

Robust Deep Ensemble Learning for Mammographic Lesion Classification on the INbreast and MIAS Datasets Using Focal Loss and Misclassification-Based Refinement

Kavita P. Shinde

College of Computer Science & Information Technology, Latur, SRTMU, Nanded, India
kavitapatil04@gmail.com (corresponding author)

Uday V. Kulkarni

Department of Computer Science and Engineering, SGGs IE & T, Nanded, Maharashtra, India
uvkulkarni@sngs.ac.in

Received: 24 July 2025 | Revised: 29 August 2025 and 14 September 2025 | Accepted: 24 September 2025

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.13615>

ABSTRACT

To increase patient survival rates, breast cancer must be detected early and, most importantly, accurately. Although mammography analysis using deep learning has advanced, generalizability is still hampered by issues including dataset variability, class imbalance, and inter-class similarity. Using pre-trained DenseNet121, InceptionV3, and EfficientNetB0 models, this study presents a unique and robust ensemble-based deep learning architecture that was independently refined on mammography images from the INbreast and Mammographic Image Analysis Society (MIAS) datasets. The pipeline incorporates Contrast Limited Adaptive Histogram Equalization (CLAHE) for preprocessing, employs focal loss to mitigate class imbalance, and utilizes a two-stage refinement strategy in which misclassified samples are reintroduced for model retraining. This approach achieves significant improvements, reaching up to 99% and 97% classification accuracy on the INbreast and MIAS datasets, respectively. The ensemble model with misclassification refinement demonstrates notable robustness and generalization capability. The same pipeline was applied independently to both datasets to perform cross-dataset validation, confirming its adaptability to diverse mammographic characteristics. Accuracy, precision, recall, F1-score, and misclassification analysis were employed to evaluate the method, making it suitable for real-world clinical applications. The proposed framework holds strong potential for deployment in clinical decision support systems and lays a solid foundation for future research on small and heterogeneous medical datasets.

Keywords- Contrast Limited Adaptive Histogram Equalization (CLAHE); focal loss; ensemble learning

I. INTRODUCTION

Breast cancer is one of the most severe and fatal diseases that affects women worldwide [1]. According to the World Health Organization, mammography screening significantly improves survival rates through early detection. Automatic deep learning-based mammography analysis has shown the potential to improve diagnostic precision and reduce variability between radiologists [2]. However, challenges such as significant class imbalance, domain variability across datasets, and low contrast in mammographic images continue to hinder real-world implementation [3]. To address these issues, recent studies have proposed domain adaptation, attention-based procedures, and ensemble learning strategies [4]. However, only a small percentage of these techniques are combined with retraining procedures on samples that were misidentified to improve generalization. The use of robust loss functions, such

as focal loss [5], in the classification of mammography lesions is also currently lacking in studies, especially when many datasets are used.

This paper addresses these limitations by proposing a robust deep ensemble learning framework that combines three advanced Convolutional Neural Networks (CNNs), namely DenseNet121, InceptionV3, and EfficientNetB0, in a bagging-based ensemble optimized by focal loss. It also incorporates image enhancement using CLAHE and a misclassification refinement stage, where erroneous predictions are reintegrated into the training set to guide further learning. The proposed method is validated using the INbreast and MIAS datasets, which contain low- and high-quality mammography images, respectively.

Deep Learning has transformed medical image processing, especially when it comes to mammogram-based breast cancer

screening. When it comes to identifying and categorizing mammogram abnormalities, conventional machine learning methods have been surpassed by CNNs [6]. Class imbalance, poor generalization across datasets, and the requirement for high interpretability in medical situations are the main issues that still exist.

Many studies used CNNs to classify breast lesions from mammograms. For instance, in [7], a CNN-based CAD system outperformed conventional feature-based methods in mammography classification. In [8], an ensemble of VGG16, DenseNet121, and InceptionV3 successfully distinguished between benign and malignant cases, achieving 90.1% accuracy on the INbreast dataset. Although these methods produce excellent results on individual datasets, domain-specific characteristics usually make it difficult for them to properly generalize to other datasets.

