A

Project Report

On

Breast Cancer Detection through Advanced Pre-processing and Deep Learning Techniques

Submitted as partial fulfilment for the award of

BACHELOR OF TECHNOLOGY DEGREE

SESSION 2024-25

In

Computer Science and Engineering

By

Sajal Bhilatia (2100290100143)

Rachit Verma (2100290100124)

Harsh Rastogi (2200290109006)

Under the supervision of

Prof. Gagan Thakral (CSE)

KIET Group of Institutions, Ghaziabad

Affiliated to

Dr. A.P.J. Abdul Kalam Technical University, Lucknow

(Formerly UPTU)

May 2025

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Signature

Name: Sajal Bhilatia

Roll no: 2100290100143

Date:

Signature

Name: Rachit Verma

Roll no: 2100290100124

Date:

Signature

Name: Harsh Rastogi

Roll no: 2200290109006

Date:

CERTIFICATE

This is to certify that Project Report entitled "Breast Cancer Detection through Advanced Preprocessing and Deep Learning Techniques" which is submitted by Sajal Bhilatia, Rachit Verma and Harsh Rastogi in partial fulfilment of the requirement for the award of degree B.Tech. in Department of Computer Science & Engineering of Dr. A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

Mr. Gagan Thakral	**Dr. Vineet Sharma**
(Assistant Professor)	**(Dean CSE)**

Date:

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the report of the B. Tech Project undertaken during B.Tech. Final Year. We owe special debt of gratitude to Prof. Gagan Thakral, Department of Computer Science & Engineering, KIET, Ghaziabad, for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavours have seen light of the day.

We also take the opportunity to acknowledge the contribution of Dr. Vineet Sharma, Dean Computer Science & Engineering, KIET, Ghaziabad, for his full support and assistance during the development of the project.

We also do not like to miss the opportunity to acknowledge the contribution of all faculty members, especially faculty/industry person/any person, of the department for their kind assistance and cooperation during the development of our project. Last but not the least, we acknowledge our friends for their contribution in the completion of the project.

Date:
Signature:
Name : Sajal Bhilatia
Roll No.: 2100290100143

Date:
Signature:
Name : Rachit Verma
Roll No.: 2100290100124

Date:
Signature:
Name : Harsh Rastogi
Roll No.: 2200290109006

ABSTRACT

Breast Cancer is one of the major causes of deaths among women, which makes early and accurate detection crucial. Mammogram image analysis provides an essential method for diagnosing breast cancer by differentiating between benign and malignant cases. This work presents a comprehensive comparison of various deep learning models applied to the classification of breast cancer using mammography scans. The dataset consists of labelled images that have been tagged as benign or malignant. The images were subjected to: rotation to adjust the image orientation; extraction of pectoral muscles using canny edge detection; contrast enhancement using CLAHE (Contrast Limited Adaptive Histogram Equalization); and morphological operations to eliminate artifacts and noise, such as text. Four architectures of deep learning were tested for performance: VGG16, ResNet50, MobileNet, and a proprietary Convolutional Neural Network (CNN). In particular, this work overcomes the challenges of generalization, at the same time, it demonstrates the potential for deep learning models to classify breast cancer by using mammograms. There is still work to be done to improve the robustness of these models and the dependability of diagnostics in real-world situations.

TABLE OF CONTENTS Page No.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Full Form
AI	Artificial Intelligence
API	Application Programming Interface
CAD	Computer-Aided Diagnosis
CC	Craniocaudal (view in mammography)
CNN	Convolutional Neural Network
CLAHE	Contrast Limited Adaptive Histogram Equalization
DL	Deep Learning
DNN	Deep Neural Network
FDA	Food and Drug Administration
FN	False Negative
FP	False Positive
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
HIS	Hospital Information System
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
LIME	Local Interpretable Model-agnostic Explanations
MLO	Mediolateral Oblique (view in mammography)
ML	Machine Learning
MRI	Magnetic Resonance Imaging
PNG	Portable Network Graphics (image format)
RNN	Recurrent Neural Network
SGD	Stochastic Gradient Descent
SHAP	SHapley Additive exPlanations
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
UI	User Interface
ViT	Vision Transformer
VGG16	Visual Geometry Group 16-layer model
XAI	Explainable Artificial Intelligence

CHAPTER 1: INTRODUCTION

1.1 Background and Motivation

Breast cancer is among the most prevalent forms of cancer globally and remains one of the leading causes of cancer-related deaths in women. According to the World Health Organization (WHO), approximately 2.3 million women were diagnosed with breast cancer in 2020, and it accounted for nearly 685,000 deaths worldwide. These statistics underscore the critical need for early detection, accurate diagnosis, and effective treatment to reduce mortality and enhance the quality of life for patients.

The earlier breast cancer is detected, the more effective the treatment, significantly improving survival rates. Traditional diagnostic methods, including physical examinations, ultrasound, and biopsies, though effective, can sometimes be invasive, subjective, or prone to human error. Mammography, a non-invasive imaging technique that utilizes low-energy X-rays, has emerged as one of the most reliable and widely used methods for breast cancer screening and detection. However, the interpretation of mammograms is complex and demands a high level of expertise. Misinterpretation can result in false positives (leading to unnecessary biopsies and anxiety) or false negatives (delaying critical treatment).

With the advent of Artificial Intelligence (AI) and deep learning technologies, the medical field is undergoing a transformative shift. Machine learning models, particularly convolutional neural networks (CNNs), have demonstrated exceptional performance in image classification tasks. Their ability to automatically learn spatial hierarchies of features from input images without the need for manual feature extraction makes them ideal for medical imaging applications. These AI models can assist radiologists by serving as a second opinion, reducing the diagnostic workload, and minimizing human error.

This project is driven by the motivation to leverage deep learning models for the early and accurate detection of breast cancer through the analysis of mammography images. By focusing on advanced pre-processing techniques and evaluating multiple deep learning architectures, the project aims to contribute to the growing body of research in AI-assisted medical diagnostics.

1.2 Breast Cancer: Medical Overview

Breast cancer originates from breast tissue, most commonly from the inner lining of milk ducts or the lobules that supply the ducts with milk. It is typically classified into two major categories:

- **Benign Tumours**: These are non-cancerous and do not spread to other parts of the body.
- **Malignant Tumours**: These are cancerous and have the potential to invade nearby tissues and metastasize to distant organs.

The stages of breast cancer range from Stage 0 (non-invasive or in situ) to Stage IV (metastatic cancer). Early detection, often through screening programs, plays a pivotal role in managing and curing the disease. Symptoms of breast cancer include a lump in the breast, change in size or shape, nipple discharge, and skin dimpling, among others.

Several diagnostic tools are available, such as:

- **Mammography**
- **Ultrasound**
- **Magnetic Resonance Imaging (MRI)**
- **Biopsy (histopathological confirmation)**

Among these, mammography remains the frontline tool for early detection.

1.3 Role of Imaging and Mammography

Mammography is a specialized medical imaging technique that uses a low-dose X-ray system for the examination of breast tissues. The goal is to detect early signs of breast cancer, often before symptoms appear. There are two types of mammograms:

1. **Screening Mammograms**: Used for women with no symptoms to detect early signs.
2. **Diagnostic Mammograms**: Used to investigate suspicious breast changes.

