Hybrid Ensemble Learning Approach for Breast Cancer Detection from Mammogram Images

C Nandini
Department of Computer Science and Engineering
Dayananda Sagar Academy Of Technology and Management
Bengaluru, India
vp@dsatm.edu.in

Weiwei Jiang
Beijing University of Posts and Telecommunications
Beijing, China
jww@bupt.edu.cn

Shashi Kant Gupta
Lincoln University College
raj2008enator@gmail.com

Manasa Sandeep
Department of Computer Science Science and Engineering
Dayananda Sagar Academy Of Technology and Management
Bengaluru, India
manasa-cs@dsatm.edu.in

***Abstract*—Breast cancer is one of the most common and deadly diseases affecting women worldwide. Early and accurate detection is crucial for improving survival rates. However, traditional diagnostic methods struggle with issues such as limited labeled data, significant class imbalance, and complex mammographic textures. These challenges often reduce the accuracy and general effectiveness of models. To overcome these obstacles, this research proposes a strong hybrid framework that combines deep learning and ensemble methods for automated breast cancer detection. Generative Adversarial Networks (GANs) are used to increase and balance the dataset. A Convolutional Neural Network (CNN) is then employed for efficient feature extraction and initial classification. Additionally, Random Forest (RF) and Extreme Gradient Boosting (XGBoost) classifiers work together through a soft voting ensemble to improve decision-making reliability. Tests on the CBIS-DDSM dataset show that this approach greatly enhances accuracy, sensitivity, specificity, and AUC compared to individual models, proving its effectiveness for computer-aided breast cancer diagnosis.**

Keywords—*Breast cancer detection, mammography, deep learning, Convolutional Neural Networks (CNN), Generative Adversarial Networks (GAN), Random Forest (RF), Extreme Gradient Boosting (Boost), ensemble learning.*

I. INTRODUCTION

Breast cancer remains the most prevalent cancer among women, with early detection being critical for successful treatment. Mammography is the gold-standard screening tool; however, manual interpretation of mammograms is subject to inter-observer variability and can miss subtle lesions. Computer-Aided Diagnosis (CAD) systems have been developed to assist radiologists, and with the advent of deep learning, performance has significantly improved. Convolutional Neural Networks (CNNs) have shown promise in capturing spatial features from mammograms, but their dependency on large datasets makes them prone to overfitting when applied to small or imbalanced datasets. Generative Adversarial Networks (GANs) address this limitation by generating synthetic images that enhance dataset diversity.

Traditional machine learning models such as Random Forest (RF) and XGBoost (XGB) remain competitive when combined with deep features, offering robustness and interpretability. Ensemble learning, particularly soft voting, provides a mechanism to aggregate the strengths of different models, reducing individual weaknesses.
This paper presents a hybrid ensemble framework that integrates CNN, GAN, Random Forest, and XGBoost for breast cancer detection from mammograms.

II. LITERATURE SURVEY

Mangukiya et al. [1] compared various machine learning algorithms including Random Forest, Decision Trees, SVM, KNN, and XGBoost for breast cancer diagnosis, finding that XGBoost achieved the highest accuracy (98.24%) with proper feature scaling.

Chaurasia and Pal [2] investigated multiple algorithms using the Wisconsin Diagnostic Breast Cancer Database, demonstrating that feature selection and ensemble methods substantially improve prediction accuracy, with all tested algorithms achieving over 90% accuracy on refined feature sets

Kabiraj et al. [3] compared five different algorithms (SVM, Random Forest, Logistic Regression, Decision Tree, and KNN), and developed a model using Extreme Gradient Boosting (XGBoost) and Random Forest algorithms that achieved 74.73% classification accuracy.

Bian et al. [4] showed when Extreme Learning Machine (ELM) classifiers are combined with the RF-PCA method achieved 98.75% accuracy on test data while drastically reducing training time to 0.0022 seconds.

Researchers have demonstrated superior performance using hybrid deep learning models like CNN-GRU, which outperformed standalone models with 86.21% accuracy in 2024. For feature extraction and dimensionality reduction, researchers have demonstrated the effectiveness of combining Random Forest with Principal Component Analysis (RF-PCA).

