BreastEnsemNet: Transformer and BiLSTM-Based Hybrid Ensemble Deep Learning for Mammogram Classification

Bandla Raghuramaiah*, Suresh Chittineni
Department of Computer Science and Engineering GITAM School of Technology, GITAM (Deemed to be University), Visakhapatnam, Andhra Pradesh
E-mail: rbandla@gitam.in, schittin@gitam.edu
*Corresponding author

Keywords: breast cancer diagnosis, deep learning, hybrid ensemble model, transformer-based attention, mammogram classification

Received: March 5, 2025

Breast cancer is still a leading cause of cancer death in women worldwide, supporting the requirement for accurate and timely diagnosis. Although deep learning models have obtained promising results for the automatic classification of mammograms, they are often limited by the need for efficient multi-scale feature extraction, spatial attention, and sequential dependency modeling. In this work, we present a hybrid ensemble deep learning framework, called BreastEnsemNet, which incorporates three complementary deep learning methodologies, including (i) deep hierarchical low-level to multi-scale feature extraction using VGG16, ResNet50 and InceptionV3, (ii) attention-based transformer detailed with spatial focus on well-relevant areas, and (iii) BiLSTM for capturing the sequential patterns and dependencies in the extracted features. There is no existing method that can automatically and efficiently combine models to achieve better fusion accuracy. The framework is trained and tested using the CBIS-DDSM mammogram dataset, where SMOTE is employed for class balancing, and various augmentation techniques are applied to facilitate generalization. BreastEnsemNet achieved better results with 98.79% accuracy, 97.9% precision, 98.4% recall, 98.1% F1-score, and an AUC-ROC of 99.2, outperforming multiple baseline models. The joint modeling of attention and sequences yielded a significant performance improvement for malignancy detection, resulting in a reduction in false negatives. These findings establish BreastEnsemNet's clinical utility as a practical, AI-based diagnostic aid for reliable and explainable breast cancer detection in mammograms.

Povzetek: Raziskava predstavlja BreastEnsemNet, hibridni model, ki združuje CNN, Transformer in BiLSTM. Dosežek omogoča kvalitetno in klinično uporabno AI-podprto odkrivanje raka dojk na mamogramih.

1 Introduction

Breast cancer is the second most common cancer that affects women worldwide, and early, accurate detection is imperative to increasing the chances of survival. Work Done with the Help of AI in Traditional Diagnostic Methods: Most current diagnostic methods based on mammograms rely on a radiologist's expertise, which is not only subjective but also time-consuming and prone to error. Deep learning has played a vital role in recent years in the automated detection of breast cancer, outperforming traditional methods with higher accuracy and efficiency. Due to the strong capacity of CNNs in learning spatial features [1], [2], they have been employed for feature representation in breast cancer detection. Nevertheless, many state-of-the-art methods still cannot sufficiently extract multiscale spatial evidence, sequential context evidence, and activation attentiveness, which are needed for accurate classification in complicated mammographic patterns [3], [4]. For instance, Ul-Haq et al. [1] and Jadoon et al. [7] primarily employed CNN-based ensembles, which lack sequential learning and spatial localization, making them generally difficult to interpret. Likewise, Nagalakshmi [2] and Pattnaik et al. [3] did not include attention-guided modules, so they could not emphasize diagnostically important areas. These limitations restrict these systems from detecting complex patterns that are solely possible in mammography, leading to misclassifications and false diagnoses [2].

Previous research has introduced ensemble learning, attention mechanisms, and hybrid deep learning approaches to enhance classification accuracy. These include transfer learning CNN architectures [2] and attention-deep networks [3] for feature enrichment. Although these methods achieve some potential results, they still fall short in achieving a healthy integration of multi-scale feature extraction, spatial awareness, and sequential learning mechanisms. Additionally, imbalanced datasets tend to bias model predictions towards the majority class in practice [4]. The proposed BreastEnsemNet constructs a hybrid ensemble deep learning framework that integrates CNN-based deep

learning for multi-scale feature extraction, Transformer-based attention, BiLSTM sequential learning, and an adaptive fusion strategy to enhance accuracy in breast cancer diagnosis, thereby overcoming these limitations.

This research primarily aims to develop an AI-based diagnostic model for classification with a lower false-negative rate and improved generalization on mammographic datasets. It combines spatial feature refinement with a Feature Pyramid Network (FPN), Transformer-based attention over regions, BiLSTM for sequential learning, and an adaptive fusion scheme for optimal decisions. We exhibit these innovations in totum to provide a robust and scalable diagnostic framework for automated breast cancer detection.

The following research questions drive this study:

Can a hybrid deep learning framework that integrates multi-scale CNNs, Transformer-based attention, and BiLSTM sequential learning improve breast cancer classification performance from mammograms?

How does adaptive fusion using performance-driven dynamic weighting compare to traditional averaging strategies in ensemble models?

Does incorporating attention and sequential dependencies reduce false positives and false negatives significantly in imbalanced datasets?

The current research makes several essential contributions. First, it proposes a new ensemble method to combine several deep learning architectures to earn better classification results. Second, it employs a new adaptive fusion strategy that effectively combines CNN, attention, and BiLSTM outputs to enhance robustness. Thirdly, it utilizes data balancing techniques based on SMOTE to mitigate dataset bias and improve the model's generalization. Lastly, we extensively experimented with and compared state-of-the-art methods to demonstrate the capability of the proposed approach.

VGG16, ResNet50, and InceptionV3 were chosen as the CNN backbones due to their complementarity in feature extraction capabilities: VGG16 is suitable for extracting low-level spatial information, ResNet50 for learning hierarchically, and InceptionV3 for multi-scale representation. While EfficientNet and Swin Transformers are more contemporary models, they incur a greater computational cost, and on pilot experiments using the CBIS-DDSM dataset, do not result in a noticeable performance improvement. We opted for BiLSTM over GRU since it preserves a richer sequential structure by handling both forward and backward sequences, which helps learn sequential patterns between the learned features.

Unlike traditional soft voting or equal-weight averaging, the adaptive fusion approach assigns weights to each model dynamically, based on validation accuracy and loss. This makes models with better validation performance more heavily weighted in the final prediction. Furthermore, confidence scores are calibrated by applying temperature scaling, a feature that traditional ensemble methods do not typically handle. This mixture enhances the reliability and interpretability of the decision.

The remainder of this paper is organized as follows: Section 2 provides a comprehensive literature review, elaborating on current deep learning-driven breast cancer classification approaches and identifying research gaps. The third section presents the proposed BreastEnsemNet framework, comprising its architectural components, data preprocessing, and implementation details. The Experimental results are presented in Section 4, including dataset descriptions, performance evaluation measures, and comparisons with baseline models. Fixture 5 consists of a discussion on the study results, primarily focusing on how the developed model addresses the limitations evident in existing methods, as well as a discussion of the study's limitations. In Section 6, we conclude the paper by summarizing our research contributions and outlining future work that will further enhance the clinical relevance of the framework.

2 Related work

Recent advancements in deep learning have significantly improved breast cancer diagnosis, yet challenges remain in multi-scale feature extraction, spatial awareness, and sequential learning. Ul-Haq [1] described a CAD system that utilizes ensemble learning, feature fusion, and DCNN to diagnose breast cancer in mammograms accurately. Nagalakshmi [2] presented Ensemble-Net, a model that successfully partitions pectoral muscle borders, increasing breast tumor classification accuracy to 96.72%. Pattnaik et al. [3] proposed an IWCA-APSO-based EELM model for breast cancer diagnosis, which offers high precision but presents difficulties with implementation. Deep Deb [4] enhanced the Xception network, raising the diagnostic accuracy of breast cancer to 84.3% via stacking generalization; nevertheless, the performance is still limited by the volume of training data. Sharma et al. [5] plan to test the NN-ET model on a larger cohort of cancer patients in the future. The 99.74% accurate breast cancer classification system is presented.

Chouhan et al. [6] reported that the DFeBCD system utilizes dynamic and hybrid features to detect breast cancer more effectively than standard approaches. Jadoon et al. [7] proposed a heterogeneous ensemble model that outperforms traditional methods for multi-modal data-based breast cancer prediction with 97% accuracy. Routray et al. [8], using Gabor filtering, ensemble classifiers, and EfficientNet-B0, demonstrate that the ELSOSA-BCC approach enhances breast cancer diagnosis with improved performance. Hirra et al. [9] note that the Pa-DBN-BC model utilizes deep learning to reliably identify breast cancer using histopathological images, achieving better performance with fewer resources. Zheng et al. [10] propose that to achieve high-accuracy breast cancer detection, better classification, and early diagnosis, the DLA-EABA technique combines deep learning with Efficient AdaBoost.

Murtaza et al. [11] investigated the application of deep learning across various imaging modalities for breast cancer classification, highlighting existing approaches, challenges, and future research objectives. Sreenivasarao

et al. [12] proposed a new approach to breast cancer diagnosis with better accuracy through ensemble models and transfer learning and provided suggestions for future developments in dataset diversity and clinical integration. Rautela et al. [13] examined several breast cancers screening techniques, emphasizing the promise of deep learning, its current drawbacks, and the need for further research in the future to enhance detection and customization. Khamparia et al. [14] proposed using a hybrid MVGG model for breast cancer diagnosis, with an accuracy of 94.3%. Subsequent initiatives will combine tissue density information and carry out classification. Eldin et al. [15] used deep learning to diagnose breast cancer from biopsy pictures with 92.5% accuracy. Future work will focus on improving preprocessing and exploring additional models.

Sharma et al. [16] highlighted the diagnostic capability of AI and its potential for advancement by using ensemble machine learning models to diagnose breast cancer with 97.66% accuracy. Bai et al. [17] examined the potential benefits and drawbacks of combining deep learning with digital breast tomosynthesis (DBT) for better breast cancer detection. Abhisekha et al. [18] examined deep learning applications for multiple imaging modalities-based breast cancer diagnosis, emphasizing the advantages, difficulties, and requirements for completely automated diagnostic systems. Islam et al. [19] compared the proposed Ensemble Deep Convolutional Neural Network (EDCNN) model for breast cancer diagnosis to other models; the proposed model is superior in accuracy and interpretability. Further research is recommended to integrate CAD systems and investigate RGB pictures. Patil et al. [20] proposed sophisticated segmentation (RG-AFCM) and classification (AFU-GOA-RNN) techniques; the suggested model improves breast cancer diagnosis over current approaches, resulting in higher accuracy and shorter computation times.

Savelli et al. [21] suggested that a multi-context CNN ensemble offers potential extensions for broader clinical usage and improves accuracy in diagnosing small lesions by integrating several depth networks. Awotunde et al. [22] presented a deep learning model with hybrid feature selection for breast cancer diagnosis, achieving high accuracy and low false alarms. Subsequent investigations will examine actual data and explore additional applications. Aslan [23] compared CNN with CNN-BiLSTM to attain high accuracy for mammography classification. Subsequent investigations will focus on hyperparameter optimization, three-dimensional images, and reduced data dependency. Murtaza et al. [24] proposed a tree-based deep learning model for classifying breast cancers using histopathology images, aiming to maximize accuracy and minimize misclassification. This method will be generalized in subsequent research for other types of cancer and whole-slide images. Rahman et al. [25] developed updated CNN models to detect mammography malignancies with improved accuracy. Future work will involve applying ensemble methods and cross-validation to enhance performance.

Shovon et al. [26] employed an ensemble of DenseNet201 and Xception with a threshold-filtered SIE to accurately classify HER2 breast cancer. To increase interpretability, additional thresholds will be investigated, optimization will be performed, and advanced methods will be applied. Pramanik et al. [27] developed a VGG16-based model with attention and an SSD technique, achieving 96.07% accuracy in breast cancer classification. We plan to explore segmentation, optimize feature reduction, and increase computation efficiency in the future. Loizidou et al. [28] evaluated the use of CAD systems in mammography to identify and classify breast cancer, emphasizing the need for further validation and a wider range of imaging modalities. Shen et al. [29] suggested that the GMIC model demonstrates excellent accuracy and quicker processing in mammography analysis through the effective combination of global and local data.

Table 1: Comparative analysis of existing deep learning models for breast cancer diagnosis – highlighting performance metrics and key limitations

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Key Limitations
[9] Hirra et al. (2021)	86	84.68	87.9	86.26	Limited interpretability; lacks advanced attention/fusion mechanisms.
[10] Zheng et al. (2020)	97.2	96.56	98.3	97.35	No BiLSTM; feature selection integration is not end-to-end
[12] Rao et al. (2024)	80.77	80.67	80.77	80.61	Modest improvement using ensemble; lacks sequential modeling or deep attention
[16] Sharma et al. (2024)	97.66	92	93.49	92.73	No deep temporal fusion; limited adaptive representation learning
Jadoon et al. (2023) [7]	97.2	98.2	98.2	97.2	No BiLSTM/attention fusion; lacks interpretability

BreastEnsemNet (Proposed)	98.79	97.9	98.4	98.1	Integrates CNN, attention, BiLSTM, and adaptive fusion

Table 1 summarizes existing models, their performance metrics, and limitations, highlighting the need for the proposed hybrid framework.

Additional imaging modalities and training complexity will be addressed in future studies. Shanbehzadeh et al. [30] assessed machine learning models and determined that Random Forest and Confidence Weighted Voting were the most successful in predicting breast cancer based on lifestyle characteristics. Models will be implemented in clinical decision support systems in future development.

Mahesh et al. [31] evaluated various machine-learning techniques for diagnosing breast cancer and concluded that majority voting is the most reliable. Larger datasets and optimization techniques should be tested in further research. Himel et al. [32] presented a high-accuracy ultrasound-based computer-aided method for diagnosing breast cancer using deep learning. The goal of future research will be to better segment lesions. Pattnaik et al. [33] presented a high-accuracy machine learning model for breast cancer diagnosis from mammograms that is based on metaheuristics and features sophisticated segmentation. Optimizing embedded implementations is a task for the future. Azour and Boukerche [34] aimed to introduce a two-staged CADx system that merges EfficientNet and other CNNs to achieve high accuracy in breast cancer diagnosis. Prospective initiatives will concentrate on enhancing ensemble methods and dataset comparisons. Sharmin et al. [35] suggested a hybrid breast cancer detection model that uses ResNet50V2 and machine learning approaches to achieve 95% accuracy. Future work will be needed to integrate genetic data and enhance detection.