Mammograms provide grayscale images that radiologists analyse for signs of abnormality such as:

- Masses
- Calcifications
- Architectural distortions
- Asymmetrical tissue densities

However, mammogram interpretation is subjective and susceptible to variability between observers. Moreover, the presence of dense breast tissue can obscure abnormalities, making detection challenging. These limitations highlight the need for computer-aided diagnosis (CAD) systems to support radiologists.

1.4 Emergence of AI and Deep Learning in Medical Imaging

Deep learning, a subfield of machine learning, has gained significant momentum in recent years, primarily due to the success of CNNs in image recognition and classification tasks. CNNs mimic the human visual cortex and are especially proficient at identifying spatial patterns and textures, making them suitable for processing medical images.

In the context of breast cancer detection, deep learning models are trained on labelled mammography datasets to classify images as benign or malignant. The following benefits are associated with the use of AI in mammography:

- **Reduction in Diagnostic Errors**: AI can catch patterns that may be missed by human eyes.

- **Improved Accuracy and Sensitivity**: Trained networks can outperform traditional machine learning methods.
- **Time Efficiency**: Models can analyse large volumes of data quickly.
- **Scalability**: Once trained, models can be deployed across healthcare settings.

Despite these advantages, challenges such as generalization, overfitting, lack of diverse datasets, and the black-box nature of deep learning models need to be addressed for widespread clinical adoption.

1.5 Objectives of the Project

The primary goal of this project is to build and evaluate a robust deep learning-based framework for breast cancer detection using mammographic images. The key objectives include:

- To collect and pre-process mammogram images to enhance data quality.
- To implement and compare multiple deep learning models (VGG16, ResNet50, MobileNet, and a custom CNN).
- To apply advanced pre-processing techniques like orientation correction, pectoral muscle removal, CLAHE, and morphological operations.
- To analyse and compare model performance based on accuracy, precision, recall, and F1-score.
- To identify challenges such as overfitting and propose solutions for real-world applicability.
- To contribute to the growing field of AI-driven diagnostic support systems in healthcare.

1.6 Scope of the Project

This project is specifically focused on mammography-based breast cancer detection using supervised deep learning techniques. The scope includes:

- Working with publicly available, labelled mammographic datasets.
- Implementing standardized image pre-processing techniques.
- Evaluating pre-trained architectures as well as designing a custom CNN.
- Conducting performance comparisons using both training and unseen test data.
- Visualizing model performance through graphs, accuracy-loss plots, and confusion matrices.

The scope does not include:

- Histopathological or ultrasound image analysis.
- Clinical trials or real-time deployment.
- Hardware acceleration or mobile deployment optimization.

1.7 Limitations

While the project offers a promising approach, it has several limitations:

- **Dataset Size and Diversity**: Limited availability of labelled mammograms, especially from varied demographics.
- **Overfitting in Deep Models**: Risk of models performing well on training data but poorly on unseen data.
- **Lack of Real-Time Clinical Evaluation**: The models are not yet tested in live hospital settings.
- **Interpretability**: Deep learning models often lack explainability, which is crucial in the medical domain.

These limitations can be addressed in future iterations by integrating advanced regularization techniques, enlarging the dataset, and collaborating with medical professionals for real-world feedback.

1.8 Project Description

The proposed system is designed to classify breast cancer through mammography image analysis using deep learning techniques. The project begins with a rigorous **pre-processing phase** to enhance image clarity and uniformity. This phase includes:

- **Orientation adjustment** to align all mammograms consistently,
- **Pectoral muscle removal** using the Canny edge detection algorithm to eliminate distractions from the region of interest,
- **Contrast enhancement** with CLAHE to highlight tissue boundaries,
- **Noise removal** using morphological operations to clean text or scan artifacts.

Post pre-processing, the refined images are fed into four different deep learning models: **a custom CNN, VGG16, ResNet50, and MobileNet**. Each model is trained to classify images as **benign or malignant**, learning from key visual features extracted during pre-processing.

The system evaluates model performance by comparing **training and testing accuracy**, with an emphasis on reducing overfitting and enhancing generalizability. The goal is to build a **trustworthy AI-based diagnostic aid** that supports radiologists by improving diagnostic precision and providing a consistent second opinion. The project also opens avenues for future exploration, including real-world testing, model optimization, and integration into clinical workflows.

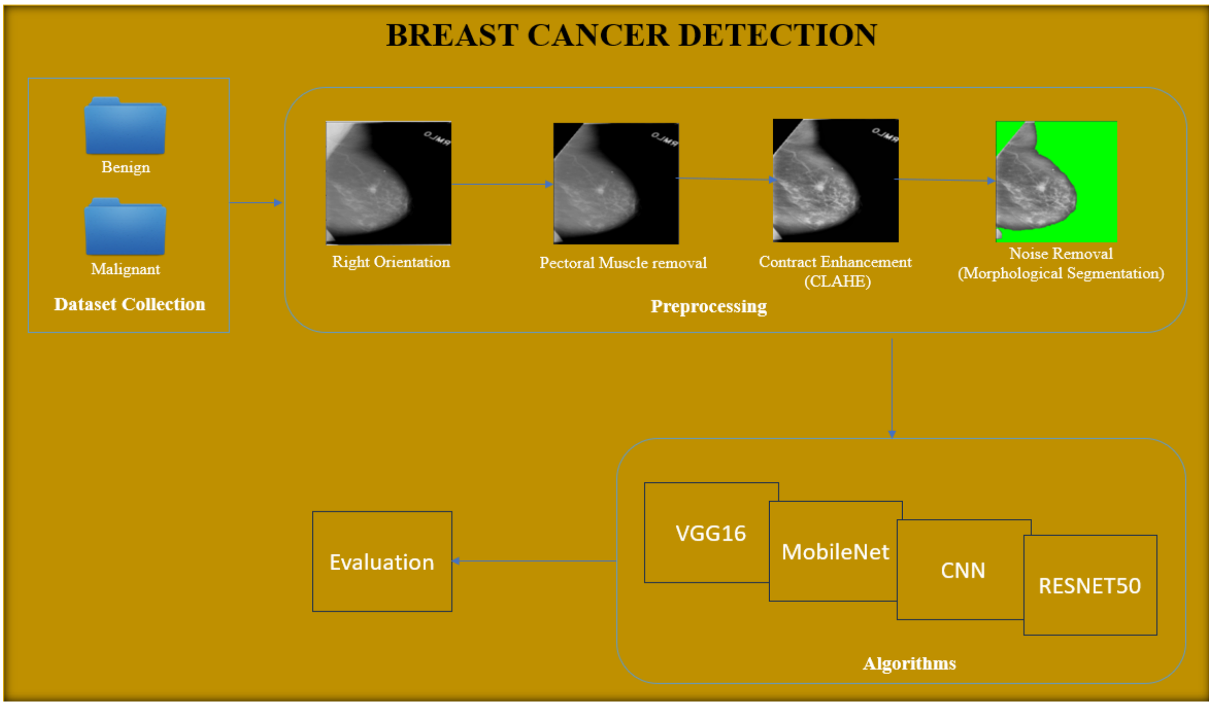

Figure 1 Architectural Diagram

1.9 Report Organization

The remainder of this report is structured as follows:

- **Chapter 2 – Literature Review**: Discusses previous work in breast cancer detection, various machine learning and deep learning techniques applied, and identifies research gaps.
- **Chapter 3 – Proposed Methodology**: Describes in detail the pre-processing pipeline, model architecture, and training procedures used.
- **Chapter 4 – Results and Discussion**: Presents a comparative analysis of all models based on performance metrics, along with observations and visualizations.
- **Chapter 5 – Conclusion and Future Scope**: Summarizes key findings and provides a roadmap for future improvements.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction to Literature Review

Breast cancer detection has been the subject of intense research across medical, imaging, and computational domains. With the growing incidence of breast cancer globally, the urgency to develop accurate, fast, and non-invasive diagnostic systems has led to the rise of computational methods, particularly those grounded in artificial intelligence (AI) and deep learning (DL). This chapter provides a critical review of the literature, covering classical approaches, the rise of deep learning in medical image analysis, and specific contributions of various deep neural network (DNN) architectures to breast cancer detection using mammography.