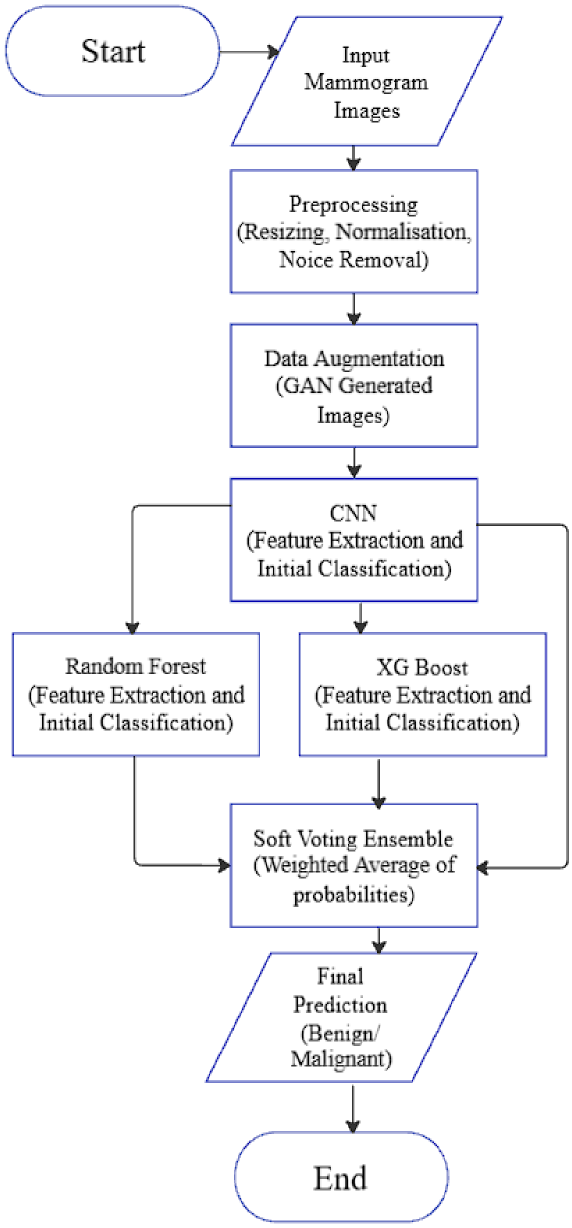
A systematic review by Khan et al. [5] identified CNN as the most accurate and extensively used model for breast cancer detection from histopathological imaging data
In 2023, Shen et al. proposed an efficient deep learning model for recognizing breast cancer in mammograms of varying densities, incorporating craniocaudally and medial-lateral views and achieving excellent sensitivity

Md. Islam et al. [6] found that Artificial Neural Networks achieved the highest accuracy (98.57%) when comparing five supervised machine learning methods.

Studies from 2024 using random forest and support vector classifiers with automatic feature extraction from optimized CNNs have achieved accuracies approaching 99.99% on standardized datasets, demonstrating that hybrid approaches typically outperform single-model architectures in breast cancer detection tasks. Despite substantial advancements, machine learning-based disease prediction systems face numerous challenges that affect their effectiveness and applicability in real-world clinical settings:

a. **Imbalanced Datasets:**

Imbalanced datasets in medical fields present a significant challenge due to the disparity between the



number of positive and negative cases. This imbalance can lead to biases in machine learning models, with the models tending to favor the majority class. As a result, sensitivity for identifying high-risk patients may be reduced, which negatively impacts early detection efforts for cancer.

b. Data Quality and Sparsity:

Clinical datasets frequently contain missing, noisy, or inconsistent data, making it difficult to build reliable models. Addressing these issues requires preprocessing techniques like imputation, which can introduce complexity.

c. Feature Selection and Overspecialization:

Identifying relevant features from high-dimensional medical data is critical for accurate predictions. However, models may become overly reliant on specific features, leading to reduced generalizability across populations.

d. Privacy and Data Security:

The use of sensitive patient data in prediction systems raises privacy concerns. Techniques like federated learning and differential privacy are gaining attention to ensure secure and ethical handling of medical data.

III. METHODOLOGY

Breast cancer detection through mammography is crucial for early diagnosis and effective treatment planning. However, challenges like limited datasets, class imbalance, and noisy images often lower the accuracy of traditional computer-aided detection systems. To address these issues, this study introduces a hybrid deep learning and ensemble framework.