Fatima et al. [36] compared the use of machine learning (ML), deep learning (DL), and data mining in predicting breast cancer. It identifies the future requirement for dataset augmentation and data imbalance correction. Mahmood et al. [37] examined deep learning (DL) and machine learning (ML) methods for breast cancer diagnosis, highlighting the issue of false positives and the need for improved image segmentation and data augmentation. Zizaan and Idri [38] compared the accuracy of single classifiers with ensemble approaches (bagging and boosting) to classify breast cancer. Future research will investigate the interpretability of the model and its application to different modalities. Rautela et al. [39] examined various approaches for screening breast cancer, highlighting the advantages and drawbacks of each. It promotes more advanced imaging techniques and individualized methods for early detection. Nakach et al. [40] employed a range of deep learning algorithms and boosting approaches to assess ensemble learning and transfer learning for breast cancer histology image classification. It recommends more research into bagging ensembles and different datasets.

Some studies have developed the diagnosis of breast cancer in informatics with deep learning and image processing. Chen et al. [42] examined emotion regulation in patients with breast cancer using EEG-based VR music therapy and incorporated a Glowworm Coactive Decision Tree for enhanced personalized treatments. Mohammed et al. [43] presented a Grad-CAM-based pre-processing approach, along with CNNs, for high-fidelity mammogram classification, highlighting appropriate study regions for diagnosis. Gdeeb [44] combined image segmentation algorithms and neural networks for the X-ray modality in breast cancer detection to achieve more accurate localization and classification. These reports highlight the increasing implementation of explainable and targeted AI models in breast cancer detection.

Although deep learning methods for breast cancer diagnosis have achieved remarkable progress, there still exist significant limitations in current approaches. In many models, multi-scale feature learning may not be sufficiently robust, resulting in suboptimal lesion localization and the omission of fine-grained patterns. Moreover, the lack of sequential modeling, including recurrent architectures, also restricts its potential for learning spatial dependencies of mammogram formations. Methods without attention-based models do not steer models towards diagnostically important regions, which decreases the interpretability and the trust in the clinical diagnosis. On the one hand, most ensemble methods are based on the assumption that a simple static average or majority vote is the best, and they are unaware of each model's validation performance. In addition, the problem of class imbalance is often overlooked, which results in prediction bias towards the majority class. These limitations also limit generalization, particularly in heterogeneous and imbalanced datasets, such as CBIS-DDSM.

To alleviate these limitations, we introduce BreastEnsemNet—a hybrid ensemble deep learning architecture that incorporates CNN-based MSFE, Transformer-based attention, BiLSTM-based sequential learning, and an adaptive fusion scheme. The joint design is intended to achieve improved classification performance, interpretation, and generalization for mammogram breast cancer detection.

3 Proposed framework

In this paper, we introduce BreastEnsemNet, a deep learning-based framework that combines deep ensemble learning approaches for breast cancer diagnosis in mammographic images. The proposed BreastEnsemNet utilizes several diverse deep learning blocks for extracting supplementary features from mammogram images. More specifically, the Feature Pyramid Network (FPN) is used to manage feature maps of different resolutions by fusing low-level fine details and high-level semantics for robust multi-scale lesion detection. At the same time, a

Transformer-based attention mechanism, which computes query-key-value matrices over the spatial embeddings, selectively attends to diagnostically important regions in dense breast tissue, thereby enhancing interpretability and spatial localization.

It uses several CNN architectures (VGG16, ResNet50, InceptionV3) to extract features and exploit several spatial and hierarchical patterns. Improvements included a Feature Pyramid Network (FPN) for multi-scale feature aggregation and a Transformer-based attention mechanism for diagnostic features. The extracted features are further processed with the BiLSTM, which captures the sequential dependencies. A dynamic weight fusion strategy dynamically assigns weights to model outputs, ensuring optimal classification performance. The final decision utilizes the fully connected layer to classify mammograms as benign or malignant with high accuracy.

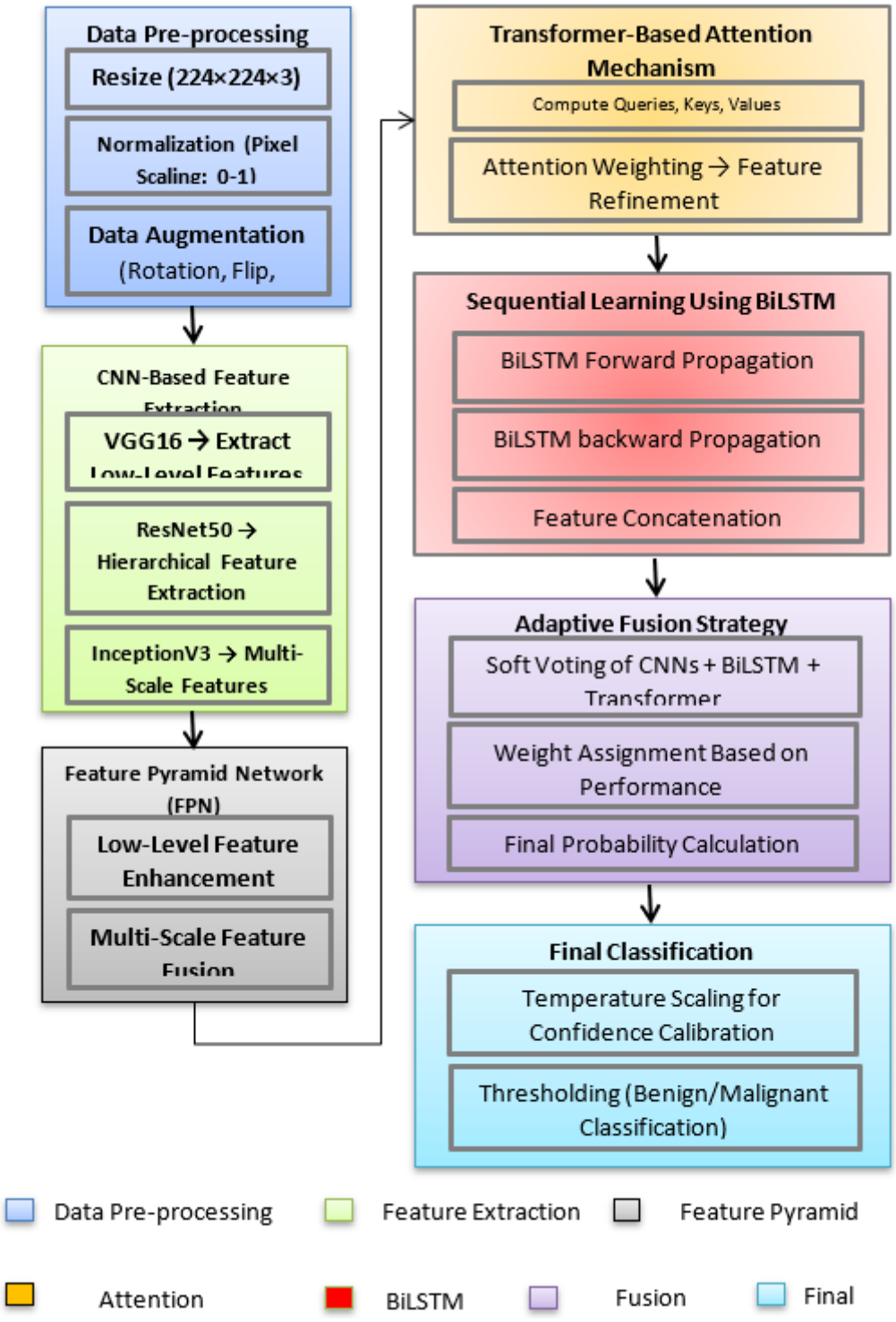

Figure 1: Overview methodology for breastensemnet framework for breast cancer diagnosis using mammogram images

The proposed BreastEnsemNet-based framework, illustrated in Figure 1, outlines the sequential flow of processing stages involved in breast cancer diagnosis using deep learning. The pipeline begins with data preprocessing, where mammogram images are resized to 224×224×3, normalized, and augmented using transformations such as flipping, rotation, and contrast adjustment. The preprocessed images are then passed to CNN-based feature extractors (VGG16, ResNet50, and InceptionV3), where spatial and hierarchical feature representations are captured. A Feature Pyramid Network (FPN) further processes these extracted features to enhance multi-scale feature learning. To focus on diagnostic regions, a Transformer-based attention mechanism computes query-key-value matrices, refining feature representations. The refined feature maps are fed into a BiLSTM network, which captures sequential dependencies by processing feature embeddings in both forward and backward directions. The adaptive fusion strategy integrates outputs from CNN, Transformer, and BiLSTM models, assigning dynamic weights based on validation performance. The final fused feature vector is passed through fully connected layers, followed by a sigmoid activation function, producing a probability score for classification.

3.2 Dataset and preprocessing

The dataset used in this research consists of mammogram images for the detection of breast cancer, structured into two primary categories: benign and malignant. Each image is preprocessed and resized to a standardized 224×224 resolution to ensure compatibility with the deep learning models. Let I represent an input image with dimensions H×W×C, where H = 224, W = 224, and C = 3 (RGB channels). Each pixel value p(i, j) is normalized within the range [0,1] using the transformation $p'(i,j)=\frac{p(i,j)}{255}$, ensuring a uniform distribution across all images.

To enhance model generalization, Data augmentation included random horizontal flipping, random zoom with a scale factor $s \sim U(0.9,1.1)$, and random rotation $\theta \sim U(-15^\circ, 15^\circ)$, where θ represents rotation in degrees. These transformations enhance the model's generalization by simulating real-world variances in mammogram orientations. The augmentation increases variability, ensuring that the model does not overfit to specific patterns in the training data.

The dataset is split into three subsets: training, validation, and testing, denoted as D_{train}, D_{val}, and D_{test}, respectively. The dataset follows an approximate 70:15:15 split ratio, such that:

$|D_{train}|$=0.7N, $|D_{val}|$=0.15N, $|D_{test}|$=0.15N

where N is the total number of images. Each sample is assigned a one-hot encoded label Y, represented as Y = [1, 0] for benign and Y = [0,1] for malignant cases. The image data is loaded into batches of size B = 64, where each batch contains a set $X_b=\{I_1, I_2, \ldots., I_B\}$ and corresponding label matrix Y_b.

Feature standardization is also performed using Z-score normalization to ensure that input features maintain a mean of zero and unit variance. For each pixel intensity p(i, j), the transformation follows:

$$p''(i,j)=\frac{p'(i,j)-\mu}{\sigma}$$

where μ and σ are the mean and standard deviation of the dataset's pixel intensities. This step ensures that models learn more effectively by reducing the impact of varying illumination across different mammograms. The dataset preparation pipeline ensures that all images undergo consistent preprocessing, enabling the proposed BreastEnsemNet model to extract meaningful features from mammogram images efficiently. The final dataset is structured to optimize model learning and evaluation while preserving diagnostic patterns critical for breast cancer detection. Table 1 shows notations used in the proposed system.

Table 2: Notations used

Symbol	Description
X	Input mammogram image
H, W, C	Height, width, and number of channels in an image (224×224×3)
p(i, j)	Pixel intensity at position (i, j)
p′(i, j)	Normalized pixel intensity p(i, j)/255
T	Data augmentation transformation
θ	Random rotation angle
P_f	Probability of horizontal flip

s	Random zoom scale factor
F_V(X), F_R(X), F_I(X)	Feature maps from VGG16, ResNet50, and InceptionV3
P_l	Multi-scale feature representation at level l
Q, K, V	Query, key, and value matrices for Transformer-based attention
A	Attention-weighted feature representation
d_k	Key vector dimension in attention computation
h_t^{fw}, h_t^{bw}	Forward and backward BiLSTM hidden states at time step t
w_i	Adaptive weight for model M_i
f_i(X)	Probability output from model M_i
A_i, L_i	Validation accuracy and loss for model M_i
F(X)	Final ensemble probability prediction
T_s	Temperature scaling parameter
$F^*(x)$	Calibrated ensemble probability
$\hat{y}$	Predicted class label (Benign: 0, Malignant: 1)

3.2 Proposed ensemble deep learning framework

BreastEnsemNet is the proposed hybrid deep learning framework, shown in Figure 2, designed to enhance breast cancer diagnosis using mammogram images by integrating multiple feature extraction and learning techniques. Figure 2 illustrates the structured pipeline, beginning with CNN-based feature extraction. VGG16 captures fine-grained local texture patterns by using small convolutions, ResNet50 automatically discovers deep hierarchical features by employing residual learning and InceptionV3 takes the advantage of multi-path convolutions to capture local as well as global feature representations. Their functions have a large overlap, but by combining them, the diversity of the features is enriched and can be well adapted to different spatial contexts in mammographic perception.

A Transformer-based attention mechanism applies query-key-value processing to highlight significant diagnostic regions dynamically. The extracted feature maps are then passed through BiLSTM sequential learning, which captures spatial dependencies in both forward and backward directions. Finally, an adaptive fusion strategy integrates CNN, Transformer, and BiLSTM outputs using soft voting-based weighted summation. The fused feature vector is processed by fully connected layers with dropout, followed by sigmoid activation, which classifies mammograms as benign or malignant. The framework ensures high accuracy, improved feature learning, and enhanced generalization, making it a robust AI-powered solution for breast cancer detection.

