the AUROC.

1.11 Thesis Organization

The thesis is organized as follows: The related work is discussed in Chapter 2. Chapter 3 discussed proposed methodology. Chapter 4 discussed evaluation and results while Chapter 5 discussed conclusion and future work.

Chapter 2

Literature Review

This chapter provides an overview of the methodologies that are closely related to our study. Recent studies are reviewed that focused on DL approaches on single, dual and tri-modality frameworks. Over the past years, there is a notable shift from single modality to multimodal architecture in BC diagnosis-based models.

2.1 Prior Work on Tri-modality

Deb et al [14] proposed trimodal diagnostic framework of MG- Digital Database for Screening Mammography(DDSM), Mammography Image Analysis Society(MIAS), InBreast, US(BUS, MBU) and HP(BreakHis). DL model with four CNN layers is employed for both MG and US, whereas, for HP additional handcrafted feature extraction technique is used with Local Binary Patterns(LBP) to capture stain colour and texture information. Data level fusion is performed, followed by a stacked ensembler classifier that includes Random Forest(RF), Extreme Gradient Boosting(XGBoost), K-Nearest Neighbour(KNN), Light Gradient Boosting Machine(LGBM) to maximise the performance. The results revealed that the multimodal CAD system outperforms the unimodal by obtaining the accuracy of 99.96%. This research introduces a diagnostic pipeline demonstrating the high accuracy up till now [20]. Author did data aggregation to maximise the accuracy.

Another tri-modality by Ishaq Pacal [21] introduced InceptionNext-Transformer, a comprehensive diagnostic tool that combines CNN with Vision Transformer(ViT).

The model applied on 7 datasets encompasses HP(BreakHis), MG (INbreast, MIAS, and DDSM) and US (BUSI and BLUID) .Binary classification predictions of HP, US and MG dataset achieved the accuracy of 100%, 92.85% and 99.97%, respectively. It uses only 40X magnification images HP dataset that limits its applicability, Further for clinical trust no explainability of predictions provided.

To cater this, U.Nawaz [15] proposed a framework of X-BCNet, an explainable CNN that predicts and interprets BC classification across multiple modalities. The dataset used are BreakHIs-400X, Inbreast, BUSI and CBIS-DDSM. To ensure datasets consistency, images are resized to 224x224x3 pixels. With data augmentation, hierarchical spatial feature extraction along with attention-based refinement is carried out where two Convolutional blocks capture low-level and high-level structural patterns. For prediction, attention module is used. Grad Cam is used for localization. The accuracy achieved 98.75% on BreakHis, 99.12% on INbreast, 98.40% on BUSI, and 99.05% on CBIS-DDSM.

The model utilizes trimodal approach with promising accuracy. But it restricts the analysis to only magnification 400X that will limit the full spectrum scope of the HP variations and this may hinder the diagnostic ability of the model in real world scenarios. Furthermore, XBCNet model does not simulate the real world clinical images which often present with varying quality, artifacts and staining variations in HP. Furthermore, MIAS, DDSM, Inbreast are the dataset of MG but, no integrated comprehensive accuracy is provided instead separate accuracies are evaluated.

2.2 Prior Work on Dual-modality

Subsequent recent studies are focused upon dual-modalities, potentially due to data limitations or resource constraints.

Kushangi Atrey et al [12] proposed dual-modal BC classification using feature-level fusion of MG and US images. The dataset collected from the Department of Radiodiagnosis, All India Institute of Medical Sciences, Raipur, consists of 43 images each of MG and US which, after augmentation, increased to 1032 images

of each modality. The results evaluated that the combined multimodal approach brought higher accuracy of 98.84% from a cubic Support Vector Machine (SVM) as compare to MG and US alone that achieved 93.41% and 91.67%, respectively. Fusion employed in [12] across modalities.

Author in [22] collected real time dataset of MG and US of same patients for fusion. US and MG are treated differently due to their inherited potential features of breast tissues.Anisotropic diffusion is used in MG whereas in US images speckle reducing anisotropic diffusion is used. Feature extraction is done through ResNet-18 with transfer learning. Feature vectors are then combined through late fusion. It is followed by ML technique of SVM in 5-fold cross achieving accuracy of 99.22%. Some researches do not use fusion for integration of modalities and prefer to train modalities separately due to dataset limitations.

Md Romzan Alom et al [23] used Deep Neural BC Detection (DNBCD) model on HP dataset(BreakHIs) and US dataset (BUSI). The proposed approach is built upon DenseNet-121, employing transfer learning to benefit from pre-trained weights. It incorporates CNN, a global average pooling layer and dense layers for final classification. Dropout layer mitigate over fitting, and Grad-CAM is utilised to enhance interpretability for the model's prediction. In binary classification the accuracy achieved was 93.97% for HP and 89.97% for US. The limitation is use of a dual-modality framework and involvement of only 400X magnification HP images. Al-Tam et al [24] use datasets of MG(MIAS, INbreast, KAU-BCMD, real private Mammograms, and CBIS-DDSM) and US(BUSI and BUS2) that has been passed through two pipelines. First Stage focused on detectable cases while the other focuses on leftover lesions. The framework employed AI detectors You Only Look Once(YOLO) version 8 which is fine-tuned and optimised the localization of the breast lesions. For classification,CNN is used with attention mechanism of ViT-ResNet50 to increase predictions capability for binary classification. The first pipeline achieved the accuracy of 97.73% while second pipeline achieved accuracy of 96.65% of merged datasets.

Another study of A. Sahu [25] on DL based ensemble classifier is proposed for detection of BC. The primary contribution lies in its significant performance on

a small dataset, achieved through the integration of three transfer learning models (ResNet, AlexNet, and MobileNetV2) for prediction. The scheme worked by achieving a prediction accuracy of 99.17% accuracy to predict abnormality and 97.75% accuracy to predict the malignancy on the mini-DDSM dataset. It correctly predicts the US dataset (BUSI) with the accuracy 96.92% and 96.62% to identify abnormality and malignancy respectively. CNNs are mostly used in the feature extraction due to their inherent properties of detecting image edges, shapes and textures.

With this, Y.-M. Wang et al [26] built a CNN based cross modality model that uses datasets from multiple sources i.e. RSNA, PAS, kaggle. Data augmentation with three models and comparative analysis is done among them. Model 1 is a pre-trained CNN model with ML classifiers, Model 2 is a transfer learning model whereas Model 3 is a custom designed CNN.It is user defined framework having 17 layers CNN that specifically capture intricate fused image features. Among all, Model 3 outperforms others, reaching an accuracy of 96.4%. With optimistic accuracy, it has a drawback that it does not perform well in dense breast tissues. Few such images down grade the model performance.

In study by Belgacem Bouallegue [27], uses private and public datasets. Multi modal approach is used that take MG, US, MRI, thermography as input dataset. This real time framework integrated Internet of Things(IoT) environment. In preprocessing, Synthetic Minority Over-sampling Technique (SMOTE) is used to balance the data. For feature extraction and prediction, different DL algorithms are used. The better the feature extraction is the better the results are. This also narrow our study to use the best CNN model in our proposed study. DL algorithms used are temporal Convolutional Networks(TCNs), Neural Factorisation Machines(NFMs), Long short-term memory(LSTMs), Recurrent Neural Networks(RNNs), Gated Recurrent Unit(GRUs) and (DKL). Among all TCN outperforms all, achieving accuracy of 97.33% on private dataset and 99.44% on public dataset.In explainable AI, SHAP is used for interpretability. SHAP although is another tool to explain interpretability, but clinicians face moderate difficulty is understanding these graphs. R. Devkota et al [28] uses a small dataset of 50 patients. The framework is based on MG, MRI and US for the evaluation of BC. In

preprocessing breast compression is used in MG and for US. For feature extraction, Breast imaging Reporting and Data System (BI-RADS) criteria is employed. The classifier used is the BI-RADS categorization system. The comparative analysis is done on individual and combined modality. It is demonstrated that US is good for differentiating solid from cystic lesions while MG was better at speculated masses and micro-calcifications. It is concluded that dual modality outperforms single modalities. US achieves 88.90%, 68.80% sensitivity and specificity respectively, whereas MG had 94.49%, 87.50% sensitivity and specificity respectively. On combining both modalities, sensitivity of diagnosing malignant reaches up to 94.4% and specificity decreases up to 31.2%. This shows that combined MG and US approach of detecting BC is more beneficial than single modality.

In extension to it, Cong J [29] also works on dual modalities (US and MG). It uses a small private dataset containing only 180 cases. Data is transformed in feature vectors using patient ids to balance feature counts per modality. Separate feature extraction technique is used. For MG, lump mobility and size changes are considered while in US acoustic shadowing, vascular patterns and calcification is focused upon. The framework employed supervised ML approach for prediction. Biopsy is used at the end for refining the model accuracy. HP although considered as the gold standard but it is mostly used after initial stages.

This is due to several reasons which also cited in [30]. In [29] classifiers such as SVM, KNN and Naive Bayesian is used. Among all SVM outperforms all by achieving accuracy of 93.3% ,96% specificity and 80% specificity. In study of Habib G, Kiryati N, Sklair-Levy M et al [13], modalities used are MG and US. In preprocessing step, noise reduction, contrast enhancement and normalisation is done. Spatial registration of MG and US followed by the splitting. Fine tune a Google-Net model is used on both modalities separately and intricate feature are extracted. These high level features are then combined which author mentioned as feature aggregation. A CNN used for evaluation of these aggregated feature of modalities. The model successfully achieves the AUC 0.94 that outperforms state of the art single modalities.

Table 2.1 shows the summary of related research on BC detection and classification that is explained in detailed.

TABLE 2.1: Summary of related research on breast cancer detection and classification.

Ref	Year	Authors	Title	Datasets	Results	Limitations
[14]	2025	D. Deb and R. Dash and D. P. Mohapatra	MMHBC-Net: A multi-modal hybrid approach for BC classification	DDSM, MIAS, InBreast, BUS, MBU, BreakHis	99.96%(combined) 88.16%(MG) 95.33%(US) 91.79%(HP)	Incorrect cross-modal matching and enforced fusion due to inconsistent patient ids across modalities
[21]	2025	Ishak Pacal and O. Attallah	InceptionNeXt-Transformer: A novel multi-scale deep feature learning architecture for multimodal BC diagnosis	BreakHIs,BUS, INbreast, DDSM, CBIS	100%(HP) 99.97%(MG) 92.85%(US)	No black box transparency
[15]	2025	U. Nawaz and Z. Saeed and H. M. UbaidUllah and F. Mirza and M. Muzzamil	Explainable Attention-Enhanced Approach for Multimodal Breast Cancer Diagnosis Across Diverse Imaging Modalities	BreakHIs-400X, Inbreast, BUSI, CBIS-DDSM	98.75%(HP) 99.12%(INBreast) 98.40%(US) 99.05% (CBIS-DDSM)	Used only cleaned data despite available raw data, limiting real-world applicability

Table 2.1 continued from previous page

Ref	Year	Authors	Title	Datasets	Results	Limitations
[23]	2025	Md Romzan Alom and Farid F. A. and M. A. Rahaman and A. Rahman and T. Debnath and Abu and S. Mansor	An explainable AI-driven deep neural network for accurate breast cancer detection from histopathological and ultrasound images	BreakHIs-400x and BUSI	93.97%(HP) 89.97% (US)	Dual-modality model, Methodological constraints
[27]	2025	Belgacem Bouallegue and Yasser M. Abd El-Latif and Hosam El-Sofany and Islam	An Enhanced Approach for Predicting Breast Cancer Using Different Deep Learning Algorithms and Explainable AI Techniques in an IoT Environment	private, public(WBCD)	97.33% (private) 99.44% WBCD	Small dataset, Dual modality model
[25]	2024	A. Sahu and P. Das and S. Meher	An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images	BUS1,BUS2, DDSM	97.75%(MG) 97.50%(US)	No black box transparency

Table 2.1 continued from previous page

Ref	Year	Authors	Title	Datasets	Results	Limitations
[26]	2024	Y.-M. Wang and others	CNN-Based Cross-Modality Fusion for Enhanced Breast Cancer Detection Using Mammography and Ultrasound	RNSA, PAS, Kaggle	96.40%(combined)	Does not work on dense breast tissues
[22]	2024	Kushangi Atrey and Bikesh Kumar Singh and Narendra Kuber Bodhey	Integration of ultrasound and mammogram for multimodal classification of breast cancer using hybrid residual neural network and machine learning	real time US-MG dataset	99.22%(combined)	Unable to perform in dense breast tissues
[12]	2023	Kushangi Atrey and Bikesh Kumar Singh and Narendra Kuber Bodhey	Multimodal classification of breast cancer using feature level fusion of mammogram and ultrasound images in machine learning paradigm	Dataset from Dept. of Radiodiagnosis, AIIMS Raipur	98.84%(combined) 93.41%(MG) 91.67%(US)	Dual modality, Methodological constraints

2.2.1 Comparison with Existing Literature

Development in medical imaging through AI has been increasing at a faster pace. In context of BC most of the current studies have focused on the analysis of dual modalities. The imaging modalities, namely, MG, US, and HP, is commonly used in the clinics to enable a unified assessment of BC tumours and to aid in the correct diagnosis. Therefore, the lack of a comprehensive tri-modal model is the glaring gap in the existing literature. Multimodal learning has the potential of increasing the diagnostic confidence through the combination of macro, meso and micro structural data obtained from MG, US and HP samples. However, the adoption of tri-modal systems is limited by methodological and practical challenges such as non black box transparency, methodological constraints, and the lack of solid ensemble strategies. Research on tri-modality has been done as cited in [14] obtained promising accuracy but enforced fusion is done through data aggregation that fails to achieve clinical admissibility. Features of different patients across modalities are fused together to achieve high accuracy. Some researches failed to predict dense breast tissues. Another study [21] does not applied XAI and no black box transparency is accommodated. Likewise, the methodology in [15] limits the HP dataset to one magnification 400X despite having having other magnifications as well. With this, other researchers also used clean data to bring high accuracy. Further, separate accuracy of same modalities are given instead of consolidated cross-dataset performance score [21]. This prevents the reflection of clinical generalizability across same modality. In order to overcome these shortcomings, we suggest to propose comprehensive multi-imaging model that ensures diagnostic accuracy and clinical trust.

To conclude, the currently BC diagnostic researches based upon single or dual modality lack the ability to operate in the real-time systems, where they are to process complementary clinical information (MG, US, and HP). Current researches on tri-modality is limited and restricted to HP magnifications, and lack of explanatory capacity, thus reducing clinical trust and generalizability. These limitations, therefore, highlight an urgent need to have a clinically admissible, explainable tri-modal DL framework that can provide BC diagnosis and build clinician trust.

Chapter 3

Proposed Methodology

This chapter outlines the proposed method for classifying BC through explainable weighted DL features of multimodal imaging (US, MG and HP). Fig is the overview of our methodology. Firstly at individual level, each modality is evaluated and then through fusion the model achieved single accuracy.

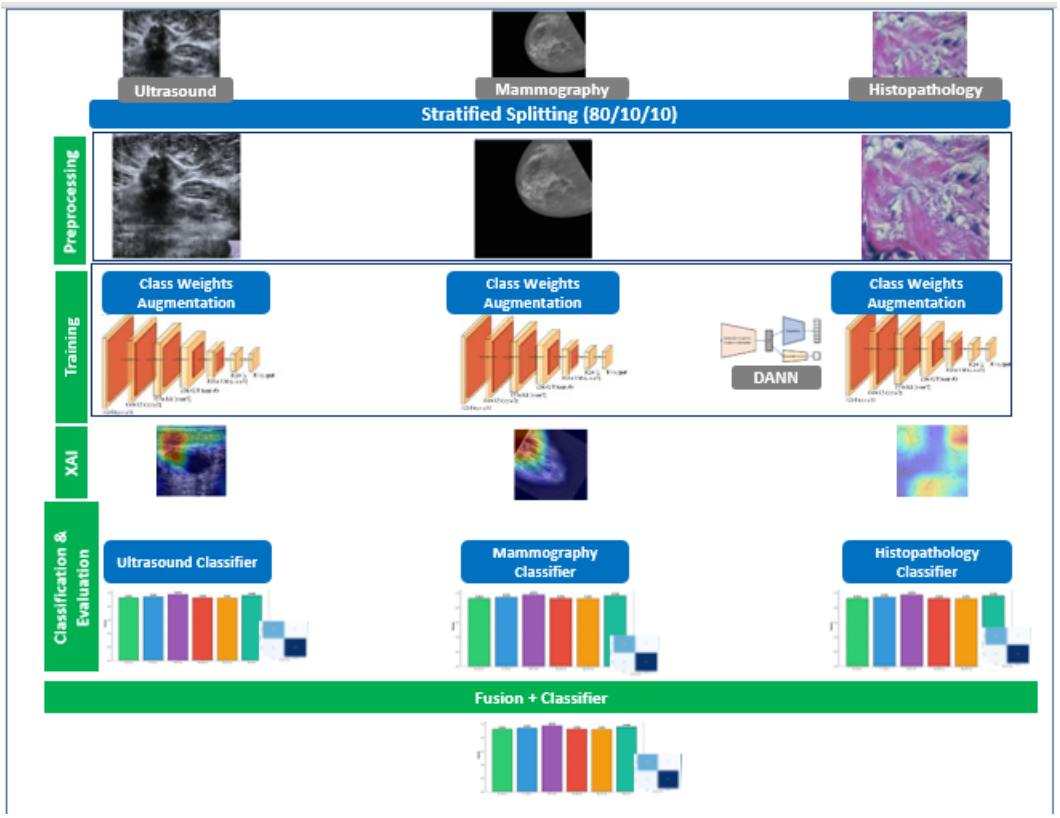


FIGURE 3.1: Overview of Proposed Methodology

3.1 Data Acquisition

Datasets used in this research are present in public directories. To investigate BC across multiple modalities (MG, US and HP), these datasets are sourced from established repositories and research publications that serve as a foundation for developing and evaluating novel diagnostics algorithms. For MG, INbreast dataset [31] that has images of full field digital mammograms, MIAS database [32] , which is most commonly used in researches. In addition, DDSM [33] and its successor, the CBIS-DDSM [34] is added to ensure the holistic representation of mammographic findings.

Mostly US are used in combination with mammograms. This is because ultrasound can detect small solid mass and cysts which cannot be found in the screening mammographic images.In this study, ultrasound images are taken from three distinct sources: Breast Ultrasound Images [35], BUS_UCLM [36] and BUS_UC [37]. These dataset offer diverse perspective of ultra-sonography of BC tissues. Lastly, for Histopathological analysis, a well known dataset is used that is BreakHis [38]. It contains microscopic images with varying magnifications of 40X, 100x, 200X and 400X. The key consideration in these modalities are the heterogeneity in the class labels.

Mammography datasets (INbreast, DDSM, MIAS, CBIS-DDSM) encompasses three class labels that are normal, benign and malignant. In ultrasound, Breast Ultrasound [35] Images, BUS_UCLM [36] uses three classification system that categorize into benign, malignant and normal while the histopathological images [38] and BUS_UC [37] uses binary classification scheme of benign and malignant. This discrepancy necessitates careful attention while evaluating the model. To develop a nuanced model, these classifications are synchronized and to make it streamline, only two are used that is benign and malignant.

