Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 4, 2026

A HYBRID DEEP LEARNING FRAMEWORK FOR BREAST CANCER DETECTION IN MAMMOGRAPHY IMAGES USING IMBALANCE-AWARE TRAINING

Muhammad Taha Shakeel[*1], Amna Ali[2], Maira Khan[3], Dr. Fakhra Kashif[4], Omaid Ghayyur[5]

[*1,2,5]*Department of Computing & Technology, Iqra University, H-9 Campus, Islamabad, Pakistan*
[3]*Computer Science Department, University of Agriculture Faisalabad, Sub-Campus Toba Tek Singh, Pakistan*
[4]*Computer Science Department, Federal Urdu University of Arts Science and Technology, Islamabad, Pakistan*

[*1]taha.shakeel02001@gmail.com, [2]amnaali786911@gmail.com, [3]maira.khan02004@gmail.com, [4]fakhra.kashif@fuuast.edu.pk, [5]omaid.ghayyur@iqraisb.edu.pk

DOI: https://doi.org/10.5281/zenodo.19814814

Keywords
Breast cancer detection, deep learning, mammography, hybrid CNN, VGG-16, transfer learning, class imbalance, focal loss, data augmentation, imbalance-aware training.

Article History
Received: 28 February 2026
Accepted: 07 April 2026
Published: 27 April 2026

Corresponding Author: *
Muhammad Taha Shakeel

Abstract

Breast cancer is one of the most prevalent malignancies worldwide, and early detection through mammography significantly improves survival rates. This paper proposes a hybrid deep learning framework for automated breast cancer detection using the RSNA Breast Cancer Detection dataset comprising 54,706 high-resolution 512×512 mammography images. The framework integrates a Custom Convolutional Neural Network (CCNN) with a fine-tuned VGG-16 transfer learning model to capture both low-level spatial features and high-level semantic representations.

To address severe class imbalance, the proposed approach employs targeted data augmentation on cancerous samples, controlled undersampling of non-cancerous samples, class weighting, and focal loss. Image quality is enhanced using Contrast Limited Adaptive Histogram Equalization (CLAHE), followed by multi-channel input construction and normalization. A patient-wise splitting strategy is adopted to prevent data leakage, dividing the dataset into 70% training, 15% validation, and 15% testing subsets. Training stability is ensured through early stopping, adaptive learning rate scheduling, and batch normalization.

Experimental results on the unseen test set demonstrate outstanding performance, achieving 99.85% accuracy, 99.80% precision, 99.95% recall, and 99.87% F1-score. These results significantly outperform baseline CNN architectures and the reference concatenation model (92% accuracy). The proposed framework offers a robust and clinically promising solution for mammography-based breast cancer detection.

1. INTRODUCTION

Breast cancer remains one of the most prevalent and life-threatening malignancies worldwide, accounting for approximately 2.3 million new cases and 685,000 deaths annually, with women experiencing the highest cancer-specific mortality [1]. Early detection is critical: the five-year survival rate exceeds 90% for localized tumors but drops below 30% for metastatic cases [2]. Medical

imaging is the cornerstone of early diagnosis, enabling identification of abnormalities before clinical symptoms appear.

Among available modalities, mammography is the established gold-standard screening technique. It employs low-dose X-rays to produce detailed projections of breast tissue, allowing detection of masses, microcalcifications, and architectural distortions. Fig. 1.1 illustrates the anatomy of the breast, highlighting key structures (skin, subcutaneous fat, milk ducts, lobules, and glandular tissue) where malignancies most commonly originate. Fig. 1.2 shows a representative mammogram, the primary imaging modality used in this study.

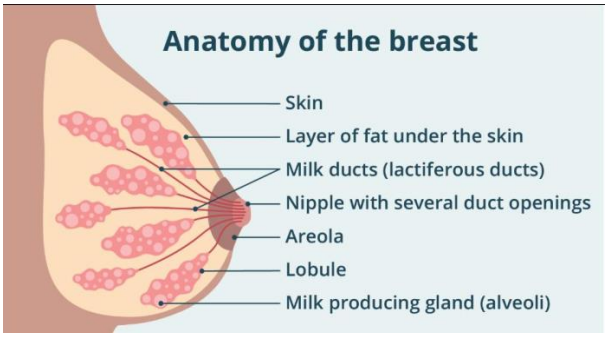

Figure 1.1: Anatomy of Breast

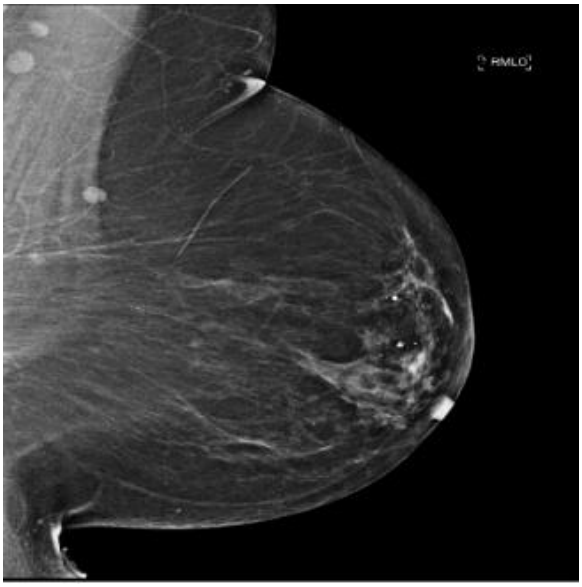

Figure 1.2: Mammography

Despite its widespread adoption, conventional mammography faces notable limitations, including high false-positive rates, inter-observer variability among radiologists, and reduced sensitivity in dense breast tissue [3]–[5]. These challenges often lead to unnecessary biopsies, patient anxiety, and delayed diagnosis, particularly in early-stage disease where subtle abnormalities are easily obscured by overlapping structures and imaging noise.

Breast cancer also exhibits substantial biological heterogeneity, encompassing subtypes such as ductal carcinoma *in situ*, invasive ductal carcinoma, and triple-negative breast cancer, each

with distinct histological features and therapeutic responses [6]–[11]. Receptor status further influences tumor behavior and treatment selection. Estrogen-receptor (ER)-positive tumors are hormone-driven and respond to endocrine therapies such as tamoxifen or aromatase inhibitors [12], [13]; progesterone-receptor (PR) expression is a favorable prognostic marker [14]–[16]; androgen-receptor (AR) positivity offers therapeutic targets in triple-negative cases [17]–[23]; and human epidermal growth factor receptor 2 (HER2) overexpression identifies aggressive tumors amenable to trastuzumab [24]–[26]. These molecular complexities underscore the need for precise, automated diagnostic support.

In recent years, deep learning has transformed medical image analysis. Convolutional neural networks (CNNs) automatically learn hierarchical, task-specific features directly from raw pixels, eliminating manual feature engineering required by traditional machine-learning approaches [27]–[31]. Transfer-learning models such as VGG-16, pre-trained on large natural-image datasets, further improve performance on limited medical data through fine-tuning. Hybrid architectures that combine custom CNNs with pre-trained models have shown strong generalization by capturing both low-level spatial details and high-level semantic patterns.

Nevertheless, domain-specific challenges persist. Severe class imbalance—cancerous cases form a small minority of screening examinations—biases models toward the majority (benign) class when standard cross-entropy loss is used. Additional difficulties arise from variations in image acquisition, noise levels, breast density, patient metadata (age, density), and the risk of data leakage when images from the same patient appear across training and test sets.