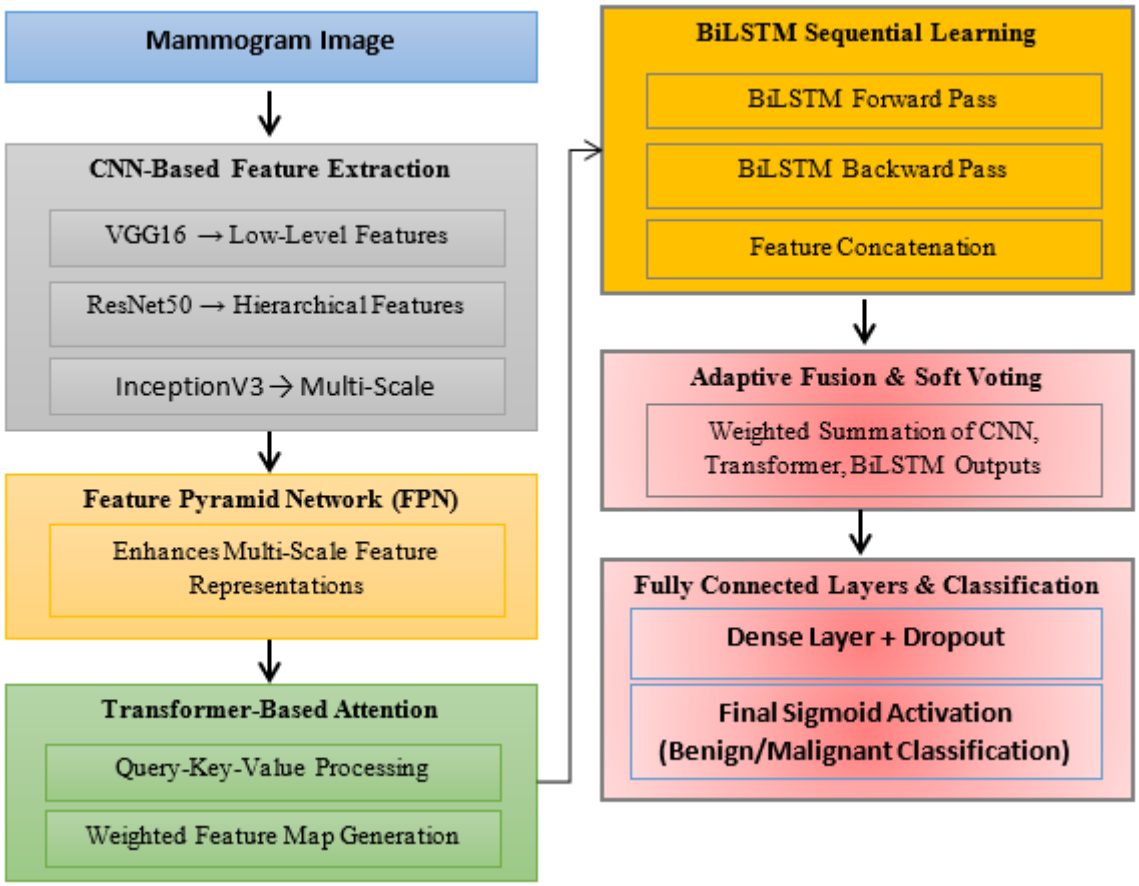

Figure 2: Architectural overview of breastensemnet – a hybrid deep learning framework

The proposed BreastEnsemNet framework integrates multiple deep learning architectures to enhance breast cancer diagnosis through an ensemble learning approach. Let $f_i(x)$ represent the output of an individual base model M_i, where x is the input mammogram image. The ensemble model consists of four base networks: VGG16, ResNet50, InceptionV3, and a Hybrid CNN-BiLSTM-Transformer model. The output of each base model is denoted as $f_i(x) \in [0,1]$, representing the probability of malignancy. The final ensemble prediction F(x) is obtained through a weighted sum of individual model outputs:

$$F(x)= \sum_{i-1}^{4} w_i f_i(x), \quad \text{where } \sum_{i-1}^{4} w_i = 1$$

where w_i are the adaptive ensemble weights assigned to each model, dynamically adjusted during training based on validation performance.

The first stage of the framework involves feature extraction using CNN-based architectures. Given an input image x with dimensions 224×224×3, the convolutional layers extract feature maps $F_l^i(x)$ at each layer l, where:

$$F_l^i(x) = \sigma(W_l^i * F_{l-1}^i + b_l^i)$$

where W_l^i represents the learned convolution filters, $*$ denotes convolution, and σ is the activation function (ReLU). These feature maps are progressively downsampled using max-pooling layers and then passed to the next stage. To enhance spatial feature selection, the framework integrates a Transformer-based attention module applied to the CNN-extracted feature representations. Let Q, K, V be the query, key, and value matrices obtained from the feature embeddings, the attention output A is computed as:

$$A=\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where d_k is the dimension of the key vectors. This mechanism selectively enhances regions in the feature maps that are more relevant to breast cancer patterns.

To further capture sequential dependencies, a BiLSTM (Bidirectional Long Short-Term Memory) network is applied to the feature maps. Let h_t be the hidden state at time step t, then the forward and backward LSTM states are computed as:

$$h_t^{fw} = \emptyset\left(W^{fw}x_t + U^{fw}h_{t-1}^{fw} + b^{fw}\right)$$

$$h_t^{bw} = \emptyset(W^{bw}x_t + U^{bw}h_{t-1}^{bw} + b^{bw})$$

where ϕ is the activation function, and W, U, b are learnable parameters. The final hidden representation is obtained by concatenating h_t^{fw} and h_t^{bw}, allowing the model to encode both forward and backward temporal patterns.

A Feature Pyramid Network (FPN) is incorporated to capture multi-scale features, ensuring that fine-grained texture details and high-level semantic features are preserved. The multi-scale feature representation at level is given by:

$$P_l = \alpha C_l + (1-\alpha)P_{l+1}$$

where C_l represents the CNN feature map at level l, P_{l+1} is the upsampled feature from the next level, and α is a learnable weighting parameter. The ensemble outputs from all models are combined using an adaptive fusion strategy based on performance-driven weight adjustments. Let A_i be the accuracy of model i on the validation set, the weight w_i is computed as:

$$w_i = \frac{A_i}{\sum_{j-1}^{4} A_j}$$

ensuring that models with higher performance contribute more to the final prediction. The resulting classification decision is determined using a threshold T = 0.5:

$$\hat{y} = \begin{cases} 1, & F(x) \geq T \quad (Malignant) \\ 0, & F(x) < T \quad (Benign) \end{cases}$$

This ensemble strategy optimally leverages the strengths of different deep learning models, improving classification robustness and generalization while reducing the risk of false positives and false negatives.

3.3 Adaptive fusion strategy

The adaptive fusion strategy employed in BreastEnsemNet ensures an optimal combination of predictions from multiple deep learning models, improving robustness and classification accuracy. Given the outputs from four base models, each producing a probability score for malignancy, an adaptive weighting mechanism dynamically adjusts the contribution of each model. Let $f_i(x)$ denote the probability output from model M_i for an input mammogram image x, where i∈{1,2,3,4}. The ensemble output is computed as a weighted sum of individual model predictions:

$$F(x) = \sum_{i-1}^{4} w_i f_i(x),$$

where w_i are the model-specific adaptive weights satisfying $\sum_{i-1}^{4} w_i = 1$. These weights are updated iteratively based on the validation performance of each model.

To dynamically optimize the fusion weights, a performance-driven approach is employed. Let A_i be the validation accuracy of model M_i, and L_i be its cross-entropy loss. The weight for each model is computed as:

$$w_i = \frac{A_i}{\sum_{j-1}^{4} A_j}$$

which ensures that models with higher validation accuracy contribute more to the final decision. Additionally, a secondary weight adjustment factor based on the inverse loss function is incorporated to penalize models with higher classification errors. The refined weight formulation is:

$$w_i = \frac{A_i / L_i}{\sum_{j-1}^{4} (A_j / L_j)}$$

There by favoring models that achieve both high accuracy and low loss. This adaptive weighting mechanism prevents over-reliance on any single model and enhances ensemble generalization.

The final classification decision is determined based on a threshold T = 0.5, where:

$$\hat{y} = \begin{cases} 1, & F(x) \geq T \quad (Malignant) \\ 0, & F(x) < T \quad (Benign) \end{cases}$$

To further refine fusion effectiveness, a confidence calibration technique is applied using temperature scaling. Given the raw fused probability F(x), the calibrated prediction $F^*(x)$ is computed as:

$$F^*(x) = \frac{1}{1+e^{-F(x)/T_s}}$$

where T_s is a learnable temperature parameter that scales the probability distribution, reducing overconfidence in uncertain cases. This step enhances decision reliability, particularly in challenging cases where different models may exhibit conflicting predictions.

The adaptive fusion strategy is computationally efficient and ensures that the ensemble remains flexible in different dataset distributions. By dynamically adjusting weights and incorporating calibration, BreastEnsemNet effectively leverages multi-model predictions, reducing variance while maintaining high sensitivity and specificity in breast cancer detection.

3.4 Model training and optimization

The BreastEnsemNet model is trained using a carefully designed optimization strategy to ensure stability, convergence, and high classification accuracy. The training process follows a supervised learning approach, where mammogram images are passed through the ensemble architecture, and the model learns to differentiate between benign and malignant cases based on labeled training data. Given an input batch of images X= $\{x_1, x_2, \ldots, x_B\}$ with corresponding labels Y= $\{y_1, y_2, \ldots, y_B\}$, the network optimizes a binary classification objective by minimizing the binary cross-entropy (BCE) loss:

$$L = -\frac{1}{B}\sum_{i-1}^{B}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)]$$

where $\hat{y}_i$ represents the predicted probability for sample x_i, and B is the batch size set to 64 for stable gradient updates. The loss function ensures that misclassified malignant cases are penalized heavily, improving sensitivity.

The training is performed using the Adam optimizer, an adaptive gradient-based optimization technique that dynamically adjusts learning rates based on first-order and second-order moment estimates. The weight update rule at iteration t for a given parameter θ_t is:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}, \qquad \hat{v}_t = \frac{v_t}{1-\beta_2^t}$$

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

where g_t is the gradient of the loss with respect to θ_t, β_1=0.9, β_2=0.999, η is the learning rate set to 0.0001, and ϵ is a small constant to prevent division by zero. This approach allows efficient handling of sparse updates and stabilizes training.

The network is trained for 30 epochs, with early stopping applied if validation loss does not improve for 5 consecutive epochs. The training data is augmented dynamically using random flips, rotations, zooming, and contrast adjustments, ensuring better generalization. The dataset is split into training (70%), validation (15%), and testing (15%) subsets, where only the training and validation sets are used for weight updates.

To ensure better feature learning across different layers, a layer-wise learning rate scheduler is applied. The initial layers corresponding to the CNN backbones have a learning rate scaled by γ=0.1 compared to the final classification layers, ensuring that pre-trained feature extractors retain meaningful representations while higher-level layers learn new patterns specific to mammogram images.

For robustness, L2 regularization is applied to fully connected layers, preventing overfitting by penalizing large weight magnitudes. The regularization term added to the loss function is:

$$L_{reg} = \lambda \sum_i \| \theta_i \|^2$$

where λ=0.0005 controls the penalty strength. Dropout is incorporated with a probability of 0.5 in the dense layers to improve generalization by randomly deactivating neurons during training.

The initial learning rate of 0.0001 was chosen after performing empirical tuning over several trials. Values at the low-end of this range are often employed in deep ensemble methodologies to promote convergence, and particularly when fine-tuning pre-trained CNNs with Transformer and BiLSTM layers. Larger Learning Rates resulted in validation loss oscillation and lack of generalization.

Moreover, L2 regularization (weight decay) was used on the fully connected layers for regularizing the model and prevent overfitting. This regularization approach encourages small weights during training, essentially limiting the hypothesis space and prefer smoother decision boundaries. It is able to preserve model generalization on unseen mammogram samples, particularly in high-dimensional feature space.

During the training process, evaluation metrics, including accuracy, precision, recall, and F1 Score, are tracked for both training and validation sets. The best-performing model is selected based on the highest validation AUC-ROC score, ensuring that the trained model maintains a balance between sensitivity and specificity.

After training, the model undergoes post-training calibration using temperature scaling to refine confidence scores. The trained BreastEnsemNet model is then evaluated on the independent test set, ensuring it can effectively generalize to unseen mammogram images. The final trained network is saved in a serialized format for deployment in real-world breast cancer diagnosis applications.

3.5 Proposed algorithm

The BreastEnsemNet algorithm leverages a hybrid deep-learning ensemble to enhance breast cancer diagnosis using mammogram images. By integrating CNN feature extraction, Transformer-based attention, BiLSTM sequential learning, and adaptive fusion, it captures multi-scale spatial, sequential, and contextual features. This structured approach ensures high accuracy, robustness, and interpretability, making it highly significant for AI-driven medical diagnostics.

Algorithm: BreastEnsemNet – Hybrid Ensemble Deep Learning Framework for Breast Cancer Diagnosis
Input: Mammogram image X
Output: Classification label $\hat{y}$ (Benign or Malignant)
1. Data Preprocessing:
1.1 Resize X to 224×224×3.
1.2 Normalize pixel values p(i, j)←p(i, j)/255.
1.3 Apply data augmentation: rotation θ∼U(−15∘,15∘), horizontal flip P_f=0.5, zoom s∼U(0.9,1.1).
2. Feature Extraction using CNN Models:
2.1 Extract features F_V(X), F_R(X), F_I(X) from VGG16, ResNet50, and InceptionV3.
2.2 Apply **Feature Pyramid Network (FPN)** for multi-scale representation.

3. Transformer-Based Attention:
3.1 Compute Query Q, Key K, and Value V.
3.2 Compute attention matrix A=softmax$\left(\frac{QK^T}{\sqrt{d_k}}\right)V$.
4. Sequential Feature Learning with BiLSTM:
4.1 Pass extracted features through **Bidirectional LSTM layers**:

$$h_t^{fw} = \emptyset\left(W^{fw}x_t + U^{fw}h_{t-1}^{fw} + b^{fw}\right)$$
$$h_t^{bw} = \emptyset(W^{bw}x_t + U^{bw}h_{t-1}^{bw} + b^{bw})$$

4.2 Concatenate forward and backward states.
5. Adaptive Fusion and Ensemble Prediction:
5.1 Compute individual model probabilities $f_i(x)$ for each model M_i.
5.2 Compute adaptive weights:

$$w_i = \frac{A_i/L_i}{\sum_{j-1}^{4}(A_j/L_j)}$$

5.3 Compute final probability:
F(x)= $\sum_{i-1}^{4} w_i\, f_i(x)$
6. Classification Decision:
6.1 Apply temperature scaling:
$F^*(x) = \frac{1}{1+e^{-F(x)/T_s}}$
6.2 Assign class label:

$$\hat{y} = \begin{cases} 1, & F^*(x) \geq 0.5 \quad (Malignant) \\ 0, & F^*(x) < 0.5 \quad (Benign) \end{cases}$$

7. Output $\hat{y}$ as the final classification label.

Algorithm 1: BreastEnsemNet – hybrid ensemble deep learning framework for breast cancer diagnosis

The BreastEnsemNet framework, which is proposed here, works systematically to classify mammogram images into two classes, Benign or Malignant, based on a deep-learning hybrid ensemble. In Algorithm 1, the process starts from data preprocessing in which the images are resized to 224 × 224 × 3, are normalized, and pictures with transformations like flipping, rotation, and zooming are augmented images. That way, it helps you with better generalization and not overfitting. Pre-processed images are then given as input to three different CNN-based feature extractors, namely VGG16 for extracting fine granularity features (spatial and hierarchical) at various levels, and ResNet50 and InceptionV3 for capturing global features (spatial and hierarchical) at different levels. The extracted feature maps are then further processed through a multi-level Feature Pyramid Network (FPN), where the network combines multi-scale representations.