Ultimately, systematic approach aims to improve the accuracy and reliability of the automated BC detection and classification at the early stage for effective treatments. It will be further preprocessed and will go through pipeline to obtain best prediction outcomes. In below sections, each dataset is separately explained with diagrams and tables. Total images obtained from these are 34,207.

3.1.1 Dataset of Ultrasound

US is clinical breast examination because it is critical in BC detection and management. It is significant in a number of major benefits. One, US does not emit any radiation and hence it is not harmful when used repeatedly especially in younger and pregnant women. Secondly, US is effective at distinguishing solid and cystic masses, a paramount distinction that MG is not able to make [32]. It is possible to prevent unnecessary biopsies of non-malignant cysts. Thirdly, it is effective especially in imaging of dense breast tissue that could conceal tumours on mammograms [39].

With thick breasts, it is able to detect the lesions, which would be obscure. In addition, it is easily accessible, comparatively cheap and able to be conducted in real-time, and enables the dynamic evaluation of the breast tissue. It directs biopsies, allowing one to target suspicious areas with pathological examination. The US can also be used in measuring the response to treatment where the clinician can determine the effectiveness of the treatment like chemotherapy. Following US datasets are enlisted that are used in our model and shown in the Fig 3.2.

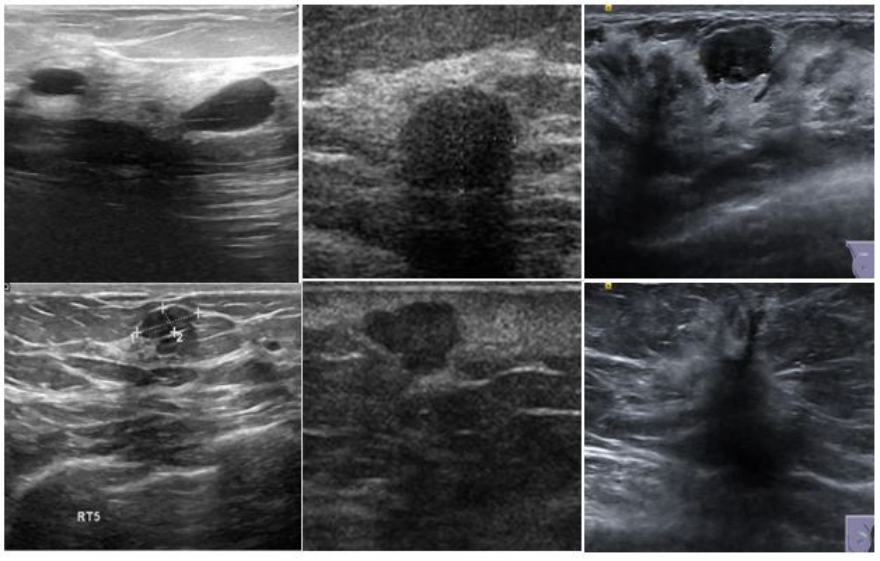


FIGURE 3.2: US Images - Top(benign) and Bottom(Malignant): Left- Breast Ultrasound Images dataset, Middle- Breast Ultrasound Dataset, Right- Breast Ultrasound Lesion Segmentation Dataset

i. Breast Ultrasound Images(BUS) [35] : The samples are collected in 2018 from 600 female patients. Total images are 780 with image size of 500 x 500

pixels in Portable Network Graphics (PNG) format. The original images are categorized into three classes which are benign (437) and malignant(210).

ii. Breast Ultrasound Dataset(BUS_UC) [37]: The collection of US images of breast tumours as either benign or malignant. All images are in 256 x 256 pixels resolution. The total 811 images are divided into benign tumour images (358) and malignant tumour images(453).

iii. Breast Ultrasound Lesion Segmentation Dataset(BUS_UCLM) [36]: The dataset is the collection of 38 patients that comprises of 683 images. It is taken from Siemens ACSUSON S2000 TM Ultrasound System in year 2022 and 2023. It is categorized into three labels as benign, malignant and normal with size of 174, 90 and 419 respectively. Mask images are also provided as RGB files.

The table 3.1 shows the summary of the ultrasound dataset.

Dataset	Benign	Malignant
BUSI[35]	437	210
BUS_UCLM[36]	174	90
BUS_UC[37]	358	453
Total	**969**	**753**

TABLE 3.1: US Dataset Distribution

3.1.2 Dataset of Mammography

MG remains the cornerstone of BC screening, delivering high-resolution images that enable the detection of masses, calcifications, and architectural distortions indicative of benign or malignant pathology [39]. Nevertheless, interpretation is challenged by the visual similarity between benign and malignant findings, varying breast density, and imaging noise or artifacts.

To support the systematic development and evaluation of CAD systems, a consolidated dataset has been created in 2024 by merging images from the public repositories that is combination of INbreast [32], MIAS [31], and DDSM [33].

This comprehensive collection has been pre-processed and enhanced for DL applications, facilitating robust BC classification studies. INbreast, MIAS, and DDSM [34]: The most recent update of the MG dataset provides a combination of imaging data that has been collected in the INbreast, MIAS, and DDSM repositories and used to train DL models. It consists of 26,602 preprocessed files, which were grouped into three types: malignant (13,710 instances), benign (10,866 instances) and normal (2,026 instances).

The typical images have been collected in the collection of DDSM thumbnail images, and contrast limited adaptive histogram equalisation (CLAHE) to improve image quality has been applied to all the images in the collection. Fig 3.3 shows combined images of MG dataset.

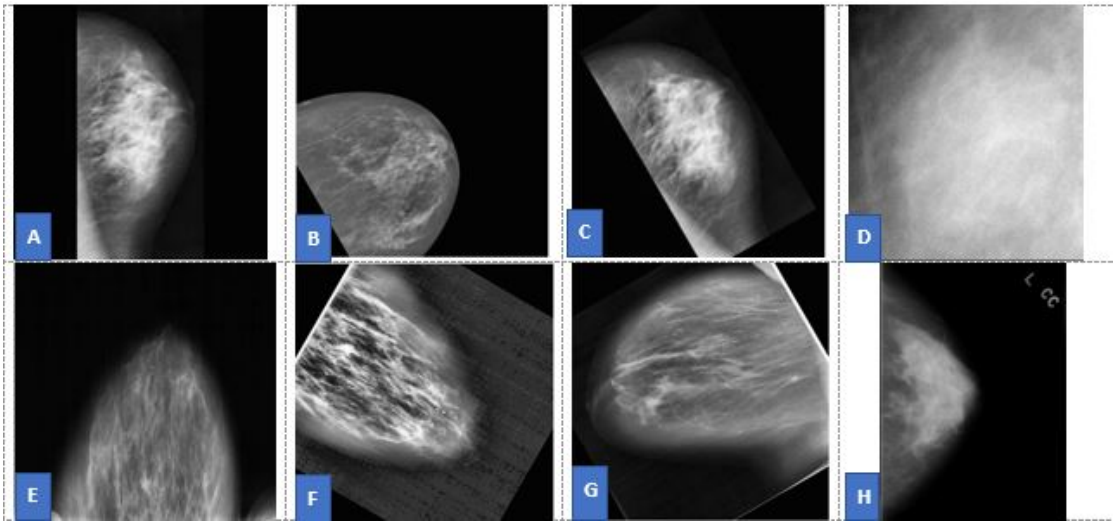


FIGURE 3.3: Preprocessed MG Images: Top Row - Benign and Bottom - Malignant

3.1.3 Dataset of Histopathology

The imaging mode of HP is a key component of the ultimate diagnosis of BC, because it allows the microscopic analysis of the structure and morphology of tissues [40]. Compared to radiological techniques like MG and US that only helps in identifying suspicious abnormalities, histopathological analysis implicates affirmative data about the existence, nature and aggressiveness of a tumour. Pathologists normally acquire tissue samples by conducting biopsy and then staining them and viewing them under a microscope.

Through this process, structural abnormalities [30], nuclear aberration, mitotic activity as well as other pathological characteristics that characterize benign growths and malignant tumours can be identified. Such that, HP is considered as the gold standard in the diagnosis of BC. The growing number of digitized histopathological slides has been used to create CAD systems that can help pathologists by alleviating manual material and enhancing diagnostic reliability.

Nevertheless, automated analysis of the HP images has a high intra-class variance, low inter-class difference, and staining variation, and more complex tissue patterns, which make analysis of HP images inherently difficult. The differences in the degree of magnification also complicate the extraction of features since cellular features can look substantially different at varying magnifications [41].

Thus, high-quality datasets with a variety of tissue samples are necessary to build a powerful DL model able to generalize the variations. The scarcity of large datasets of public histopathology can be explained by the extreme complexity of obtaining, maintaining and distributing such medical data. In contrast to non-invasive radiology, in histopathology, physical samples of tissues are necessary in a biopsy or operation [42], which limits volume.

The resulting whole-slide images are huge gigapixel files which are very expensive to store and manipulate, and manual annotation by pathologists is very time-consuming and expensive. Moreover, strict patient privacy laws (such as HIPAA/GDPR) and intricate data de-identification procedures, and institutional regulations regarding data ownership present a high legal and ethical impediment to making data public.

BreakHis dataset [38]: The dataset is built with collaboration with P & D Laboratoy-Pathological Anatomy and Cytopathology, Brazil. It consists of 9109 microscopic images of breast tumour tissues collected from 82 patients. It uses different magnifying factors of 40X, 100X, 200X and 400X. Table shows distribution of dataset.

Robust DL models therefore demand extensive, varied, high-quality datasets, which are scarce.

This scarcity stems from the difficulty of obtaining, storing, and annotating large gigapixel scans, coupled with stringent privacy regulations (HIPAA/GDPR) and

TABLE 3.2: BreakHis Dataset: Benign Subtypes Distribution by View

View	Adenosis	Fibroadenoma	Phyllodes Tumor	Tubular Adenoma	Total
40X	114	253	109	149	625
100X	113	260	121	150	644
200X	111	264	108	140	623
400X	106	237	115	130	588
Total	**444**	**1,014**	**453**	**569**	**2,480**

TABLE 3.3: BreakHis Dataset: Malignant Subtypes Distribution by View

View	Ductal Carcinoma	Lobular Carcinoma	Mucinous Carcinoma	Papillary Carcinoma	Total
40X	864	156	205	145	1,370
100X	903	170	222	142	1,437
200X	896	163	196	135	1,390
400X	788	137	169	138	1,232
Total	**3,451**	**626**	**792**	**560**	**5,429**

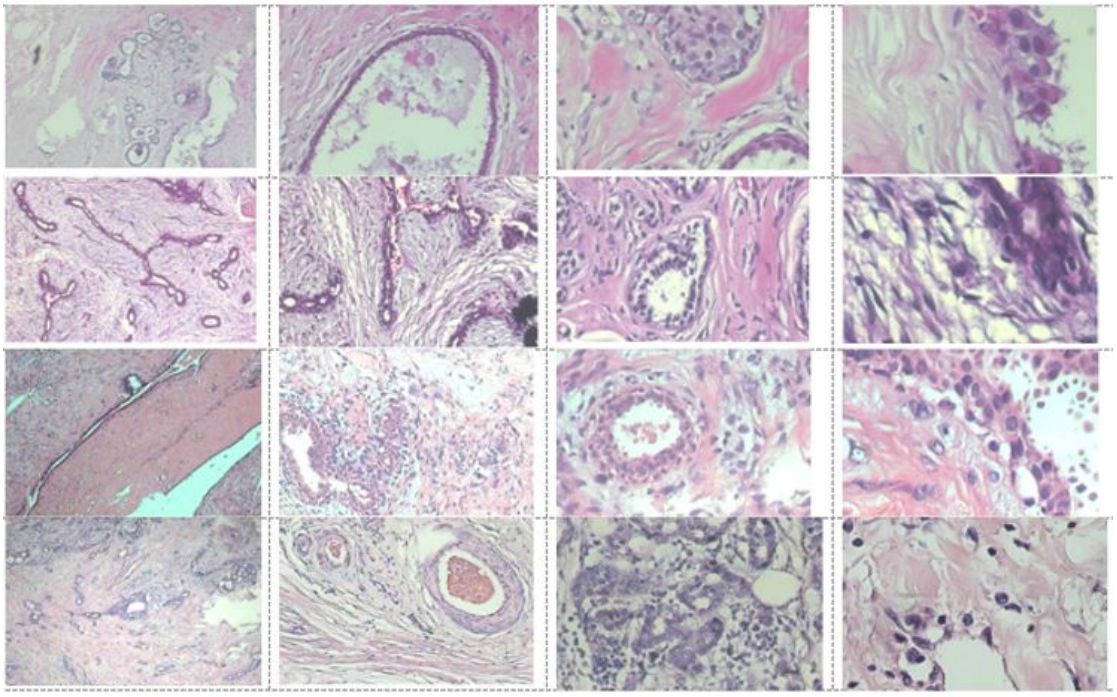


FIGURE 3.4: Histopathology Images - Benign : From Left to right - 1st Column: Adenosis, 2nd Column: Fibroadenoma ,3rd Column: Phyllodes Tumor and 4th Column: Tubular Adenona

institutional ownership constraints [42]. Unlike MG and US, only available dataset for HP is BreakHis [38]. The dataset is built with collaboration with P & D

Laboratoy-Pathological Anatomy and Cytopathology, Brazil. It consists of 7,909 microscopic images of breast tumour tissues collected from 82 patients. It uses different magnifying factors of 40X, 100X, 200X and 400X with four levels of benign and malignant tissue types.

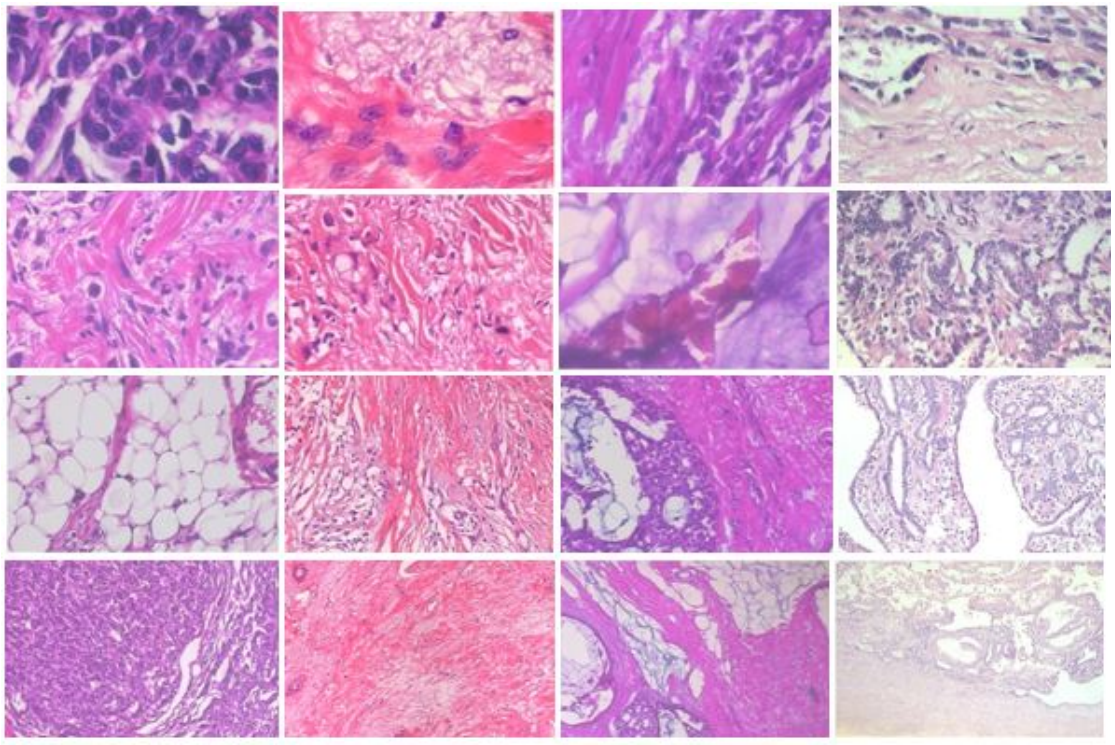


FIGURE 3.5: Histopathology Images - Malignant: 1st Column: Carcinoma ,2nd Column: Lobular Carcinoma , 3rd Column: Mucinous Carcinoma and 4th Column: Papillary Carcinoma

Table 3.2 and 3.3 shows distribution of dataset. Lesions are collected after biopsy and then sent for microscopic visualization.

The use of increased magnification in the breasts imaging multiplied the images according to the patient by nearly four times leading to a significant number increase; benign cases were increased by approximately 2,480, and malignancy cases were increased by approximately 5,429. Fig 3.4 and Fig 3.5 shows images benign and malignant images of HP respectively at different magnification level.

To simplify the analysis of multimodal data and to expedite the training-validation partitioning, the three modality-specific data sets were created into one representation. Each single workbook in the excel sheet, that pertained to US, MG and HP respectively were saved to ensure traceability and the contents of the workbooks were consolidated into a master sheet that contains the fields of `patient_id`,

`path`, `modality` and `label`. This centralized model allows stratified division of modalities and standard preprocessing when developing models.

3.2 Data Splitting

The consolidated dataset shows high disparities across modalities and classes (Table 3.4). This extreme heterogeneity—both in modality representation and class balance—creates significant challenges for data splitting. The proposed methodology followed the data partitioning technique that follows three fundamental requirements:(1) Conserve the allocations of modalities to every partition. (2) Preserving the labels ratio within each modality across splits (3) Make sure that separation is at the patient level to prevent accidental data leakage.

TABLE 3.4: Consolidated dataset distribution across modalities and labels

Modality	**Benign**	**Malignant**	**Total**
Histopathology	2,480	5,429	7,909
Mammography	10,866	13,710	24,576
Ultrasound	969	753	1,722

The correct division of the datasets is a key to the creation of the objective and generalizable DL models. Much of the literature is based on the major part on K-Fold Cross-Validation [25] or Shuffle Split / Random Subsampling [26].

The random partitioning may cause over representation of specific labels in the training sample and under representation in the validation or the test sample, which results in the overly optimistic performance estimates. Similarly, despite the advantages of cross-validation in smaller cohorts, it is more expensive in terms of computational costs, and may still be insufficient in terms of guaranteeing proportional representation in multimodal data [43].