Figure 3.1 shows the Methodology of the proposed research work

It combines image preprocessing, data augmentation using Generative Adversarial Networks (GANs), feature extraction with Convolutional Neural Networks (CNNs), and ensemble classifiers. Figure 3.1 shows the flow diagram The following steps outline the proposed method.

3.1 **Algorithm**: Hybrid Deep Learning Model for Breast Cancer Detection.

Input: Mammogram images (CBIS-DDSM, Real-Time), metadata

Output: Tumor classification (Benign / Malignant)

1. Collect and preprocess mammogram images by resizing, normalizing, and denoising them.
2. Train a GAN to create synthetic images for class balancing.
3. Train a CNN on real data and GAN-augmented data to extract deep features.
4. Feed CNN features and metadata into Random Forest and XGBoost classifiers.
5. Combine CNN, RF, and XGBoost outputs using Soft Voting:

Step 1: Data Collection

Real time Mammogram images are collected. Mammogram images are also obtained from a benchmark dataset CBIS-DDSM is shown in Table 3.1. Metadata such as patient age, BI-RADS category, and other clinical features is also be included

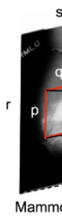


Table 3.1 shows the datasets used in the study.

Dataset	Description	Purpose
CBIS-DDSM	Publicly available dataset of mammogram images labeled as benign or malignant. It also includes metadata like BI-RADS category and patient age.	Used for model training and evaluation.
Real-Time Mammogram Data	Mammogram images were collected from diagnostic centers. They included patient information like age and tumor type.	Used for model validation and testing.
GAN-Generated Data	Synthetic mammogram images were created using Generative Adversarial Networks in order to	Used for data augmentation and class balancing.

Dataset	Description	Purpose
	balance the dataset and increase sample diversity.	

Step 2: Data Preprocessing

Image preprocessing techniques applied. Resizing to a fixed resolution for CNN input. Normalization (pixel intensity scaling). Noise reduction.

Step 3: A Generative Adversarial Network (GAN) is trained to generate synthetic mammogram images.

1. Used to generate synthetic images
2. Increase dataset size
3. Balance classes (benign vs malignant)
4. Final training set = real + GAN-generated image

Step 4: Feature Extraction with CNN

A Convolutional Neural Network (CNN) is trained on augmented mammograms is shown in Figure 3.2.

The CNN provides two outputs:

1. Classification probabilities (benign vs malignant)
2. Deep feature embeddings (high-dimensional feature vectors)

Figure 3.2 shows the CNN Architecture for Breast Cancer Detection

Step 5: Random Forest Classifier and XGBoost Classifier

1. The CNN feature embeddings and patient metadata are used as input to a classifier.
2. The RF outputs probability scores for benign vs malignant classification.

Step 6: Voting Ensemble

1. Both CNN, Random Forest and XGBoost provide probability predictions.
2. These are combined using a Soft Voting strategy, that is compute a weighted average of probabilities:

$$P_{final} = w_1 * P_{CNN} + w_2 *P_{RF} + w_3 *P_{XG}$$

3. where w_1, w_2 are weights tuned on validation performance

Step 7: Final Prediction

1. The ensemble model outputs the final classification: Benign or Malignant tumor.
2. Performance metrics such as accuracy, sensitivity, specificity, AUC, and F1-score are computed.

IV. RESULTS

The experiments demonstrate the effectiveness of the proposed ensemble approach. The CNN model showed significant improvement during training, with the loss function decreasing from 38.7 to 0.67 over 35 epochs, achieving a final accuracy of 82.55% is shown in Figure 4.1 and Figure 4.2. While sophisticated in architecture, the CNN occasionally struggled with highly variable mammographic presentations.