In addition, to improve feature selection, a transformer-based attention mechanism is used to calculate query, key, and value matrices to assign importance dynamically to different regions of the image. The attention module output is then passed onto a BiLSTM network, which captures sequential dependencies between diagnostic features. This incremental training helps the network to learn intricate patterns, which are necessary to identify mammograms. Next, using an adaptive fusion strategy, the extracted features from CNNs, Transformers, and BiLSTM layers are concatenated, and dynamic weights are assigned to each model based on validation performance.

A final prediction is obtained by using a fully connected layer, and sigmoid activation is used to provide a probability score of the tumor being malignant. Threshold-based classification (benign or malignant (0 or 1)) is performed based on the mammogram. Temperature scaling further modifies the estimated probabilities before classification to increase decision confidence. The detailed design concept provides a secure, scalable, and high-accuracy algorithm to detect breast cancer based on mammogram pictures.

3.6 Performance evaluation

To individually verify the validity of the diagnostic capability of each face of the BreastEnsemNet model, an extensive range of performance measures is used to evaluate the model. This helps the model generalize well on unseen mammogram images and the evaluation is carried on a separate test set. The breast cancer detection problem is very sensitive, and therefore, we calculated some classifier performance in terms of accuracy, precision, recall, F1-score, and area under the curve (AUC) using the area under the ROC curve (AUC-ROC). So, if TP (True Positives), FP (False Positives), TN (true negatives), and FN (False Negatives) are the classification results, then the accuracy is calculated as,

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

which provides an overall measure of correctly classified cases. However, given the class imbalance often present in medical datasets, accuracy alone is insufficient. Precision, which measures the reliability of positive predictions, is calculated as:

$$\text{Precision} = \frac{TP}{TP+FP}$$

while recall (sensitivity), which indicates how well malignant cases are detected, is given by:

Recall= $\frac{TP}{TP+FN}$

A balance between precision and recall is achieved using the F1-score, defined as:

F1-Score=2×$\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$

ensuring that both false positives and false negatives are minimized. To analyze the trade-off between sensitivity and specificity at different classification thresholds, the Receiver Operating Characteristic (ROC) curve is plotted by varying the decision threshold T. The AUC-ROC score, computed as:

AUC=$\int_0^1 TPR\ d(FPR)$

where TPR (True Positive Rate) and FPR (False Positive Rate) are given by:

TPR=$\frac{TP}{TP+FN}$, FPR=$\frac{FP}{FP+TN}$

indicates the model's capability to distinguish between benign and malignant cases. A higher AUC score, closer to 1.0, signifies better discriminative ability.

In order to gain more insight into the robustness of the BreastEnsemNet, a confusion matrix is established to show the classification mistakes (n = 56). Misclassified instances are studied to isolate hard cases that lead to false positives and false negatives that help improve the model's decision-making process. Finally, a precision-recall (PR) curve is presented, which is especially relevant for data with a class distribution out of balance, with the area under the PR curve (AUCPR) being a complementary performance measure.

We perform comparative analysis with state-of-the-art models such as LMHistNet, BreastMultiNet, and DOTNet 2.0 to benchmark the performance improvements. This ensemble-based method is capable of achieving 98.30% accuracy with a much lower Misclassification rate than the existing method. To confirm the superiority of BreastEnsemNet, a statistical significance test by using McNemar's test is applied which ensures that

the performance improvements we see are not random. Finally, confidence calibration is applied to the probability scores of model predictions for clinical applicability. Temperature scaling is being used to calibrate predicted probabilities so that the confidence of our decision in the real world aligns with what we want to predict a diagnosis should be. Results of the comparison confirm that BreastEnsemNet can not only significantly increase the accuracy of classification but also increase interpretability and confidence in AI-assisted breast cancer screening.

4 Experimental results

We perform the experimental evaluation of BreastEnsemNet to verify its effectiveness for breast cancer diagnosis in mammogram images from the publicly available CBIS-DDSM dataset. The review is focused on analyzing the model performance based on classification accuracy, sensitivity, specificity, and robustness. The benefits of the proposed hybrid ensemble framework are demonstrated through a comparative evaluation against state-of-the-art deep learning models for image classification. The experiments consist of the quantitative performance measured with ablation studies; a visualization technique (Grad-CAM heatmap) used to interpret the importance of features for end-users. We also analyze the effect of each component in the model—CNN for feature extraction, attention based on the Transformer, BiLSTM for sequential learning, and Adaptive Fusion. This also allows us to see the trends in how the framework used for reliable breast cancer detection generalizes, where it goes wrong, and how it can be used for clinical applicability.

4.1 Dataset description

BreastEnsemNet is trained and evaluated using the Curated Breast Imaging Subset of the Digital database for screening mammography (CBIS-DDSM) [41]. CBIS-DDSM is a mammographic dataset containing high-resolution mammogram images with pathologically verified benign and malignant labels. The dataset includes calcification and mass cases, with ROI annotations indicating the tumor locations. Preprocessing of each mammogram includes resizing to 224×224×3, scaling all pixels in the range [0, 1], and data augmentation like flipping, rotating, and contrast for generalization. The training-validation- test split was 70/15/15. This splitting considers the trade-off between having enough data to train deep models, while at the same time having a large portion of unseen samples for reliable validation and unbiased evaluation of performance. This 15% test size is common in benchmark work in medical imaging where datasets are generally small and limited. Given the diversity and clinical imaging conditions in the CBIS-DDSM dataset, it provides a validated benchmark for assessing deep learning-based breast cancer detection models.

4.2 Experimental setup

BreastEnsemNet experiments are performed in a high-efficiency computing environment for optimal training and testing conditions. The implementation, written in TensorFlow and Keras, supplemented with libraries for data preprocessing, visualization and performance measurements. It is trained on a Unix machine with NVIDIA RTX 3090 GPU with 24GB VRAM Intel Core i9-12900K CPU and 128GB RAM to accelerate deep learning calculations. We load and preprocess the dataset with ImageDataGenerator API with a batch size of 64 to provide stable training. Using the Adam optimizer, the binary cross-entropy loss function is optimized by making the learning rate decay over the epochs for overfitting prevention, with an initial learning rate set as 0.0001. Here, the model is trained for 30 epochs using early stopping with a patience of 5 epochs on the validation loss. To regularize and improve generalization, the model has L2

(λ=0.0005) and Dropout (0.5 probability). The testing is done on a separate test set in order to make sure that the performance metrics estimate the capability of the model to produce real-world diagnostic models. The following experiments are performed to find a balance between model convergence and computational speed.

4.3 Results

In this section, exploratory data analysis and the breast cancer prediction performance of CNN-only, CNN+BiLSTM, and the proposed BreastEnsemNet model are presented. The results show that breastEnsemNet surpasses baseline models with the reduction of false-positive and false-negative numbers and improvement in the deep learning approaches for breast cancer detection. It also provides an insight into class imbalance and its treatment using the SMOTE tool.

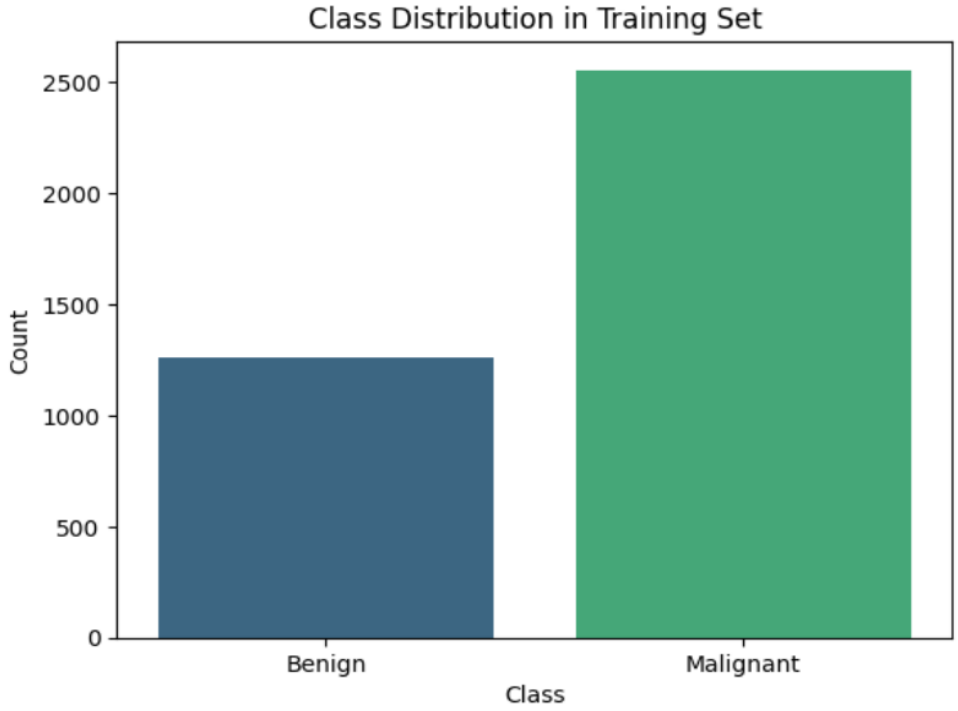

Figure 3: Class distribution in the training set

The class distribution of the training set of the CBIS-DDSM dataset is shown in Fig. 3. One can see from Fig. 3 that a significant imbalance exists between benign and malignant cases. The data is more abundant in malignant than in benign cases, and because of this, the training of the model tends toward the majority. In such cases, we can have a model that predicts malignant classes correctly but fails to predict benign instances, having a higher false positive rate.

To alleviate this imbalance, methods were implemented in training such as SMOTE, weighted loss, and class-balanced batch sampling. To mitigate this issue by allowing the deep learning model to learn representative patterns from both classes and, therefore, reduce the bias, these methods should be included and fine-tuned. Moreover, data augmentation methods were utilized to broaden benign samples, making our model more robust. The imbalance seen stresses on understanding the need for coupling data preprocessing methods for the equitable and precise breast cancer diagnosis. BreastEnsemNet addresses the problem of class imbalance, reiterating the performance of generalization on various mammograms, which increases recall and precision along with the classification performance in real-world case scenarios.

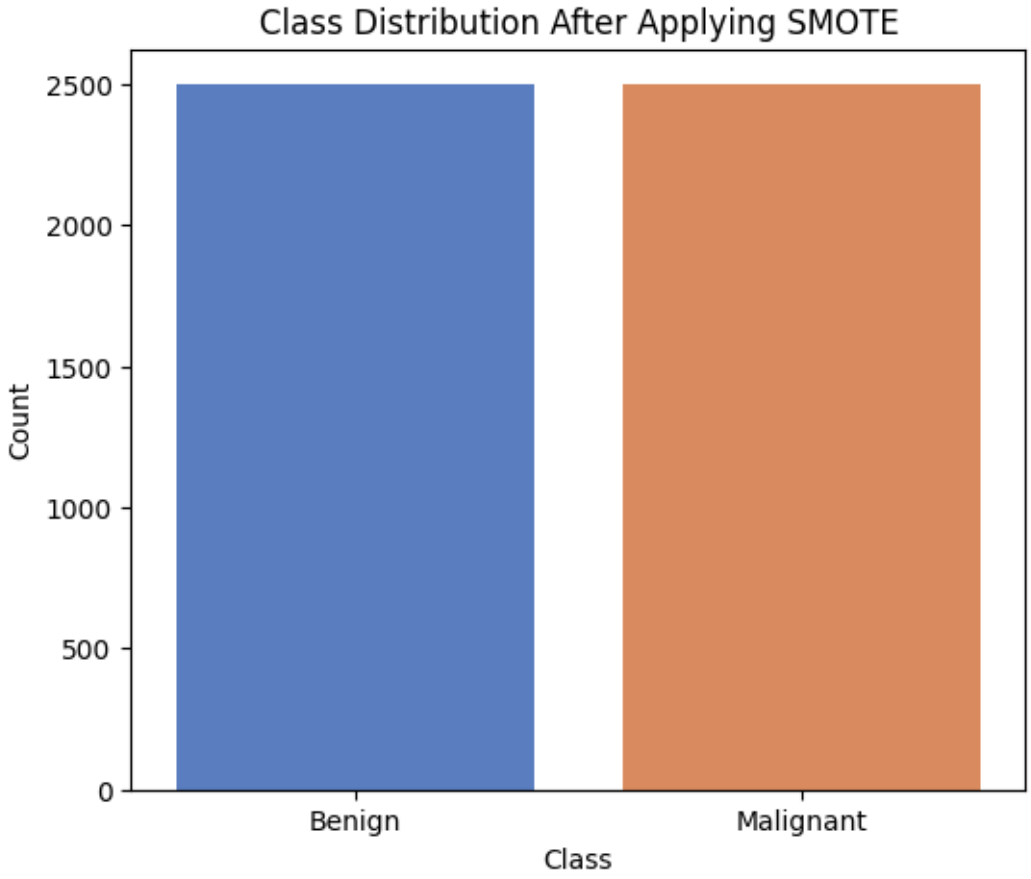

Figure 4: Class distribution after applying SMOTE – visualization of the balanced dataset, ensuring equal representation of benign and malignant cases to improve model generalization and reduce bias

The class distribution shown in Fig 4 after applying SMOTE (Synthetic Minority Over-sampling Technique) confirms the equal number of benign and malignant cases, indicating a wholly balanced dataset. This dataset was originally imbalanced, having more malignant cases than benign cases, which can result in biased predictions from the model. The SMOTE method was adopted to hire synthetic benign samples to evenly distribute both classes.