In order to address these weaknesses, the research embraces multi level stratified splitting that ensures every partition; training, validation and test maintains the class proportions found in the original set. Firstly, images are grouped by modality and label. Within each modality-label group, stratify by patient. This ensures that it is not biased towards the majority classes and that the performance is

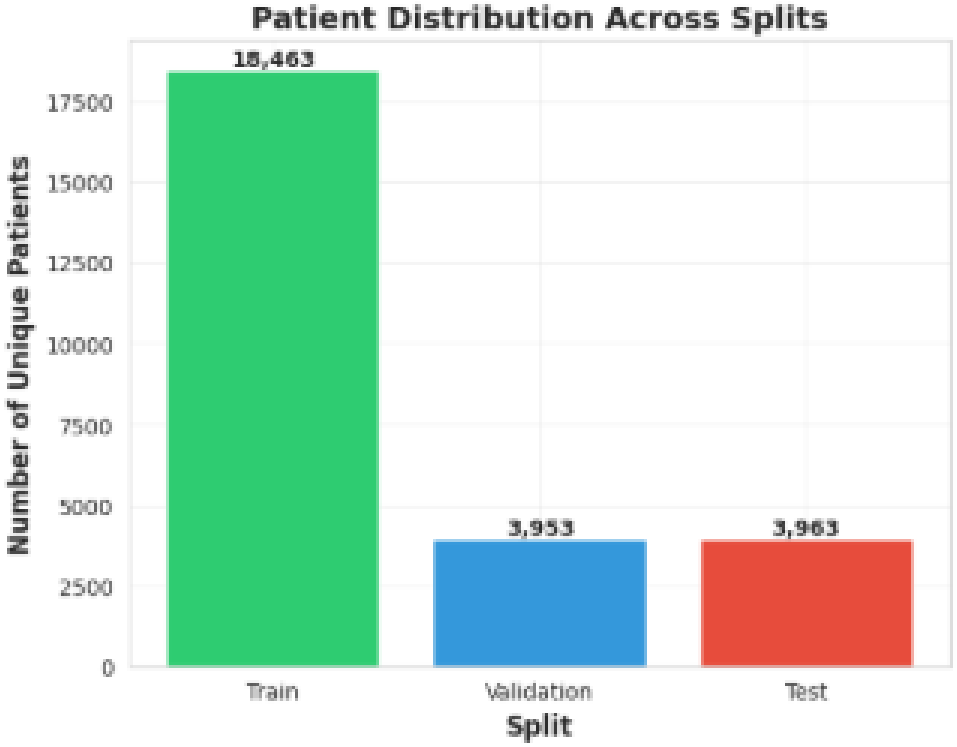


FIGURE 3.6: Patient Distribution across Splits

assessed evenly among the diagnostic categories while supporting the consistency and repeatability of the experimental outcomes.

In our dataset, 6 modality-label combinations are found:(1) histopathology_benign: 2,480 images from 24 patients(2) histopathology_malignant: 5,429 images from 57 patients(3) mammography_benign: 10,866 images from 10866 patients(4) mammography_malignant: 13,710 images from 13710 patients(5) ultrasound_benign: 969 images from 969 patients(6) ultrasound_malignant: 753 images from 753 patients Comparison is drawn between three different split ratios (80/10/10, 70/15/15 and 60/20/20) as demonstrated in Table 3.5 through employing hierarchical patient-level stratification in order to find the most appropriate split ratio for developing and evaluating the model.

The main selection criterion was that modality distributions do not change between all splits, which were tested with the chi-square test of homogeneity 3.1. Some secondary considerations were full representation of all modality -label combinations in both the validation and test set, and large sample sizes to demonstrate reliable performance estimation.

TABLE 3.5: Evaluation of Different Train-Validation-Test Split Ratios with Statistical Validation

Split Ratio (T/V/Te)	**T**	**V**	**Te**	**Total**	**T %**	**V %**	**Te %**	**Chi-Sq χ^2**
80/10/10	27,526	3,228	3,453	34,207	80.47%	9.44%	10.09%	3.179
70/15/15	24,153	4,900	5,154	34,207	70.61%	14.32%	15.07%	6.007
60/20/20	20,859	6,539	6,809	34,207	60.98%	19.12%	19.91%	18.12

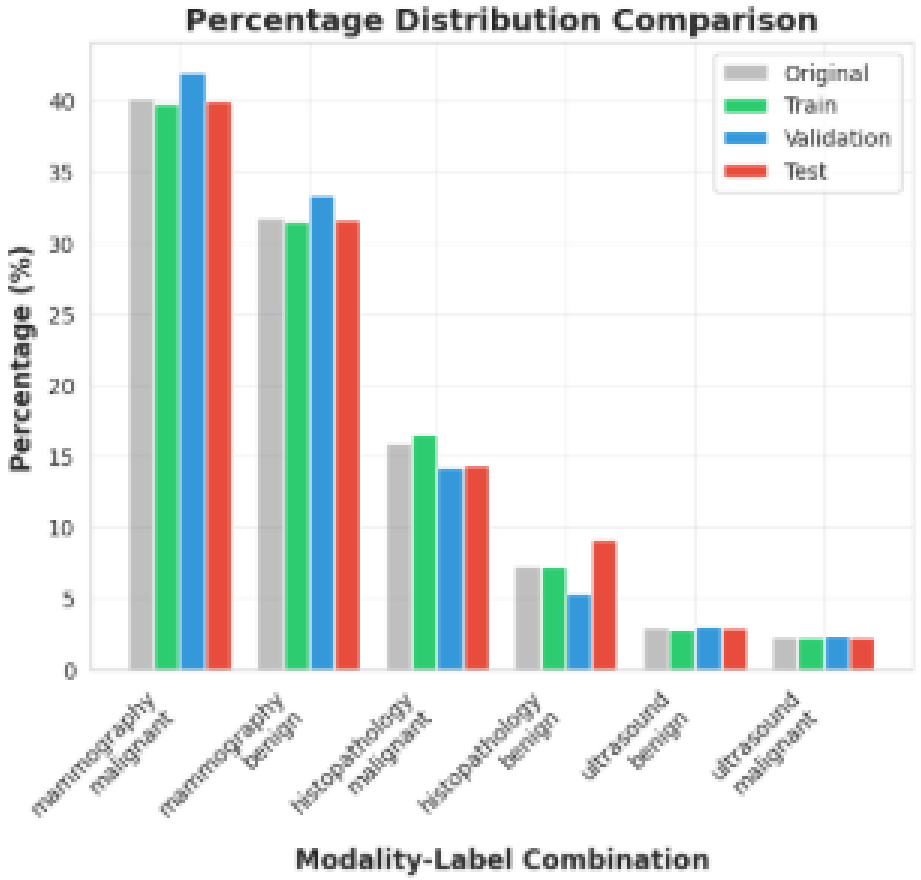


FIGURE 3.7: Percentage Distribution in Modality-Label Combination

Although all three splits retained full representation of all modalities-label combinations, the 70/15/15 split had come close to statistical significance and 60/20/20 split had significant deviation on the original distribution. The results of these experiments allowed choosing the optimal proportion of 80/10/10 split as the starting point of all the further experiments. The purpose of this choice is that the validation set and test set are representative of the underlying data distribution, and thus give a more reliable and generalizable performance estimate. As consistency is even more important in a tri-modal framework, where disproportionate modality representation can bias the decision-making process. Fig 3.7 .

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{3.1}$$

The equation 3.1 is statistical validation of the difference in both original and splits. Chi-square test evaluates whether the modality distribution in each split differs significantly from the original dataset distribution. Division by expected value shows that if expected value is smaller the result has a large value. It means how much the data deviated from the expected value. So smaller it is the smaller are the distances are.

3.3 Preprocessing

In pre-processing, minimal steps are adopted like de-noise, contrast lightening, resizing, channel conversion and standard normalization.

Moreover, methods like region of interest (ROI), tissue segmentation, masking has side effects of biased observer region, loss of anatomical region etc. Each modality undergoes its own preprocessing pipeline which is also mentioned in their respective section.

3.3.1 Preprocessing of Ultrasound

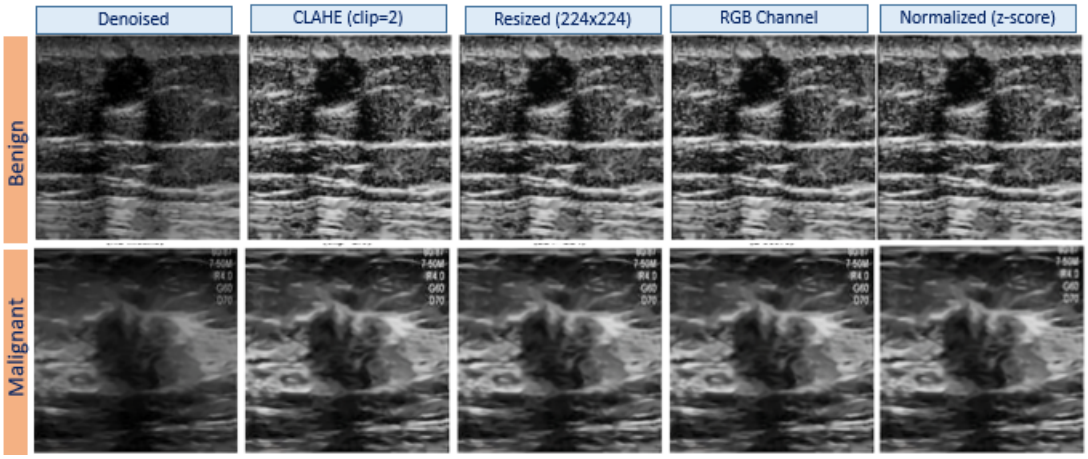


FIGURE 3.8: Ultrasound Preprocessing Pipeline

The vast use of US imaging in the diagnosis of BC has been attributed to its ability to differentiate between cystic and solid masses, and ability to give real time

visualisation of soft tissues. However, speckle noise, low contrast, intensity inhomogeneity, and the effect of operator-dependent acquisition settings all adversely affect ultrasound images[44].

Such issues often cause difficult lesion boundaries and textures of the lesions to be hard to detect and classify with fully automated systems. As a result, a preprocessing pipeline has been established to improve the quality of images and to ensure that the features which are of diagnostics value are well preserved to be used in the later training of the model.

In US, speckle noise and low contrast imaging is often the obscure to identify the lesions boundaries. To cater it, median blur of k=3 is incorporated to suppress the effect of the high frequency speckle noise while preserving the sharp edges of the crucial masses, which are critical for assessment of BI-RADS[45].

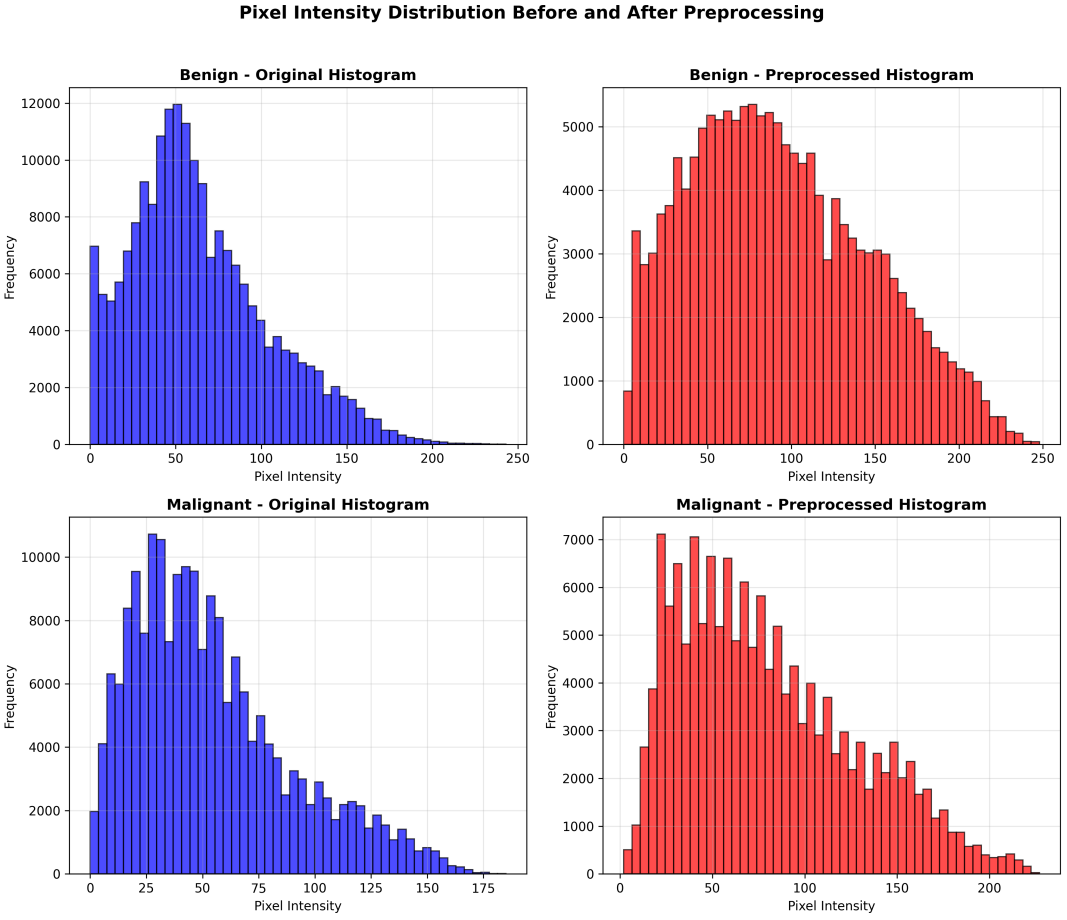


FIGURE 3.9: Ultrasound Histogram Comparison

Then, noise reduction is followed by adding Contrast Limited Adaptive Histogram Equalisation (CLAHE) to enhance the local contrast and boost the small changes

in tissue echogenicity. As US images are often obtained on various machines at different gain levels, CLAHE makes it much easier to normalise variations in illumination but avoids over-enhancement of noise. To retain natural anatomical features, image is resized to 22 x 224 x 3 RGB channels. Normalisation of z-score on a global basis is then carried out to normalize pixel intensity distributions to have a zero-mean unit-variance intensity distribution along with clipping and color space transformation. Fig 3.8 shows step wise preprocessing of US images. Together, these preprocessing methods are used to alleviate artifacts, increase lesion appearance, equalize cross image variation, and increase the dependability of automated BC detection in US imaging.

The pipeline in Fig 3.8 starts with median blur, where pixels of images are sorted and then the mid value is selected. This smoothens the features, further CLAHE is implemented that increased its contrast as shown in the Fig 3.8 and Fig 3.9. After it, it is resized to 224x224 and three RGB channels. Initially the image is in 1D form, replicating it three times like R=G=B , makes it 3D. Finally z scoring applied on each channel.

Global statistics is average pixel value of approximately 88.8 to 90.2.

3.3.2 Preprocessing of Mammography

MG continues to be the most common imaging method used in early diagnosis of BC and particularly in detection of micro-calcium, architectural anomalies, and unequal tissue patterns.

Despite their clinical significance, mammographic images have many weaknesses such as low signal-to-noise ratio, overlapping tissues, and inconsistency in the thickness of the breast and X-ray exposure [46].Therefore, a powerful preprocessing pipeline is required to improve the quality of images.

The dataset used in this thesis is already pre-processed with CLAHE, making our subsequent steps confined to de-noising, resizing, conversion and normalization. Median blur is often used to reduce the number of impulses without damaging the edges, which improves the image quality that must be provided to outline thin anatomical features.

Maintenance of the sharp and well defined edges will play a major role in preventing misidentification of lesion margins and micro-cavity. Many studies have focused on removing the pectoral muscle in the MLO [47] scans to simplify the analysis processes; this approach is not feasible when used in normal clinical practice due to the existence of anatomical differences. By contrast, the methodology in question focuses on global standardization of intensities, so that the distributions of pixel intensity across examinations are normalized. The pipeline of preprocessing of MG is shown in Fig 3.10. The latter approach helps to reduce sensitivity to changes in breast thickness, compression, and exposure, thereby, enabling the creation of more generalized and flexible learning models.

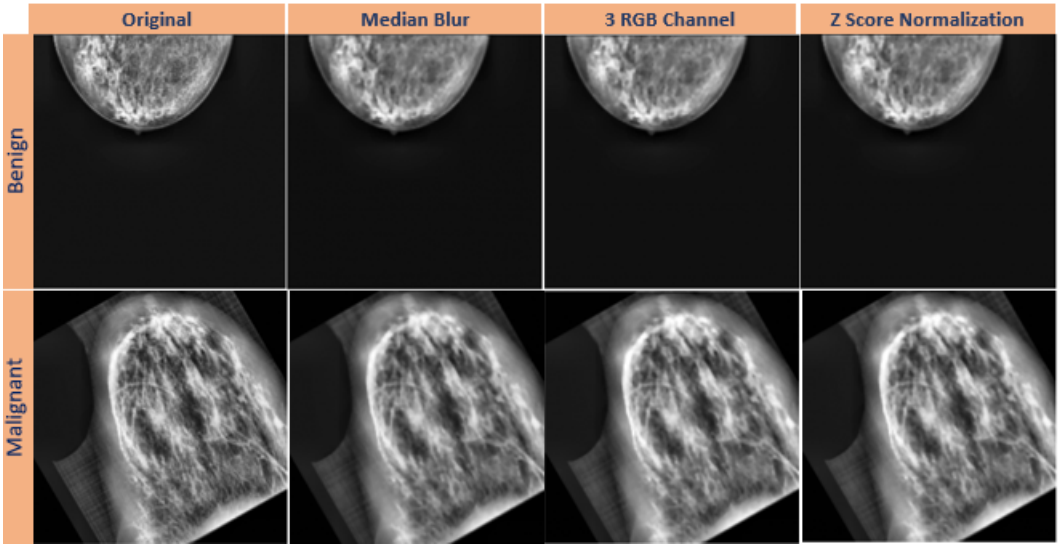


FIGURE 3.10: Mammography Preprocessing Pipeline

It is also also resized to 224 x 224 x 3 RGB channels, where R=G=B is in 3D format, replicating 1d grayscale image.

3.3.3 Preprocessing of Histopathology

The HP images are microscopic views of the breast tissue that retain cellular structures, nuclear position and tissue architecture which are all essential in determining the final diagnosis oncologic diagnosis. Unlike radiological modalities, HP is very much dependent on colour composition based on H&E staining with the hematoxylin emphasizing the nuclei whereas eosin emphasizing the cytoplasmic and stromal elements.

Traditional approaches to normalisation of stains e.g. handcrafted techniques, LBP, the Macenko and Reinhard methods, attempt to normalise these variations but can potentially induce artifacts, mask tissue textures, or hide quiet diagnostic information which has clinical meaning [48]. These transformations may be biased when they affect the natural biological patterns of the original slides as it trains deep learning models.A minimalist preprocessing strategy is implemented in the study in order to maintain the originality of the microscopic structures.

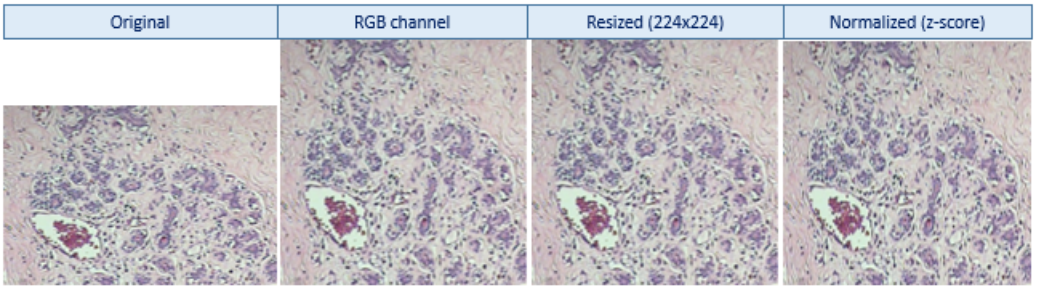


FIGURE 3.11: Histopathology Preprocessing Pipeline

Raw histopathological images are not aggressively manipulated in terms of colour distributions, rather they are stored at their original size, but resized to patches of 224, 224 pixels that will make model compatible with the DenseNet121 model. More significantly, patch-based learning allows the model to concentrate on localised cellular structures which are major tumour markers in tumour grading. It is then converted to global z -score normalisation to standardise distributions of pixel intensity across images. This normalisation stabilizes the training process because the feature activations are kept in a standard numerical range thus enhancing gradient flow as well as speeding up the convergence.Fig 3.11 shows pipeline of HP preprocessing. This minimalist approach, is designed to do as less interference with DANN [49] that will be used with DenseNet for feature extraction [50].