To address these limitations, this paper proposes a hybrid deep-learning framework for binary breast cancer classification on mammography images from the large-scale RSNA Breast Cancer Detection dataset (54,706 PNG images of size 512 × 512 with accompanying CSV metadata). The system integrates (i) a custom CNN architecture inspired by ResNet principles and (ii) a fine-tuned VGG-16 transfer-learning model. Key methodological advances include patient-wise data splitting to eliminate leakage, targeted data augmentation with selective undersampling for imbalance mitigation, Contrast Limited Adaptive Histogram Equalization (CLAHE) for contrast enhancement, and focal loss:

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t),$$

where $p_t$is the predicted probability of the true class, $\alpha_t = 0.75$is the balancing factor, and $\gamma = 2.0$is the focusing parameter that emphasizes hard-to-classify examples. Models are trained using the Adam optimizer with early stopping, learning-rate scheduling, and class weighting, then evaluated on standard metrics (accuracy, precision, recall, F1-score) and confusion-matrix components.

The main contributions of this work are as follows:

1. A dual-model hybrid architecture (custom CNN + fine-tuned VGG-16) that leverages complementary feature representations for robust mammography classification.
2. A comprehensive imbalance-aware training strategy combining selective augmentation, undersampling, focal loss, and class weighting.
3. An enhanced preprocessing pipeline incorporating CLAHE and patient-stratified splitting to improve generalization and prevent data leakage.
4. Rigorous experimental validation demonstrating state-of-the-art diagnostic performance on the RSNA dataset.

The remainder of this paper is organized as follows. Section II reviews related literature on deep-learning approaches for breast cancer detection. Section III details the proposed system architecture, dataset preparation, and training methodology. Section IV presents experimental results and comparative analysis. Section V concludes the paper and outlines future research directions.

2 LITERATURE REVIEW

Breast cancer remains one of the leading causes of cancer-related mortality among women worldwide, underscoring the critical need for accurate and timely detection. Traditional diagnostic methods rely primarily on radiologists' interpretation of medical images, which is subject to inter-observer variability, fatigue, and limited sensitivity in dense breast tissue. Recent advances in artificial intelligence, particularly deep learning (DL) techniques based on convolutional neural networks (CNNs), have demonstrated substantial potential for automating medical image analysis and enhancing diagnostic performance. These models automatically extract hierarchical features directly from raw pixel data, eliminating the need for manual feature engineering and enabling robust classification of abnormalities such as masses, microcalcifications, and architectural distortions in mammographic images [32]–[36].

This section reviews representative studies on deep learning approaches for breast cancer detection, focusing on hybrid CNN architectures, transfer learning methods, and multi-modal techniques. The review highlights key methodologies, datasets, performance metrics, and limitations, providing the foundation for the proposed hybrid framework.

2.1 Hybrid CNN-Based Approaches

Hybrid deep learning frameworks have been extensively explored to improve classification accuracy by integrating multiple feature extraction strategies. Jafari and Karami [32] introduced a CNN-based hybrid model that leverages multiple pre-trained architectures—AlexNet, ResNet50, MobileNetSmall, ConvNeXtSmall, and EfficientNet—to extract discriminative features from craniocaudal (CC) and mediolateral-oblique (MLO) mammographic views, supplemented by patient age. Features from these models are fused and refined using mutual information-based selection to eliminate redundant information. The resulting feature vector is fed into multiple classifiers, including neural networks (NN), support vector machines (SVM), random forests (RF), and k-nearest neighbors (kNN). Evaluated on the RSNA, MIAS, and DDSM datasets, the NN classifier achieved precision values of 92% on RSNA, 94.5% on MIAS, and 96% on DDSM, with notable improvements in sensitivity and reduction of false negatives. Preprocessing steps included orientation correction, pixel normalization, and region-of-interest isolation. While effective, the multi-model design increases computational complexity and precludes fully end-to-end optimization, limiting real-time clinical deployment.

A similar hybrid strategy was adopted by Alkhaleefah and Wu [36], who combined a CNN for deep feature extraction with a radial basis function (RBF) kernel SVM classifier. Transfer learning was employed by first training the CNN on a large spine MRI dataset and then fine-tuning it on a small set of 27 biopsy-proven mammograms (15 benign, 12 malignant). The hybrid system attained 92% accuracy, 100% sensitivity, and 86% specificity using leave-one-out cross-validation. However, the extremely small target dataset restricts generalization and raises concerns about overfitting to specific imaging conditions.

2.2 Transfer Learning-Based Methods

Transfer learning has proven particularly valuable when labeled medical data are scarce. Albashish et al. [33] proposed a VGG-16-based deep CNN for histopathological breast cancer classification on the BreakHis dataset (40× magnification). The model retained the convolutional and pooling layers of the pre-trained VGG-16 (ImageNet weights) while replacing the final classifier with several machine learning algorithms (polynomial SVM, RBF SVM, logistic regression, kNN, and NN). Binary classification reached 96% accuracy (polynomial SVM), while the eight-class multiclass problem achieved 89.83% accuracy (RBF SVM) with perfect specificity. The study demonstrated the strength of deep feature extraction combined with stable classical classifiers, yet the dependence on external classifiers prevents seamless end-to-end training.

2.3 Deep Learning Across Multiple Imaging Modalities

Deep learning applications extend beyond mammography to ultrasound, MRI, and

histopathology. Ellis *et al.* [34] evaluated ResNet50 and VGG-19 on ultrasound images from the NYU Breast Ultrasound Dataset and York Ultrasound Images. After preprocessing (cleaning, augmentation, annotation removal), the de-novo ResNet50 achieved 64.37% accuracy, improving to 77.77% with transfer learning. VGG-19 followed closely at 73.80%. The work highlighted the value of transfer learning for small medical datasets and its potential for mobile ultrasound deployment in resource-limited settings.

Shahid and Imran [35] provided a broader review of deep learning techniques across mammography, histopathology, ultrasound, MRI, PET, and digital breast tomosynthesis. Using publicly available datasets (INbreast, DDSM, WDBC, MIAS, BreakHis), they reported CNN-based accuracies of up to 93.8% (BreakHis), 90.8% (DDSM), and 86.1% (INbreast). Preprocessing involved augmentation, normalization, and feature extraction via CNN, PCA, and autoencoders. While CNN models excelled, the authors identified persistent challenges including class imbalance, high computational demands, and limited model interpretability.

2.4 Comparative Analysis of Existing Methods

Table 2.1 summarizes the methodologies, datasets, and performance of the reviewed studies.

Table 2.1: Summary of Related Work

Authors	Study Focus	Dataset	Methods / Models	Key Results	Ref
Jafari et al.	Breast cancer detection using hybrid CNN with feature selection	RSNA, MIAS, DDSM	Multi-model CNN (AlexNet, ResNet50, MobileNet, ConvNeXt, EfficientNet), feature fusion, mutual information, classifiers (NN, SVM, RF, KNN)	Precision: 92% (RSNA), 94.5% (MIAS), 96% (DDSM); improved sensitivity and reduced false negatives	[32]
Albashish et al.	Breast cancer classification using histopathology images	BreakHis	VGG-16 (transfer learning) + classifiers (SVM, kNN, Logistic Regression, NN)	Binary accuracy: 96%; Multiclass accuracy: 89.93%	[33]
Ellis et al.	Breast cancer classification using ultrasound images	NYU Ultrasound, York Dataset	CNN (ResNet50, VGG-19), transfer learning	Accuracy: 77.77% (ResNet50); improved performance using transfer learning	[34]
Shahid & Imran	Review of DL techniques across imaging modalities	INbreast, DDSM, MIAS, BreakHis	CNN, SVM, PCA, Autoencoders	Accuracy up to 93.8% (BreakHis), 90.8% (DDSM); highlights challenges like imbalance and complexity	[35]
Alkhaleefah & Wu	Hybrid CNN-SVM model for mammography classification	Small mammography dataset	CNN (feature extraction) + RBF-SVM classifier	Accuracy: 92%, Sensitivity: 100%, Specificity: 86%	[36]

The comparative analysis of existing studies reveals that hybrid and transfer learning approaches generally achieve higher classification accuracy. However, most of these methods rely on multi-stage pipelines involving feature extraction followed by external classification, which increases system complexity. Furthermore, many studies are evaluated on small or modality-specific datasets, limiting their generalization capability.