Preprocessing techniques such as CLAHE have been widely used to improve contrast and highlight lesions in mammograms. In [9], CLAHE preprocessing improved feature extraction quality and increased the classification accuracy of CNN-based models. Additionally, image augmentation via albumentations has been employed to improve model generalization, especially in low-data regimes.

In breast cancer databases, which often suffer from class imbalance, malignant samples are underrepresented. Traditional cross-entropy loss does not adequately handle this. The focal loss approach developed in [10] reduces the contribution of well-classified examples and focuses training on cases that are hard to classify. Recent studies that incorporated focal loss into medical image classification have revealed improved results in unbalanced conditions [11].

The use of ensemble learning to improve robustness and generalization has increased. In [12], it was shown that merging multiple deep CNNs with soft voting enhances diagnostic performance. Using a variety of architectures in ensemble settings, such as ResNet, Exception, and DenseNet, has been found to increase accuracy by capturing complementary properties [13]. However, most ensemble approaches are inflexible and cannot adapt to evolving error patterns, such as frequent misclassifications.

Systematic retraining on misclassified samples is infrequently explored, despite its proven usefulness in error correction and model building. Misclassification refinement has been studied in facial recognition and medical imaging to increase model resilience. A very new and untested method for classifying breast cancer is to include a retraining phase based on samples that were mistakenly classified. The proposed method leverages this concept to enhance recollection and progressively reduce classification errors, especially in significant malignant cases. To our knowledge, no prior work combined CLAHE preprocessing, focal loss, ensemble learning, and misclassification-based retraining in a single pipeline for the INbreast and MIAS datasets, despite advances in the aforementioned disciplines. This work aims to bridge this gap and offer a sound, widely applicable solution.

II. MATERIALS AND METHODS

Figure 1 illustrates the proposed ensemble architecture with a detailed layer-wise CNN design for all models, developed on INbreast and cross-validated on MIAS.

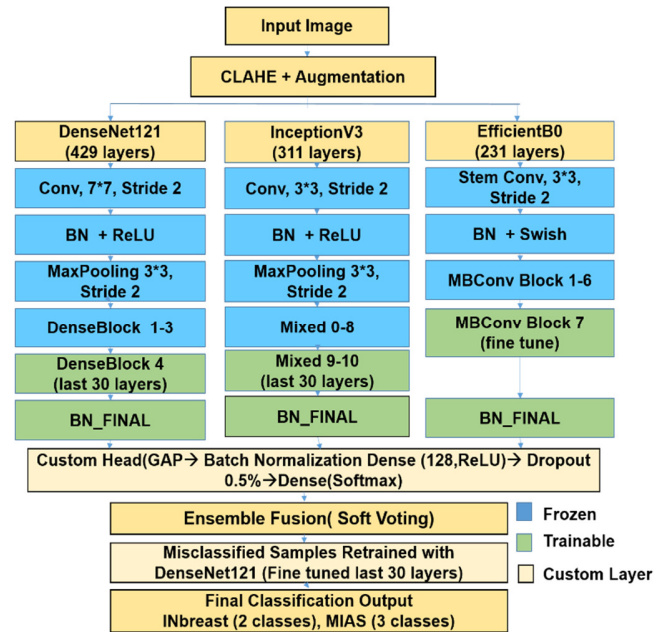


Fig. 1. Proposed ensemble classification pipeline with detailed layer-wise CNN architectures. The same pipeline was trained on INbreast (2-class) and subsequently applied for cross-dataset validation on MIAS (3-class).

A. Dataset Description

This study used two publicly accessible benchmark datasets for breast cancer, MIAS and INbreast. The INbreast dataset, obtained from Kaggle [14], comprises 410 DICOM full-field digital mammograms (CC/MLO) from 115 patients with thorough annotations and BI-RADS scores classified as Normal and Malignant [15]. The MIAS dataset, obtained from the official repository [16], comprises 322 PGM mammograms (1024×1024 pixels) with annotated lesions classified as Normal, Benign, and Malignant [17].