Instead, changing the visual representation that may have damaged the significant cells and morphology, in our proposed methodology, we have trained model in a way that captures intricate features while discriminating the dependent domain variants. This interaction between low preprocessing and adversarial feature learning is one that allows the proposed framework to generalise better with the

diversity of histopathological data, and in the end boost the classification reliability. After preprocessing, data splitting is carried out that is explained in next section.

Global mean (μ) and standard deviation(σ) for each modality in training is computed through Welford's Online Algorithm to maintain memory efficiency. Every image is z-score normalised following formula shown in equation 3.2. It is implemented in all modalities.

$$z = \frac{x - \mu_{\text{modality}}}{\sigma_{\text{modality}}} \tag{3.2}$$

3.4 Feature Extraction

For this study, DenseNet-121 to act as the feature-extractor in the BC pipeline due to its dense connectivity that would enable feature reuse in all the previous layers [51]. This architectural scheme retains low level spatial detail, such as edges, textures and micro-structure, which is vital in detecting subtle lesions and micro-calcifications, whilst providing a large scale gradient flow which helps address the vanishing-gradient problem [52]; this is especially beneficial during training on small medical datasets. The model naturally provides multi scale representation by combining local patterns with more semantics (higher) thus allowing the detection of small scale abnormalities (e.g. calcifications in MG) and larger lesions (e.g. masses in US or in HP).

Further, the effective parametrization of DenseNet -121, as well as its transfer-learning qualities - due to pre-training on large generic datasets - enables much faster convergence and improves generalisation across BC datasets [53]. We therefore use slightly modified versions of DenseNet-121 as feature extractor in each imaging modality that we are going to be implement in our system. Standard DenseNet framework is shown in the Fig 3.12.

It has four blocks with transition layer in between. Each block has densely connected For feature extraction DenseNet121 is used. It is especially effective in feature extraction of BC images since the architectural nature presents an explicit

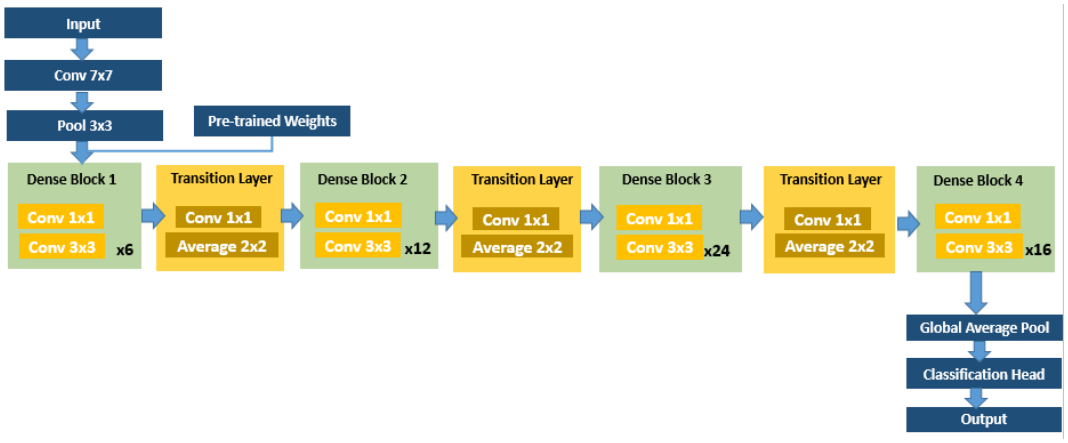


FIGURE 3.12: Standard DenseNet-121 Framework

solution to most of the difficulties encountered in medical images such as small lesion manifestation, small datasets, and fine-tuning feature extraction at multiple scales. Standard DenseNet framework is shown in the 3.13.

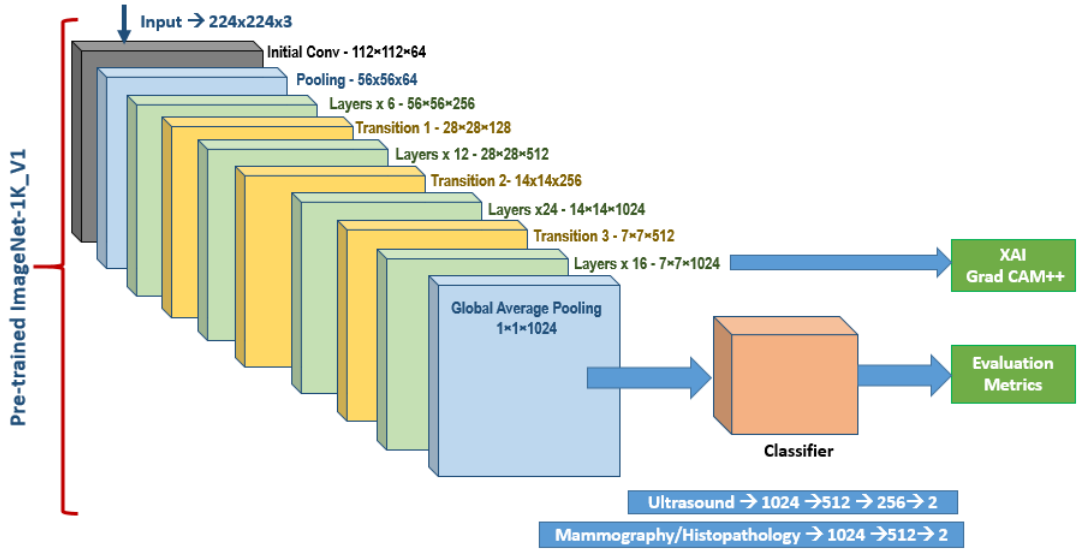


FIGURE 3.13: Architecture of the Proposed Model

Pre trained imageNet is used across each modality providing trained features such as edges and shapes. The entire feature extraction backbone including the initial convolution layer initialized with imageNEt-1k v1 weights. The input dimension of 224x224x3 is pre-requisite to achieve compatible with pre-trained backbone.

After training, the last dense block that captures major discriminative patterns and produces feature maps of 7x7x1024. These are used for localizing the region in the image and the heat map is generated from this layer using GradCam. Each modality is explained separately in further sub sections.

3.4.0.1 Ultrasound Specialist Model

The US specialist model uses standard DenseNet-121 structure after necessary balancing. The class weights which are calculated in the preprocess phase are used during training along with augmentation to alleviate the impact of imbalance in the classes. The proposed model will use a class-weighting mechanism based on the empirical frequency distribution of the classes in the training set, that is, benign instances will be weighted with about 0.88 weight and the malignant instances will have a weight of about 1.144; this is to ensure that the learning algorithm will give enough attention

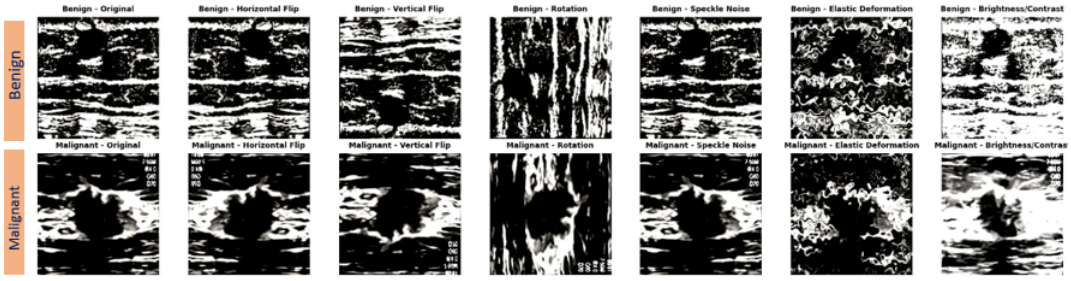


FIGURE 3.14: Ultrasound Augmentation

to the minority malignant class. Geometric augmentation of rotation in 90 degree, 180 degree and 270 degrees and intensity/speckle noise augmentation is implemented .It is also demonstrated in Fig 3.14. After passing by dense layers, global average pooling turns 7x7x1024 feature maps are turned into 1x1x1024. This architecture provides a multi-scale feature- representation which is computationally efficient in nature. The model is classified with a custom classifier head for binary classification.

The custom classifier is the major part where it is specially designed to overcome the US characteristic of speckle noise and variety of contrast. The custom classifier has also multiple layers, Starting from 1024 features , 512 ,256 till 2 binary classification. Training curves also show that Fig 3.17. The comprehensive model behaviour over 20 epochs shows the training/val accuracy that is steadily increasing which shows it learning pattern and generalizing without over-fitting. Both sets loss is decreased over time. Although validation loss fluctuates more but lowers than training loss, which suggests the model is not memorizing but capturing

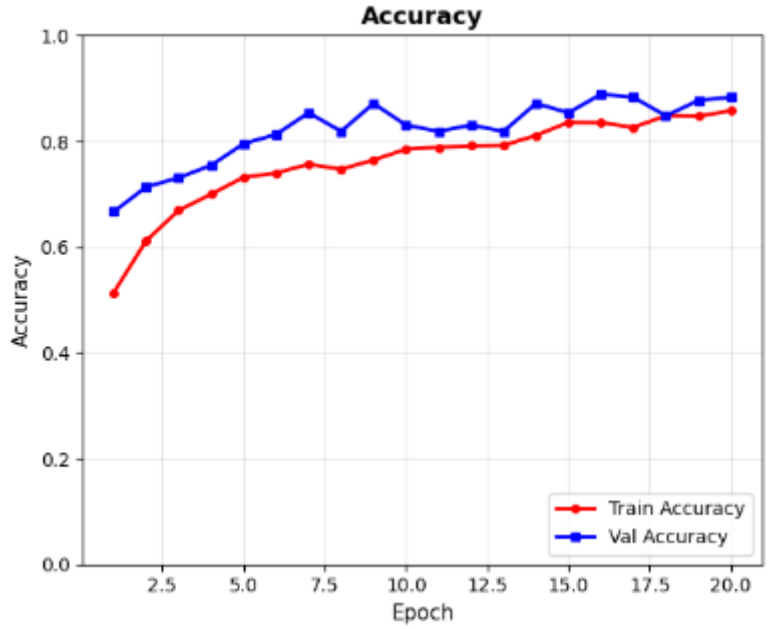


FIGURE 3.15: Ultrasound Training Curves - Accuracy Curve

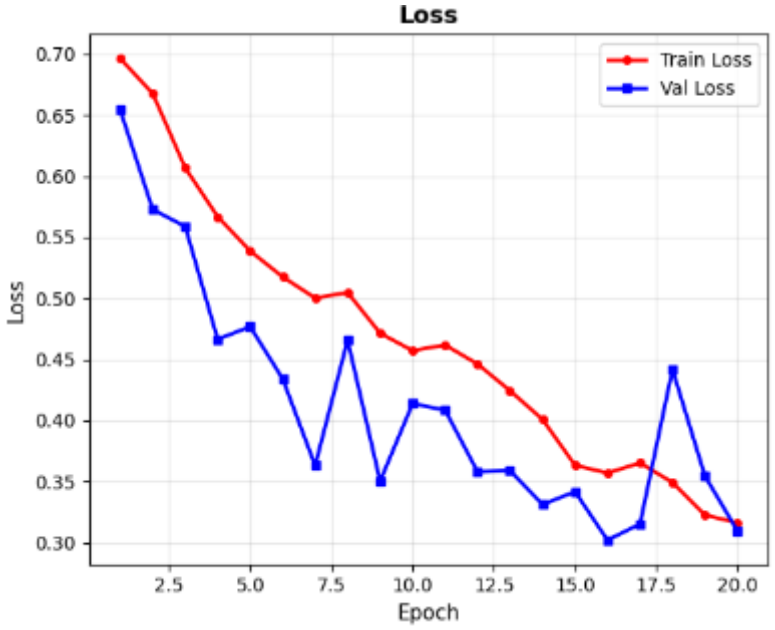


FIGURE 3.16: Ultrasound Training Curves - Loss Curves

the meaning patterns. The F1 score and ROC-AUC curve shows improvement in both classes. Learning rate remain stable until epoch 12 and then drops to $5 \times 10^{\neg}5$. The batch size used is 32. This step-down helps stabilize training and fine-tune the weights.The custom classifier takes 1024-dimensional feature vector and applies dropout regularisation and linear transformation to compress it into 512 and 256. Each compression if followed by a RELU activation function to ensure non-linear feature learning while maintaining robustness against over fitting.

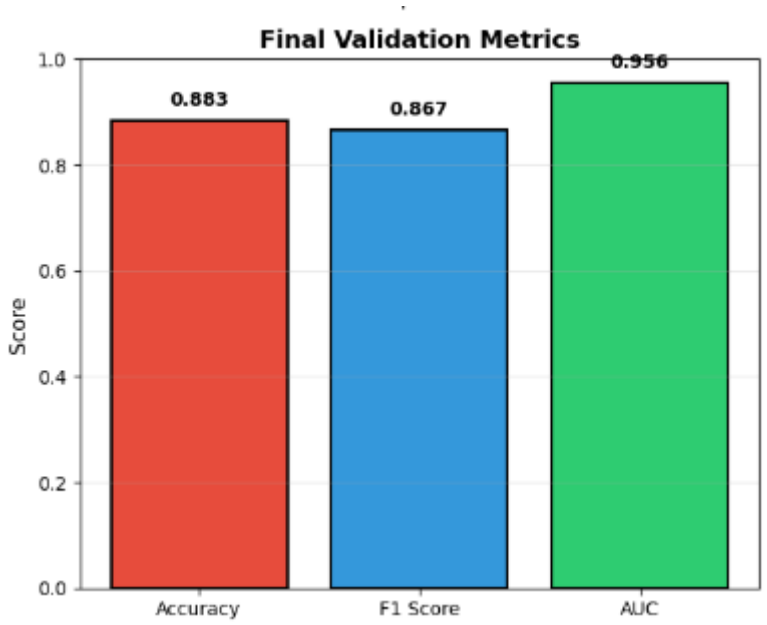


FIGURE 3.17: Ultrasound Training Curves - Validation Evaluation Metrics

The purpose is to retain the most discriminative primary features that contribute to final decision and eradicate redundant features. The final layer maps 256 feature vector to 2 binary output nodes (benign and malignant).

3.4.0.2 Mammography Specialist Model

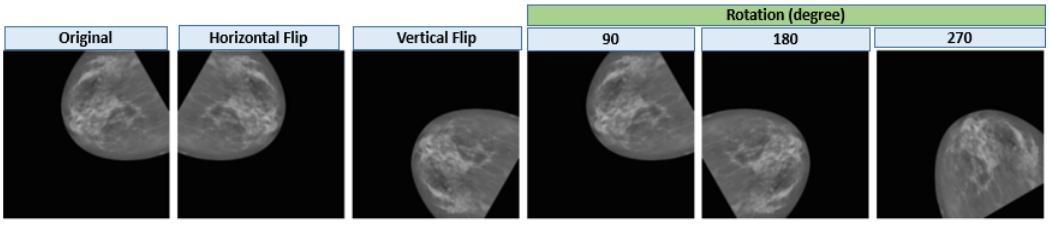


FIGURE 3.18: Mammography Minimal Geometric Augmentation

The MG specialist model is based on the DenseNet-121 architecture, incorporating into itself a focused approach to solve specific problems that come with the analysis of mammograms. Input images are passed sequentially through the backbone. This setup helps extracting multi scale features relevant to mammographic features, like, for example, tissue density and micro-calcifications, increasing the discriminatory capacity of the model.

To reduce the problem of class imbalance, the model has an associated class weighting approach according to the empirical class distribution in the training cohort.

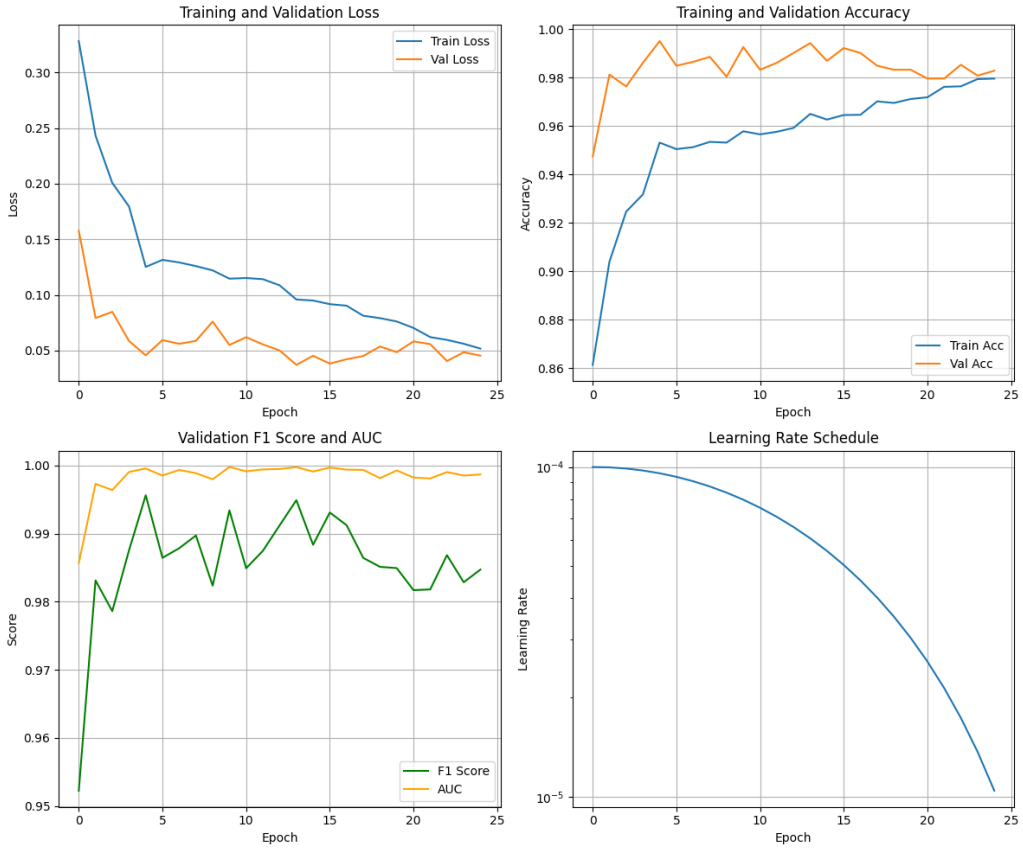


FIGURE 3.19: Mammography Training Curves

Benign observations are weighted at around 1.131 while malignant observations are weighted at around 0.896, thus ensuring the learning algorithm gives due importance to the rare, but clinically relevant malignant class. The training regime uses only geometric augmentations (i.e., random horizontal flips , vertical flips, 90/180/270- degree rotations) to maintain important intensity values and structural integrity of the mammographic images as shown in the Fig 3.18

The training curves of specialist model is shown in Fig 3.19. The curves show nearly perfect classification. Training and validation loss are constantly decreasing that shows model is generalising well, learning meaningful patterns and not over-fitting. Accuracy on both sets increasing. Validation set accuracy is more than a training accuracy that suggests the model will perform well on unseen data. Another metrics of F1 score and AUC remained high through the training which demonstrated excellent balance between sensitivity, specificity and strong distinguishing ability between benign and malignant. The sub classifier has a strong regularization scheme that consists of Dropout layer with probability 0.5, followed

by linear transformation changing the dimensionality from 1024 to 512 features, followed by ReLU activation function, Dropout layer with probability equal to 0.4 and the final linear mapping of two outputs (benign vs. malignant). This architecture is expressly designed to handle the high resolution, complex textural patterns of mammographic imagery. The Adam optimizer with a starting learning rate of $1 \times 10^{-}4$ was used to optimize the model, with a weight decay coefficient of $1 \times 10^{-}4$. A cosine annealing learning rate schedule was utilized and a 10 epoch early stopping criterion was used. The training trajectory revealed that there was effective learning in the training since the performance metrics at the first epoch were significantly higher than at the last epoch. The batch size used is 16.