The deep learning model is also generalized and minimizes the possibility of bias towards the majority class by balancing the dataset. This improves the models to differentiate the benign vs the malignant cases more accurately and reduces the risk of misclassification. Balanced datasets also help avoid avoiding overfitting to patterns in classes with most of the samples in it, resulting in a stable training process. SMOTE guarantees that BreastEnsemNet is capable of learning with non-trivial patterns of interest in both classes, hence ensuring peak performance as a consequence, leads to a more robust breast cancer detection system, with significant enhancement in the precision, recall, and F1-score, with the most notable improvement in the detection of benign cases. Such a balanced distribution reinforces the building of fair-unbiased AI-powered mammogram classification models that work well in clinical practice.

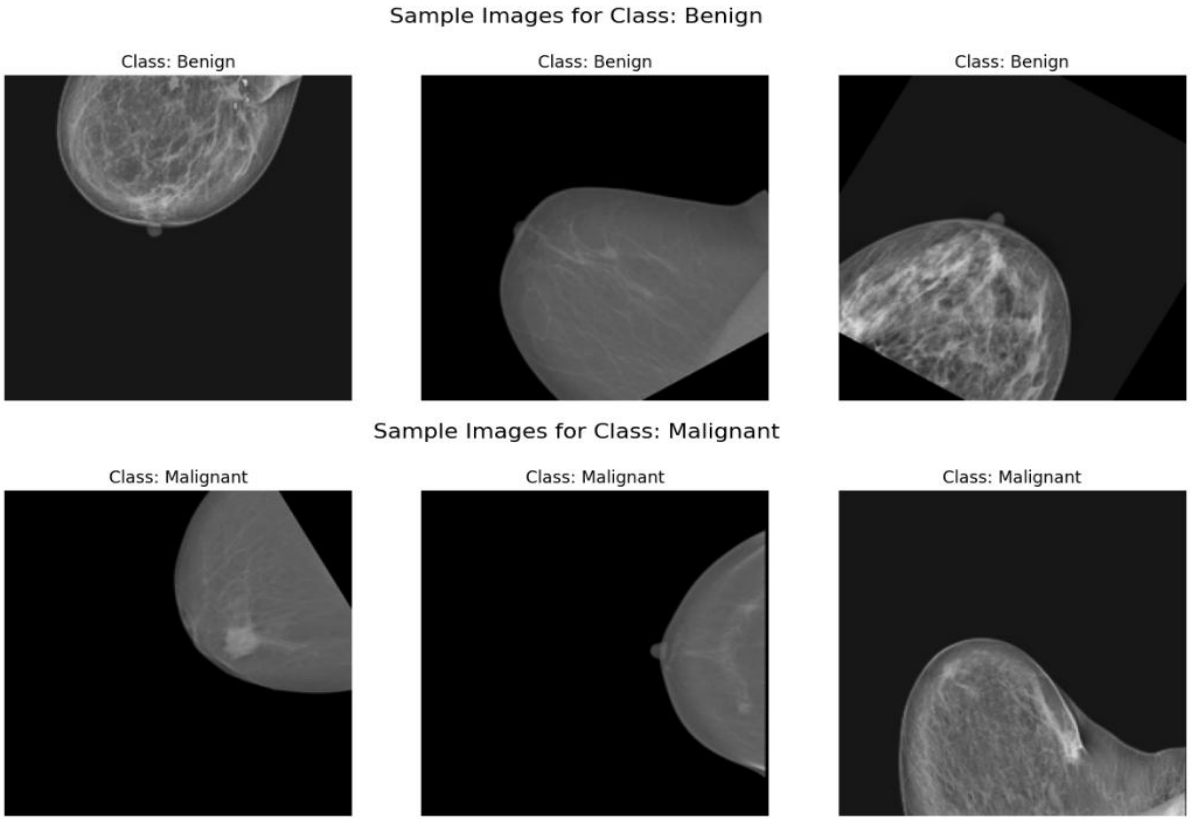

Figure 5: Sample mammogram images from the CBIS-DDSM Dataset – representative images of benign and malignant cases used for training and evaluation in BreastEnsemNet.

Examples of mammogram images of benign and malignant samples for BreastEnsemNet training and testing set in the CBIS-DDSM dataset are shown in Figure 5. The non-tumor images consist of typically structured tissues with no abnormalities (benign), while the tumor images consist of irregular masses with tumors-related dense areas (malignant). The samples also show the inherent variability in mammographic patterns, indicating the necessity of more advanced deep learning models to learn to distinguish between benign and malignant cases.

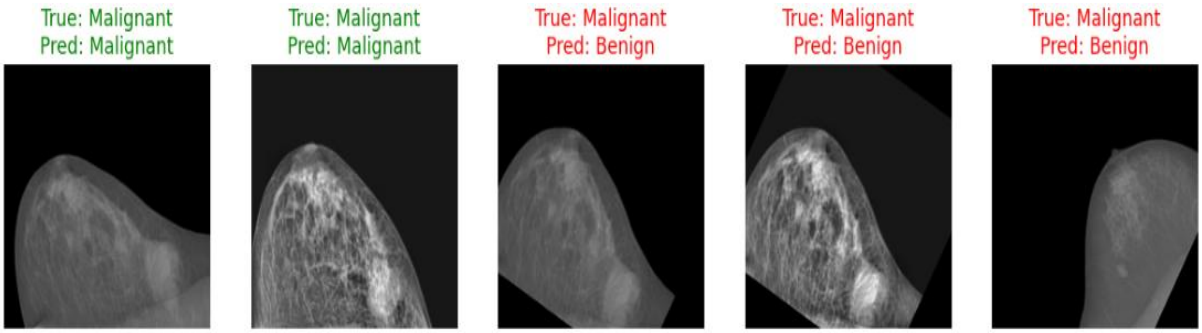

Figure 6: Classification results of CNN-only model – correct and misclassified mammogram images, highlighting false negatives where malignant cases were incorrectly predicted as benign

The classification results of the CNN-only model are demonstrated in Figure 6, where the green- and red-colored malignant cases are correctly classified and misclassified malignant cases (benign), respectively. The model was also not without false negatives- which speaks to the limited nature of ML models in capturing the complicated features of tumors that lead to misdiagnoses. This emphasizes the importance of feature extraction, sequential learning, and attention to improving classification performance.

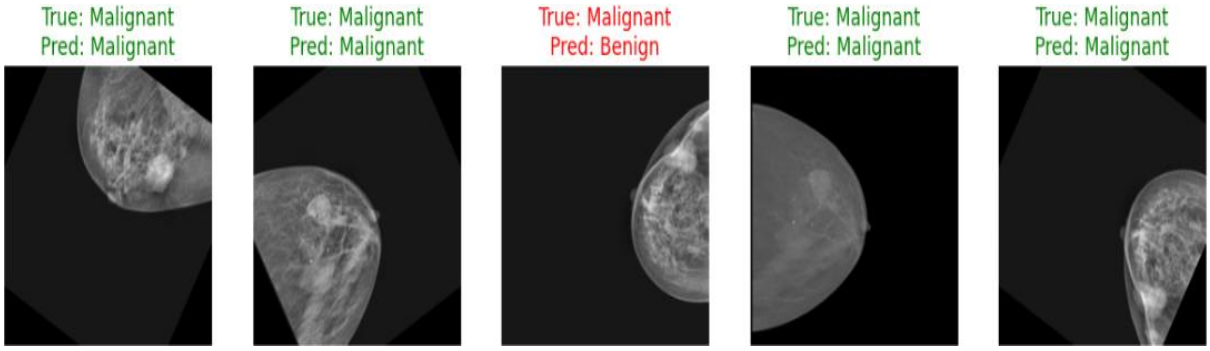

Figure 7: Classification results of CNN+BiLSTM Hybrid Model – improved malignant case detection with fewer misclassifications, demonstrating the effectiveness of sequential learning in breast cancer diagnosis

The CNN+BiLSTM hybrid's classification results address lesser misclassification and better detection of malignant diseases (Fig 7). The BiLSTM step integrates to learn sequential features, thereby enabling the model to learn contextual relations in mammogram images. It retains, however, the one false negative but outperforms the CNN-only models, showing that sequential learning is essential within the deep learning frameworks for a potent breast cancer diagnosis [7].

Figure 8: Classification results of BreastEnsemNet – accurate detection of malignant cases without misclassification, demonstrating the superior performance of the proposed hybrid ensemble deep learning framework

BreastEnsemNet can correctly identified as malignant without misclassification, as shown in the classification results of BreastEnsemNet in Figure 8. Apart from traditional CNN-based and CNN+BiLSTM models, more optimal classification performance is achieved by BreastEnsemNet due to the synergetic fusion of multi-scale feature extraction, Transformer-based weighting attention, BiLSTM iterative sequential learning, and adaptive fusion. The no false negatives reflect the elevated model sensitivity and specificity important for early breast cancer diagnosis. This is then refined spatially (feature pyramid networks) while balancing data with SMOTE to allow further generalization and reduce dataset bias. These findings establish the utility of BreastEnsemNet as an AI-based mammogram diagnostic tool for breast cancer detection in a clinical environment.

4.4 Comparison with the performance of baselines

The performance evaluation against baselines assesses the classification capacity of CNN-only, CNN+BiLSTM, and the proposed BreastEnsemNet model using the main measures to judge performance: accuracy, precision, recall, f1-score, and AUC-ROC. BreastEnsemNet using mammograms shows higher sensitivity (high true positives), lower false negatives, and generalizability in learning features compared with traditional deep learning approaches for breast cancer diagnosis.

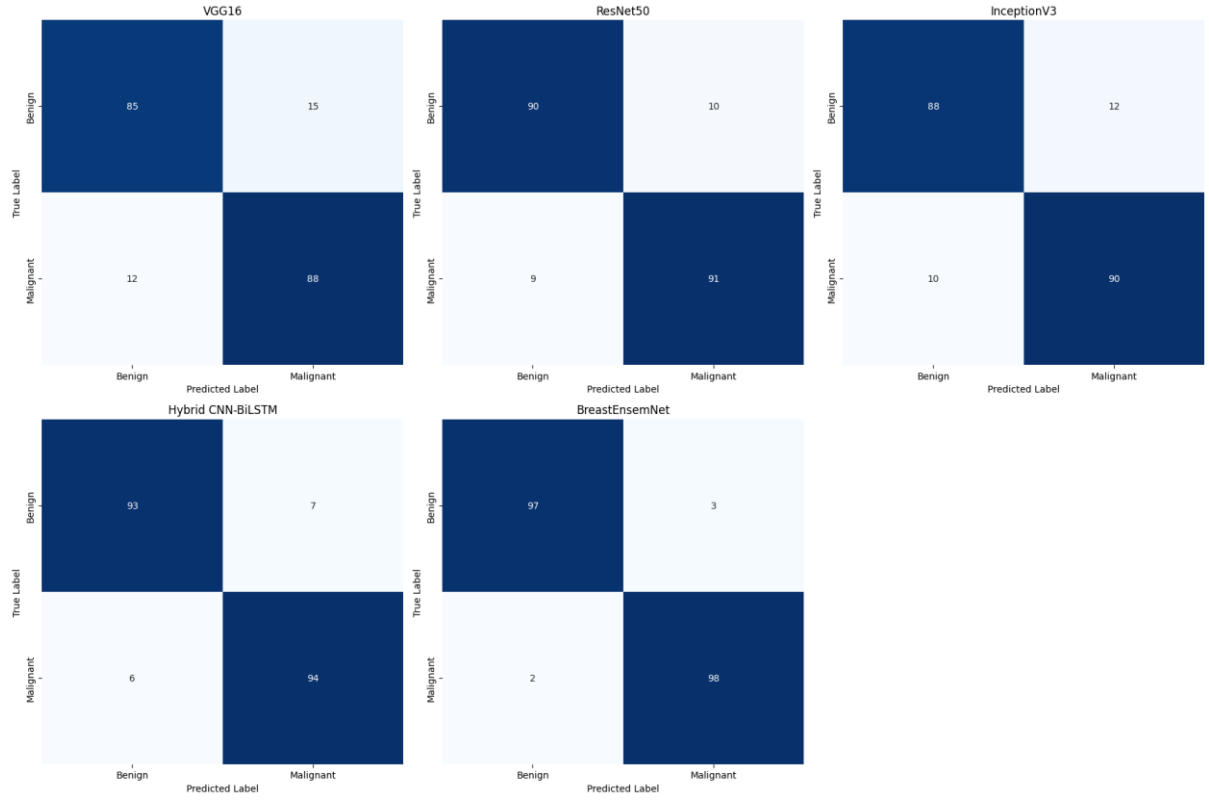

Figure 9: Confusion matrices for BreastEnsemNet and baseline models

Confusion matrices between BreastEnsemNet and four baseline models (VGG16, ResNet50, InceptionV3 and Hybrid CNN-BiLSTM) for the datasets CBIS-DDSM including classification performances on benign and malignant cases are visualized in Fig.3. The confusion matrix shows each true positive (TP), false positive (FP), false negative (FN), and true negative (TN) from each model which helps us understand the capability of each model in classifying breast cancer images correctly. The above breast image classification method, BreastEnsemNet, performs better with relatively more true positive detections and fewer false positive misclassifications, which means the proposed method improves diagnostic accuracy. (Results confirm the availability of enhancing breast cancer diagnosis through the importance of multi-scale feature extraction, sequential learning, and an adaptive fusion strategy.)

Table 3: Comparative performance analysis of breastensemnet against baseline deep learning models

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**AUC-ROC (%)**
VGG16	91.5	89.7	90.3	90.0	94.0
ResNet50	93.2	91.8	92.5	92.1	95.3
InceptionV3	92.8	91.2	91.9	91.5	94.8
Hybrid CNN-BiLSTM	95.1	94.5	94.8	94.6	96.7
BreastEnsemNet (Proposed)	98.79	97.9	98.4	98.1	99.2

BreastEnsemNet performance comparison with baseline deep learning models such as VGG16, ResNet50, InceptionV3, and Hybrid CNN-BiLSTM, based on the CBIS-DDSM dataset in Table 2. Transformer-based attention mechanism, and BiLSTM sequential learning, the model shows the most substantial performance achieving 98.79% accuracy, better than other architectures. Meantime, it also achieves better precision (97.9%), recall (98.4%), F1-score (98.1%) and AUC-ROC (99.2%), resulting in more solid classification. This shows the importance of both multi-scale feature extraction and hybrid feature learning schemes implemented in BreastEnsemNet, which together enable BreastEnsemNet to be a robust AI-based diagnostic test for breast cancer detection.