The goal of this review is to understand the evolution of techniques, identify existing research gaps, and highlight how the current project contributes meaningfully to this domain.

2.2 Evolution of Computer-Aided Diagnosis (CAD) in Breast Cancer

Early systems for computer-aided diagnosis (CAD) were rule-based and statistical, relying on handcrafted features like shape, texture, and intensity extracted from medical images. These features were fed into classifiers such as Support Vector Machines (SVM), Decision Trees, or K-Nearest Neighbours (KNN). While these methods showed promise, they often suffered from limitations due to their dependence on domain expertise, limited generalizability, and inability to capture complex patterns.

By the early 2010s, the paradigm began shifting toward **machine learning (ML)**, where algorithms could learn patterns from data. However, these techniques still relied on manual feature engineering. It wasn't until the success of deep learning—especially Convolutional Neural Networks (CNNs)—in large-scale image recognition tasks (e.g., ImageNet) that AI-based medical image analysis saw a breakthrough.

2.3 Pioneering Work in Deep Learning for Image Classification

AlexNet (Krizhevsky et al., 2012)

One of the most influential papers in modern deep learning, Krizhevsky et al. (2012) proposed **AlexNet**, a deep CNN that won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) with a top-5 error rate of 15.3%, nearly 10% lower than the next best. This work inspired the adoption of deep CNNs across various domains, including medical imaging.

Though not designed for medical data, AlexNet's success led to adaptations for mammography, laying the groundwork for architectures like VGG and ResNet.

2.4 Use of Deep Learning in Mammography

Simonyan and Zisserman's VGGNet (2014)

The **VGG family of networks**, especially **VGG16**, brought forward the concept of using very small (3x3) convolutional filters stacked in depth. VGG16 has been widely adopted in medical imaging because of its simplicity, depth, and performance.

In breast cancer classification tasks, VGG16 has shown effectiveness in extracting fine-grained features from mammograms. However, it is also known for high computational cost and sensitivity to overfitting if not properly regularized.

He et al.'s ResNet (2015)

The **ResNet architecture**, particularly **ResNet50**, introduced **residual learning** to address vanishing gradient issues in deep networks. This enabled the training of networks with more than 100 layers, a feat not possible with traditional CNNs at the time.

In mammogram analysis, ResNet50 is particularly useful for learning deep hierarchical patterns that might not be evident in shallower networks. It is especially helpful in datasets with high visual complexity but relatively limited sample sizes.

Howard et al.'s MobileNet (2017)

MobileNet is a lightweight CNN designed for mobile and resource-constrained environments. It utilizes **depthwise separable convolutions** to significantly reduce computational complexity without compromising much on accuracy.

Its application in medical imaging, especially in real-time analysis or deployment in rural clinics with limited infrastructure, is promising. Studies have shown that MobileNet can provide reliable classification performance with fewer parameters and faster inference time.

2.5 Review of Domain-Specific Studies in Breast Cancer Detection

Shen et al. (2018)

This study applied CNN-based feature extraction to classify mammograms into benign or malignant categories. They reported that pre-trained networks fine-tuned on medical datasets significantly outperformed models trained from scratch, due to better generalization and reduced training time.

Litjens et al. (2017)

A seminal survey that reviewed over 300 papers, demonstrating the depth and breadth of deep learning in medical image analysis. It emphasized the transformative potential of CNNs in diagnostics, including breast cancer detection.

Esteva et al. (2017)

While not directly related to breast cancer, this landmark study achieved dermatologist-level performance in skin cancer classification using CNNs. It established that deep learning models can match or exceed expert performance given sufficient training data.

These studies collectively reinforce the effectiveness of deep learning models in medical diagnostics when high-quality labelled datasets are available and proper pre-processing is performed.

2.6 Challenges in Deep Learning-Based Breast Cancer Detection

Despite promising results, the literature also reveals several **challenges**:

1. Dataset Size and Diversity

Medical image datasets are often small, imbalanced, or lack diversity in terms of age, ethnicity, and imaging devices. Overfitting remains a major problem, especially when models perform extremely well on training data but poorly on unseen samples.

2. Interpretability

Medical professionals are hesitant to adopt "black-box" models. Techniques such as **Grad-CAM** or **SHAP** are proposed to interpret decisions made by CNNs, but they're still under active research.

3. Generalization Across Institutions

Models trained on one dataset may not generalize to others due to variations in resolution, labelling, and equipment. The **need for cross-institutional validation** is a critical next step.

4. Integration into Clinical Workflow

For real-world use, models must meet regulatory standards (e.g., FDA approval), integrate with hospital IT systems, and provide actionable results.

2.7 Recent Advancements and Alternative Approaches (2020–2025)

1. Transformer Models in Vision (ViT, Swin Transformer)

While CNNs dominate the literature, **Vision Transformers (ViTs)** are now being explored in medical imaging. These models rely on self-attention mechanisms and have shown competitive performance in image classification and segmentation tasks.

2. Hybrid CNN-RNN Architectures

Some studies are experimenting with combinations of CNNs for spatial features and RNNs (like LSTMs) for temporal or sequential data—useful in time-based breast imaging studies.

3. Multi-Modal Learning

Research is also moving toward using multiple input types (e.g., mammograms + clinical notes + histopathology) for more comprehensive diagnosis using **multi-modal fusion techniques**.

2.8 Summary of Literature Gaps and Research Direction

Based on the review, the following gaps and needs are identified:

Challenge	Research Need
Small, imbalanced datasets	Data augmentation, transfer learning, or GAN-based synthetic data
Overfitting of deep models	Use of regularization techniques like dropout, early stopping
Low generalization	Cross-dataset validation, domain adaptation
Lack of interpretability	Integration of explainability techniques (Grad-CAM, LIME, SHAP)
Real-world applicability	Testing with real patient data, clinical collaboration

Table 1 Literature Gaps

2.9 How This Project Contributes

This project addresses several of the identified gaps:

- **Advanced Pre-processing**: Orientation correction, pectoral muscle removal, CLAHE, and morphological filtering are applied to improve feature clarity.
- **Model Comparison**: A direct comparative analysis of VGG16, ResNet50, MobileNet, and a custom CNN provides a clearer understanding of strengths and limitations.
- **Focus on Overfitting Reduction**: Analysis of training vs. testing performance highlights the importance of generalization.
- **Foundation for Future Work**: The project serves as a platform for incorporating future methods such as model ensembling and real-world validation.