3.4.0.3 Histopathology DANN Specialist Model(Dual-Branch Head)

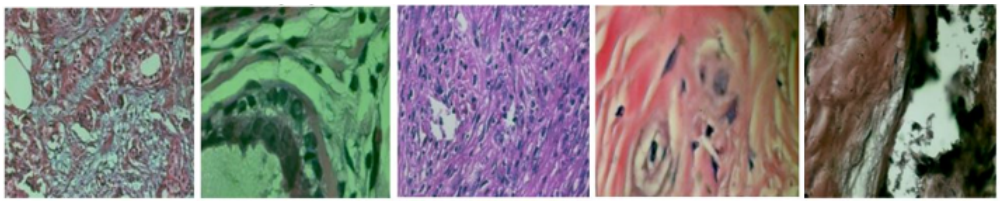


FIGURE 3.20: Histopathology Images

HP presents a significant level of variability in tissue samples collected in different hospitals and laboratories in terms of staining strategies, fixation steps, scanning methods, and magnification factors [54] as also shown in Fig 3.20 collected from our dataset, which also causes domain shift and deteriorates the performance of standard DL models.

High performance systems use DANN to provide reliable predictions of data that has never been observed before. The DANN model is trained to learn to discriminate between malignant and benign tissue but breaks a domain discriminator, where the network is trained to learn domain-invariant [55], [56], morphology-based discriminative features, like nuclear atypia, mitotic activity, and glandular architecture, as opposed to superficial features due to staining or scanner artifact [55]. This approach corresponds to clinical reasoning available in the literature and improves generalization on non-homogeneous data significantly [61]. Mathematically, DANN is divided into three different components: feature extractor / feature

generator, G_f ; label classifier/ label generator, G_y; and a domain discriminator/ domain generator, G_d. To prevent the gradient obstructing back-propagation, the gradient-reversal layer (GRL) [57] multiplies the gradients effecting on G_f with a negative factor lambda $(-\lambda)$ thus encouraging the feature extractor to produce the most confounding representations to the domain discriminator whilst retaining the ability to classify inputs with the label prediction task as shown in Equation 3.3.

$$\Delta\theta_f = \frac{\partial L_y}{\partial \theta_f} - \lambda \frac{\partial L_d}{\partial \theta_f} \tag{3.3}$$

The combined goal in this case reduces the diagnostic loss, L_y, and maximizes the domain loss, L_d, producing the resultant update rule of the parameters of the feature extractor as in Equation 3.4

$$(G_f, G_y, G_d) = \sum_{i=1}^{N} L_y(G_y(G_f(x_i)), y_i) - \lambda \sum_{i=1}^{N} L_d(G_d(G_f(x_i)), d_i) \tag{3.4}$$

In proposed methodology, the HP images are subjected to geometric augmentation as shown in Fig 3.21 along with class weights. For benign has 1.613 while malignant has 0.725.

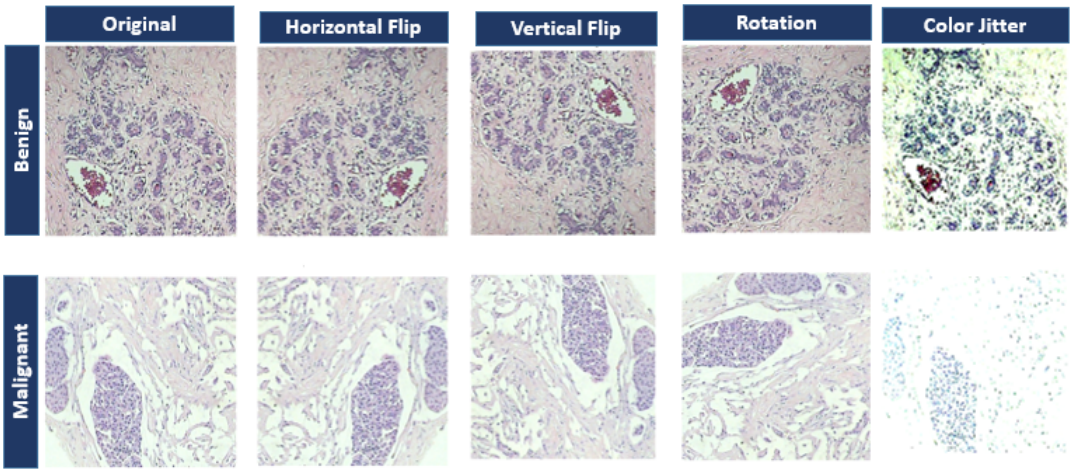


FIGURE 3.21: Histopathology Moderate Geometric Augmentation with colour jitter

The concept of DANN examines stain invariance by getting it learned into the extractor of features instead of using traditional processing of the images.

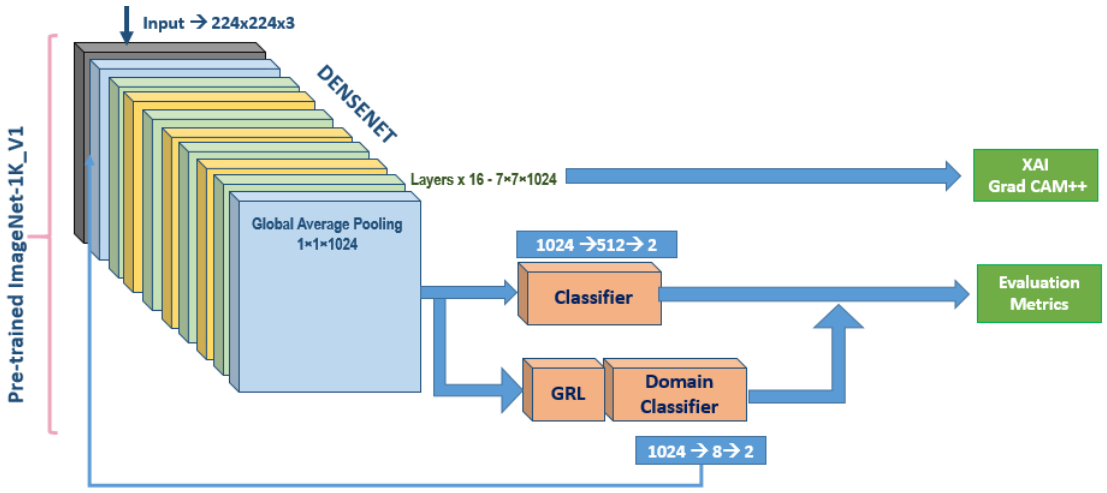


FIGURE 3.22: Working of DANN on Histopathology

The use of unprocessed H&E images eliminates the possibility of eliminating the diagnostically important features that could be induced by standard methods of normalization like Macenko or Reinhard [58] techniques that is why it maintains the authenticity of the data.

This strategy is enforced with DenseNet architecture: its densely connected pathways reuse both low- and high-level features across layers so that the DANN is provided with a holistic representation rather than a fragmented one which enables better gradient flow to overcome the vanishing-gradient problem [58] and parameter set curbs over-fitting to scanner- or stain-specific artefacts.

As a result, DenseNetDANN combined entity encourages the development of domain invariant, clinically significant features across a range of heterogeneous histopathological samples. In our proposed HP specialist model uses DANN with a DenseNet-121 as a backbone to reduce the domain shift due to staining variability. The network take input images of size 224x224x3. Along with pre-trained weights, each image go through the densenet121 model. The last densenet block 4 is responsible for generating 7x7x1024 feature map that localizes the primary regions that are influential to label classifier predictions.

Further global average pooling reduces these maps to a 1024-dimensional vector, which then branches into two parallel heads that is also demonstrated in the Fig 3.22. (i) a label classifier ($1024 \rightarrow 512 \rightarrow 2$) trained to predict benign vs malignant, and (ii) a domain classifier ($1024 \rightarrow$ GRL $\rightarrow 8 \rightarrow 2$) trained adversarially to

distinguish source vs target patient domains. Domain classifier generate 2 domain-specific outputs (source vs target).

This setup creates an adversarial training regime where the feature extractor learns domain-invariant representations, while the label classifier distinguishes between benign and malignant cases. The Adam optimizer with a starting learning rate of $1 \times 10^{-}4$ was used to optimize the model with batch size of 8.

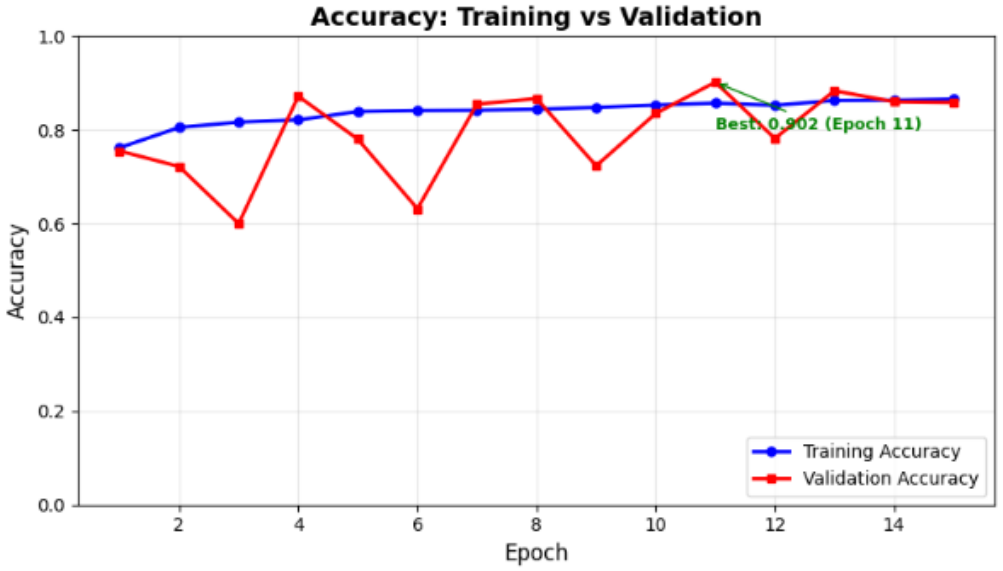


FIGURE 3.23: Histopathology Training Curves - Accuracy Curves

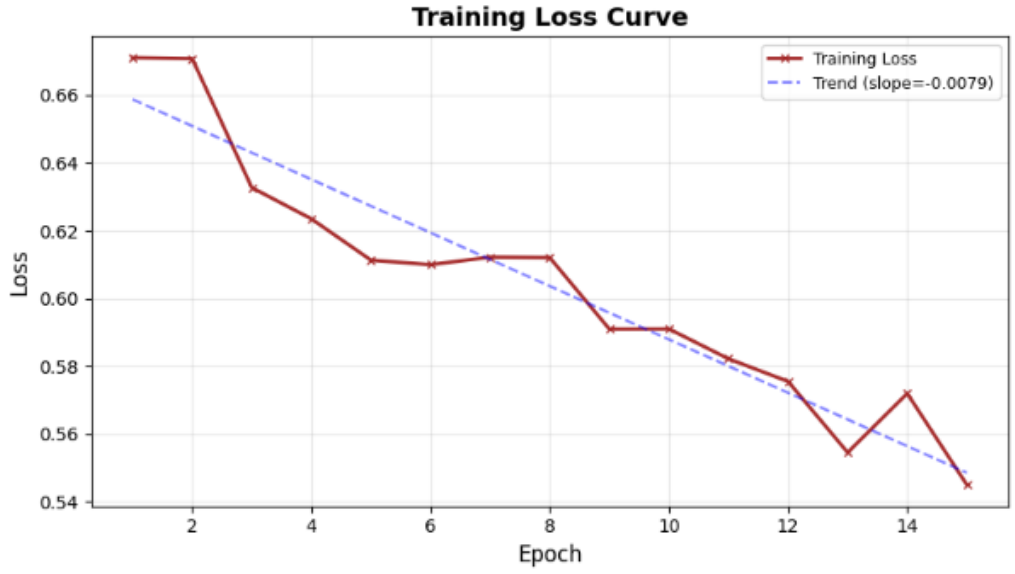


FIGURE 3.24: Histopathology Training Curves - Training Loss Curves

The training curves in the Fig Fig 3.25 demonstrated stable convergence without evidence of over fitting.

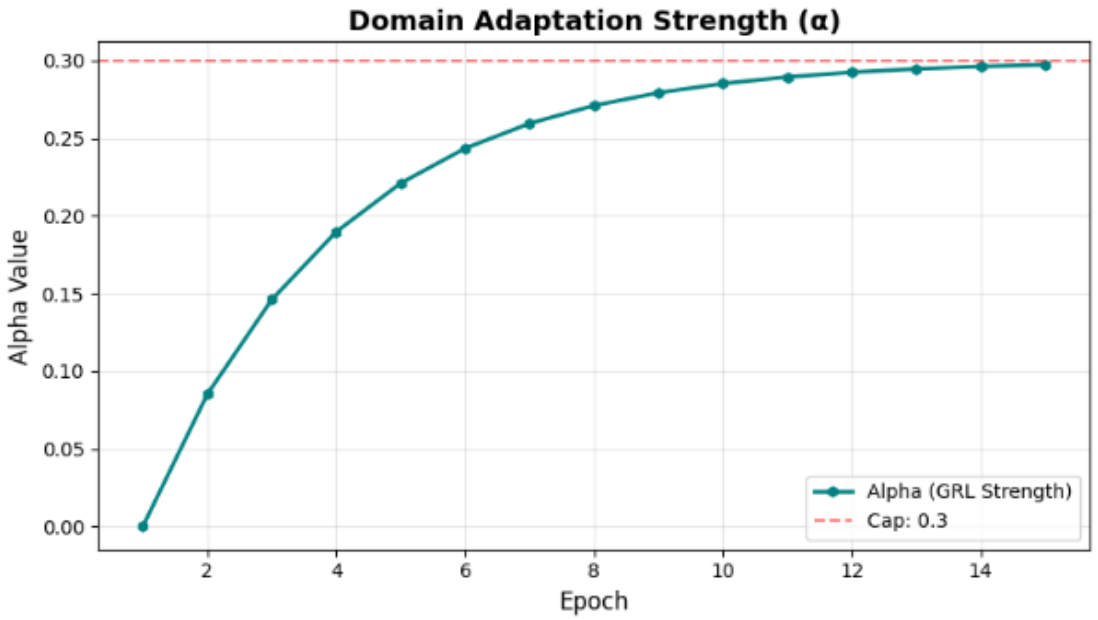


FIGURE 3.25: Histopathology Training Curves - Domain Adaption Curve

Training and validation accuracies shows continuous improvement and stable convergence. The training loss was large in the start but decreased as the training proceeds confirms that meaningful patterns are captured early in the optimization process.

Validation metrics, including accuracy, F1 score, and ROC-AUC, show consistently strong and promising results, suggesting reliable performance on unseen data. Finally, the domain adaptation schedule ramps conservatively to its cap of 0.3, ensuring that the domain classifier contributes to learning domain-invariant representations while maintaining stability in label classification.

3.4.1 Breasct Cancer Interpretability

To facilitate clinical trust and transparency, the classification head would be complemented by an XAI module [59] in our pipeline. Since the identification of BC requires explainable decisions [60], we use Grad-CAM++ to produce class-discriminative heat maps to identify the most significant areas.

DenseNet -121 learns features on mammographic, ultrasonographic, and histopathological images; and achieved the spatial feature map of 7 x 7 x 1024 on last dense block 4. Each feature map is converted to a 1024-dimensional vector by a terminal

global-average - pooling layer that represents every feature map with a capture of texture, structure and morphology.

Grad-CAM bridges this gap and puts an emphasis on masses [61], calcifications, distortion of glands or nuclear density [8] as a suitable per-modality, yet has a uniform interpretability framework across all data types.

Last dense block 4 having spatial information with feature map is responsible for the localization in the image.

3.4.1.1 Ultrasound Explainability

The Grad-CAM images in Fig 3.26 shows that the US specialist model has clinically relevant distributions of attention, as well as a very high classification effect of benign and cancerous lesions in the breast.

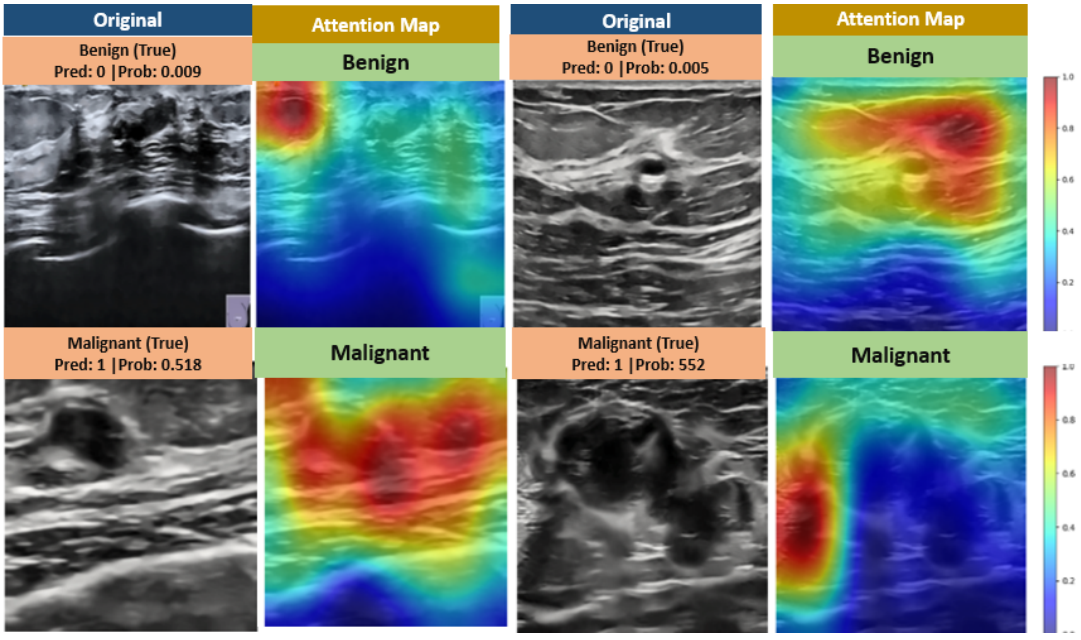


FIGURE 3.26: US Original vs XAI heat map

The symbolic cases in Fig 3.26 is showing a case of benign strength where the model has a high rate of predictive confidence of 0.95; the heat maps display focused activity in the homogeneous internal echo texture and along the well-defined margins, which are the hallmarks of benign pathology.