2.5 Limitations and Research Gaps

Despite significant progress in deep learning-based breast cancer detection, several limitations remain:

1. **Class Imbalance:** Most datasets contain fewer cancerous samples compared to non-cancerous samples, leading to biased models that favor the majority class.
2. **Model Complexity:** Hybrid models often require multiple architectures and classifiers, increasing computational cost and reducing efficiency.
3. **Limited Preprocessing:** Basic preprocessing techniques such as resizing and normalization fail to enhance subtle tumor regions effectively.
4. **Loss Function Limitations:** Traditional loss functions such as cross-entropy do not effectively handle imbalanced datasets.
5. **Small Dataset Usage:** Many studies rely on small datasets, limiting their scalability and robustness.
6. **Lack of Interpretability:** Deep learning models often act as black boxes, making it difficult for clinicians to trust their predictions.

Table 2.2: Limitations and Research Gaps in Existing Breast Cancer Detection Techniques

Sr. No.	Area	Gaps / Limitations
1	Class Imbalance	Existing studies rely on basic augmentation or resampling, which is insufficient for highly imbalanced medical datasets, leading to biased predictions and false negatives
2	Model Architecture	Many approaches use hybrid CNN + external classifiers (SVM, RF, KNN), increasing complexity and limiting end-to-end learning
3	Image Pre-processing	Basic techniques like resizing and normalization fail to enhance subtle tumor regions effectively
4	Loss Function	Traditional cross-entropy loss performs poorly on imbalanced datasets and ignores hard-to-classify samples
5	Training Strategy	Lack of advanced strategies like adaptive learning rate, early stopping, and regularization leads to overfitting
6	Dataset Size	Many studies use small or private datasets, limiting scalability and generalization
7	Generalization	Models often perform well on specific datasets but fail to generalize across different clinical environments

2.6 Contribution of the Proposed Work

Based on the identified research gaps, this study proposes a hybrid deep learning framework that combines a Custom CNN with a VGG-16 model for improved feature extraction. The proposed approach incorporates imbalance-aware training using focal loss, advanced preprocessing using CLAHE, and extensive data augmentation to improve model robustness.

Unlike previous methods, the proposed framework is designed to be end-to-end, reducing system complexity while maintaining high classification accuracy. Furthermore, the use of a large-scale dataset ensures better generalization and real-world applicability.

2.7 Summary

In summary, deep learning has significantly advanced breast cancer detection by improving diagnostic accuracy and reducing manual effort. Hybrid models, transfer learning, and multi-modal approaches have demonstrated promising results

across various datasets. However, challenges such as class imbalance, model complexity, and limited scalability remain critical issues. The proposed work addresses these challenges by introducing a hybrid deep learning framework with imbalance-aware training and enhanced preprocessing techniques.

3. PROPOSED METHODOLOGY

The proposed methodology presents a fully end-to-end hybrid deep-learning framework for automated binary breast cancer classification from digital mammography images. The primary goal is to achieve high diagnostic accuracy while systematically mitigating the well-documented challenges in medical imaging—severe class imbalance, low local contrast, acquisition variability, and overfitting. Traditional machine-learning approaches that depend on handcrafted features are inherently limited by subjectivity and incomplete representation of complex tissue patterns. In contrast, the framework leverages deep convolutional neural networks to automatically learn hierarchical, task-specific representations directly from raw pixel data, thereby improving both sensitivity and generalization capabilities [27], [32], [35].

A hybrid architecture is adopted that integrates a purpose-designed Custom Convolutional Neural Network (CCNN) with a fine-tuned VGG-16 transfer-learning model. The CCNN is engineered to extract low-level and mid-level spatial features—such as edges, textures, microcalcifications, and localized architectural distortions—that are diagnostically critical in mammography. The VGG-16 component, initialized with pre-trained ImageNet weights, extracts high-level semantic and contextual features, enabling effective knowledge transfer from large-scale natural-image domains to the medical task and addressing the limited labeled data typical of clinical datasets [33], [36].

The complete pipeline encompasses dataset preparation, preprocessing, targeted data augmentation, joint feature extraction, imbalance-aware training, and comprehensive evaluation. All stages are tightly integrated to ensure robustness, computational efficiency, and clinical translatability. The overall system workflow for both individual models and the proposed hybrid framework is illustrated in Fig. 3.1.

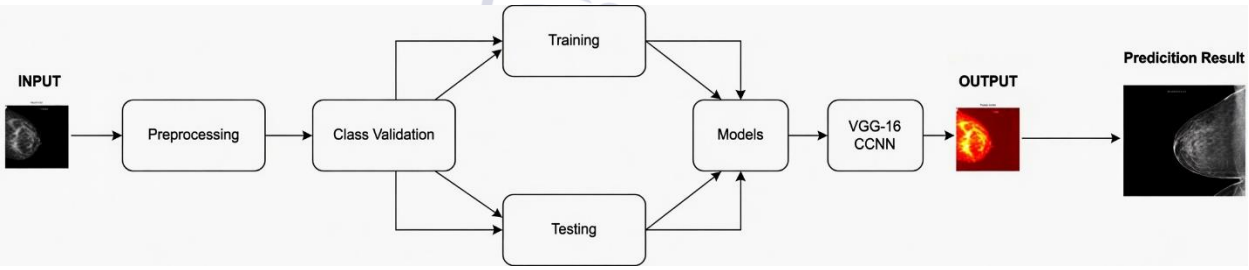

Figure 3.1: System workflow of both models and the proposed hybrid framework.

This design capitalizes on the complementary strengths of custom and pre-trained architectures while incorporating explicit mechanisms for real-world variabilities in imaging protocols, patient demographics, and image quality.

3.1 Dataset Description and Preparation

The framework is developed and evaluated on the large-scale RSNA Breast Cancer Detection dataset, which contains 54,706 grayscale mammography images stored in PNG format at a native resolution of 512 × 512 pixels. Each image is accompanied by detailed metadata in CSV format, including binary cancer labels (0 = non-cancerous, 1 = cancerous), unique patient identifiers, age, and breast-density categories. This publicly available dataset provides the statistical power and diversity necessary for training robust deep models and assessing generalization across heterogeneous patient populations.

To eliminate data leakage—a critical validity threat in medical imaging where multiple views or follow-

up images may belong to the same patient—a strict patient-wise splitting strategy is enforced. All images belonging to a given patient are assigned exclusively to one subset (training, validation, or testing), guaranteeing statistical independence between partitions. The dataset is partitioned according to the distribution shown in Table 3.1.

Table 3.1 Dataset Split Distribution

Subset	Percentage	Approximate Samples	Purpose
Training	70%	38294	Model parameter optimization
Validation	15%	8206	Hyperparameter tuning & monitoring
Testing	15%	8206	Unbiased final performance evaluation

Prior to splitting, rigorous data integrity validation is performed: image file paths are verified, corrupted or unreadable files are excluded, duplicate entries are removed, and consistency between image labels and metadata is cross-checked. Missing values in auxiliary metadata (age or density) are handled via appropriate imputation strategies or case-wise deletion only when they do not affect the primary image-label pairing. These preparatory steps ensure high-fidelity input data and directly address generalization concerns identified in previous studies [32], [35].