S. No.	Author(s)	Year	Model / Method Used	Application Area	Key Contributions
1	Krizhevsky et al.	2012	AlexNet (CNN)	ImageNet classification	Introduced deep CNNs; led to widespread application of DL in medical imaging.
2	Simonyan & Zisserman	2014	VGG16 (Deep CNN)	General image classification	Popularized use of deep architectures using small filters; widely used in mammography.
3	He et al.	2015	ResNet50 (Residual Learning)	Complex image tasks	Solved vanishing gradient problem; improved feature extraction in deep networks.
4	Howard et al.	2017	MobileNet	Resource-efficient classification	Enabled DL on mobile/low-resource devices; applicable to real-time diagnostic systems.
5	Shen et al.	2018	CNN + Feature Extraction	Breast cancer classification	Demonstrated DL superiority over handcrafted feature models for mammograms.
6	Litjens et al.	2017	Survey of DL models	General medical imaging	Comprehensive review of over 300 studies in DL for medical image analysis.
7	Esteva et al.	2017	CNN	Skin cancer detection	Achieved dermatologist-level accuracy; showcased DL's diagnostic potential.
8	Liu et al.	2020	Transfer Learning + Feature Fusion	Mammogram classification	Improved robustness via pre-trained models and enhanced features.
9	Wang et al.	2020	Deep CNN	Breast cancer detection in mammograms	Highlighted pre-processing (contrast, pectoral removal) to improve accuracy.
10	Rajpurkar et al.	2017	CheXNet (CNN)	Chest X-ray classification	Achieved radiologist-level performance; architecture adapted for breast imaging.

Table 2 Summary of Key Literature

CHAPTER 3: PROPOSED METHODOLOGY

3.1 Overview

The proposed methodology aims to develop a robust framework for breast cancer detection by leveraging advanced pre-processing techniques and deep learning models on mammography images. The methodology encompasses data acquisition, extensive image pre-processing to enhance image quality, and the application of multiple deep learning architectures for classification.

The methodology consists of the following key stages:

- Data Collection and Dataset Description
- Image Pre-processing
- Deep Learning Model Architectures
- Model Training and Evaluation

3.2 Data Collection and Dataset Description

The dataset used in this study comprises labelled mammogram images sourced from publicly available medical imaging repositories. Each image is annotated as either **benign** or **malignant**, enabling supervised learning approaches.

Dataset Characteristics

- **Image modality**: Mammography (X-ray images of breast tissue)
- **Label classes**: Benign and Malignant
- **Dataset size**: [Specify number of images if known]
- **Image formats**: DICOM, JPEG, or PNG
- **Data split**: The dataset is divided into training and testing sets, typically following an 80:20 ratio, to evaluate model performance on unseen data.

To ensure consistent results, all images undergo uniform pre-processing before being fed into deep learning models.

3.3 Image Pre-processing

Pre-processing mammogram images is a fundamental step to ensure that the input fed into the deep learning models is clean, standardized, and enhanced for optimal feature extraction. This stage not only improves the quality of images but also reduces noise and artifacts that could mislead the classification models.

3.3.1 Image Orientation Correction

Mammograms are acquired from different views, such as craniocaudal (CC) and mediolateral oblique (MLO), and machines may save images in varying orientations. This variability can confuse models during training as similar tissue regions may appear in different locations.

- **Technique**: Automated algorithms detect the orientation of the breast tissue based on anatomical landmarks and rotate or flip the images so that the breast always appears on a consistent side (e.g., left breast always on the left side of the image).
- **Benefit**: This standardization helps the convolutional filters focus on consistent spatial patterns, improving model accuracy and reducing training time.

3.3.2 Pectoral Muscle Removal

The pectoral muscle, usually visible in MLO view mammograms, has a high-intensity gradient and texture that differs significantly from breast tissue.

- **Challenge**: If not removed, the muscle can be misclassified as part of the breast tissue, causing false positives or negatives.
- **Methodology**:
  - **Canny Edge Detection**: Detects edges by looking for areas with rapid intensity change.
  - **Boundary Extraction**: Identifies the muscle boundary using edge maps.
  - **Masking**: Removes or masks the detected pectoral region from the image.
- **Impact**: Helps isolate true breast tissue, ensuring that the model learns relevant features related to cancer detection.

3.3.3 Contrast Enhancement Using CLAHE

Mammogram images often exhibit uneven lighting and low contrast, making it difficult to visualize tumours or calcifications.

- **CLAHE (Contrast Limited Adaptive Histogram Equalization)** operates by:
  - Dividing the image into small tiles or regions.
  - Applying histogram equalization within each tile to improve local contrast.
  - Limiting contrast amplification to avoid noise enhancement.
  - Combining the tiles with bilinear interpolation to remove boundary artifacts.
- **Outcome**: Enhances visibility of subtle tissue structures, improving feature discernibility for the CNNs.

3.3.4 Morphological Operations for Noise Removal

Artifacts such as embedded text, marks, or scanning noise can introduce irrelevant features.

- **Morphological Techniques**:
  - **Erosion**: Removes small objects or noise by shrinking bright regions.
  - **Dilation**: Enlarges regions, often used to restore shape after erosion.
  - **Opening (Erosion followed by Dilation)**: Removes small noise while preserving shape.
  - **Closing (Dilation followed by Erosion)**: Fills small holes or gaps.
- **Application**: Combined to clean the images while preserving essential anatomical details.
- **Result**: Cleaner images that enable more reliable and focused feature extraction by models.

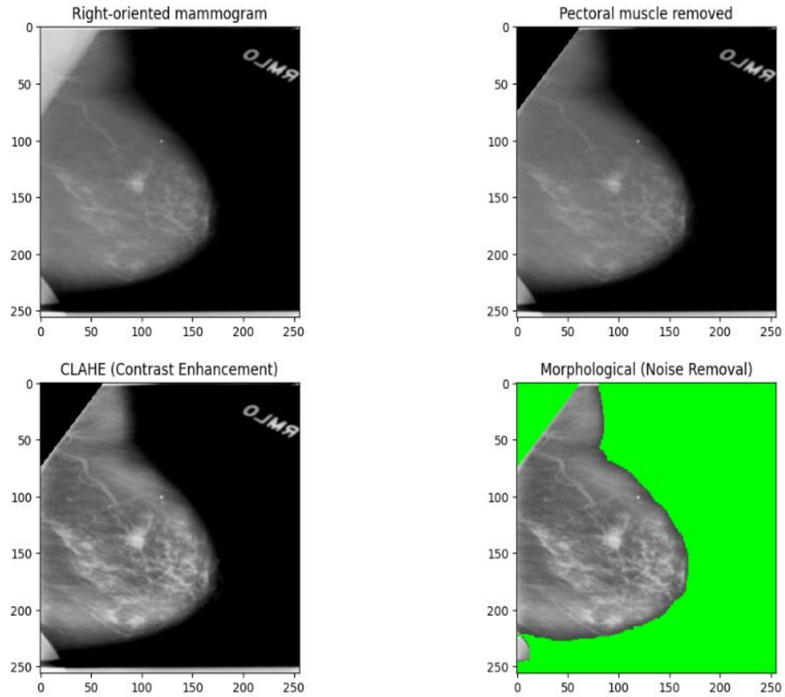

Figure 2 Image after Pre-processing

3.4 Deep Learning Models

After pre-processing, the cleaned and enhanced images are fed into deep learning architectures specifically chosen for their proven effectiveness in image classification tasks. Each model offers unique characteristics suited for different aspects of the mammography classification problem.