On the other hand, in the same Fig, malignant case with a high predictive confidence is shown. As a group, these Gradient-weighted Class Activation Mapping

(Grad Cam) visualizations support the view that the US expert framework incorporates clinically significant attributes that comply with established radiological standards, i.e. margin delineation, internal echogenicity patterns, and posterior acoustic properties that are individually helpful in distinguishing between benign and malignant breast lesions.

3.4.1.2 Mammography Explainability

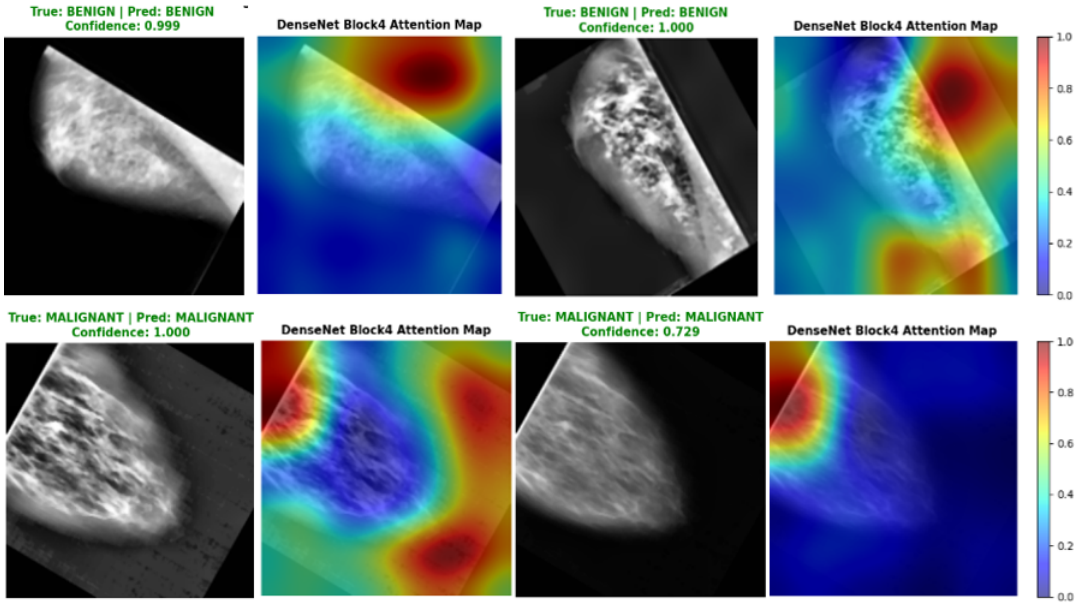


FIGURE 3.27: MG GradCam XAI heat map

The images evidenced reveal that an AI model can be used to analyse mammographic images to identify benign and malignant outcomes using the Grad-CAM method. Fig 3.27 shows original mammograms with heat maps.

The original images predicted as benign and malignant having confidence of more than 0.5 probability.

Taken together, such images demonstrate the potential of the AI to effectively categorize breast lesions and provide reasoning with why it is making a particular decision, which may eventually support radiologists in their interpretation of the clinical images and increase their confidence and accuracy in their diagnosis.

The Grad-Cam that provides visual explanation of the CNN model. Here through heat map shows the particular areas on the influential region that the AI is focusing

on to devise its decision with the warmer colors in the heat map denoting higher focus.

3.4.1.3 Histopathology Explainability

The histopathological images that are categorized as either benign or malignant, draw out refined differences in tissue structure and cell morphology. The cells in benign specimen are relatively homogenous in size and morphology, organized into a tissue structure, and have distinct boundaries, which are reminiscent of a non-invasive, low-growing pathology.

Nuclear sizes and staining appearance are normal and mitotic figures are rare or non-existent.

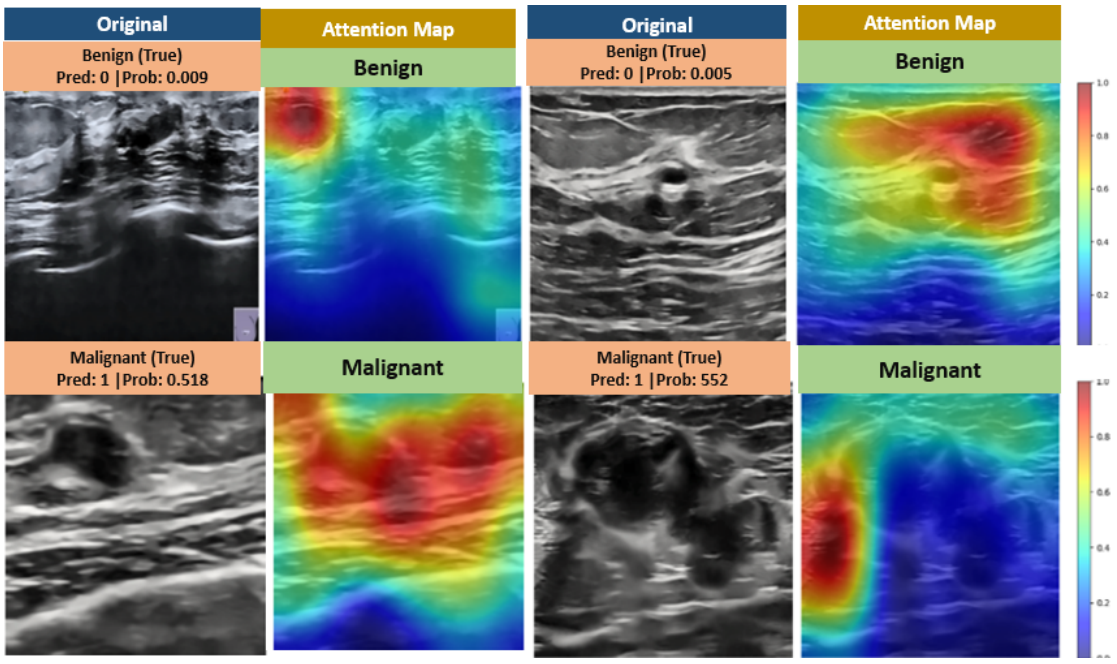


FIGURE 3.28: HP original vs XAI Heatmaps

On the contrary, the malignant specimen is characterized by high level of disorganization; the cells are characterized by pleomorphism, hyperchromatic nucleus and high nuclear to cytoplasmic ratio. Mitotic activity is increased, and there may be abnormal divisions, and tissue structure is lost, which is a sign of an aggressive and invasive phenotype inherent to malignancy. These visual differences in Fig 3.28 are critical in the histopathological diagnosis, which allow distinction of non-neoplastic and neoplastic disorders.

The heat maps overlay shows strong attention concentrated on cellular regions and irregular grandular structure. These attention distributions histograms further quantify these areas shows models assigns higher importance score to biologically meaningful regions rather than background artifacts. This assures that model decision making process is not only accurate but also interpretable, providing transparency in predictions.

3.4.2 Feature Level Fusion With Attention

In previous section, each modality undergoes its specified pipeline of preprocessing and feature extraction. As patients ids in each modality is difference so feature level fusion is done based at the modality level for benchmarking modality contribution. This represents that how much each modality helps in improving classification performance in a controlled research setting. Each modality is first counted and code trims all to same count, minimum wise so they can be aligned index wise. As per modality they are grouped index wise and made into parallel feature streams. This creates a synthetic multimodal representation to support classification. Each modality contributes to its own feature of 1024 vector, that collectively makes 3072 features. The attention mechanism learns the weight. In prior study [14], data level concatenation is conducted where blindly merging of features of different patients create a clinically invalid dataset. The proposed methodology uses feature level fusion with attention in the controlled setting. The learnt weights are US: 0.173, MG: 0.397, Histo: 0.430 that demonstrates that HP and MG provides the strongest contribution while US contributes the least.

Although, it is not clinically feasible due to synthetic fusion, but it offers valuable insights into the relative importance of modalities and guides the design of future patient-aligned multimodal datasets. Training curves shows the rapid and stable convergence of the model where training accuracy reached to 100 percent while validation reached to 93 percent. The gap in the loss curves indicates effective generalization. High F1 score yields balanced classification across classes. These results confirm that attention-based feature fusion yields a robust model by learning modality importance, in contrast to prior data-level concatenation approaches

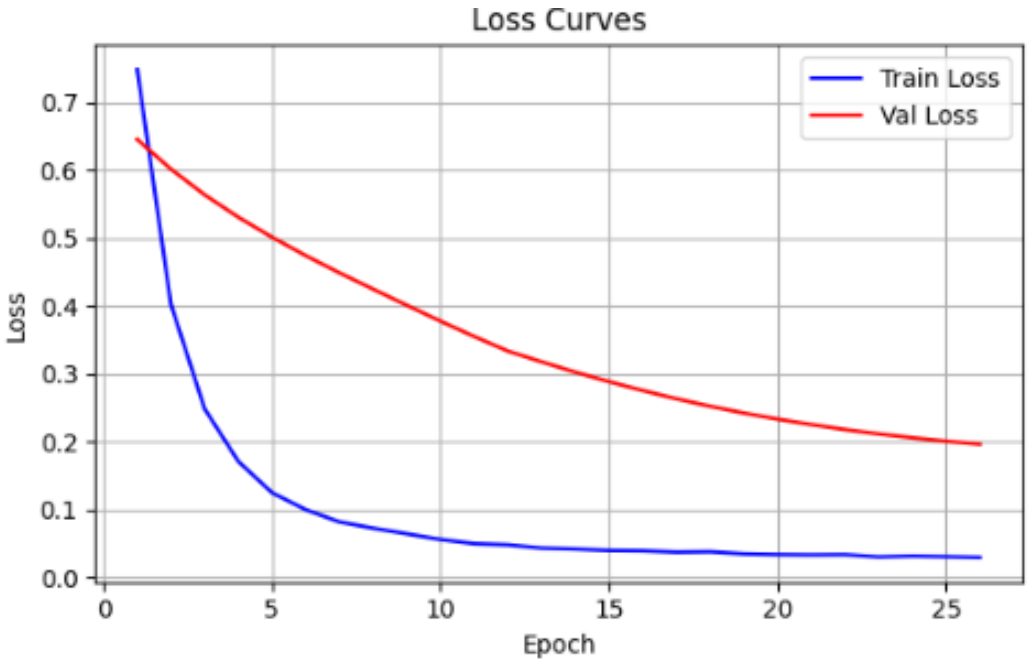


FIGURE 3.29: Feature Level Fusion Training Curves - Loss Curve

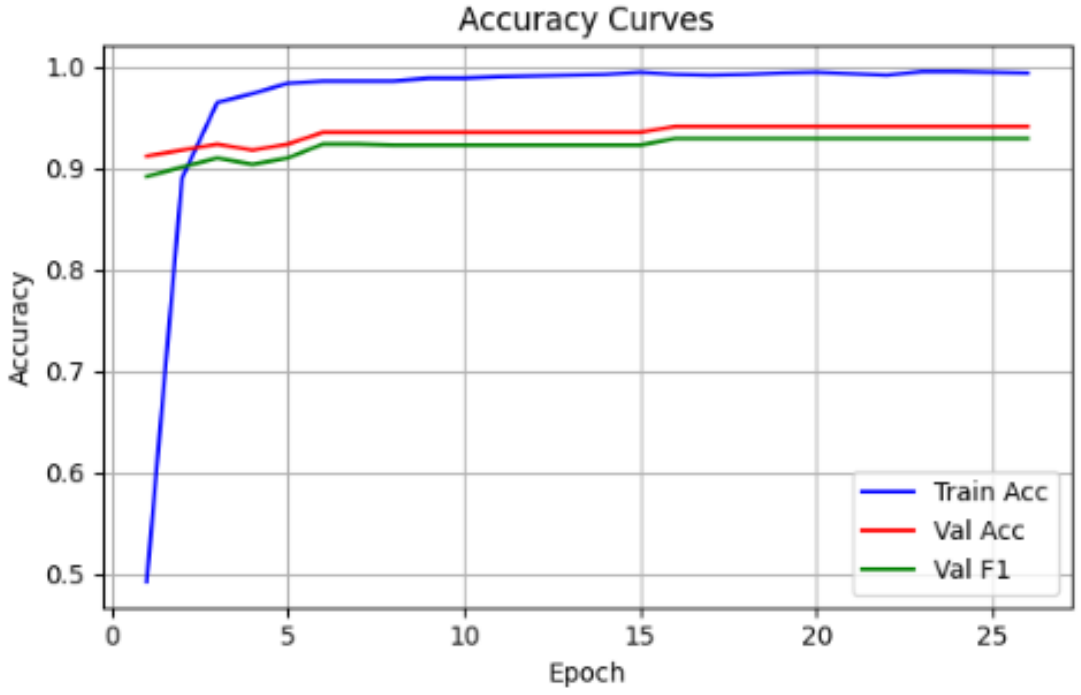


FIGURE 3.30: Feature Level Fusion Training Curves - Accuracy Curve

that risk over-fitting and clinical invalidity. For classification, feed forward neural network also MLP, is employed where the features are linear reduced from $3072 \longrightarrow 1024 \longrightarrow 512$ and then 2 binary classification after dropout and Relu activation function.

Considering both data level concatenation cited in and intermediate level fusion

(feature level fusion) is both not acceptable in this research due to unmatched patient ids. It is not recommended in clinical practices that different patients having different modalities are fused together. For these fusions, it is necessary to have same patient ids across modalities.

3.4.3 Decision Level Fusion

Instead of forceful and clinical inadmissible fusion, population-level fusion is employed due to different patient ids. Since patient ids across modalities does not match, feature level fusion is not clinically admissible, instead decision level fusion is employed, that synthesizes independent modality prediction at the population level. Although the modalities were evaluated on different cohorts, but their complementary strengths are integrated statistically.

This mirrors clinical workflows where independent reports from different modalities are synthesized into a final decision. The implementation works by first loading the features and generating the prediction per modality. US has 171 labels, MG has 2457 labels and HP has 600 labels. Each modality is first balanced per label(benign and malignant). Balanced sampling of benign sampling are 96 while malignant 75 resulting into 171 total. Fusion strategies are applied as mentioned below

i. Weighted voting: combine predictions with fixed weights. Each model's prediction is multiplied by its assigned weight.

ii. Majority voting: final decision if 2 out of 3 modalities agree. It depends on majority votes.

iii. Smart weighted fusion: dynamic weights that are learned on each modality's accuracy

This pipeline is clinically acceptable because it respects the independence of each modality, avoids trimming ,information loss and mismatched patient ids, and fuses predictions using strategies that mirror real clinical workflows. The combined accuracy reflects how well these independent reports can be synthesized into a final decision. Final decision is based upon the fused probabilities just like a board

integrated multiple specialist's inputs for tumour decision. In result section, the results are further described.

3.5 Experimental Results and Discussion

This section outlines the evaluation metrics, results and analysis of the proposed model in classifying BC using the various imaging modalities. In addition, the model is compared to the recent published models in the later part of the section. Google colab was used to run simulation experiments, and the results were saved in Google drive. The experiments used to assess the effectiveness of the proposed model used a 12.7 GB RAM setup, an NVIDIA A100 GPU, and were done in Python on the PyTorch API.

The training and assessment procedures are based on a single architecture in order to ensure clinical validity and feasibility. Making a diagnostic decision based on the feature extractions by the DenseNet-121 backbone, which are then integrated with the classifiers. Each specialist model has the classifier that gives binary output. In the case of MG, structural and textural signals are separated by the head, in the case of the US, echogenicity and lesion morphology, and in the HP-DANN framework, the cellular structure domain-invariant structures are separated. Their training and validation results are described in each of their section above. These classification head are used in the testing otherwise it would break the vital relationship between the extracted evidence and the final predictive outcome, thus an evaluation of a different system to that which was optimised invalidating the reported performance metrics.

Therefore, validation and testing of fairness need the same evaluator to be used in each stage, in order to represent the end-to-end performance when it is deployed. Assessment of accuracy, sensitivity, and specificity is made on the set of validation prior to the test set which is to provide real-world applicability. Modalities (MG, US, HP) are tested separately to provide a fair comparison of modalities that is clinically relevant and yet maintains the same end-to-end architecture throughout experimental execution.

Measures of quantitative performance are based on metrics derived using confusion-matrices accuracy, specificity, sensitivity, and area under the receiver-operator characteristic curve (AUROC). The findings explain the diagnostic effectiveness of the suggested framework in different imaging modalities. The confusion matrix, which is a basic machine-learning classification tool, is used to infer the model parameters. The matrix will include true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). TP is the correct prediction of positive cases; TN is the correct prediction of negative cases; FP is the negative cases that are classified as positive; FN is the positive cases that are classified as negative.

These totals make it possible to calculate key performance indices. Accuracy , sensitivity , specificity), precision and F1-score are derived based on the confusion matrix. All the metrics are derived based on TP, TN, FP, and FN using the test images. The expressed formulations are as follows. Accuracy: The general accuracy of the model. Equation shown in

3.5.1 Results of Ultrasound

Figure 3.31 presents the key evaluation metrics for the ultrasound-based breast lesion classification model. US model will attain a good overall accuracy of 90.8% in which 158 out of 174 test cases are correctly classified using the final DenseNet head ($1024 \longrightarrow 512 \longrightarrow 256 \longrightarrow 2$). Its sensitivity is 88.15% which is a great capacity to detect the malignant lesions correctly, when 67 out of 76 cancer cases were accurately detected. The sensitivity of 92.8% indicates the performance of good results in terms of detecting benign lesions since 91 of 98 non-cancer cases were correctly classified. The accuracy of 90.8 percent shows that mine out of ten times every time the model predicts malignancy it is the correct choice. A balance is ensured between recall and precision as shown by the F1 score of 0.900 which means that the model performance is robust and stable.