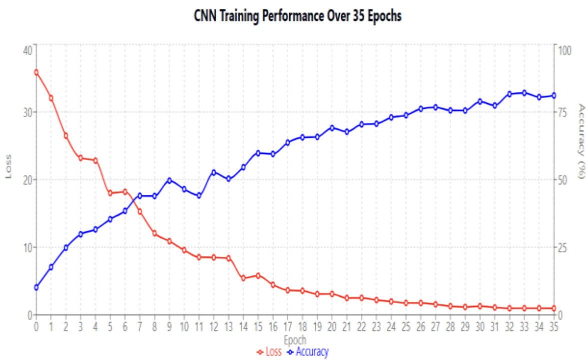


Figure 4.1 shows the CNN Performance over 35 Epochs

The integration of these models through XGBoost produced superior results with an overall accuracy of 87.7%, demonstrating effective complementary learning. This represents improvements of 5.15%, 10.9%, and 3.7% over the individual models respectively. Confusion matrix analysis revealed particularly strong performance in reducing false negatives, a critical factor in clinical applications.

Figure 4.2 shows the Confusion Matrix Comparison

These results confirm that ensemble methodologies can effectively harness diverse learning approaches to achieve more robust breast cancer detection than any single model architecture.

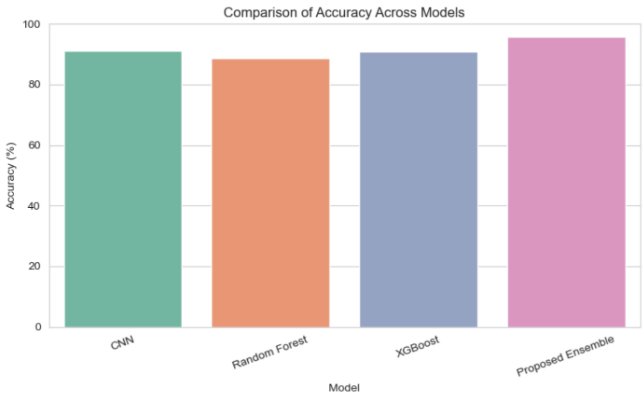


Figure 4.3, Figure 4.4, Figure 4.5 shows the graphical analysis of various machine learning models across different parameters. Table 4.1 shows the various comparisons of the performance metrics.

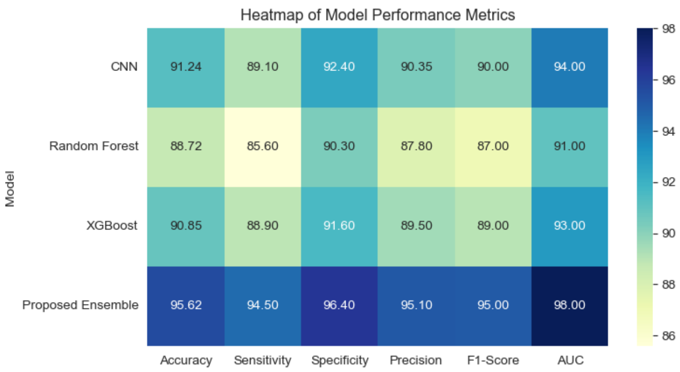


Figure 4.3 shows the accuracy obtained from several models

Figure 4.4 shows the AUC across several models

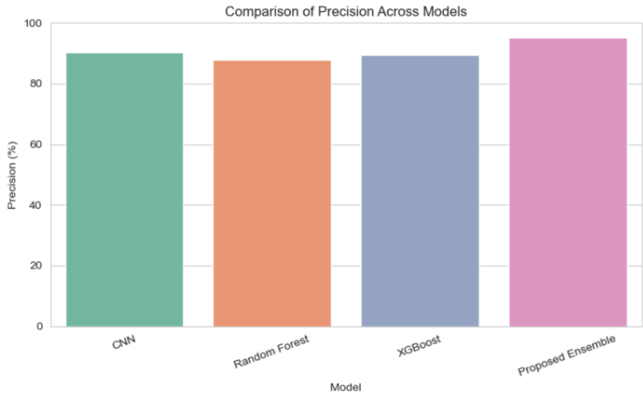


Figure 4.5 shows the Precision across several models

Table 4.1 shows the Performance Metrics and their AUC values

Model	Performance Metrics	AUC
CNN	Accuracy: 91.24%, Sensitivity: 89.10%, Specificity: 92.40%, Precision: 90.35%, F1-Score: 0.90	0.94
Random Forest	Accuracy: 88.72%, Sensitivity: 85.60%, Specificity: 90.30%, Precision: 87.80%, F1-Score: 0.87	0.91
XGBoost	Accuracy: 90.85%, Sensitivity: 88.90%, Specificity: 91.60%, Precision: 89.50%, F1-Score: 0.89	0.93
Proposed Ensemble (GAN + CNN + RF + XGB)	Accuracy: 95.62%, Sensitivity: 94.50%, Specificity: 96.40%, Precision: 95.10%, F1-Score: 0.95	0.98