B. Preprocessing

The same preprocessing pipeline was applied to the INbreast and MIAS datasets to ensure consistency and optimize convergence during training. Both datasets' images were converted to JPG format and then transformed to RGB, as these three models accept input in this format. CLAHE was used for contrast enhancement with an 8×8 tile grid size and a clip limit of 3.0, which improved local contrast while preserving edge information. The images were scaled to 224×224 pixels and adjusted to the [0, 1] range. The datasets were divided into training (80%) and validation (20%) sets.

C. Data Augmentation and Class Balancing

Similar augmentation techniques were used on both datasets to address class imbalance and improve model robustness. These included random rotations ($\pm 15^\circ$), width and height shifts (10%), zooming (10%), and horizontal flipping.

These transformations increased data diversity, reduced overfitting, and improved generalization without compromising lesion integrity [18].

D. Model Design and Training

For transfer learning, three pre-trained CNNs were used, namely DenseNet121, InceptionV3, and EfficientNetB0, all trained on the ImageNet dataset.

1) DenseNet121

The feature maps of every prior layer in the same dense block are fed into each layer of DenseNet121. This design mitigates the issue of vanishing gradients and promotes feature reuse. Thus, more compact and effective models can be created without sacrificing performance. Since lesion boundaries and textures are frequently delicate and spatially localized, this hierarchical structure ensures improved feature propagation and retention of low-level spatial information, which is essential in medical imaging tasks [19].

2) InceptionV3

InceptionV3 uses a modular structure made up of parallel convolutional branches to effectively capture multiscale features. The network's initiation modules handle the same input at various receptive field sizes, enabling the model to identify lesions of various shapes and sizes. This architecture works especially well for heterogeneous breast imaging datasets [20].

3) EfficientNetB0

By consistently balancing network depth, width, and resolution, EfficientNetB0 presents a novel scaling strategy that allows the model to reach state-of-the-art accuracy with a greatly reduced number of parameters and computational cost [21]. Recent research has demonstrated its effectiveness in diagnosing breast cancer using a variety of modalities, particularly when paired with transfer learning and data augmentation [22].

For each model, the final classification layers were swapped out for a custom head that included a global average pooling layer, batch normalization, a dense layer with 128 ReLU activated units, a 50% dropout layer, and a final dense output layer with softmax activation (2 units for Inbreast—cancer, normal—and 3 units for MIAS—benign, malignant, normal). For fine-tuning, the final 30 layers were unfrozen to adjust the models to mammographic features [23]. To improve class imbalance, class weights were incorporated into the Adam optimizer with categorical cross-entropy loss throughout the training process for all models. Training efficiency was improved and overfitting was reduced by using a learning rate scheduler and early stopping (patience = 5).

4) Training Setup

The models were trained using a Tesla T4 GPU in Google Colab. In the training setup, a maximum number of 50 epochs and a batch size of 32 were used. If the validation loss did not improve after five epochs, training was terminated early. A factor of 0.5 was used to reduce the learning rate (patience = 3 epochs). During training, model checkpoints were saved based on the optimal validation loss.

E. Ensemble Learning with Majority Soft Voting and Retraining

Ensemble learning has been widely proven to be an effective strategy in medical image classification tasks involving breast cancer diagnosis [24]. By integrating the three basic models (DenseNet121, InceptionV3, and EfficientNetB0) and utilizing soft voting to average their projected class probabilities, an ensemble learning approach was used to improve classification performance [25].

The training set of each dataset was expanded to include validation samples that the ensemble had misclassified to increase model robustness and decrease misclassification. To enhance overall accuracy and generalization, a DenseNet121 model was retrained using this enhanced training data. This allowed the network to learn from examples that were challenging to classify.

Table I lists the essential architectural elements that are common to all of the suggested ensemble models. A uniform classification head consisting of global pooling, normalization, and dense layers follows each backbone (DenseNet121, InceptionV3, and EfficientNetB0) to provide uniformity in feature extraction and classification across all models. Table II shows the parameter distribution and feature vector dimensionality of each backbone model utilized in the ensemble are compiled in. Each model's representational capability and computational depth are reflected in the GAP output size, which ranges from 1024 to 2048 units.