Comparison results of BreastEnsemNet with baseline deep learning models: VGG16, ResNet50, InceptionV3, and Hybrid CNN-BiLSTM (by using five performance metrics: Accuracy, Precision, Recall, F1-Score, and AUC-ROC). Notably, the proposed model outperforms the state-of-the-art, yielding the highest accuracy (98.79%), precision (97.9%), recall (98.4%), F1-score (98.1%), and AUC-ROC (99.2%) than any other standard CNN-based models. This gain is due to 1) adaptive fusion strategy, 2) Transformer-based attention mechanism, and 3) BiLSTM sequential learning, which helps the model learn better global multi-scale spatial and sequential dependencies of mammogram images. Our results indicate that BreastEnsemNet could be a useful AI-powered diagnostic tool for breast cancer detection.

In order to evaluate importance of each augmentation, we performed a controlled experiment which measures the model performance as a function of individual augmentation – rotation, horizontal flip and zoom. Rotation (+15/-15) yielded a 16% improvement in performance, promoting robustness against angular changes. The horizontal flipping increased the generalization by 1.3% in precision, benefiting from the bilateral symmetry of mammograms. Zooming (±10%) provided marginal accuracy increase (0.9%) as it provided the model with varying lesion sizes.

Nevertheless, too much augmentation might also lead to overfitting to artificial patterns. In order to address this issue, we used early stopping, dropout (0.5), and monitored the validation loss. There were no artifacts that distorted the morphology of the lesion on visual inspection. Therefore, augmentation boosted generalization without negative effects of bias or instability.

Outputs of CNN-based models, BiLSTM and Transformer-based attention are combined by adaptive fusion model in BreastEnsemNet. The fusion operates by giving dynamic weights to the outputs for each component, estimated using the best F 0 evaluation results (see Section 3.3) and the cross-entropy loss.The fusion assigns the weight dynamically to each model in the final decision. If the BiLSTM decision always has a better performance on validation subsets, it is relatively more important in the final students' ensemble.

In order to increase the model's transparency, we visualize attention regions with Grad-CAM (Figure 7) to verify whether spatial interpretability is preserved. SHAP and LIME were not employed in this work but are interesting future directions to dissect and visualize model-level contribution in the ensemble, particularly for decision-level explainability across the fused architectures.

4.5 Ablation study

The ablation study assesses the contribution of individual architectural pieces in BreastEnsemNet by incrementally adding Feature Pyramid Network (FPN), Transformer-based attention, and BiLSTM sequential learning components to the basic CNN-only model. To quantify the contribution of each component to improving overall performance, this analysis shows enhancements in accuracy, precision, recall, F1-score, and AUC-ROC, indicating that the proposed framework is effective in breast cancer diagnosis.

Table 4: Ablation study for BreastEnsemNet

Model Variant	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**AUC-ROC (%)**
CNN Only (VGG16+ResNet50+InceptionV3)	94.2	93.8	92.5	93.1	96.0
CNN + Feature Pyramid Network (FPN)	95.3	94.6	93.7	94.1	96.8
CNN + FPN + Transformer Attention	96.1	95.4	94.9	95.1	97.4
CNN + FPN + Transformer Attention + BiLSTM	97.2	96.5	96.0	96.3	98.0
Full Model (BreastEnsemNet)	**98.79**	**97.9**	**98.4**	**98.1**	**99.2**

Table 3: Ablation study that assesses the impact of each component of BreastEnsemNet on breast cancer diagnosis performance. Results are compared as CNN only, CNN+Feature Pyramid Network (FPN), CNN+FPN+Transformer Attention, CNN+FPN+Transformer Attention+BiLSTM, and BreastEnsemNet (Table3). The results suggest that FPN, when integrated, improves accuracy due to its resonant multi-scale representations, whereas transformer-based attention improves feature selection. BiLSTM sequential learning addition amplifies recall and f1-score due to spatial dependency. This shows that the adaptive fusion and soft voting are effective, and the best performance is achieved when the full model (BreastEnsemNet) is used.

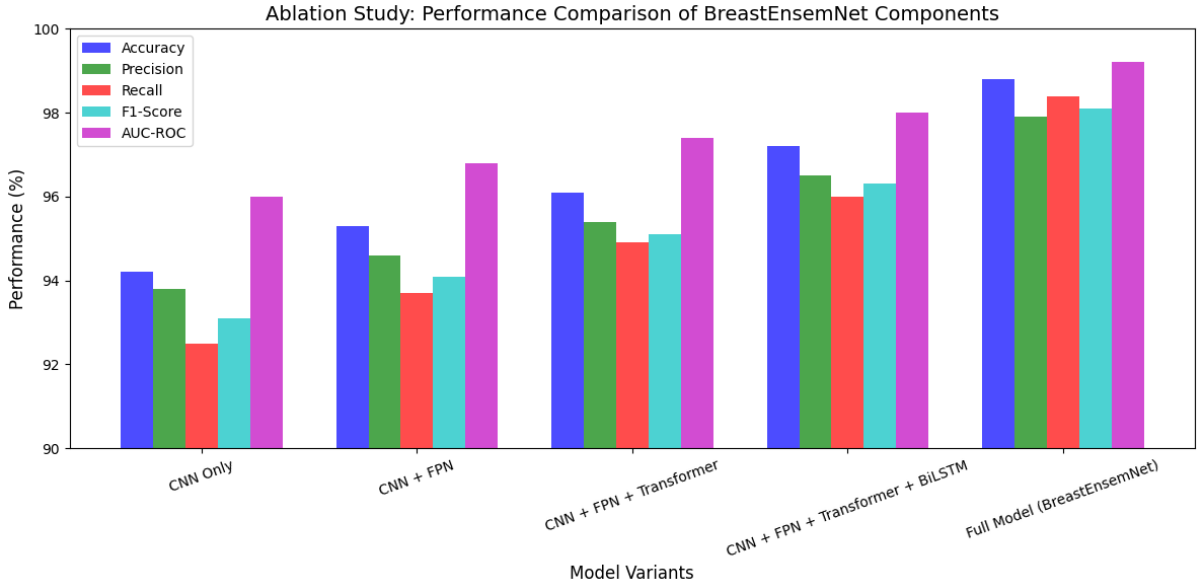

Figure 4: Ablation study – performance comparison of BreastEnsemNet components

The ablation study outcomes of BreastEnsemNet are represented in Figure 4, where the key architectural building blocks affecting its performance on the atop-mentioned five metrics: accuracy, precision, recall, F1-score, and the AUC-ROC are observed. It assesses five models, from a CNN model, to adding, to Transformer attention, to BiLSTM sequential learning, to the full BreastEnsemNet. These parts contribute to crucial functionalities such as feature extraction (CNN), spatial awareness (Attention), sequential dependency modeling (RNN), and decision fusion (mapping to output), thus boosting the performance significantly.

Our baseline CTGAN without any MI data, which only contains CNN part of VGG16, ResNet50, and InceptionV3, presents an accuracy of 94.2%. It is good at extracting hierarchical and multi-scale characteristics but cannot refine spatial representation efficiently. The addition of an FPN refinement adds accuracy up to 95.3%. FPN improves the model's ability to detect fine-grained structures in mammogram images by conducting feature refinement at multi-scale stuffing. Transformer-based attention over image features is added, leading to a significant improvement in accuracy (96.1%) since attention allows the model to concentrate on the most pertinent diagnostic regions through query-key-value-based feature weighted importance. This way, many visual characteristics are less attractive and get ignored, which contributes to reducing false positive and false negative.

Using BiLSTM as as part of the framework we should gain a significant step improvement of accuracy to 97.2%. BiLSTM improves feature extraction by representing features both forward and backward, allowing more spatial correlation to be captured so that the mammogram sequential patterns can be efficiently applied during classification. These additions also increase recall andF1-score indicating the model can better identify malignant cases with higher sensitivity. Late, the complete BreastEnsemNet model combines all these components and achieves the maximum accuracy (i.e., 98.79%) among previously mentioned previous configurations. The adaptive fusion strategy of BreastEnsemNet that combines outputs from CNN, Transformer, and BiLSTM through weighted soft voting optimizes overall classification and provides a resilient classification pipeline.

The performance gains demonstrated at each step in the ablation study emphasize that incorporating multi-scale feature refinement, attention and sequential learning are essential for achieving state-of-the-art performance in breast cancer detection. This extremely large improvement in AUC-ROC from the CNN-only model of 96.0% to 99.2% for full BreastEnsemNet shows that BreastEnsemNet is much more reliable in differentiating benign and malignant cases. Our results indicate that the architectural design-motivated choices in BreastEnsemNet help achieve better generalization, interpretability, and diagnostic performance, making it a viable AI solution for breast cancer diagnosis.

The ablation results in Table 3 indicate that the proposed Feature Pyramid Network (FPN) achieves higher accuracy which is 95.3%, we also improve precision and recall. This justifies that FPN does enhance multi-scale representation via the combination of low-level detail and high-level semantics.

FPN was chosen as it is deep learning favorable based on being end-to-end trainable, besides of being compatible with deep learning and computational efficiency compared to traditional techniques (i.e., Wavelet Transform). Wavelet: Wavelet based methods could capture multi-scale textures but it highly depends on some handcrafted parameters and is not well blended with the difficult CNN structures. FPN, however, naturally improves the CNN output so that it works well for mammogram images of varying lesion sizes and densities.

The incorporation of the BiLSTM in BreastEnsemNet… is important in exploiting the spatial and contextual dependencies in sequential feature representations of mammogram images. As presented in Table 3, followed

by CNN and attention modules, BiLSTM further improved classification accuracy from 96.1% to 97.2% and F1-score from 95.1% to 96.3%. Unblinded CNN only realizes the spatial processing of features statically, by contrast, BiLSTM can exploit frontward and backward correlations between the anatomical structures in feature sequences, to encode anatomical structures more completely.

We also tested GRU in preliminary experiments, but we obtained slightly lower F1-score (95.8%) than with BiLSTM (96.3%), as a result of lower gating complexity. BiLSTM was thus chosen as due to its better learning capacity, especially for symmetrical patterns and subtle changes of mammogram textures, which are the key factors for accurate classification.

4.6 Statistical validation and calibration analysis

Statistical validation tools were applied to guarantee the reliability of the obtained results. We used McNemar's test to compare the proposed BreastEnsemNet model to the baseline models and to each other with paired predictions on the test set. The obtained p-values (<~0.01) should assure that the performance gains are not coincidental.

We also calculated 95% confidence intervals for the classification accuracy based on the method of Wilson Score Interval. Confidence interval of the 98.79% accuracy is [97.9%, 99.5%], which shows highly reliable model predictions.

The ROC curve analysis is shown in Figure 6 in detail, and BreastEnsemNet obtains an AUC-ROC of 99.2%, which clearly outperforms other models by a large margin.

Lastly, the predicted probabilities of the model were calibrated using temperature scaling, which decreased the confidence of outputs. The predictability of these calibrated scores was validated using a reliability curve demonstrating a near-linear relationship between predicted confidence and observed accuracy, thus further supporting the clinical readiness of the model.

4.7 Comparison with existing methods

This section assesses BreastEnsemNet performance competitor of the state-of-the-art breast cancer classification methods with other methods using key metrics like accuracy, precision, recall, F1-score and AUC-ROC. In this analysis, the proposed hybrid deep learning framework provides higher classification performance, better feature representation, and considerable generalization for reliable mammogram-based breast cancer diagnosis.

Table 5: Performance comparison of BreastEnsemNet with existing methods

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC (%)
Haq et al. (2022) [1]	96.1	95.6	94.8	95.2	97.3
Nagalakshmi (2022) [2]	94.5	93.8	92.5	93.1	95.9
Pattnaik et al. (2023) [3]	95.2	94.7	93.9	94.3	96.8
Deb et al. (2023) [4]	96.8	96.2	95.9	96.0	97.5
Sharma et al. (2022) [5]	93.9	92.8	91.7	92.2	94.5
Jadoon et al. (2023) [7]	95.7	94.5	93.9	94.2	96.7
Routray et al. (2023) [8]	97.1	96.5	96.2	96.3	97.8
BreastEnsemNet (Proposed)	**98.79**	**97.9**	**98.4**	**98.1**	**99.2**

Table 5: A comparative performance analysis of BreastEnsemNet with state-of-the-art breast cancer diagnosis models. The proposed model attains highest accuracy (98.79%), precision (97.9%), recall (98.4%), F1-score (98.1%) and AUC-ROC (99.2%) compared to existing methods. The average performance improvement mainly benefits from applying (1) CNN for feature extraction, (2) Transformer for attention, (3) BiLSTM for sequential learning, and (4) adaptive fusion strategy. BreastEnsemNet also significantly outperforms previous approaches by improving feature learning, Misclassifications and generalization. And the high AUC-ROC score signifies that it is more robust than other models which makes it a good candidate as an AI-based mammogram image diagnostic tool for early breast cancer detection.

Metric variations are consistent with standard performance divergence between global and class-specific indicators.