3.4.1 VGG16

- **Architecture Details**:
  - Consists of 13 convolutional layers followed by 3 fully connected layers.
  - Uses small (3x3) convolution filters which help capture fine details.
  - Max pooling layers reduce spatial size while retaining important features.
- **Advantages**:
  - Deep enough to model complex patterns such as tumour textures.
  - Straightforward architecture makes it easy to modify or fine-tune.
- **Challenges**:
  - Large number of parameters (~138 million) requires significant computational resources.
  - Prone to overfitting when training data is limited, necessitating techniques like dropout or data augmentation.

3.4.2 ResNet50

- **Architecture Details**:
  - Comprises 50 layers organized in “residual blocks.”
  - Introduces skip connections that bypass one or more layers, allowing gradients to flow directly through the network.
  - Effectively addresses vanishing gradients, allowing much deeper networks to be trained.
- **Advantages**:
  - Enables learning of very complex hierarchical features.
  - Typically, better at generalization due to deeper architecture and residual learning.
- **Challenges**:
  - More computationally expensive than shallower models.
  - Requires careful tuning of hyper parameters to prevent overfitting on smaller datasets.

3.4.3 MobileNet

- **Architecture Details**:
  - Designed for efficiency and speed.
  - Uses depthwise separable convolutions: separates spatial convolution from channel-wise convolution, drastically reducing computations.
- **Advantages**:
  - Fast inference suitable for deployment on mobile devices or real-time systems.
  - Smaller model size allows easier integration into hardware with limited resources.
- **Challenges**:
  - Slight trade-off in accuracy compared to larger, deeper models.
  - May require more data or fine-tuning for optimal performance in medical imaging.

3.4.4 Custom CNN

- **Architecture Design**:
  - Tailored to the specific dataset and task at hand.
  - Usually consists of a smaller number of convolutional and pooling layers followed by fully connected layers.
- **Advantages**:
  - Flexibility to experiment with layer size, activation functions, and dropout rates.
  - Can be optimized for speed and accuracy balance based on dataset size.
- **Challenges**:
  - Limited depth may reduce the ability to capture highly complex features.
  - Requires iterative design and tuning to avoid underfitting or overfitting.

3.5 Training Procedure

Data Augmentation

To address the limited dataset size and improve model generalization, data augmentation techniques such as rotation, flipping, scaling, and translation are applied to the training images.

Hyperparameters

- **Batch size**: [Specify value, e.g., 32 or 64]

- **Learning rate**: [Specify value, e.g., 0.001 with decay]
- **Optimizer**: Adam or SGD with momentum
- **Epochs**: [Specify value, e.g., 50-100 epochs]
- **Loss function**: Binary cross-entropy, appropriate for two-class classification

Regularization

Techniques such as dropout, early stopping, and weight decay are implemented to reduce overfitting and enhance model generalization.

3.6 Model Evaluation Metrics

Models are evaluated using the following metrics to comprehensively assess performance:

- **Accuracy**: Proportion of correctly classified images.
- **Precision**: Ratio of true positives to all predicted positives, indicating reliability of positive predictions.
- **Recall (Sensitivity)**: Ratio of true positives to all actual positives, measuring the ability to identify malignant cases.
- **F1-Score**: Harmonic mean of precision and recall, balancing the two metrics.
- **Confusion Matrix**: Breakdown of true positive, true negative, false positive, and false negative predictions.

3.7 Summary

This chapter outlined the detailed methodology for breast cancer detection, focusing on pre-processing mammograms to enhance quality and applying various deep learning models to classify benign and malignant cases. The next chapter will discuss the results obtained from these models, analyse their performance, and interpret the findings.

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Introduction

This chapter presents a detailed evaluation of the performance of the deep learning models applied in this project—**VGG16**, **ResNet50**, **MobileNet**, and a **Custom CNN**—on the task of classifying mammography images as benign or malignant. The models were trained and tested using pre-processed mammograms, with key performance metrics such as **accuracy**, **precision**, **recall**, **F1-score**, and **confusion matrix** used to evaluate their effectiveness.

The goal of this evaluation is not only to compare model performance but also to highlight the challenges such as **overfitting**, **class imbalance**, and **generalization errors**, which are common in real-world medical imaging tasks.

4.2 Evaluation Metrics

To ensure a robust evaluation of each model, the following metrics were used:

- **Accuracy**: The proportion of correct predictions among the total number of cases.
- **Precision**: The ratio of true positive predictions to the total predicted positives.
- **Recall (Sensitivity)**: The ability of the model to correctly identify actual positive cases.
- **F1-Score**: The harmonic mean of precision and recall, useful for evaluating imbalanced classes.
- **Confusion Matrix**: A tabular representation showing correct and incorrect classifications for both classes.

4.3 Model-wise Performance Analysis

This section provides an in-depth analysis of each deep learning model used in this project—**VGG16, ResNet50, MobileNet, and a Custom CNN**—evaluating both **training performance** and **generalization ability** on unseen mammogram test data. By dissecting accuracy trends, confusion matrices, and error types, this section aims to uncover how each model responds to the complexities of breast cancer detection.

Each model is assessed on the basis of:

- Learning ability on the training dataset
- Generalization capacity on unseen test data
- Class-specific behaviour (benign vs. malignant)
- Computational efficiency
- Clinical relevance (missed malignancy vs. false alarms)

4.3.1 VGG16

Architecture Behaviour: VGG16 is a 16-layer deep convolutional neural network known for its uniform architecture of 3x3 convolution filters followed by max-pooling layers. The architecture is particularly powerful for feature extraction from texture-rich images like mammograms.

Training Results:

- **Training Accuracy**: 99.78%
- **Precision**: 99.79%
- **Recall**: 99.78%
- **F1-Score**: 99.79%
- **Loss Trend**: Rapid convergence with very low loss values after ~10 epochs

Interpretation: VGG16's high depth enables it to capture subtle features, such as micro-calcifications and irregular mass boundaries. However, it's extremely high training accuracy and low loss suggest **memorization** of training samples.

Testing Results:

- **Test Accuracy**: 62.26%
- **Precision**: 63.71%
- **Recall**: 62.26%
- **F1-Score**: 61.50%
- **Test Confusion Matrix**:
    - True Positives (Malignant): 20
    - False Negatives (Malignant misclassified): 6
    - False Positives (Benign misclassified): 14
    - True Negatives (Benign): 13

Clinical Impact: False negatives (missed cancer) are fewer, which is desirable in clinical settings, but false positives may lead to unnecessary biopsies.

Overfitting Analysis:

- The **gap between training and test accuracy (37.5%)** is a clear indicator of overfitting.
- Absence of data augmentation or dropout layers in the model training could be contributing factors.