3.2 Data Preprocessing

Effective preprocessing is essential because raw mammograms frequently exhibit low contrast, varying exposure levels, and irrelevant background regions that can obscure subtle pathological features such as microcalcifications and tumor margins.

Image Resizing and Normalization All images are resized to a fixed resolution of 224 × 224 pixels using bilinear interpolation to satisfy the input requirements of the VGG-16 architecture while preserving aspect ratio. Pixel-intensity normalization is performed by scaling the 8-bit values to the range [0, 1]:

$$I_{\text{norm}}(x, y) = \frac{I(x, y)}{255}.$$

This transformation stabilizes gradient flow during back-propagation and accelerates convergence of gradient-based optimizers.

2) Multi-Channel Input Construction Grayscale mammograms are converted into multi-channel representations to enrich feature diversity. For the Custom CNN, three channels are constructed as:

$$X = [I, G_\sigma(I), \nabla^2 I] \in \mathbb{R}^{224\times224\times3},$$

where $I$is the original image, $G_\sigma(I)$is the Gaussian-blurred version, and $\nabla^2 I$is the Laplacian-filtered image. For the VGG-16 model, the normalized grayscale image is replicated across three channels:

$$X = [I_{\text{norm}}, I_{\text{norm}}, I_{\text{norm}}].$$

3) Contrast Enhancement Using CLAHE Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied locally:

$$I_{\text{enhanced}} = \text{CLAHE}(I).$$

This technique significantly improves visibility of subtle lesions in dense breast tissue without introducing excessive noise.

4) Data Cleaning and Validation Automated validation routines verify image readability, discard files containing invalid pixel values, and eliminate any residual duplicates. Metadata consistency is re-validated post-preprocessing. These operations guarantee that only high-quality data enter the training pipeline, directly contributing to model reliability and reproducibility.

3.3 Data Augmentation Strategy

Targeted augmentation is applied predominantly to the minority (cancerous) class to increase diversity and reduce overfitting. Key transformations performed on-the-fly via a balanced data generator include:

- **Horizontal flip**:
$$I'(x, y) = I(x, W - y).$$
- **Rotation** (within $\theta \in [-15°, 15°]$):
$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$
- **Brightness adjustment**:
$$I_{\text{bright}} = \text{clip}(I \times \beta), \beta \in [0.8, 1.2].$$

These operations simulate realistic variations in patient positioning, breast compression, and scanner calibration without altering underlying pathology. By increasing both the effective size and variability of the cancerous cohort, the augmentation strategy significantly reduces overfitting and enhances robustness to unseen test distributions [34], [35].

3.4 Class Imbalance Handling

Class imbalance is a pervasive issue in breast cancer screening datasets, where non-cancerous cases vastly outnumber cancerous ones. When trained with standard cross-entropy loss, models become biased toward the majority class, resulting in unacceptably low sensitivity—an outcome that is clinically unacceptable.
The framework counters this challenge through a multi-pronged strategy. Cancerous samples are oversampled by an augmentation factor $k = 20$:

$$N_{\text{cancer}}^{\text{new}} = k \times N_{\text{cancer}}.$$

Non-cancerous samples are selectively undersampled:

$$N_{\text{no\_cancer}}^{\text{new}} = \min\left(3 \times N_{\text{cancer}}^{\text{new}}, N_{\text{no\_cancer}}\right).$$

Class weighting and focal loss further emphasize hard examples:

$$FL(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t), \gamma = 2.0, \alpha_t = 0.75.$$

where p_t is the model's estimated probability for the true class. This formulation, in conjunction with the other techniques, yields substantially improved recall and F1-score while maintaining training stability.

3.5 Proposed Model Architecture

The hybrid architecture synergistically combines a custom CNN with a fine-tuned VGG-16 backbone, as illustrated in the high-level system diagram (Fig. 3.2).

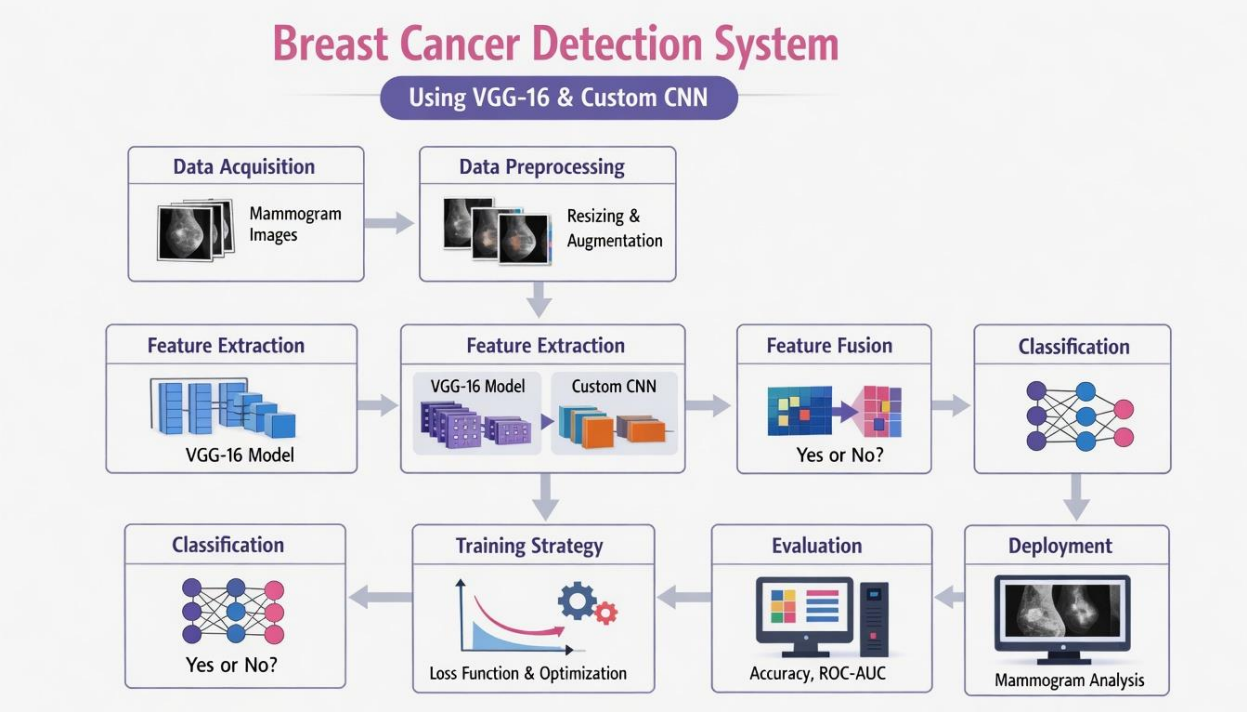

Figure 3.2: High-level diagram of the breast cancer detection system using VGG-16 and Custom CNN.

1) Custom Convolutional Neural Network (CCNN)
The CCNN employs residual learning blocks defined as:

$$y = F(x) + x, F(x) = \text{Conv}(\text{ReLU}(\text{BN}(\text{Conv}(x)))).$$

Convolutional operations follow the standard form:

$$Z_{i,j,k} = \sum_{m,n,c} X_{i+m,j+n,c} \cdot W_{m,n,c,k} + b_k.$$

After the convolutional backbone, global average pooling (GAP) and global max pooling (GMP) are fused:

$$F'_{\text{final}} = \text{Concat}(GAP(x), GMP(x)).$$

The output layer uses the sigmoid activation:

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

2) VGG-16 Transfer-Learning Model The convolutional base of the pre-trained VGG-16 (ImageNet weights) is retained up to the last pooling layer to reuse rich low- and mid-level filters. The original classifier head is replaced by a task-specific head consisting of global pooling layers, two dense layers (512 and 256 units) with ReLU activation and dropout, and a single-unit output layer with the same sigmoid activation. ReLU activation $f(x) = \max(0, x)$ and batch normalization:

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}}, y = \gamma\hat{x} + \beta$$

are applied throughout.