TABLE I. MODEL ARCHITECTURE SUMMARY WITH CUSTOM CLASSIFICATION HEAD

Component	O/P Shape	Parameters	Train	Description
Input Layer	224,224,3	0	No	Post CLAHE and norm.
DenseNet121	7, 7, 1024	7,037,504	Partial	Pretrained on ImageNet
InceptionV3	5, 5, 2048	21,802,784	Partial	Pretrained on ImageNet
Efficient Net B0	7, 7, 1280	5,330,564	Partial	Pretrained on ImageNet
GAP	1024/2048/1280	0	Yes	Converts spatial features into vector
Batch Norm.	1024/2048/1280	4,096	Yes	Improves training stability
Dense (128 units, ReLU)	128	~131K – 262K	Yes	No. varies due to units
Dropout (rate = 0.5)	128	0	Yes	Prevents overfitting
Dense (3 units, Softmax)	3	387- 3,843	Yes	Final classifier for 3 classes

TABLE II. TOTAL, TRAINABLE, AND NON-TRAINABLE PARAMETERS PER MODEL

Model	Total params (M)	Trainable params (M)	Non-trainable params (M)	GAP O/P (Units)
DenseNet121	7,202,187	~1.04	~6.16	1024
InceptionV3	22,070,723	~2.10	~19.9	2048
EfficientNetB0	5,596,947	~1.00	~4.59	1280

F. Evaluation Metrics

Model performance was evaluated using the following metrics [26].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \quad (4)$$

III. RESULTS

A. Results on INbreast Dataset

Table III shows all three models' overall accuracy as well as their class-wise performance metrics. InceptionV3 outperformed DenseNet121 and EfficientNetB0 in terms of accuracy (93.50%) and balanced F1-scores between the two classes. Figure 2 shows that the best validation accuracy was achieved by InceptionV3 (90.74%), closely followed by DenseNet121 (89.81%).

TABLE III. PERFORMANCE METRICS OF INDIVIDUAL MODELS ON INBREAST DATASET

Model	P (N)	R (N)	F (N)	P (C)	R (C)	F (C)	A (%)
DenseNet121	0.82	1.00	0.90	1.00	0.87	0.93	92.00
InceptionV3	0.85	0.98	0.91	0.98	0.90	0.94	93.50
EfficientNetB0	0.60	0.80	0.69	0.85	0.65	0.74	73.00

Abbreviations: P=Precision, R=Recall, F=F1-score, A= Accuracy, N=Normal, C=Cancer.

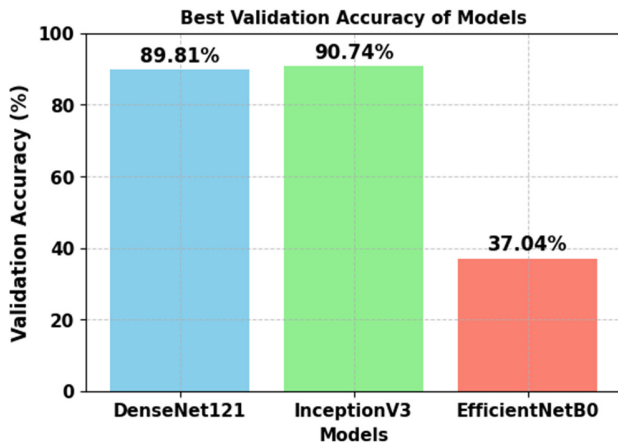


Fig. 2. Validation accuracy of DenseNet121, Inception, and EfficientNetB0 on InBreast.

Table IV shows the ensemble model's performance on the INbreast dataset before refinement. It obtained 92% accuracy, excellent recall in normal situations, and high precision in cancer detection. The weighted average and balanced macro results show robust and reliable classification in both classes.