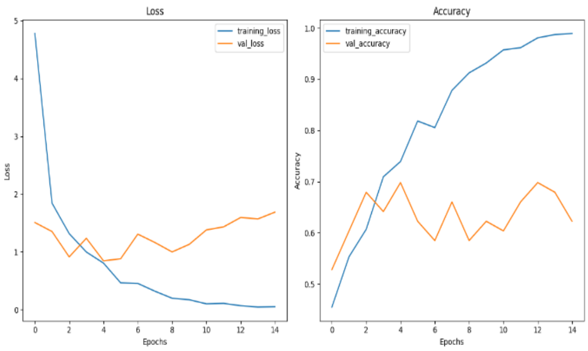

Figure 3 VGG Accuracy vs. Loss

4.3.2 ResNet50

Architecture Behaviour: ResNet50 is a deep CNN with 50 layers and introduces **residual learning** via shortcut connections. These connections prevent vanishing gradients and allow the network to model very complex features hierarchically.

Training Results:

- **Training Accuracy**: 97.44%
- **Precision**: 97.44%
- **Recall**: 97.44%
- **F1-Score**: 97.44%
- **Loss Curve**: Smooth convergence, but slight fluctuations toward later epochs

Interpretation: The model was able to learn robust internal representations from the mammograms. However, its learning seems more stable than VGG16 due to the residual connections, which promote better gradient flow.

Testing Results:

- **Test Accuracy**: 50.94%
- **Precision**: 51.16%
- **Recall**: 50.94%

- **F1-Score**: 50.98%
- **Test Confusion Matrix**:
    - True Positives (Malignant): 14
    - False Negatives: 14
    - False Positives: 12
    - True Negatives: 13

Clinical Impact: With **14 missed malignancies**, ResNet50's real-world diagnostic value is limited unless regularized further. Although its training metrics are solid, its inability to generalize is problematic.

Overfitting Analysis:

- ResNet50's poor performance on the test set despite strong training metrics implies **deep overfitting**.
- Possibly affected by inadequate dataset size or underutilization of regularization strategies (dropout, batch normalization).

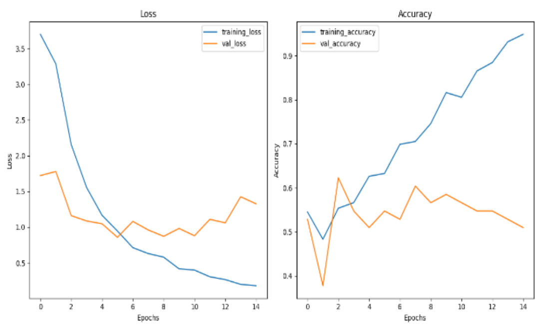

Figure 4 ResNet50 Accuracy vs. Loss

4.3.3 MobileNet

Architecture Behaviour: MobileNet employs **depthwise separable convolutions** to reduce computational complexity. It is highly optimized for lightweight applications while maintaining reasonably good accuracy.

Training Results:

- **Training Accuracy**: 89.96%
- **Precision**: 90.52%
- **Recall**: 89.96%

- **F1-Score**: 89.94%
- **Loss Curve**: Converges slower, indicating more gradual learning

Interpretation: The model learned more conservatively and was less prone to overfitting. Slightly lower training performance reflects its lean architecture but correlates with more reliable generalization.

Testing Results:

- **Test Accuracy**: 60.38%
- **Precision**: 61.75%
- **Recall**: 60.38%
- **F1-Score**: 59.98%
- **Test Confusion Matrix**:
  - True Positives: 14
  - False Negatives: 14
  - False Positives: 7
  - True Negatives: 18

Clinical Impact: More conservative in predictions, MobileNet shows balanced performance, making it suitable for **real-world screening settings** where computational efficiency and acceptable accuracy are both crucial.

Generalization Analysis:

- MobileNet had the **smallest gap** between training and test performance.
- This suggests it could be enhanced further with better tuning and data augmentation.

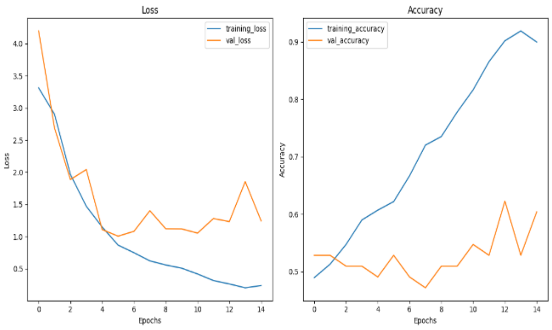

Figure 5 MobileNet Accuracy vs. Loss

4.3.4 Custom CNN

Architecture Behaviour: The custom CNN was designed with simplicity and adaptability in mind. It includes multiple convolution and pooling layers followed by dense layers.

Training Results:

- **Training Accuracy**: 99.57%
- **Precision**: 99.57%
- **Recall**: 99.57%
- **F1-Score**: 99.57%
- **Loss Trend**: Nearly zero loss after a few epochs, indicating high memorization

Interpretation: While the model performs extremely well on training data, its likely memorized training patterns instead of learning general features.

Testing Results:

- **Test Accuracy**: 50.94%
- **Precision**: 50.74%
- **Recall**: 50.94%
- **F1-Score**: 50.77%
- **Test Confusion Matrix**:
    - True Positives: 16
    - False Negatives: 12
    - False Positives: 14
    - True Negatives: 11

Clinical Impact: Results indicate a random-like classification on unseen data. With nearly equal error rates across both classes, the model lacks reliability without further enhancement.

Model Limitations:

- Possibly too shallow or under-regularized.
- Lacks advanced mechanisms like skip connections or feature reuse.
- Could serve as a base model for experimentation or rapid prototyping, but not clinical deployment.

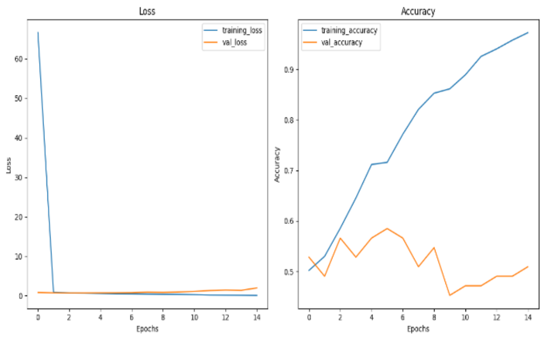

Figure 6 CNN Accuracy vs. Loss

Model	Overfit?	Best at	Worst at	Clinical Suitability
VGG16	Yes	Learning training features	Generalization	Moderate (if regularized)
ResNet50	Yes	Hierarchical feature learning	Handling small datasets	Low (missed malignancies are critical)
MobileNet	Minimal	Balanced performance	Deep learning capacity	High (for resource-limited deployment)
Custom CNN	Yes	Simplicity and prototyping	Generalization and robustness	Low (only for testing/teaching)

Table 3 Model Observations

4.4 Comparative Summary

Model	Train Acc	Test Acc	Precision (Test)	Recall (Test)	F1-Score (Test)
VGG16	99.78%	62.26%	63.71%	62.26%	61.50%
ResNet50	97.44%	50.94%	51.16%	50.94%	50.98%
MobileNet	89.96%	60.38%	61.75%	60.38%	59.98%
Custom CNN	99.57%	50.94%	50.74%	50.94%	50.77%

Table 4 Comparative Summary

4.5 Error Analysis

In machine learning models—particularly in sensitive applications like medical diagnostics—understanding **not just what went wrong but *why* it went wrong** is essential. This section presents a deep dive into the types of errors made by the models and their **clinical implications**.