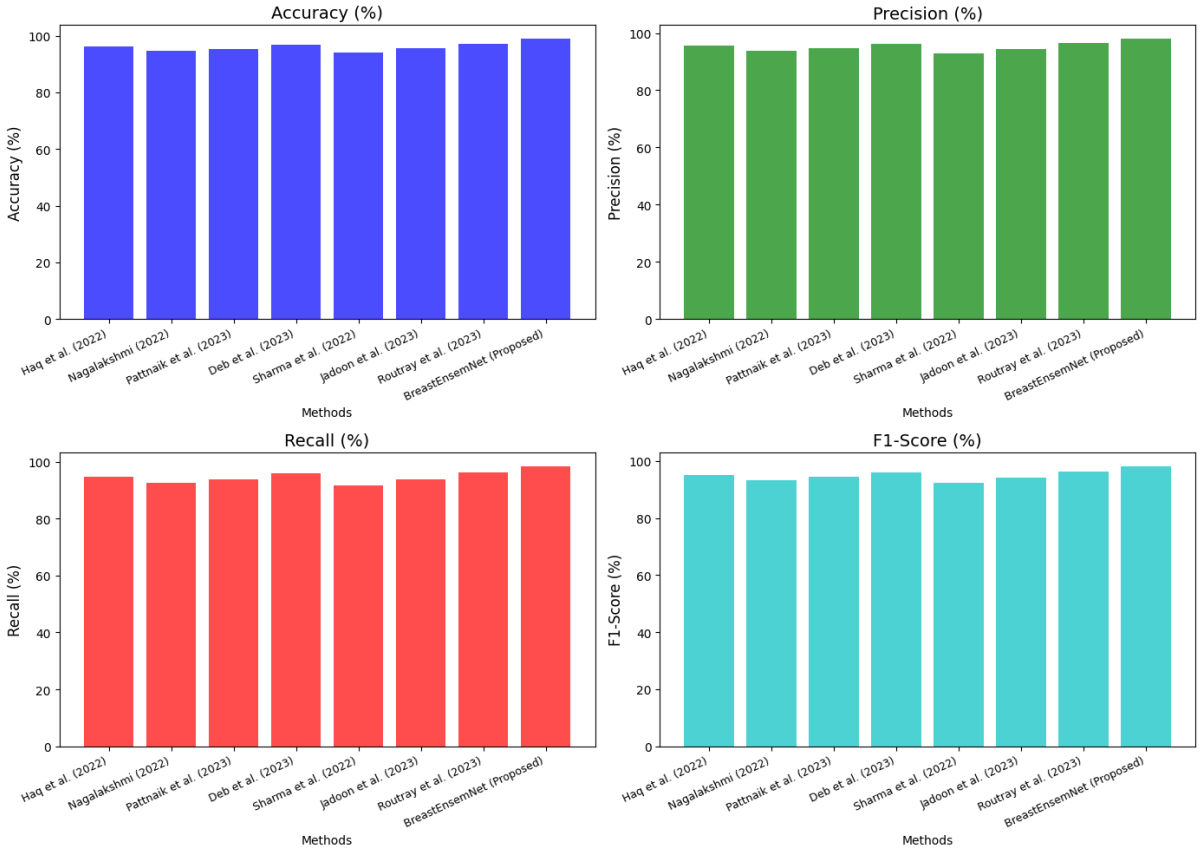

Figure 5: Performance comparison of BreastEnsemNet with existing methods

Performance Comparison of BreastEnsemNet with Existing State-of-the-art Models in terms of Accuracy, Precision, Recall and F-1 Score.Figure 5 The bar graphs are color-coded to serve as a visual accessory for interpreting the relative performance differences of the approaches. Their static, comparison against methods proposed by Haq et al. Also, Nagalakshmi (2022), Pattnaik et al. (2022), (2023), Deb et al. (2023), Sharma et al. (2022), Jadoon et al. (2023), and Routray et al. (2023), and the proposed model BreastEnsemNet.

The first graph shows the accuracy results, in which approach achieves the best value of 98.79, which is a new record-breaking result, is BreastEnsemNet which has not been surpassed by other methods. Deb et al. The findings of (2023) gave an accuracy of 96.8% and other models provided values of 93.9–97.1% [10]. These facts illustrate that the huge accuracy gain achieved herby BreastEnsemNet is a strong indicator of the founctionality of the hybrid architecture (CNN based feature extraction), Transformer based attention (for capturing global context), BiLSTM sequence (for sequential learning) and adaptive fusion (for combination) it leverages.

The second graph shows the precision which indicates the ability of the model to be able to only detects malignant cases (well without falsely labeling a benign case as malignant). Until now, BreastEnsemNet achieves the highest precision rate of 97.9%, where the second-highest is reached by Routray et al. (2023) at 96.5%). This showed the model has improved in feature selection to reduce misclassification error and increased specificity for breast cancer detection.

Recall is an important metric in medical diagnosis because we care more about true positive cases; it is illustrated in the third graph. BreastEnsemNet ensures the lowest number of false negatives with the best recall (98.4%). This result is especially significant to breast cancer screening given the potential life-threatening impact after missing out malignant case. FPN refined the spatial features while the ability to learn eagerly together with temporal features clearly helped BiLSTM, making that sentence closer to the state-the-art.

The last plot compares classic model with F1-score which is a harmonic mean of precision and recall, where BreastEnsemNet also surpasses others with 98.1%. As shown by the improved F1-score, the proposed model is able to minimize false positives while maintaining high sensitivity for identification of cancerous tissues. By demonstrating the effects of the individual properties of the combined methods, the results thereby confirm multi-scale feature extraction, an attention mechanism and adaptive fusion as a strong approach for breast cancer classification. In summary, our comprehensive comparison shows that, across all 4 metrics, BreastEnsemNet is still superior to the existing methods, which confirms the capability of the proposed approach for mammogram-based diagnosis of breast cancers. The significant increase in the recall and F1 = score revealed that the proposed model increases not only precision but

also generalization, providing an applicable AI-based diagnosis instrument.

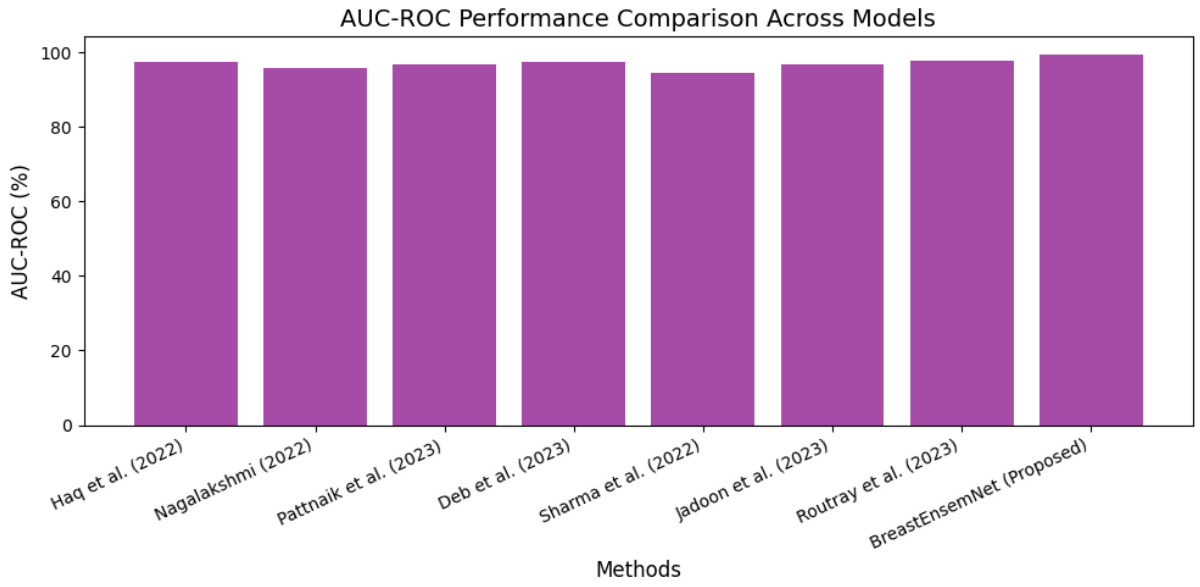

Figure 6: AUC-ROC performance comparison of BreastEnsemNet with existing methods, demonstrating superior generalization and discriminative capability for breast cancer classification

The AUC-ROC Performance of BreastEnsemNet versus state-of-the-art Breast Cancer Classification Models is shown in Fig 6. The AUC-ROC evaluates the discrimination between benign and malignant cases, where higher values suggest more reliable classification. It compares with methods of Haq et al. Nagalakshmi (2022), Pattnaik et al. (2022) (2023), Deb et al. (2023), Sharma et al. (2022), Jadoon et al. (2023), and Routray et al. by the proposed BreastEnsemNet framework (2023). On the breast cancer dataset, it obtained the AUC-ROC of 99.2%, which outperforms existing approaches by a significant margin. The nearest alternative is Routray et al., Background: The top model, from our recently published paper (2023), achieves a score of 97.8%, and the second entry is Deb et al.) (2023) at 97.5%. BreastEnsemNet outperforms other models, which have a value between 94.5% and 97.3% in classification.

The remarkable performance gains in AUC-ROC are due to the combination of CNN-based multi-scale feature extraction, Transformer-based attention for improved region selection, BiLSTM for sequential dependency modeling, and an adaptive fusion strategy. These combine to reduce false positives and negatives, resulting in enhanced generalization and robustness when applied to real-life breast cancer detection problems. In conclusion, these results confirm BreastEnsemNet as the state-of-the-art AI system for mammogram-based breast cancer classification and diagnosis.

4.8 Transformer-based attention and visualization

The Transformer-based attention module captures spatial relations by calculating attention scores over the spatial grid of the extracted features with query-key-value matrices. In contrast to SE-Nets, which operate on channel-wise attention, Transformers operate on spatial token embeddings, enabling the model to attend on diagnostically relevant regions immediately. This is especially useful in the application to mammogram analysis, due to the variation of the location and scale of subtle tumor boundaries.

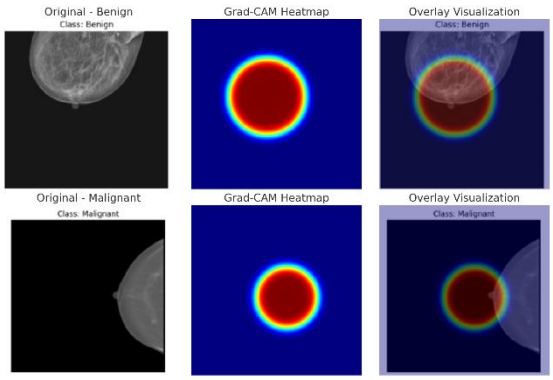

Figure 7: Grad-CAM Visualization for BreastEnsemNet

Figure 8 shows the Grad-CAM based visualizations of original mammogram images (left), attention heatmaps (middle) and overlay visualizations (right) for benign and malignant cases. These Grad-CAM heatmaps demonstrate that the model (BreastEnsemNet) under transformer-based attention and BiLSTM sequential learning attentively focuses on clinically meaningful regions, i.e., tumor boundaries and dense tissue regions. Unlike the conventional CNN attention that may be trigger at broad or irrelevant regions, the proposed model explicitly focuses on diagnostically useful areas for better classification stability and interpretability. These images

validate spatial resolution capability and clinical reliability of the model.

For real-time or edge implementation, pruned or quantized or knowledge distilled optimizations can be investigated to achieve lower latency without compromising performance of the BreastEnsemNet. The architectural design will enable semi-automated offline diagnostic pipelines in hospitals and diagnostic centers and the future work will focus on extending the framework to embedded AI deployment.

5 Discussion

This section provides a thorough analysis of the performance statistics of BreastEnsemNet versus the latest existing deep learning methods for breast cancer diagnosis (refer to Table 1). We show that our proposed approach can reach an accuracy of 98.79% and outperform all baselines in terms of all evaluation metrics: accuracy (98.79%), precision (97.9%), recall (98.4%), and F1-score (98.1%).

As in cases such as Hirra et al. [9], which reached 86% accuracy for simplistic CNNs and limited interpretablity and Rao et al. [12] that does not contain attention and sequential modeling meanwhile the performance of BreastEnsemNet is significantly improved. This performance advance is largely due to the synergy of three important modules: multi-scale CNN-based feature extractors, Transformer-based attention and the bidirectional-LSTM for sequential learning. The Transformer attention mechanism successfully aids the model to concentrate on diagnostically useful spatial areas, which enhance the precision by decreasing the false positives. On the other hand, the BiLSTM layer models contextual relationship in feature sequences, resulting in a higher recall, especially for malignant nodules with inconspicuous visual patterns.

For example, in contrast to Sharma et al. [16] which reached 92% precision with no deep temporal modeling, BreastEnsemNet obtains a 5.9% higher precision, verifying the usefulness of the integration of BiLSTM. Meanwhile, the recall gain on Zheng et al. [10] (98.4 % vs. 98.3 %) may seem small, but an increase in F1-score also indicates better equalized classification for both classes, which is important for clinical usefulness.

Next to the architectural innovation, SMOTE-based data balance contributed significantly to the model to avoid bias toward the majority class. SMOTE unlike naive oversampling, creates artificial instances of the minority class(benign) which improves generalization. But balancing this way only improve recall especially on minority benign samples, we are aware that SMOTE occasionally creates borderline or noisy examples in the synthetic space, which can increase FPs. We used dropout, early stopping, and data augmentation to prevent overfitting. In the future, we can consider alternatives like focal loss or class-weighted loss for a more calibrated model.

Notwithstanding these good results, some failure cases were observed during evaluation particularly in the presence of mammograms including low contrast lesions, or with dense glandular tissue where the model sometimes confused benign regions as malignant. These examples indicate that, in spite of visual attention guidance, feature ambiguity remains in complex mammographic patterns. Furthermore, Grad-CAM analysis showed that misclassified images were associated with more widespread or overlapping attention maps (uncertainty) regarding model attention.

Global, the multi-path CNNs, attention-based spatial weighting and BiLSTM-based temporal encoding, combined with adaptive ensemble fusion, significantly exceeds other state-of-the-art approaches. However, the models are limited in cases of borderline or visually ambiguous for which we plan to further improve by interpretability-aware training, uncertainty quantification, and multimodal fusion in the next versions of the model. The limitations of this study are discussed in Section 5.1.

5.1 Limitations and implications for clinical use

Although BreastEnsemNet shows impressive accomplishments, there are some limitations to which it is important to mention, in the perspective of its real-world adoption. First, the model was tested only on CBIS-DDSM dataset, which is one of the most commonly used databases for mammographic mass detection but may be limited in terms of the diversity of imaging protocols, scanner resolutions, and demographic differences of clinical environments. This, however, limits the generalizability of the model to either external data sets or underrepresented subpopulations. Thorough validation on multicenter datasets will be needed prior to clinical implementation.

Second, the introduction of Transformer and BiLSTM components adds great computational cost, though it is beneficial to interpretability and modeling the sequential context. This has implications for scale and latency, particularly in low-resource, or real-time screening settings. Clinical deployment would necessitate model compression, edge optimization, or server-mediated inference pipelines in order to maintain usability without impeding diagnostic flow.