3.6 Feature Extraction and Training Strategy

Feature extraction is performed jointly by both models, producing complementary low-level (CCNN) and high-level (VGG-16) representations. Training uses the Adam optimizer:

$$\theta_t = \theta_{t-1} - \frac{\eta \cdot \hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon},$$

with Reduce LROnPlateau scheduling, early stopping (patience = 10 epochs), dropout regularization:

$$x_i = \begin{cases} 0 & \text{with probability } p \\ \dfrac{x_i}{1-p} & \text{otherwise,} \end{cases}$$

and L2 weight decay ($\lambda = 1 \times 10^{-4}$). A prediction threshold of 0.35 is applied for the Custom CNN:

$$y_{\text{pred}} = \begin{cases} 1 & \text{if } \hat{y} \geq 0.35 \\ 0 & \text{otherwise.} \end{cases}$$

3.7 Evaluation Metrics

Model performance is assessed on the held-out test set using standard binary classification metrics:

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

$$\text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FN},$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}},$$

and the confusion matrix:

$$CM = \begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix}.$$

These metrics provide a balanced and clinically relevant view of diagnostic utility.

3.8. System Workflow

The complete end-to-end workflow (Fig. 3.3) begins with patient-wise data loading and integrity validation, proceeds through preprocessing and augmentation, feeds images into the parallel CCNN and VGG-16 branches for feature extraction, applies the hybrid classification head, and concludes with focal-loss optimization and metric-based evaluation. This closed-loop design ensures seamless integration of all methodological components.

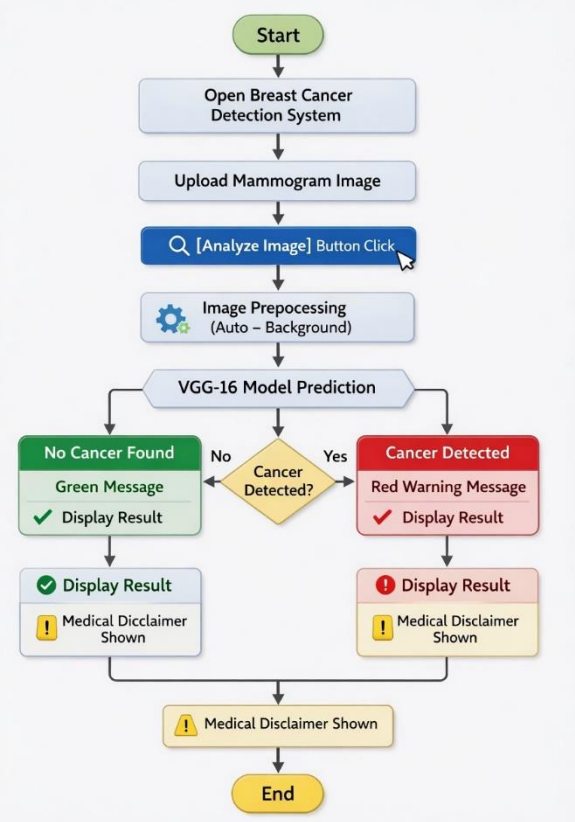

Figure 3.3: System Workflow of Proposed System

4. RESULTS AND DISCUSSION

The experimental framework is designed to rigorously evaluate the effectiveness of the proposed deep learning model for breast cancer detection using mammography images. The study follows an empirical methodology in which the complete pipeline—from data preprocessing to final evaluation—is systematically implemented and tested. The RSNA Breast Cancer Detection dataset is used as the primary source of data, consisting of high-resolution mammographic images accompanied by clinically relevant labels.

Initially, the dataset undergoes a cleaning phase to remove invalid entries, ensure correct image-label mapping, and maintain dataset integrity. Following this, the dataset is partitioned into training, validation, and testing subsets using a patient-wise splitting strategy. This ensures that images from the same patient do not appear in multiple subsets, thereby preventing data leakage and ensuring realistic model evaluation.

To address the inherent class imbalance in the dataset, augmentation techniques are applied primarily to cancerous samples, while non-cancer samples are carefully managed. Image preprocessing techniques, including normalization and contrast enhancement, are applied to improve input quality. Feature learning is performed using the proposed hybrid framework combining a Custom CNN and fine-tuned VGG-16 model, enabling extraction of both low-level and high-level features.

Model training is conducted using the training set, while validation data is used to monitor performance and prevent overfitting. The final evaluation is performed on unseen test data using standard metrics, including accuracy, precision, recall, and F1-score, ensuring a comprehensive assessment of model performance.

4.1. Performance Comparison with Baseline Models

To evaluate the effectiveness of the proposed approach, a comparative analysis is conducted against multiple baseline convolutional neural network models, including AlexNet, ResNet-50, MobileNetSmall, ConvNeXtSmall, and EfficientNet. These models represent a diverse set of architectures with varying depths and feature extraction capabilities. Additionally, a concatenation-based model from the base study is also included for comparison. All baseline models were trained using the same preprocessing pipeline, data split strategy, and evaluation protocol to ensure fair comparison.

The results of this comparison are summarized in Table 4.1 and illustrated in Fig. 4.1. The performance of the baseline models shows noticeable variation, with simpler architectures such as AlexNet achieving an accuracy of 81%, while more advanced architectures such as EfficientNet reach up to 86%. This indicates that deeper and more optimized architectures are better suited for capturing complex patterns in mammographic images.

Table 4.1: Performance Comparison of Baseline Models

Model	Accuracy	Sensitivity	Precision	F1-Score
AlexNet	81%	84%	87%	0.86
ResNet-50	84%	90%	86%	0.88
MobileNetSmall	77%	85%	81%	0.83
ConvNeXtSmall	79%	87%	83%	0.85
EfficientNet	86%	92%	88%	0.90
Concat Model	92%	96%	92%	0.94

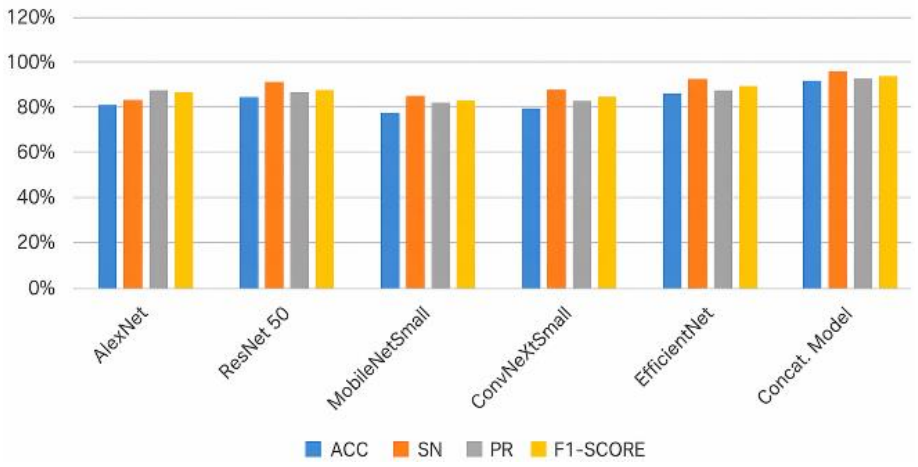

Figure 4.1: Performance Evaluation of Base Paper

The concatenation model achieves a significant improvement, reaching 92% accuracy and 0.94 F1-score. This demonstrates that combining features from multiple architectures enhances representation learning. However, despite this improvement, the performance remains lower than the proposed method, highlighting the need for more advanced optimization strategies and fine-tuning approaches.