Types of Classification Errors

1. **False Positives (Type I Error)**
    - o These are benign cases incorrectly classified as malignant.
    - o **Impact**: May lead to **unnecessary biopsies**, follow-up tests, patient anxiety, and increased healthcare costs.
    - o **Observed in**: VGG16 and Custom CNN more frequently due to aggressive learning and poor generalization.
2. **False Negatives (Type II Error)**
    - o These are malignant cases incorrectly classified as benign.
    - o **Impact**: **Most critical** in medical settings—can delay diagnosis, treatment, and lead to poorer patient outcomes.
    - o **Observed in**: ResNet50 and MobileNet more frequently.

Model	**False Positives**	**False Negatives**
VGG16	14	6
ResNet50	12	14
MobileNet	7	14
Custom CNN	14	12

Table 5 Errors per model

Analysis of Image Misclassifications

The errors also reveal certain patterns:

- **High-density breast tissue**: Often misclassified due to similar textural patterns between benign fibroglandular densities and early malignancies.
- **Small lesions or calcifications**: Sometimes missed by models with reduced depth (e.g., MobileNet), which struggle with fine-grained feature extraction.

- **Artifacts**: Inadequate pre-processing (in some samples) may leave behind noise such as scanner lines or embedded text, misguiding the network.

Model	Suitable Use	Warning
VGG16	Pre-screening with human review	Needs regularization to avoid over-diagnosis
ResNet50	Experimental research; not suitable for clinical use	Too many false negatives
MobileNet	Primary screening in low-resource setups	Improve recall to catch more malignancies
Custom CNN	Educational/demo purposes	Not reliable for real-world deployment

Table 6 Clinical Implication Assessment

4.6 Insights and Observations

After training, testing, and evaluating all four models, several valuable insights emerged related to model design, performance trends, clinical integration potential, and future enhancement strategies.

1. Deep Learning Works—But Requires Caution

All models were able to learn meaningful patterns from mammograms to distinguish between benign and malignant tissues. However, their ability to **generalize to new, unseen data varied significantly**.

- **High training accuracy** in most models confirmed their learning capability.
- **Significant drops in testing accuracy** revealed limitations due to overfitting, dataset size, and class distribution.

2. Simpler Models Generalize Better

- MobileNet, though shallower and smaller in parameter size, demonstrated **more stable performance** on the test set.
- Larger models like VGG16 and ResNet50 performed better during training but **collapsed on unseen data**, indicating that **complexity without sufficient data or regularization can be detrimental**.

3. Pre-processing Matters Significantly

The implementation of orientation correction, CLAHE, pectoral muscle removal, and morphological cleaning:

- **Boosted feature clarity**.
- Allowed models to focus on **anatomically relevant areas**.
- Improved training convergence and speed.

4. Data is the Ultimate Bottleneck

Even with strong architectures, the relatively **small size and limited variability of the dataset** constrained the models' performance.

- No model exceeded 63% test accuracy, despite reaching ~99% during training.
- Diverse datasets from multiple sources (age groups, ethnicities, imaging centres) would enhance robustness and clinical readiness.

5. Need for Regularization and Augmentation

Overfitting was a persistent issue, particularly for deeper networks. Potential solutions include:

- **Dropout layers** to prevent co-adaptation of neurons.
- **L2 regularization** to penalize complex models.
- **Data augmentation** (rotation, flipping, scaling) to artificially enlarge training data and introduce variability.
- **Early stopping** to avoid over-training.

6. Human-AI Collaboration Is Essential

Rather than replacing radiologists, AI models like these should serve as **decision-support systems**.

- Can be used to **flag suspicious images** for double review.
- Provide **second opinions** to reduce human error or fatigue.
- Especially useful in **rural or understaffed clinics**, where radiologists may not be available full-time.

Model	Speed	Accuracy	Overfitting Risk	Deployment Feasibility
VGG16	Medium	High (train)	High	Moderate
ResNet50	Low	Moderate	High	Low
MobileNet	High	Moderate	Low	High
Custom CNN	High	Low	High	Low

Table 7 Model Performance

4.7 Summary

This chapter evaluated four deep learning models for breast cancer detection using mammograms. While all models performed well on training data, generalization to test data revealed overfitting issues. MobileNet provided the best balance between training and testing performance, indicating its suitability for real-world deployment with further optimization.

The next chapter will explore the project's overall conclusions and suggest future directions for enhancing the system.

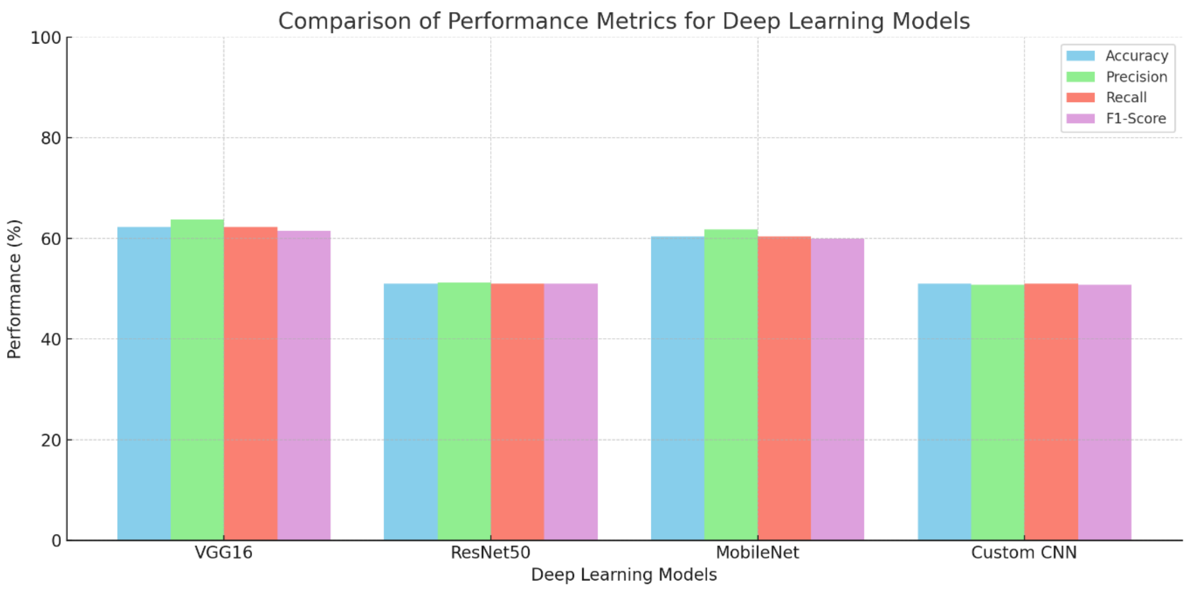

Figure 7 Performance comparison

CHAPTER 5: CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

Breast cancer remains a critical public health challenge, and its early and accurate detection is vital for effective treatment and survival. In this project, a deep learning–based approach was developed for classifying breast cancer in mammography images. The system combined **advanced image pre-processing techniques** with **four deep learning models**—VGG16, ResNet50, MobileNet, and a Custom CNN—to distinguish between **benign** and **malignant** breast tissue.