The confusion matrix in 3.31 gives a close breakdown of the classification. The model has accurately classified 91 benign and 61 malignant cases as true negatives and true positives respectively. But there were 7 false positives of benign cases

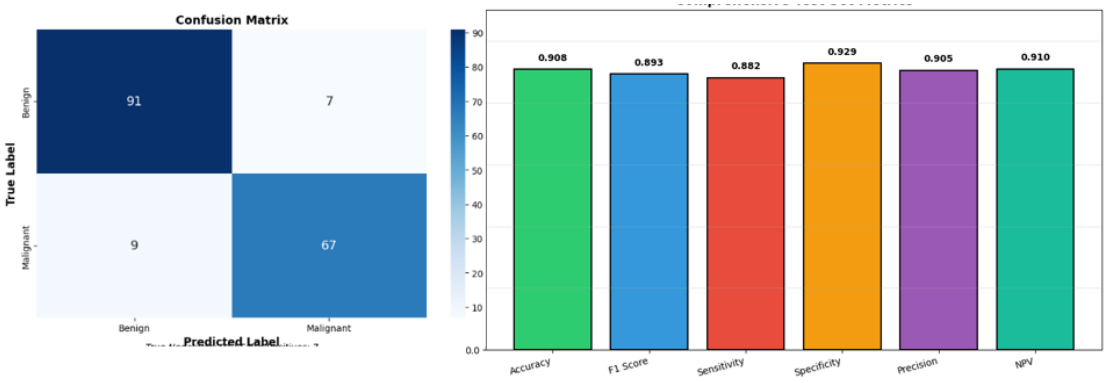


FIGURE 3.31: Confusion Matrix and Bar Chart for US Model

(not malignant) indicating false positives that comprised 7.14 percent of the benign cases and there were 9 false negatives which were actually malignant cases but not detected (false negatives). This tendency illustrates a clinically desirable sensitivity versus specificity bias, i.e., the model is slightly more likely to place ambiguous cases in the malignant category, i.e. using the conservative model in which minimal risk of not detecting the actual cancers is borne at the expense of potentially unnecessary follow-up testing. The respective evaluation metrics is also shown in Fig 3.31 right side that shows F1 score of 89%, senstivity 88.2%, specificity 92.9%, precision 90% and NPV 91%. NPV is the proportion of negative predictions that are actually correct. This validates the high discriminative power of the model to differentiate between benign and malignant lesions at the different decision thresholds. All these measurements confirm the prospect of the model to be a dependable computer-aided diagnostic tool in ultrasound breast imaging, and high sensitivity is particularly useful in screening where the identification of malignancies is mostly needed.

3.5.2 Results of Mammography

The first test of the classification accuracy of the model was made using the confusion matrix, which provided a detailed breakdown of TP, TN, FP and FN per class. Such a representation enables the determination of the predictive behaviour of classes and sheds light on future misclassifying patterns across classes of diagnosis. Mammography attained Accuracy 96% , Sensitivity 96% ,Specificity 95.60%

and ROC-AUC 97%. The confusion matrix and evaluation metrics is shown in the Fig 3.32.

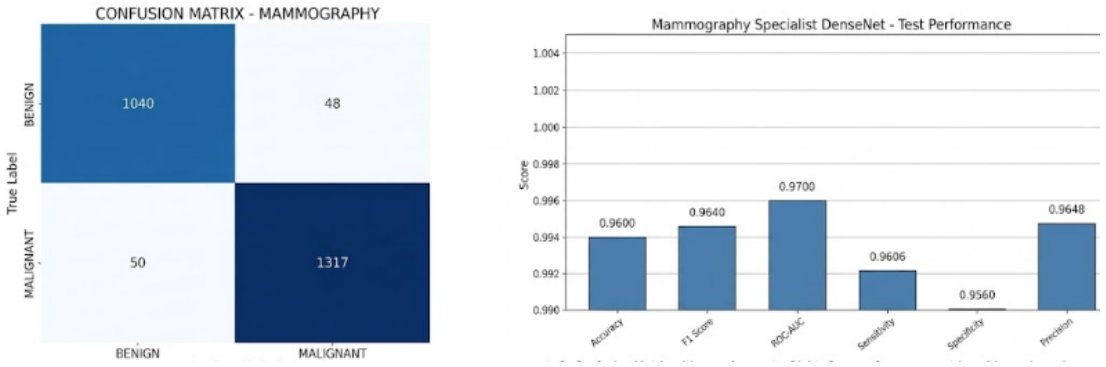


FIGURE 3.32: Confusion Matrix and Evaluation metrics of MG Model on Test Set

As shown, out of 2459 cases, 2357 were correctly classified, including 1,040 benign lesions and 1,317 malignant lesion. High percentage of sensitivity and specificity reflects correct detection of malignant and bening cases. Precision reached to 0.96 that reflects balanced trade-off between recall and precision. The ROC-AUC of 0.97 further validated the model's high discriminative power across thresholds. The confusion matrix shows only 48 FP and 50 FN, highlighting a clinically desirable bias toward minimizing missed cancers while maintaining strong specificity. These results confirm that the mammography head provides robust and generalizable performance, making it a dependable component of the multimodal diagnostic framework.

3.5.3 Results of Histopathology

The HP model has an outstanding diagnostic performance in the classification of BC with an overall accuracy of 94.8% with an outstanding AUROC value of 0.992, which signifies the almost perfect capability to classify malignant and benign tissue samples. Figure 3.34 of the appendix displays the confusion matrix, which shows that there were 258 benign cases correctly identified as such (true negatives) and 519 malignant cases correctly identified as such (true positives), and only 24 benign cases were falsely identified as such (false positive) and 19 malignant cases were falsely identified as not (false negative).

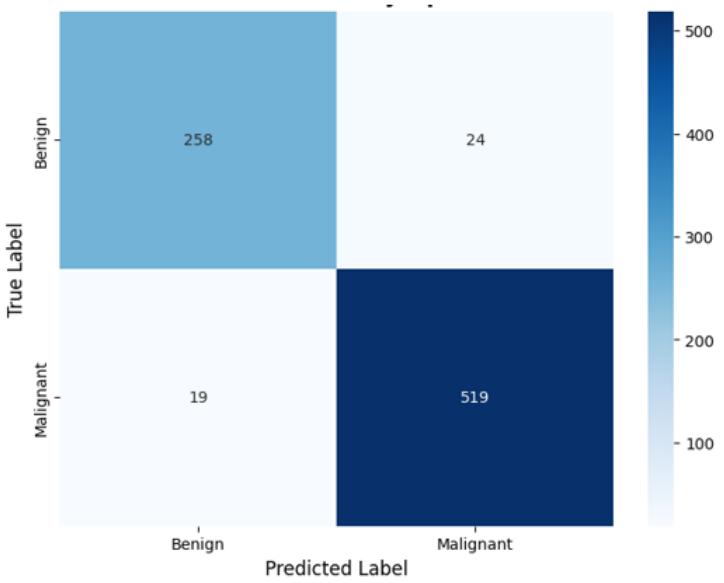


FIGURE 3.33: Confusion matrix HP on Test Set

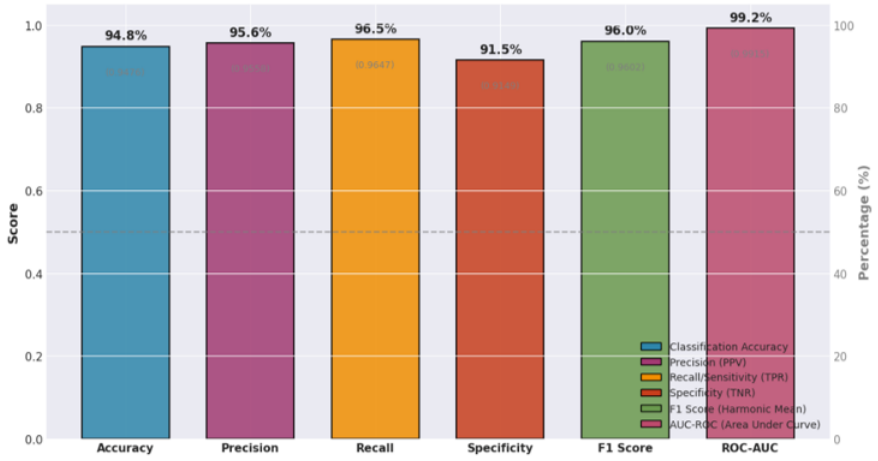


FIGURE 3.34: HP Evaluation Metrics on Test Set

The performance measures summarized in Figure **??** indicate that the model is well balanced with sensitivity of 96.4, specificity of 91.4 and precision of 95.8 indicating that the model is accurate across both classes. The good F1 score of 0.960 validates the high degree of accuracy and recall accuracy and the ROC that demonstrates that the model is highly classifying with an AUC ROC of 0.992, well above a random classifier. The fact that the precision-recall indicates a steady high precision score at higher recall rates. All this together makes the histopathology model a very dependable diagnostic tool in breast cancer and its tremendous AUC-ROC and sensitivity-specificity aspect that requires both false positive and false negative to have very serious clinical ramifications makes it especially a very useful tool

in confirmed diagnostic uses. The end-to-end use of DenseNet-121 from feature extraction to classification ensures consistency that only the suggested architecture can account for the reported performance, therefore, making it easier to make meaningful comparisons to the existing body of knowledge, as well as providing a reliable benchmark on which clinical reasoning can be made. Finally, the uniformity of the architectural construct directly enables the main clinical purpose of the framework, which consists of making autonomous and reliable diagnoses that will be based on a single imaging modality. Furthermore, when applied to HP, DANN ensure the features obtained stay unchanged when changing the staining and acquisition domains. This type of domain invariance is especially important to clinical use, where it is essential that there is consistency across laboratories and staining procedures. The results from evaluation metrics for each modality are promising with additional feature of interpretability. The table 3.35 demonstrates that the accuracy results of HP while using DenseNet 121 with DANN and without DANN. The minimal difference shows that dataset acquired have less domain variance but in broader perspective it plays an important role in generalizability and cross modal implementation at multi institutional level.

MODEL	Test set Accuracy
DenseNET without DANN	92% plus jittery training graphs
DenseNET with DANN	94 %

FIGURE 3.35: Results of HP on Test Dataset from DenseNET121 backbone with DANN/Without DANN

3.5.4 Fusion Results

i. Feature Level Fusion With Attention: Feature level fusion with attention achieved 94% accuracy that shows high performance. As shown in the Fig 3.36, out of 96 benign cases, 95 were correctly classified and in malignant cases, 66 are correctly classified and 9 are misclassified as benign. This demonstrates strong specificity i.e. 99% while sensitivity is good i.e. 88% with attention mechanisms effectively weighting modality contributions to improve malignant detection.

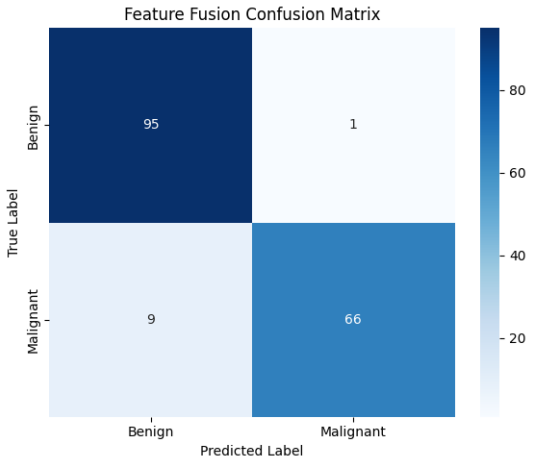


FIGURE 3.36: Feature Level Fusion Confusion Matrix

ii. Decision Level Fusion: In terms of fusion strategies, decision level fusion is clinical acceptable as it respects the individual modality independence without forceful data aggregation and feature level fusion on different patient ids. Different strategies such as weighted voting, majority voting and smart weighted fusion is evaluated. In weighted voting, different weights are evaluated as shown in Table 3.6.

TABLE 3.6: Performance of Weighted Fusion Strategies

US	MG	HP	Accuracy	F1-Score	AUC
0.33	0.34	0.33	0.9842	0.9833	0.9985
0.30	0.40	0.30	0.9842	0.9833	0.9985
0.40	0.30	0.30	0.9825	0.9796	0.9999
0.20	0.30	0.50	0.9825	0.9796	1.0000
0.50	0.30	0.20	0.9825	0.9799	0.9985
0.20	0.50	0.30	0.9890	0.9889	1.0000
0.60	0.20	0.20	0.9240	0.9103	0.9918
0.20	0.20	0.60	0.9766	0.9726	1.0000

It achieved near perfect accuracy 98.90% when MG was given higher weight. At the decision level, independent modalities produce their own predictions, and these decisions are then combined using a fusion rule to produce a final, unified output. This demonstrates that weighting modalities according to their reliabilities improves performance. Another strategy is majority voting, where each modality shares its probability of the decision and majority is selected. Overall, accuracy achieved through this is 98.2%. It is effective but slightly less than a weighted voting. As shown in the Fig 3.39.

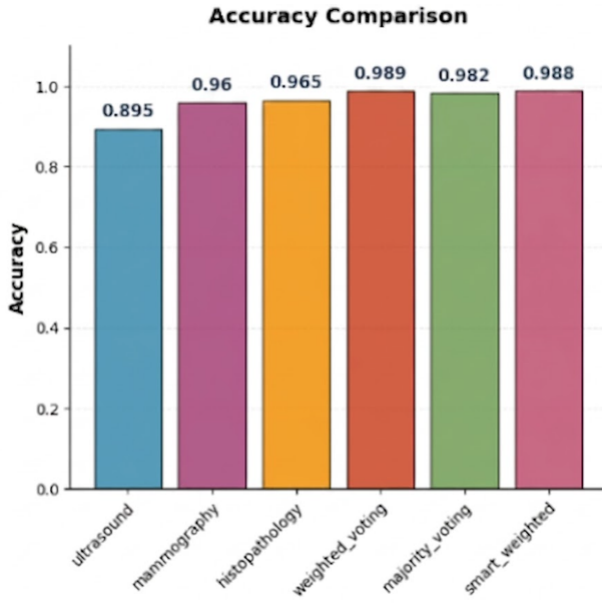


FIGURE 3.37: Decision Level Bar chart - Accuracy Comparison

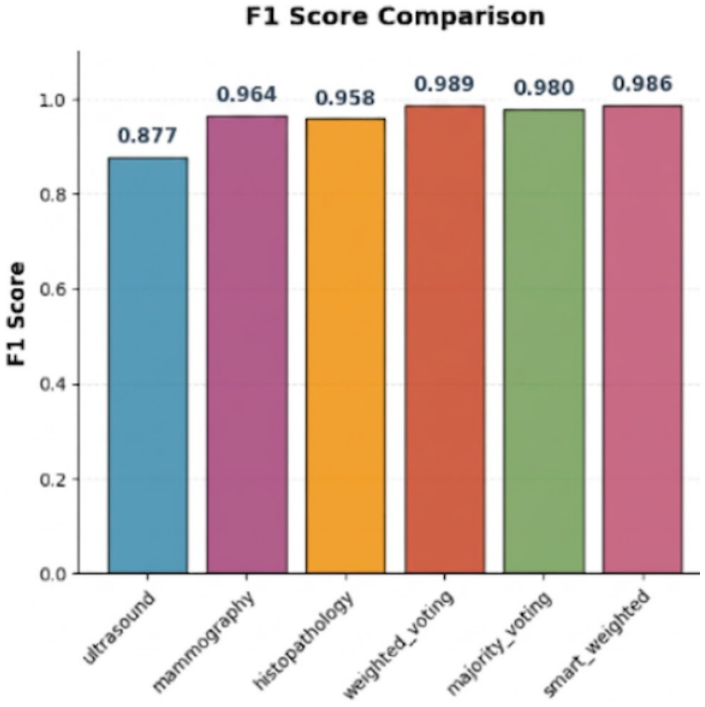


FIGURE 3.38: Decision Level Bar chart - F1 Score Comparison

As compare to per modality evaluation metrics, fusion strategy especially weighted fusion surpass others. It is also evident through confusion matrix, where in weighted summary accuracy is 98.90%, in majority voting it is 98.42% where specificity is 100% but sensitivity is 96%, in smart weighted fusion strategy that is between than majority weighted but less then weighted voting.It has accuracy of 98.4%, specificity is 100% and sensitivity is 97.33%.

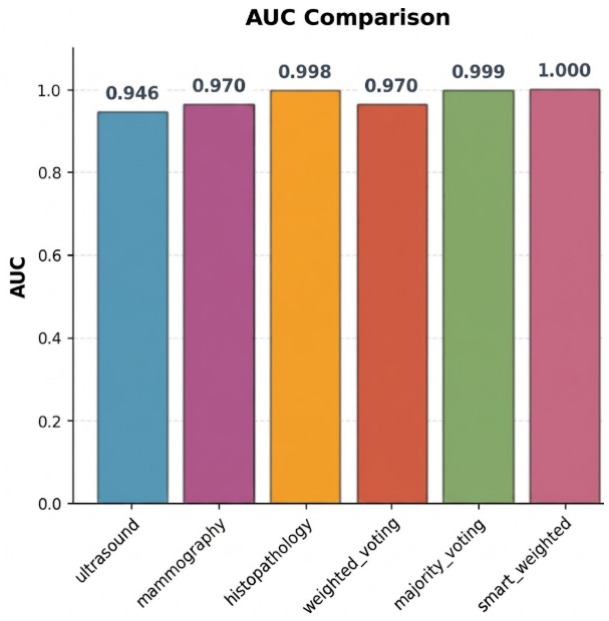


FIGURE 3.39: Decision Level Bar chart - AUC Comparison

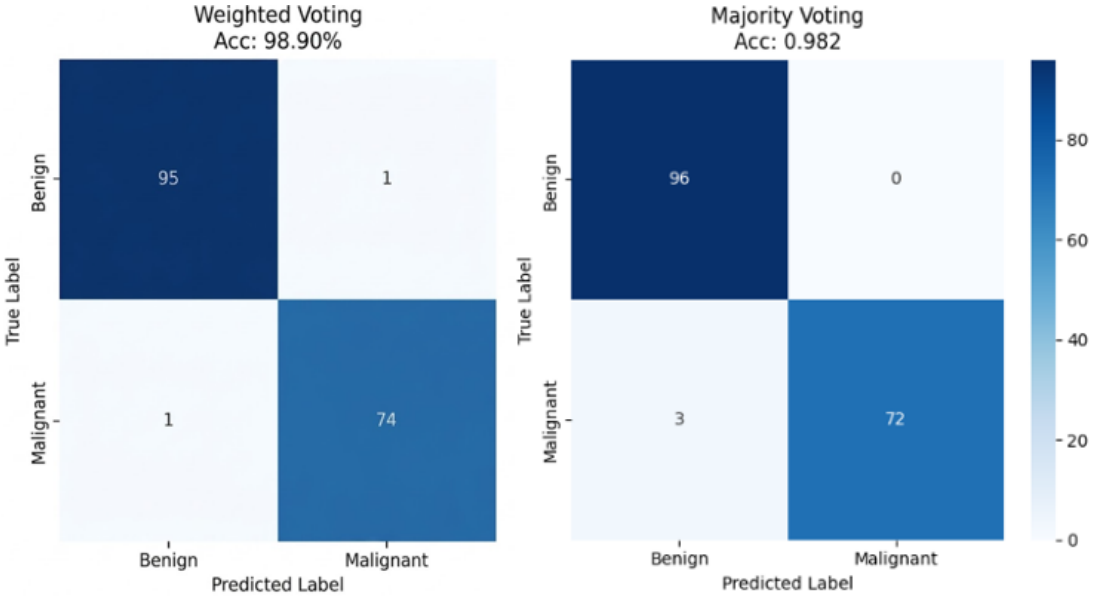


FIGURE 3.40: Decision Level Confusion Matrix - Weighted Voting and Majority Voting

It is also shown in the visualization of confusion matrix shown in Fig 3.41.