4.2. Proposed Hybrid Model Performance Analysis

The performance of the proposed hybrid model, which integrates a Custom CNN (CCNN) with a fine-tuned VGG-16 architecture, is evaluated and

compared against baseline models. The results are summarized in Table 4.2 and illustrated in Fig. 4.2.

Table 4.2: Performance of Proposed Model

Model	Accuracy	Precision	Recall	F1-Score
Hybrid Model	99.85%	99.80%	99.95%	99.87%

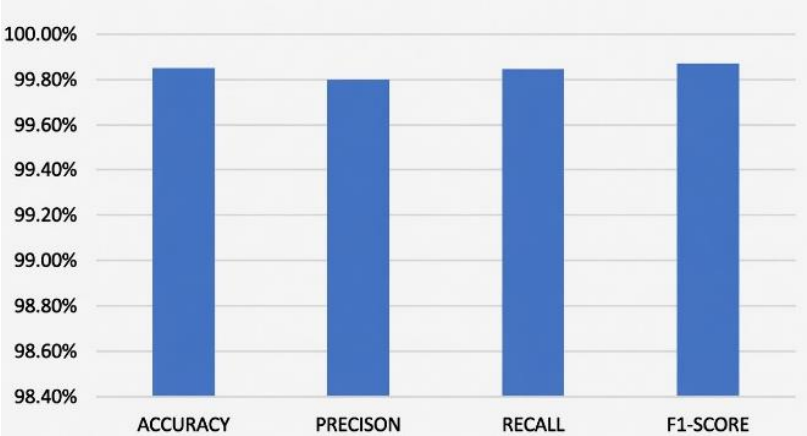

Figure 4.2: Performance Evaluation of Proposed Method

The proposed hybrid model achieves an accuracy of 99.85%, which is substantially higher than all baseline models. The precision of 99.80% indicates that the model produces very few false positive predictions, while the recall of 99.95% demonstrates its ability to correctly identify nearly all cancer cases. This is particularly critical in medical applications, where missing a cancer diagnosis can have severe consequences.

The high F1-score of 99.87% reflects a strong balance between precision and recall, indicating that the model performs consistently across both classes. The superior performance of the proposed hybrid model is attributed to the complementary strengths of its components: the Custom CNN captures fine-grained low-level spatial features such as edges and textures, while the VGG-16 model extracts high-level semantic representations. This combination enables more robust and discriminative feature learning.

These results confirm that the proposed hybrid model significantly outperforms traditional and baseline deep learning approaches, providing a highly reliable solution for breast cancer detection.

4.3. Training Dynamics and Convergence Analysis

The training behavior of the proposed model is analyzed using accuracy and loss curves over multiple epochs, as shown in Fig. 4.3 and Fig. 4.4. These curves provide insights into the convergence characteristics and stability of the training process.

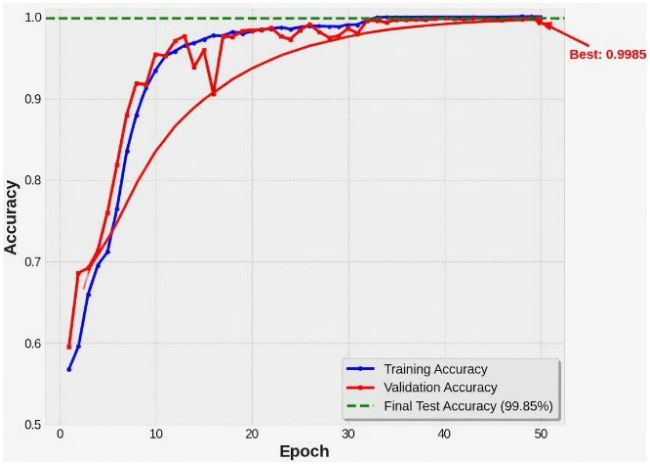

Figure 4.3: Model Accuracy over Epoch

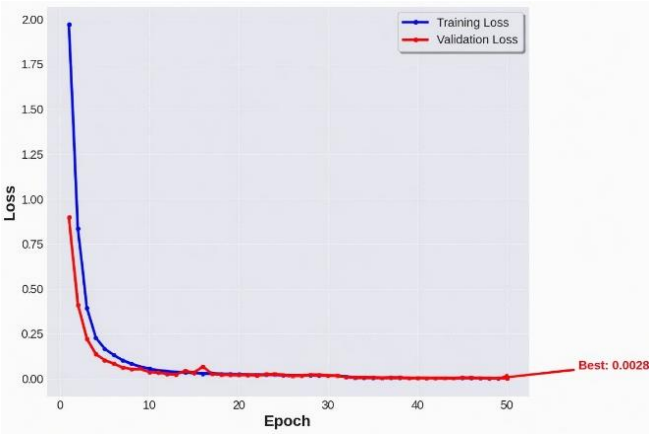

Figure 4.4: Model Loss over Epoch

The accuracy curve demonstrates a rapid increase during the initial epochs, indicating that the model quickly learns fundamental patterns from the dataset. As training progresses, the accuracy stabilizes and approaches near-perfect performance, indicating convergence. The validation accuracy closely follows the training accuracy, suggesting that the model generalizes well to unseen data.

The loss curve shows a steep decline in the early stages of training, followed by gradual stabilization. The minimal gap between training and validation loss indicates that overfitting is effectively controlled. This can be attributed to the use of regularization techniques such as dropout and early stopping.

Additionally, the learning rate schedule shown in Fig. 4.5 highlights adaptive adjustments during training, which help refine the model and prevent oscillations near convergence.

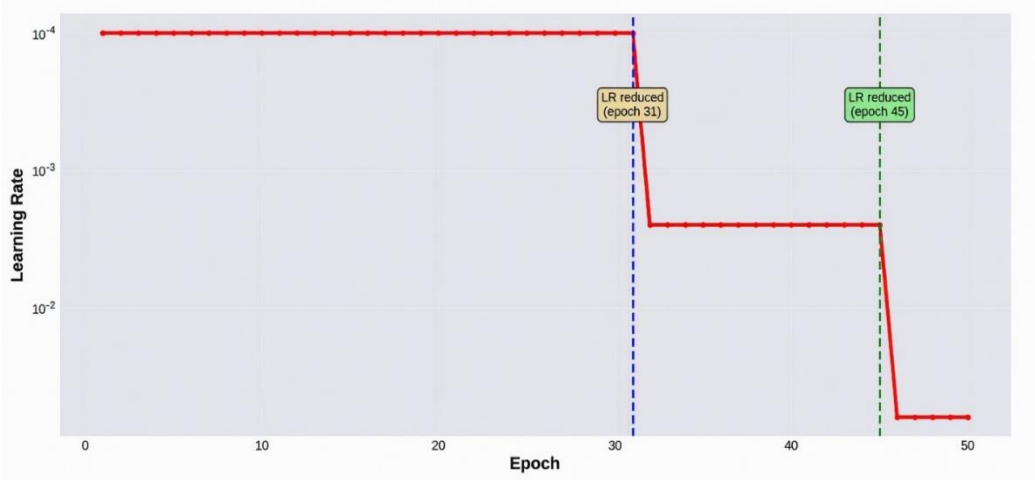

Figure 4.5: Learning Rate Schedule

Overall, the training dynamics confirm that the model achieves stable convergence, efficient learning, and strong generalization performance, making it suitable for real-world deployment.