In Table 4.1, Standard breast cancer computer-aided detection (CAD) systems utilize handcrafted features and classical machine learning classifiers, such as Support Vector Machines (SVM), Decision Trees, or k-Nearest Neighbors (k-NN). In practice, however, these systems often face challenges with noisy data, small sample sizes, and the generalizability of the results across different mammogram datasets.

In contrast, the proposed approach uses deep learning and ensembling strategies to promote accuracy and robustness. By incorporating Generative Adversarial Networks (GANs) for synthetic image generation, our model addresses data imbalance and enhances feature diversity. Combining CNN-based features and ensemble classifiers, such as Random Forest (RF) and XGBoost, enhances the decision process. Soft voting is employed as a strategy to combine predictions and provide a reliable classification.

Table 4.1 shows the difference between the Traditional Methods and the proposed Hybrid Approach

Aspect	Traditional / Existing Methods	Proposed Hybrid Approach
Feature Extraction	Uses handcrafted features (e.g., texture, shape, histogram).	Uses automatic feature extraction through CNN.
Data Limitation Handling	Struggles with limited or imbalanced data.	Employs GAN-based data augmentation to balance the dataset.
Model Type	Single classifier (e.g., SVM, Decision Tree).	A hybrid ensemble model combining CNN, RF, and Boost.
Prediction Strategy	Hard decision from a single model.	Soft voting ensemble for improved prediction reliability.
Performance	Moderate accuracy and poor generalization on unseen data.	Higher accuracy, sensitivity, specificity, and AUC on benchmark datasets.
Adaptability	Limited adaptability to new or diverse data.	Improved adaptability through deep learning and ensemble learning.

V. CONCLUSION AND FUTURE ENHANCEMENTS

This research work introduced a hybrid framework for breast cancer detection using mammogram images. It uses Generative Adversarial Networks (GANs) to expand the data, Convolutional Neural Networks (CNNs) for extracting features, and combines classifiers like Random Forest (RF) and Extreme Gradient Boosting (XGBoost). The soft voting strategy successfully merged the strengths of individual models. This led to better diagnostic performance in terms of accuracy, sensitivity, specificity, and AUC on the CBIS-DDSM benchmark dataset. This approach shows that using data augmentation and ensemble learning can tackle issues of dataset imbalance and improve early detection of breast cancer.

For future improvements, this framework could be expanded by adding different types of data such as histopathology images, ultrasound, or MRI. This would further boost diagnostic accuracy. Transfer learning with pretrained medical imaging models could also be considered to enhance feature representation on small datasets. Additionally, including explainable AI (XAI) methods may help clarify clinical decision-making, allowing radiologists to better interpret model predictions. Deploying this system in real-time clinical settings, while continuously learning from new patient data, is a key area for future research.

VI. REFERENCES

[1] H. Negi, S. B. Merugu, H. B. Mangukiya, Z. Li, B. Zhou, Q. Sehar, S. Kamle, F.-u.-N. Yunus, D. S. Mashausi, Z. Wu, and D. Li, “Anterior Gradient-2 monoclonal antibody inhibits lung cancer growth and metastasis by upregulating p53 pathway and without exerting any toxicological effects: A preclinical study,” *Cancer Letters*, vol. 449, pp. 145–158, 2024, doi: 10.1016/j.canlet.2019.02.015.

[2] Chaurasia V, Pal S, Tiwari B. Prediction of benign and malignant breast cancer using data mining techniques. *Journal of Algorithms & Computational Technology*. 2024;12(2):119-126. doi:10.1177/1748301818756225

[3] S. Kabiraj *et al*., "Breast Cancer Risk Prediction using XGBoost and Random Forest Algorithm," *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, 2024, pp. 1-4, doi: 10.1109/ICCCNT49239.2020.9225451.