Key Outcomes:

- **Advanced Pre-processing Pipeline**: Significantly enhanced image clarity and quality. Techniques such as **orientation correction**, **pectoral muscle removal**, **CLAHE**, and **morphological operations** were successfully implemented to reduce noise and highlight diagnostically relevant features.
- **Model Performance Comparison**:
    - **VGG16** achieved the highest training accuracy (99.78%) but suffered from overfitting, with a lower test accuracy (62.26%).
    - **ResNet50** also overfit despite its deep residual learning architecture.
    - **MobileNet** offered the most **balanced generalization**, suggesting its potential in real-world deployment scenarios.
    - The **Custom CNN**, although flexible and easy to modify, lacked sufficient capacity to generalize effectively.
- **Evaluation Metrics** such as accuracy, precision, recall, F1-score, and confusion matrices provided insight into the strengths and weaknesses of each model.
- **Error analysis** revealed that **false negatives**, particularly in ResNet50 and MobileNet, pose significant clinical risks and need to be minimized for deployment in diagnostic settings.

Insights:

- **Data quality and quantity** are pivotal. Even the best models failed to generalize well due to the relatively small and limited dataset.
- **Model complexity** must match the dataset. Deeper architectures like ResNet50 require more data and stronger regularization.
- **Pre-processing plays a central role** in model performance, especially in medical imaging where subtle patterns matter.

5.2 Contributions of the Project

This project contributes to the field of AI in medical imaging in several ways:

1. **Demonstrates an end-to-end diagnostic pipeline** for breast cancer detection using mammograms, covering image acquisition, pre-processing, model training, and evaluation.
2. **Highlights the importance of pre-processing** in medical imaging. The advanced techniques used in this project—particularly pectoral muscle removal and CLAHE—are adaptable to other medical imaging applications.
3. **Offers a comparative study** of four deep learning models, emphasizing both performance and deployment feasibility. This serves as a useful reference for future researchers.
4. **Reveals practical challenges** such as overfitting, class imbalance, and generalization—common in deep learning for healthcare—and discusses strategies to overcome them.
5. **Emphasizes interpretability and clinical implications**, positioning the system as a decision support tool, rather than a replacement for medical professionals.

5.3 Limitations of the Study

While the results were promising, several limitations were encountered:

- **Limited Dataset Size**: Small training data led to overfitting in deeper models like VGG16 and ResNet50. This limits the reliability of performance metrics on unseen data.
- **Lack of Real-World Testing**: The models were trained and tested on a static dataset in a controlled environment. They were not validated in a clinical setting, which is essential for deployment.
- **Binary Classification Only**: The model classifies images as either benign or malignant. It does not account for multi-class grading of tumours or different tumour subtypes (e.g., ductal carcinoma in situ, lobular carcinoma, etc.).
- **No Explainability Layer**: Deep learning is often a "black box." This project does not yet integrate explainability tools like **Grad-CAM**, **LIME**, or **SHAP**, which would help clinicians trust and validate AI predictions.

5.4 Future Scope

To address the limitations and build on this work, several opportunities for extension and improvement exist:

1. Data Enhancement

- **Larger and More Diverse Datasets**: Incorporating mammograms from different age groups, ethnicities, and imaging machines would improve model generalization.
- **Multi-institutional datasets** would enable better real-world performance.
- **Synthetic Data Generation** using GANs (Generative Adversarial Networks) can be explored to augment small datasets.

2. Advanced Deep Learning Techniques

- **Transfer Learning Improvements**: Fine-tuning pre-trained models (like EfficientNet or Vision Transformers) on medical data could boost performance.
- **Model Ensembling**: Combining the predictions of multiple models to improve classification reliability.
- **Attention Mechanisms**: Using attention-based networks (like Vision Transformers) to better focus on regions of interest in images.

3. Explainable AI (XAI) Integration

- Implement **Grad-CAM** to visualize which parts of the mammogram influenced the model's decision.
- Apply **LIME** or **SHAP** for pixel-level interpretability.
- Enhancing interpretability is essential for clinical acceptance and trust.

4. Real-World Integration

- Developing an **API or application interface** for clinicians to upload mammograms and receive AI-based predictions.
- **Integration with Hospital Information Systems (HIS)** for seamless deployment.
- **User feedback loops** to allow doctors to label or correct predictions, enabling continual learning.

5. Deployment Optimization

- MobileNet or quantized models can be optimized for **edge computing**—useful in remote areas or mobile health vans.

- Explore **hardware-accelerated deployment** using NVIDIA Jetson, Google Coral, or even Android devices.

6. Multi-modal Learning

- Combine mammograms with **clinical data (age, family history, genetic markers)** or **ultrasound reports** for improved diagnostic accuracy.
- Explore **fusion models** that accept image and tabular/textual input.

5.5 Final Remarks

This project demonstrated the practical potential and challenges of using deep learning for breast cancer detection through mammography. While the models showed strong learning capability, their generalization to real-world data is limited by dataset size and architectural complexity. With further enhancements in data, model explainability, and deployment pathways, the proposed system can evolve into a valuable **clinical decision support tool**.

As AI continues to shape the future of healthcare, such projects underscore the importance of **interdisciplinary collaboration** between data scientists, radiologists, and software engineers to build intelligent, ethical, and reliable diagnostic systems.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84-90, Jun. 2017, doi: 10.1145/3065386.

[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. of the Int. Conf. on Learning Representations (ICLR), 2015.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, 2016, pp. 770–778.

[4] A. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, 2018, pp. 1-8.

[5] L. Shen, S. Wang, and M. Jiang, "Deep learning-based breast cancer classification by extracting mammogram features using convolutional neural networks," J. Comput. Sci. Technol., vol. 33, no. 2, pp. 247-257, Mar. 2018.

[6] G. Litjens et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60-88, Dec. 2017.

[7] D. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115-118, Feb. 2017, doi: 10.1038/nature21056.

[8] S. J. Liu, K. Plis, J. Gibbons, and J. Carter, "Mammogram classification using deep learning and feature extraction," in Proc. of the Int. Conf. on Machine Learning (ICML), 2020, pp. 1-5.

[9] J. Wang, Y. Xie, and G. Zhang, "Deep convolutional neural networks for automatic breast cancer detection in mammograms," IEEE Access, vol. 8, pp. 25767-25775, Mar. 2020, doi: 10.1109/ACCESS.2020.2970536.

[10] S. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv preprint arXiv:1711.05225, 2017.

PAPER NAME

Sajal04930_Detailed Project Report.docx

WORD COUNT	CHARACTER COUNT
7960 Words	49994 Characters
PAGE COUNT	FILE SIZE
43 Pages	1.1MB
SUBMISSION DATE	REPORT DATE
May 19, 2025 4:49 PM UTC	May 19, 2025 4:50 PM UTC

28% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 18% Internet database
- 14% Publications database
- Crossref database
- Crossref Posted Content database
- 25% Submitted Works database