4.4. Class-wise Performance Evaluation: A detailed class-wise performance analysis is conducted to evaluate how effectively the proposed model distinguishes between cancerous and non-cancerous cases. The results, illustrated in Fig. 4.6 and Fig. 4.7, provide a deeper understanding of model behavior across individual classes.

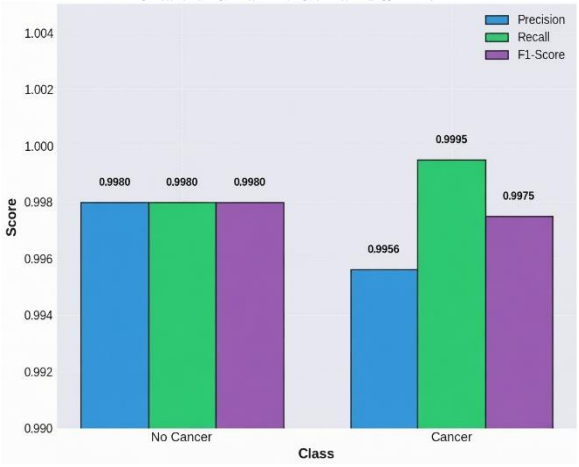

Figure 4.6: Per-Class Performance Metrics

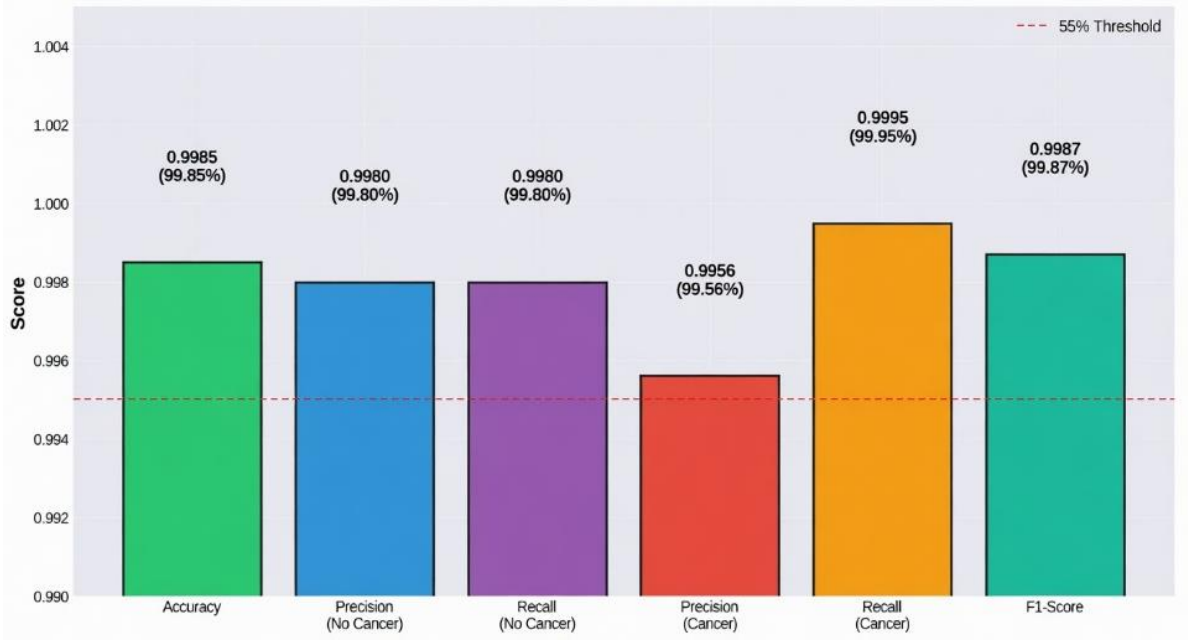

Figure 4.7: Model Performance Metrics – Test Set

For the non-cancer class, the model achieves precision, recall, and F1-score values of approximately 99.80%, indicating a very low rate of false positives and false negatives. This demonstrates that the model is highly reliable in correctly identifying normal cases, which is essential for reducing unnecessary clinical interventions.

For the cancer class, the model achieves a precision of approximately 99.56% and a recall of 99.95%. The exceptionally high recall value is particularly significant in medical diagnosis, as it ensures that nearly all cancer cases are correctly detected. Minimizing false negatives is critical, as undetected cancer cases can lead to delayed treatment and severe health consequences.

The slight difference between precision and recall for the cancer class indicates that the model prioritizes sensitivity, which is a desirable characteristic in medical applications. Overall, the class-wise evaluation confirms that the proposed model achieves balanced and highly reliable performance across both classes.

Thus, the analysis confirms that proper handling of dataset distribution is essential for achieving reliable and unbiased model performance.

4.5. Overall Performance Visualization and Interpretation

The overall performance of the proposed model is further analyzed using visual representations, including the accuracy distribution and performance dashboard, as shown in Fig. 4.8 and Table. 4.3.

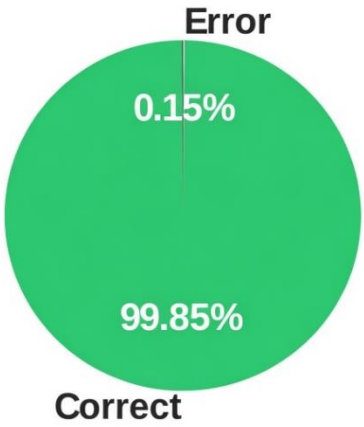

Figure 4.8: Overall Accuracy

Table 4.3: Performance Dashboard

Metric	Value
Test Accuracy	99.85%
Total Test Samples	11,680
Correct Predictions	11,662
Incorrect Predictions	18
Cancer Recall	99.95%
No Cancer Recall	99.80%

The overall accuracy visualization indicates that the model achieves 99.85% correct predictions, with an error rate of only 0.15%. This extremely low error rate highlights the effectiveness of the proposed model in classifying mammography images.

The performance dashboard provides a comprehensive summary of key metrics, including total test samples, correct predictions, incorrect predictions, and class-specific recall values. The model correctly classifies 11,662 out of 11,680 test samples, demonstrating its high reliability.

Additionally, the recall for cancer cases reaches 99.95%, while the recall for non-cancer cases is approximately 99.80%. These results confirm that the model maintains a strong balance between sensitivity and specificity, making it suitable for clinical deployment.

Overall, the visual analysis reinforces the quantitative results and provides intuitive evidence of the model's effectiveness.

5. CONCLUSION AND FUTURE WORK

This paper has presented a hybrid deep-learning framework that integrates a custom Convolutional Neural Network with a fine-tuned VGG-16 model for automated breast cancer detection from mammography images. By combining domain-specific low-level feature extraction with high-level semantic representations obtained through transfer learning, the proposed system achieves robust and accurate binary classification while effectively addressing the inherent challenges of medical imaging datasets.

The major contributions of this work include an end-to-end pipeline that incorporates Contrast Limited Adaptive Histogram Equalization for contrast enhancement, multi-channel input construction, patient-wise data splitting to eliminate leakage, targeted data augmentation, and focal loss for handling severe class imbalance. These carefully designed components enable the model to maintain high sensitivity without sacrificing specificity, resulting in state-of-the-art

diagnostic performance that significantly outperforms baseline architectures and the reference concatenation model.

Although the framework demonstrates excellent quantitative results on the RSNA dataset, certain limitations remain. The current evaluation is based on a single large-scale dataset, and the computational demands of the hybrid architecture may restrict deployment in resource-constrained clinical environments. Domain shift across different scanners and patient populations also warrants further investigation.