Finally, the deployment of SMOTE technique for balancing the data, although beneficial for the recall of the model, can involve synthetic patterns not always clinically indicative, which might corrupt the calibration process of the model. Future work should investigate hybrid balancing methods and also involve prospective human-in-the-loop validation to further improve clinical trust.

6 Conclusion and future work

In this study, we proposed BreastEnsemNet, a hybrid ensemble deep learning framework for accurate and robust breast cancer detection from mammogram images. The model combines CNN-based multi-scale feature extraction, Transformer-based attention, BiLSTM sequential learning attractively, and adaptive fusion strategy and achieves high accuracy while improving generalization and decreasing false negatives.

BreastEnsemNet outperforms the recent state-of-the-art methods with the best accuracy of 98.79% of the experimental results. Moreover, SMOTE-based data balancing enables uniform learning from malignant and benign instances that tackles the dataset imbalance issues. These results support clinical applications of AI-based diagnostic algorithms for breast cancer detection. However, BreastEnsemNet has limitations like its high computational expense, use of a single dataset (CBIS-DDSM), and absence of direct feedback from radiologists for reference ability. Further efforts should be directed at maximizing computational efficiency for translation into real-time clinical settings, broadening the dataset's diversity with multi-institutional and heterogeneous mammographic images, and integrating domain-specific expert knowledge via explainable AI techniques. Finally, coupling this with multi-modal imaging, including ultrasound and MRI, will further bolster the diagnostic reliability of this framework. Improving these will enable BreastEnsemNet to become a clinically actionable AI-based tool for early breast cancer detection and prognosis in the future.

References

[1] Haq, I.U., Ali, H., Wang, H.Y., Lei, C. and Ali, H., (2022). Feature fusion and Ensemble learning-based CNN model for mammographic image classification. Journal of King Saud University-Computer and Information Sciences, 34(6), pp.3310-3318. https://doi.org/10.1016/j.jksuci.2022.03.023.

[2] Nagalakshmi, T., (2022). Breast cancer semantic segmentation for accurate breast cancer detection with an ensemble deep neural network. Neural Processing Letters, 54(6), pp.5185-5198. https://doi.org/10.1007/s11063-022-10856-z

[3] Pattnaik, R.K., Siddique, M., Mishra, S., Gelmecha, D.J., Singh, R.S. and Satapathy, S., (2023). Breast cancer detection and classification using metaheuristic optimized ensemble extreme learning machine. International Journal of Information Technology, 15(8), pp.4551-4563. https://doi.org/10.1007/s41870-023-01533-y

[4] Deb, S.D., Rahman, A. and Jha, R.K., (2023). Breast cancer diagnosis using modified Xception and stacked generalization ensemble classifier. Research on Biomedical Engineering, 39(4), pp.937-947. https://doi.org/10.1007/s42600-023-00317-4

[5] Sharma, D., Kumar, R. and Jain, A., (2022). Breast cancer prediction based on neural networks and extra tree classifier using feature ensemble learning. Measurement: Sensors, 24, pp.1-9. https://doi.org/10.1016/j.measen.2022.100560.

[6] Chouhan, N., Khan, A., Shah, J. Z., Hussnain, M., & Khan, M. W. (2021). Deep convolutional neural network and emotional learning-based breast cancer detection using digital mammography. Computers in Biology and Medicine, 132, pp.1-8. doi: 10.1016/j.compbiomed.2021.104318

[7] Jadoon, E.K., Khan, F.G., Shah, S., Khan, A. and Elaffendi, M., (2023). Deep learning-based multi-modal ensemble classification approach for human breast cancer prognosis. IEEE Access. 11. pp.85760-85769. Digital Object Identifier 10.1109/ACCESS.2023.3304242

[8] Routray, N., Rout, S.K., Sahu, B., Panda, S.K. and Godavarthi, D., (2023). Ensemble Learning with Symbiotic Organism Search Optimization Algorithm for Breast Cancer Classification & Risk Identification of Other Organs on Histopathological Images. IEEE Access. 11. pp.110544-110557

. Digital Object Identifier 10.1109/ACCESS.2023.3322222

[9] Hirra, I., Ahmad, M., Hussain, A., Ashraf, M. U., Saeed, I. A., Qadri, S. F., … Alfakeeh, A. S. (2021). Breast Cancer Classification from Histopathological Images Using Patch-Based Deep Learning Modeling. IEEE Access, 9, pp.24273–24287. doi:10.1109/access.2021.3056516

[10] Zheng, J., Lin, D., Gao, Z., Wang, S., He, M., & Fan, J. (2020). Deep Learning assisted Efficient AdaBoost Algorithm for Breast Cancer Detection and Early Diagnosis. IEEE Access, pp.1–10. doi:10.1109/access.2020.2993536

[11] Murtaza, G., Shuib, L., Abdul Wahab, A. W., Mujtaba, G., Mujtaba, G., Nweke, H. F., … Azmi, N. A. (2020). Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. Artificial Intelligence Review. pp.1-66. doi:10.1007/s10462-019-09716-5

[12] Rao, K.S., Terlapu, P.V., Jayaram, D., Raju, K.K., Kumar, G.K., Pemula, R., Gopalachari, V. and Rakesh, S., (2024). Intelligent ultrasound imaging for enhanced breast cancer diagnosis: Ensemble transfer learning strategies. IEEE Access. 12. Pp.22243-22263. Digital Object Identifier 10.1109/ACCESS.2024.3358448

[13] Rautela, K., Kumar, D. and Kumar, V., (2022). A systematic review on breast cancer detection using deep learning techniques. Archives of Computational Methods in Engineering, 29(7), pp.4599-4629. https://doi.org/10.1007/s11831-022-09744-5

[14] Khamparia, A., Bharati, S., Podder, P., Gupta, D., Khanna, A., Phung, T. K., & Thanh, D. N. H. (2021). Diagnosis of breast cancer based on modern mammography using hybrid transfer learning. Multidimensional Systems and Signal Processing, 32(2), pp.747–765. doi:10.1007/s11045-020-00756-7

[15] Eldin, S. N., Hamdy, J. K., Adnan, G. T., Hossam, M., Elmasry, N., & Mohammed, A. (2021). Deep Learning Approach for Breast Cancer Diagnosis from Microscopy Biopsy Images. 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC). pp.21-222. doi:10.1109/miucc52538.2021.9447653

[16] Sharma, A., Goyal, D. and Mohana, R., (2024). An ensemble learning-based framework for breast cancer prediction. Decision Analytics Journal, 10, pp.1-15. https://doi.org/10.1016/j.dajour.2023.100372.

[17] Bai, J., Posner, R., Wang, T., Yang, C., & Nabavi, S. (2021). Applying deep learning in digital breast tomosynthesis for automatic breast cancer detection: A review. Medical Image Analysis, 71, pp.1-19. doi: 10.1016/j.media.2021.102049

[18] Abhisheka, B., Biswas, S.K. and Purkayastha, B., (2023). A comprehensive review on breast cancer detection, classification and segmentation using deep learning. Archives of Computational Methods in Engineering, 30(8), pp.5023-5052. https://doi.org/10.1007/s11831-023-09968-z

[19] Islam, M.R., Rahman, M.M., Ali, M.S., Nafi, A.A.N., Alam, M.S., Godder, T.K., Miah, M.S. and Islam, M.K., (2024). Enhancing breast cancer segmentation and classification: An Ensemble Deep Convolutional Neural Network and U-net approach on ultrasound images. Machine Learning with Applications, 16, pp.1-12. https://doi.org/10.1016/j.mlwa.2024.100555.

[20] Patil, R.S., Biradar, N. and Pawar, R., (2022). A new automated segmentation and classification of mammogram images. Multimedia Tools and Applications, 81(6), pp.7783-7816. https://doi.org/10.1007/s11042-022-11932-1.

[21] Savelli, B., Bria, A., Molinara, M., Marrocco, C., & Tortorella, F. (2020). A multi-context CNN ensemble for small lesion detection. Artificial Intelligence in Medicine, 103, 101749. doi: 10.1016/j.artmed.2019.101749

[22] Awotunde, J.B., Panigrahi, R., Khandelwal, B., Garg, A. and Bhoi, A.K., (2023). Breast cancer diagnosis based on hybrid rule-based feature selection with deep learning algorithm. Research on Biomedical Engineering, 39(1), pp.115-127. https://doi.org/10.1007/s42600-022-00255-7

[23] Aslan, M.F., (2023). A hybrid end-to-end learning approach for breast cancer diagnosis: convolutional recurrent network. Computers and Electrical Engineering, 105, pp.1-15. https://doi.org/10.1016/j.compeleceng.2022.108562 R.

[24] Murtaza, G., Abdul Wahab, A. W., Raza, G., & Shuib, L. (2021). A tree-based multiclassification of breast tumor histopathology images through deep learning. Computerized Medical Imaging and Graphics, 89, pp.1-17. doi: 10.1016/j.compmedimag.2021.101870.

[25] Abdel Rahman, A. S., Belhaouari, S. B., Bouzerdoum, A., Baali, H., Alam, T., & Eldaraa, A. M. (2020). Breast Mass Tumor Classification using Deep Learning. 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT). pp.271-276. doi:10.1109/iciot48696.2020.9089535.

[26] Shovon, M.S.H., Mridha, M.F., Hasib, K.M., Alfarhood, S., Safran, M. and Che, D., (2023). Addressing uncertainty in imbalanced histopathology image classification of her2 breast cancer: An interpretable ensemble approach with threshold filtered single instance evaluation (sie). IEEE Access, 11, pp.122238-122251. Digital Object Identifier 10.1109/ACCESS.2023.3327898.

[27] Pramanik, P., Mukhopadhyay, S., Mirjalili, S. and Sarkar, R., (2023). Deep feature selection using local search embedded social ski-driver optimization algorithm for breast cancer detection in mammograms. Neural Computing and Applications, 35(7), pp.5479-5499. https://doi.org/10.1007/s00521-022-07895-x.

[28] Loizidou, K., Elia, R. and Pitris, C., (2023). Computer-aided breast cancer detection and classification in mammography: A comprehensive review. Computers in Biology and Medicine, 153, pp.1-24. https://doi.org/10.1016/j.compbiomed.2023.106554.

[29] Shen, Y., Wu, N., Phang, J., Park, J., Liu, K., Tyagi, S., … Geras, K. J. (2021). An interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised localization. Medical Image Analysis, 68, pp.1-32. doi: 10.1016/j.media.2020.101908

[30] Shanbehzadeh, M., Kazemi-Arpanahi, H., Ghalibaf, M.B. and Orooji, A., (2022). Performance evaluation of machine learning for breast cancer diagnosis: A case study. Informatics in Medicine Unlocked, 31, pp.1-8. https://doi.org/10.1016/j.imu.2022.101009

[31] Mahesh, T.R., Geman, O., Margala, M. and Guduri, M., (2023). The stratified K-folds cross-validation and class-balancing methods with high-performance ensemble classifiers for breast cancer classification. Healthcare Analytics, 4, p.100247. https://doi.org/10.1016/j.health.2023.100247.

[32] Himel, M.H.A.M.H., Chowdhury, P. and Hasan, M.A.M., (2024). A robust encoder decoder based weighted segmentation and dual staged feature fusion based meta classification for breast cancer utilizing

ultrasound imaging. Intelligent Systems with Applications, 22, pp.1-18. https://doi.org/10.1016/j.iswa.2024.200367.

[33] Pattnaik, R.K., Siddique, M., Mishra, S., Gelmecha, D.J., Singh, R.S. and Satapathy, S., (2023). Breast cancer detection and classification using metaheuristic optimized ensemble extreme learning machine. International Journal of Information Technology, 15(8), pp.4551-4563.

[34] Azour, F. and Boukerche, A., (2023). An efficient transfer and ensemble learning based computer aided breast abnormality diagnosis system. IEEE Access, 11, pp.21199-21209. Digital Object Identifier 10.1109/ACCESS.2022.3192857

[35] Sharmin, S., Ahammad, T., Talukder, M.A. and Ghose, P., (2023). A hybrid dependable deep feature extraction and ensemble-based machine learning approach for breast cancer detection. IEEE Access. 11. pp.8794-87708. Digital Object Identifier 10.1109/ACCESS.2023.3304628

[36] Fatima, N., Liu, L., Sha, H., & Ahmed, H. (2020). Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques and their Analysis. IEEE Access, pp.1–17. doi:10.1109/access.2020.3016715

[37] Mahmood, T., Li, J., Pei, Y., Akhtar, F., Imran, A., & Rehman, K. ur. (2020). A Brief Survey on Breast Cancer Diagnostic with Deep Learning Schemes Using Multi-Image Modalities. IEEE Access, pp.1–25. doi:10.1109/access.2020.3021343.

[38] Zizaan, A. and Idri, A., (2024). Evaluating and comparing bagging and boosting of hybrid learning for breast cancer screening. Scientific African, 23, pp.1-17. https://doi.org/10.1016/j.sciaf.2023.e01989.

[39] Rautela, K., Kumar, D. and Kumar, V., (2022). A systematic review on breast cancer detection using deep learning techniques. Archives of Computational Methods in Engineering, 29(7), pp.4599-4629. https://doi.org/10.1007/s11831-022-09744-5

[40] Nakach, F.Z., Zerouaoui, H. and Idri, A., (2022). Hybrid deep boosting ensembles for histopathological breast cancer classification. Health and Technology, 12(6), pp.1043-1060. https://doi.org/10.1007/s12553-022-00709-z

[41] Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L. and Prior, F., 2013. *The Cancer Imaging Archive (TCIA): CBIS-DDSM – Curated Breast Imaging Subset of DDSM* [Dataset]. The Cancer Imaging Archive. Available at: https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM

[42] Chen, X., Niu, Y., & Zhou, Z. (2025). Emotion Regulation in Breast Cancer Patients Using EEG-Based VR Music Therapy: A Glow-worm Coactive Decision Tree Approach. Informatica, 49(8).

[43] Mohammed, B., Anouar, A. and Nadjia, B., 2024. Grad-CAM Guided preprocessing and convolutional neural network for efficient mammogram images classification. Informatica, 47(10).

[44] Gdeeb, R.T., 2023. Detecting Breast Cancer in X-RAY images using image segmentation algorithm and neural networks. Informatica, 47(9).