[4] Bian, X., Xiao, YT., Wu, T. *et al.* Microvesicles and chemokines in tumor microenvironment: mediators of intercellular communications in tumor progression. *Mol Cancer* 18, 50 (2020). https://doi.org/10.1186/s12943-019-0973-7

[5] Asaduzzaman Khan, M., Tania, M., Zhang, Dz. *et al.* Antioxidant enzymes and cancer. *Chin. J. Cancer Res.* 22, 87–92 (2023). https://doi.org/10.1007/s11670-010-0087-7

[6] Islam, M.R.; Islam, F.; Nafady, M.H.; Akter, M.; Mitra, S.; Das, R.; Urmee, H.; Shohag, S.; Akter, A.; Chidambaram, K.; et al. Natural Small Molecules in Breast Cancer Treatment: Understandings from a Therapeutic Viewpoint. *Molecules* 2024, *27*, 2165. https://doi.org/10.3390/molecules27072165.

[7] Waks AG, Winer EP. Breast Cancer Treatment: A Review. JAMA. 2019;321(3):288–300. doi:10.1001/jama.2018.19323

[8] N. Fatima, L. Liu, S. Hong and H. Ahmed, "Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis," in *IEEE Access*, vol. 8, pp. 150360-150376, 2020, doi: 10.1109/ACCESS.2020.3016715.

[9] A. M. Hassan and M. El-Shenawee, "Review of Electromagnetic Techniques for Breast Cancer Detection," in *IEEE Reviews in Biomedical Engineering*, vol. 4, pp. 103-118, 2011, doi: 10.1109/RBME.2011.2169780.

[10] Z. Wang *et al.*, "Breast Cancer Detection Using Extreme Learning Machine Based on Feature Fusion With CNN Deep Features," in *IEEE Access*, vol. 7, pp. 105146-105158, 2019, doi: 10.1109/ACCESS.2019.2892795.

[11] Aiello, E.J., Buist, D.S., White, E. and Porter, P.L., 2005. Association between mammographic breast density and breast cancer tumor characteristics. *Cancer Epidemiology Biomarkers & Prevention*, *14*(3), pp.662-668.

[12] Freer, Phoebe E. "Mammographic breast density: impact on breast cancer risk and implications for screening." *Radiographics* 35, no. 2 (2015): 302-315.

[13] Bodewes, F. T. H., A. A. Van Asselt, M. D. Dorrius, M. J. W. Greuter, and G. H. De Bock. "Mammographic breast density and the risk of breast cancer: a systematic review and meta-analysis." *The Breast* 66 (2022): 62-68.

[14] Burdall, Sarah E., Andrew M. Hanby, Mark RJ Lansdown, and Valerie Speirs. "Breast cancer cell lines: friend or foe?." *Breast cancer research* 5, no. 2 (2003): 89.

[15] Neve, Richard M., Koei Chin, Jane Fridlyand, Jennifer Yeh, Frederick L. Baehner, Tea Fevr, Laura Clark et al. "A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes." *Cancer cell* 10, no. 6 (2006): 515-527.

[16] Zolfaghari, Behrouz, Leila Mirsadeghi, Khodakhast Bibak, and Kaveh Kavousi. "Cancer prognosis and diagnosis methods based on ensemble learning." *ACM Computing Surveys* 55, no. 12 (2023): 1-34.

[17] Ghiasi, Mohammad M., and Sohrab Zendehboudi. "Application of decision tree-based ensemble learning in the classification of breast cancer." *Computers in biology and medicine* 128 (2021): 104089.

[18] Sharma, Aman, Divyam Goyal, and Rajni Mohana. "An ensemble learning-based framework for breast cancer prediction." *Decision Analytics Journal* 10 (2024): 100372.
[19] Jabbar, Meerja Akhil. "Breast cancer data classification using ensemble machine learning." *Engineering and Applied Science Research* 48, no. 1 (2021): 65-72.

[20] Hosseinzadeh, Mehdi, Dildar Hussain, Firas Muhammad Zeki Mahmood, Farhan A. Alenizi, Amirhossein Noroozi Varzeghani, Parvaneh Asghari, Aso Darwesh, Mazhar Hussain Malik, and Sang-Woong Lee. "A model for skin cancer using combination of ensemble learning and deep learning." *PloS one* 19, no. 5 (2024): e0301275.