3.5.5 Comparison with Prior Work

In first study, where data aggregation of different patient ids across modalities is done shows promising accuracy, but it is clinically in admissible due to enforced fusion strategy and posing a risk of cross modal model fusion with different patients

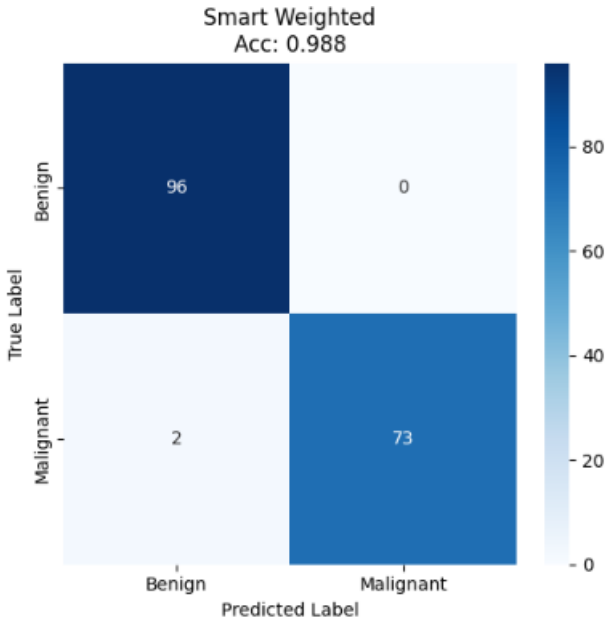


FIGURE 3.41: Decision Level Confusion Matrix - Smart Weighted Voting

TABLE 3.7: Comparative Analysis of Prior Research with the Proposed Method

Ref	HP	MG	US	Comb.	Proposed Method
[14]	91.79	88.16	95.33	99.96	Avoids enforced fusion
[21]	100	99.97	92.85	–	Grad-CAM++ based localisation
[15]	98.75	99.12	98.40	–	Combined accuracy; all HP magnifications
[23]	93.97	–	89.97	–	Tri-modality; all HP magnifications
[27]	–	97.33	99.44	–	Moderate data; tri-modality
[25]	–	97.75	97.50	–	Grad-CAM++ localisation
[26]	–	–	–	96.40	Robust to dense breast tissues
[22]	–	–	–	99.22	Moderate data handled by DenseNet121
[12]	–	93.41	91.67	98.84	Tri-modality with moderate dataset
[13]	–	–	–	94.00	Moderate data with Grad-CAM++
[29]	–	–	–	93.30	Moderate data; tri-modality
Model	**96.4**	**99.4**	**90.8**	100	Tri-modality, explainability

across modalities. Table 3.7 is the comparison of prior work where it highlights methodological/data limitations [15], no black box transparency [21] and clinical inadmissibility [14] . Other researches [23] - [29] are based on dual modalities. All these characteristics make our approach to the problem a more transparent, reliable, and clinically relevant solution in comparison with the prior work .

Chapter 4

Conclusion and Future Work

4.1 Limitations and Future Work

This study is also limited by several flaws even though it is showing promising performance. The framework is also limited to imaging modalities and does not include molecular or biomarkers that would be vital in the complete characterization of BC. Retrospective model evaluation was done, which needs to be prospectively validated by multi-institutional in order to ensure real world generalization. Moreover, the system is static in its training paradigm and does not have continuous learning skills to follow the changing clinical environment.

Although Grad-CAM offers meaningful visual explanations per modality level to overcome black box transparency. With individual modality evaluation metrics, fusion at feature and decision level is also employed and their pros and cons are discussed which is a novel approach. As per clinical admissibility, decision level fusion is incorporated which prior studies have failed to implement. Future developments of the project introduce later stage diagnostic platforms such as genomic data including estrogen receptor (ER), progesterone receptor (PR), and HER2 markers.

In addition, the framework might be turned into a life-long-learning model that is both safe to engage new, institution-specific information after deployment, thus, adjusting to the local patients and imaging guidelines.

4.2 Conclusion

Conclusively, the thesis has formulated justified explainable, tri-modal DL architecture of Breast cancer classification that fills the major gaps in the existing Computer-Aided Design system. Prior studies have limitations of methodological constraints, no black box transparency, use of only cleaned data and enforced fusion despite of different patient ids.

Proposed methodology overcome these limitations by introducing individual level prediction of each modality. Separate pipeline of each modality not only predicts the cancerous and non-cancerous cells but also localize the important region of the image that influenced the prediction of the modality. It also used different fusion strategies like feature level fusion with attention and decision level fusion and comparison is drawn between them and latter is preferred due to it clinical admissibility.

This model enables each of the imaging modalities to capture macro-, real-time tissue characterisation, and microscopic cellular evidence in order to offer a diagnostic evaluation compared to single-, or dual-modality systems. The pipeline starts with minimalist preprocessing approach per modality followed by stratified level splitting based upon patient id, modality and label.

The feature extraction carried by DenseNet121 along with transfer learning. For histopathology, domain adversial neural network is also employed along with DenseNet121. After extracting features, multi layer perceptron is employed for classification. Each modality perceptron vary differently as per the modality features and dataset size. Explainable Artificial Intelligence is incorporated for visualization of black box transparency and build clinical trust. For Explainable Artificial Intelligence, GradCam is used for localization of important regions in the images. At the end, fusion is employed. Both feature level and decision level fusion is evaluated and based upon clinical admissibility, decision level fusion is preferred. With clinical admissibility, promising accuracies are achieved from the model. Ultrasound has 90.8%, Mammography has 99.4%, Histopathology has 96.4% and combined accuracy through decision level fusion is 100%.

Bibliography

[1] F. Bray *et al.*, "Global cancer statistics 2022: Global cancer society estimates of incidence and mortality worldwide for thirty six cancers in one hundred and eighty five countries," *CA: A Cancer Journal for Clinicians*, vol. 74, pp. 229–263, 2024.

[2] American Cancer Society, United States of America, "Breast cancer facts & figures on united states of america," https://www.cancer.org/research/cancer-facts-statistics/breast-cancer-facts-figures.html, 2024.

[3] National Breast Cancer Foundation, "Breast cancer facts & statistics 2024," https://www.nationalbreastcancer.org/breast-cancer-facts/, 2024, accessed: 08-08-2024.

[4] Y. Zhang, Y. Ji, S. Liu *et al.*, "Global burden of female breast cancer," *Journal of the National Cancer Center*, 2025.

[5] R. Gerami, S. S. Joni, N. Akhondi *et al.*, "A literature review on the imaging methods for breast cancer - mammography, histopathology, ultrasound, mri," *International Journal of Physiology, Pathophysiology and Pharmacology*, vol. 14, no. 3, p. 171, 2022. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9301184/

[6] L.-J. Tseng, A. Matsuyama, and V. MacDonald-Dickinson, "Histology/histoapthology: The gold standard for diagnosisvfor breast cancer," *The Canadian Veterinary Medical Association Journal - Revue vétérinaire canadienne*, vol. 64, no. 4, p. 389, 2023. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10031787/

[7] M. A. Rahman *et al.*, "Advancements in breast cancer detection: A review of global trends, risk factors, imaging modalities, machine learning, and deep learning approaches," *BioMedInformatics*, vol. 5, no. 3, p. 46, 2025.

[8] T. Arravalli, K. Chadaga, H. Muralikrishna *et al.*, "Detection of breast cancer using machine learning and explainable artificial intelligence," *Scientific Reports*, vol. 15, no. 1, 2025.

[9] Q. Lin, W.-M. Tan, J.-Y. Ge, Y. Huang, Q. Xiao, Y.-Y. Xu, Y.-T. Jin, Z.-M. Shao, Y.-J. Gu, B. Yan, and K.-D. Yu, "Artificial intelligence-based diagnosis of breast cancer by mammography microcalcification," *Fundamental Research*, vol. 5, no. 2, pp. 880–889, 2025.

[10] H. Rehman, I. Ahmad, S. Rashid, M. Mukhtar, A. A. Khan, and H. Khaliq, "Comparison of diagnostic accuracy of ultrasound and mammography in detecting breast cancer in radiographically dense breasts," *Cureus*, 2025.

[11] R. Tenajas, D. Miraut, C. I. Illana, R. Alonso-Gonzalez, F. Arias-Valcayo, and J. L. Herraiz, "Recent advances in artificial intelligence-assisted ultrasound scanning," *Applied Sciences*, vol. 13, no. 6, p. 3693, 2023.

[12] K. Atrey, B. K. Singh, and N. K. Bodhey, "Multimodal classification of breast cancer using feature level fusion of mammogram and ultrasound images in machine learning paradigm," *Multimedia Tools and Applications*, vol. 83, no. 7, pp. 21 347–21 368, 2023.

[13] G. Habib, N. Kiryati, M. Sklair-Levy *et al.*, "Automatic breast lesion classification by joint neural analysis of mammography and ultrasound," in *Multimodal Learning for Clinical Decision Support and Clinical Image-Based Procedures*. Cham: Springer, 2020, pp. 125–135.

[14] D. Deb, R. Dash, and D. P. Mohapatra, "Mmhbc-net: A multimodal hybrid approach for breast cancer classification," *Neural Computing and Applications*, 2025.

[15] U. Nawaz, Z. Saeed, H. M. UbaidUllah, F. Mirza, and M. Muzzamil, "Explainable attention-enhanced approach for multimodal breast cancer diagnosis

across diverse imaging modalities," *International Journal of Imaging Systems and Technology*, vol. 35, no. 6, 2025.

[16] C. L. Srinidhi, O. Ciga, and A. L. Martel, "Deep neural network models for computational histopathology: A survey," *Medical Image Analysis*, vol. 67, p. 101813, 2021.

[17] X. Yao, X. Wang, S.-H. Wang, and Y.-D. Zhang, "A comprehensive survey on convolutional neural network in medical image analysis," *Multimedia Tools and Applications*, vol. 81, no. 29, pp. 41 361–41 405, 2020.

[18] T. Aitazaz, A. Tubaishat, F. Al-Obeidat, B. Shah, T. Zia, and A. Tariq, "Transfer learning for histopathology images: an empirical study," *Neural Computing and Applications*, vol. 35, no. 11, pp. 7963–7974, 2022.

[19] K. Raghavan, S. B, and K. V, "Attention guided grad-cam: an improved explainable artificial intelligence model for infrared breast cancer detection," *Multimedia Tools and Applications*, vol. 83, pp. 57 551–57 578, 2024.

[20] K. Gadzicki, R. Khamsehashari, and C. Zetzsche, "Early vs late fusion in multimodal convolutional neural networks," in *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*. IEEE, 2020, pp. 1–6.

[21] I. Pacal and O. Attallah, "Inceptionnext-transformer: A novel multi-scale deep feature learning architecture for multimodal breast cancer diagnosis," *Biomedical Signal Processing and Control*, vol. 110, p. 108116, 2025.

[22] K. Atrey, B. K. Singh, and N. K. Bodhey, "Integration of ultrasound and mammogram for multimodal classification of breast cancer using hybrid residual neural network and machine learning," *Image and Vision Computing*, p. 104987, 2024.

[23] M. R. Alom, F. F. A., M. A. Rahaman, A. Rahman, T. Debnath, Abu, and S. Mansor, "An explainable ai-driven deep neural network for accurate breast cancer detection from histopathological and ultrasound images," *Scientific Reports*, vol. 15, no. 1, 2025.

[24] R. M. Al-Tam, A. M. Al-Hejri, S. S. Alshamrani, M. A. Al-Antari, and S. M. Narangale, "Multimodal breast cancer hybrid explainable computer-aided diagnosis using medical mammograms and ultrasound images," *Journal of Applied Biomedicine*, vol. 44, no. 3, pp. 731–758, 2024.

[25] A. Sahu, P. Das, and S. Meher, "An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images," *Biomedical Signal Processing and Control*, vol. 87, p. 105377, 2024.

[26] Y.-M. Wang *et al.*, "Cnn-based cross-modality fusion for enhanced breast cancer detection using mammography and ultrasound," *Tomography*, vol. 10, no. 12, pp. 2038–2057, 2024.

[27] B. Bouallegue, Y. M. A. El-Latif, H. El-Sofany, and Islam, "An enhanced approach for predicting breast cancer using different deep learning algorithms and explainable ai techniques in an iot environment," *International Journal of Intelligent Systems*, vol. 2025, no. 1, 2025.

[28] R. Devkota *et al.*, "Evaluation of breast mass by mammography and ultrasonography with histopathological correlation," *Journal of Nepal Health Research Council*, vol. 19, no. 3, pp. 487–493, 2021.

[29] J. Cong, B. Wei, Y. He, Y. Yin, and Y. Zheng, "A selective ensemble classification method combining mammography images with ultrasound images for breast cancer diagnosis," *Computational and Mathematical Methods in Medicine*, vol. 2017, pp. 1–7, 2017.

[30] A.-R. Kinnunen, R. Sironen, and P. Sipola, "Magnetic resonance imaging characteristics in patients with histopathologically proven fibrous dysplasia: A systematic review," *Skeletal Radiology*, vol. 49, no. 6, pp. 837–845, 2020.

[31] J. J. Choi, S. H. Kim, B. J. Kang, B. G. Choi, B. Song, and H. Jung, "Mammographic artifacts on full-field digital mammography," *Journal of Digital Imaging*, vol. 27, no. 2, pp. 231–236, 2013.

[32] T. W. Stephens, "Breast cancer screening with imaging: Recommendations from the society of breast imaging and the acr on the use of mammography,

breast mri, breast ultrasound, and other technologies for the detection of clinically occult breast cancer," *Yearbook of Diagnostic Radiology*, vol. 2011, pp. 46–47, 2011.

[33] M. Heath *et al.*, "Current status of the digital database for screening mammography," in *Computational Imaging and Vision*, 1998, pp. 457–460.

[34] "Cbis-ddsm breast cancer image dataset," https://www.kaggle.com/datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset, accessed: 2026.

[35] "Breast ultrasound images dataset," https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset, accessed: 2026.

[36] "Bus uclm breast ultrasound dataset," https://www.kaggle.com/datasets/orvile/bus-uclm-breast-ultrasound-dataset, accessed: 2026.

[37] "Bus uc breast ultrasound dataset," https://www.kaggle.com/datasets/orvile/bus-uc-breast-ultrasound, accessed: 2026.

[38] "Breakhis dataset," https://www.kaggle.com/datasets/ambarish/breakhis, accessed: 2026.

[39] E. Devolli-Disha, S. Manxhuka-Kerliu, H. Ymeri, and A. Kutllovci, "Comparative accuracy of mammography and ultrasound in women with breast symptoms according to age and breast density," *Bosnian Journal of Basic Medical Sciences*, vol. 9, no. 2, pp. 131–136, 2009.

[40] M. Cooper, Z. Ji, and R. G. Krishnan, "Machine learning in computational histopathology: Challenges and opportunities," *Genes, Chromosomes and Cancer*, vol. 62, no. 9, pp. 540–556, 2023.

[41] M. Veta, J. P. Pluim, P. J. Van Diest, and M. A. Viergever, "Breast cancer histopathology image analysis: A review," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 5, pp. 1400–1411, 2014.

[42] A. M. Hanby, "The pathology of breast cancer and the role of the histopathology laboratory," *Clinical Oncology*, vol. 17, no. 4, pp. 234–239, 2005.

[43] F. T. Hatoum, R. M. Tomek, H. M. Whitney, and M. L. Giger, "Importance of stratified sampling for use in the development of training and test sets: medical imaging ai applications," in *Medical Imaging 2025: Computer-Aided Diagnosis*, vol. 13407. SPIE, 2025, pp. 210–217.

[44] Z. Sajjadnia, R. Khayami, and M. R. Moosavi, "Preprocessing breast cancer data to improve the data quality, diagnosis procedure, and medical care services," *Cancer Informatics*, vol. 19, p. 1176935120917955, 2020.

[45] S. Chucherd and S. S. Makhanov, "Sparse phase portrait analysis for preprocessing and segmentation of ultrasound images of breast cancer," *IAENG International Journal of Computer Science*, vol. 38, no. 2, 2011. [Online]. Available: https://www.iaeng.org/IJCS/issues_v38/issue_2/IJCS_38_2_09.pdf

[46] M. J. George and D. A. S. Dhas, "Preprocessing filters for mammogram images: a review," in *2017 Conference on Emerging Devices and Smart Systems (ICEDSS)*. IEEE, March 2017, pp. 1–7.

[47] R. Ramani, N. S. Vanitha, and S. Valarmathy, "The pre-processing techniques for breast cancer detection in mammography images," *International Journal of Image, Graphics and Signal Processing*, vol. 5, no. 5, p. 47, 2013.

[48] M. Dabass and J. Dabass, "Preprocessing techniques for colon histopathology images," 2019, pp. 1121–1138.

[49] Y. L. K. F. Cruz, A. F. M. Silva, E. E. C. Santana, and D. G. Costa, "Generative adversarial networks in histological image segmentation: A systematic literature review," *Applied Sciences*, vol. 15, no. 14, p. 7802, 2025.

[50] Y. Zhang, H. Chen, and Y. Wei, "Deep microscopy adaptation network for histopathology cancer image classification," in *International Conference on Medical Image Computing*. Springer International Publishing, 2019.

[51] T. Pavlović, T. Popović, and S. Čakić, "Breast cancer detection using resnet and densenet," in *29th International Conference on IT*. IEEE, 2025.

[52] D. Bhatt, "Cnn variants for computer vision," *Electronics*, vol. 10, no. 20, p. 2470, 2021.

[53] M. Chhabra and R. Kumar, "A smart healthcare system based on classifier densenet 121 , booktitle = Mobile Radio Communications and 5G Networks, doi = 10.1007/978-981-19-8367-$4_2$4."

[54] G. Orchard and B. Nation, *Histopathology*. USA: Oxford University Press, 2012.

[55] A. Chaddad and Y. Wu, "A practical simulation for domain adaptation models," in *2023 IEEE International Conference on E-health Networking, Application & Services (Healthcom)*. IEEE, December 2023, pp. 203–204.

[56] H. Ajakan and P. Germain, "Domain-adversarial neural networks," *arXiv preprint*, 2014. [Online]. Available: https://arxiv.org/abs/1412.4446

[57] J. Jang and B. Na, "Unknown-aware domain adversarial learning," in *Advances in Neural Information Processing Systems*, vol. 35, 2022, pp. 16 755–16 767. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/file/abcdf123e4567890-Paper-Conference.pdf

[58] M. W. Lafarge, "Domain-adversarial neural networks," in *International Workshop on Deep Learning in Medical Image Analysis*. Cham: Springer International Publishing, September 2017, pp. 83–91.

[59] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, others, and R. Ranjan, "Explainable ai (xai): Core ideas, techniques, and solutions," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–33, 2023.

[60] P. P. Angelov and E. A. Soares, "Explainable artificial intelligence: an analytical review," *Wiley Interdisciplinary Reviews*, vol. 11, no. 5, p. e1424, 2021.

[61] M. A. Talukder, "An improved xai-based densenet model for breast cancer detection using reconstruction and fine-tuning," *Results in Engineering*, vol. 26, p. 104802, 2025.