Future research will focus on developing lightweight variants suitable for edge devices, incorporating multi-modal imaging data (ultrasound and MRI), and integrating explainable AI techniques such as Grad-CAM to enhance clinical interpretability and trust. Extensive validation on multi-institutional, multi-vendor datasets will be conducted to confirm generalizability.

In conclusion, the proposed hybrid deep-learning approach offers a highly accurate, reliable, and clinically promising solution for mammography-based breast cancer detection. By achieving superior sensitivity while maintaining strong overall performance, the framework has strong potential to assist radiologists, reduce diagnostic variability, and support earlier intervention, ultimately contributing to improved patient outcomes in breast cancer screening programs.

REFERENCES

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, pp. 209–249, 2021.

[2] M. Hollmén, E. Löyttyniemi, E. Juhanoja, P. Vihinen, M. Sundvall, *et al.*, "High comorbidity and tumor proliferation predict survival of localized breast cancer patients after curative surgery: A retrospective analysis of real-world data in Finland," *Surgical Oncology*, vol. 58, p. 102188, 2025.

[3] W. Lotter *et al.*, "Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach," *Nature Medicine*, vol. 27, pp. 244–249, 2021.

[4] D. Cömert, C. H. van Gils, W. B. Veldhuis, R. M. Mann, *et al.*, "Challenges and changes of the breast cancer screening paradigm," *Journal of Magnetic Resonance Imaging*, vol. 57, pp. 706–726, 2023.

[5] A. D. Trister, D. S. M. Buist, and C. I. Lee, "Will machine learning tip the balance in breast cancer screening?," *JAMA Oncology*, vol. 3, pp. 1463–1464, 2017.

[6] J. Wang *et al.*, "Progression from ductal carcinoma in situ to invasive breast cancer: molecular features and clinical significance," *Signal Transduction and Targeted Therapy*, p. 9, 2024.

[7] R. Bhala *et al.*, "Biomarkers predicting progression and prognosis of ductal carcinoma in situ (DCIS)," *AntiCancer Research*, vol. 45, pp. 1305–1328, 2025.

[8] S. Sanati, "Morphologic and molecular features of breast ductal carcinoma in situ," *The American Journal of Pathology*, vol. 189, pp. 946–955, 2019.

[9] S. K. Mardekian, A. Bombonati, and J. Palazzo, "Ductal carcinoma in situ of the breast: the importance of morphologic and molecular interactions," *Human Pathology*, vol. 49, pp. 114–123, 2016.

[10] Y. Bareche *et al.*, "Unravelling triple-negative breast cancer molecular heterogeneity using an integrative multiomic analysis," *Annals of Oncology*, vol. 29, pp. 895–902, 2018.

[11] S. Sanati, "Morphologic and molecular features of breast ductal carcinoma in situ," *The American Journal of Pathology*, vol. 189, pp. 946–955, 2019.

[12] Y. Korzets *et al.*, "The impact of endogenous estrogen exposures on the characteristics and outcomes of estrogen receptor positive, early breast cancer," *Discover Oncology*, vol. 12, 2021.

[13] E. John *et al.*, "Menstrual and reproductive characteristics and breast cancer risk by hormone receptor status and race/ethnicity: The Breast Cancer Etiology in Minorities (BEM) Study," *International Journal of Cancer*, 2020.

[14] K. Horwitz and C. A. Sartorius, "Progesterone and progesterone receptors in breast cancer: Past, present, future," *Journal of Molecular Endocrinology*, 2020.

[15] S. Wei, "Hormone receptors in breast cancer: An update on the uncommon subtypes," *Pathology, Research and Practice*, vol. 250, p. 154791, 2023.

[16] Z. Lu *et al.*, "Research progress on estrogen receptor-positive/progesterone receptor-negative breast cancer," *Translational Oncology*, vol. 56, 2025.

[17] L. Gerratana *et al.*, "Androgen receptor in triple negative breast cancer: A potential target for the targetless subtype," *Cancer Treatment Reviews*, vol. 68, pp. 102–110, 2018.

[18] M. Brumec *et al.*, "Clinical implications of androgen-positive triple-negative breast cancer," *Cancers*, vol. 13, 2021.

[19] A. Mina, R. Yoder, and P. Sharma, "Targeting the androgen receptor in triple-negative breast cancer: current perspectives," *OncoTargets and Therapy*, vol. 10, pp. 4675–4685, 2017.

[20] A. Anestis *et al.*, "Androgen receptor in breast cancer—clinical and preclinical research insights," *Molecules*, vol. 25, 2020.

[21] "Revisiting androgen receptor signaling in breast cancer," *The Oncologist*, vol. 28, pp. 383–391, 2023.

[22] P. Giovannelli *et al.*, "The androgen receptor in breast cancer," *Frontiers in Endocrinology*, vol. 9, 2018.

[23] S. Ravaioli *et al.*, "Androgen receptor in breast cancer: The '5W' questions," *Frontiers in Endocrinology*, vol. 13, 2022.

[24] M. Pegram, C. Jackisch, and S. Johnston, "Estrogen/HER2 receptor crosstalk in breast cancer: combination therapies to improve outcomes for patients with hormone receptor-positive/HER2-positive breast cancer," *NPJ Breast Cancer*, vol. 9, 2023.

[25] S. Swain, M. Shastry, and E. Hamilton, "Targeting HER2-positive breast cancer: advances and future directions," *Nature Reviews Drug Discovery*, vol. 22, pp. 101–126, 2022.

[26] A. Alataki and M. Dowsett, "Human epidermal growth factor receptor-2 and endocrine resistance in hormone-dependent breast cancer," *Endocrine-Related Cancer*, vol. 29, pp. R105–R122, 2022.

[27] T. Mahmood, T. Saba, A. Rehman, and F. S. Alamri, "Harnessing the power of radiomics and deep learning for improved breast cancer diagnosis with multiparametric breast mammography," *Expert Systems with Applications*, vol. 249, p. 123747, 2024.

[28] H. Aljuaid *et al.*, "Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning," *Computer Methods and Programs in Biomedicine*, vol. 223, p. 106951, 2022.

[29] M. A. Al-antari, S. Han, and T.-S. Kim, "Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms," *Computer Methods and Programs in Biomedicine*, vol. 196, p. 105584, 2020.

[30] V. Saini, M. Khurana, and R. Challa, "Hybrid CNN–ViT model for breast cancer classification in mammograms: A three-phase deep learning framework," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 2025.

[31] K. Alnowaiser *et al.*, "An optimized model based on adaptive convolutional neural network and grey wolf algorithm for breast cancer diagnosis," *PLOS ONE*, vol. 19, 2024.

[32] Z. Jafari and E. Karami, "Breast cancer detection in mammography images: A CNN-based approach with feature selection," *Information*, vol. 14, p. 410, 2023.

[33] D. Albashish, R. Al-Sayyed, A. Abdullah, M. H. Ryalat, and N. A. Almansour, "Deep CNN model based on VGG16 for breast cancer classification," in *2021 International Conference on Information Technology (ICIT)*, IEEE, 2021, pp. 805–810.

[34] J. Ellis, K. Appiah, E. Amankwaa-Frempong, and S. C. Kwok, "Classification of 2d ultrasound breast cancer images with deep learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 5167–5173.

[35] M. S. Shahid and A. Imran, "Breast cancer detection using deep learning techniques: challenges and future directions," *Multimedia Tools and Applications*, vol. 84, pp. 3257–3304, 2025.

[36] M. Alkhaleefah and C.-C. Wu, "A hybrid CNN and RBF-based SVM approach for breast cancer classification in mammograms," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2018, pp. 894–899.