TABLE IV. CLASSIFICATION REPORT FOR ENSEMBLE MODEL ON INBREAST DATASET BEFORE REFINEMENT

	Precision	Recall	F1-Score	Support
Normal	0.82	1.00	0.90	40
Cancer	1.00	0.87	0.93	68
Accuracy			0.92	108
Macro Avg	0.91	0.93	0.91	108
Weighted Avg	0.93	0.92	0.92	108

Figure 3 shows that 59 cancer and 40 normal instances were accurately classified, and there were zero normal cases misdiagnosed as cancer. Nine cancer cases were misclassified as normal. Table V lists all nine cancer cases that were misclassified as Normal.

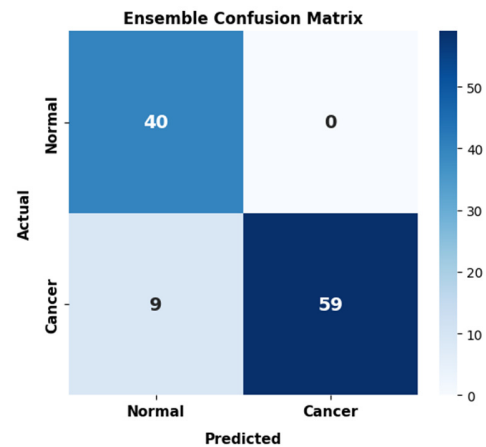


Fig. 3. Ensemble confusion matrix on INbreast dataset.

TABLE V. MISCLASSIFIED CANCER CASES OF THE ENSEMBLE MODEL

Index	True label	Predicted label
1,30,54,60,62,69,78,81,82	Cancer	Normal

Table VI shows that the retrained ensemble model achieved an overall 99% accuracy with macro and weighted averages of 0.99. The model shows balanced performance with 100% precision for cancer, and strong recall and F1 scores for both classes.

TABLE VI. CLASSIFICATION REPORT FOR THE RETRAINED ENSEMBLE MODEL

	Prec.	Recall	F1-score	Support
Normal	0.98	1.00	0.99	40
Cancer	1.00	0.99	0.99	68
Accuracy			0.99	108
Macro Avg.	0.99	0.99	0.99	108
Weighted Avg.	0.99	0.99	0.99	108

Figure 4 shows that the original ensemble model achieved 91.0% accuracy, while the retrained model with misclassified samples achieves a higher accuracy of 99%, illustrating the efficiency of the retraining technique.

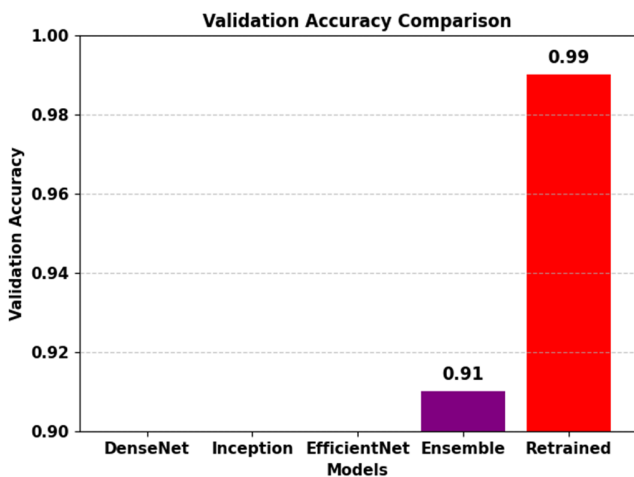


Fig. 4. Validation accuracy of the ensemble and the retrained ensemble model on InBreast.

B. Results on MIAS Dataset

Table VII shows the classification report for the ensemble model on the MIAS dataset. To determine generalizability, the proposed ensemble model was first trained on INbreast and then on the MIAS dataset without change. On INbreast, it achieved 92% accuracy, F1-scores of 0.90 (Normal) and 0.93 (Cancer), and 100% precision for cancer. On the MIAS dataset, the model performed well, with 90% accuracy, including an F1-score of 1.00 for Malignant and solid scores for Benign (0.85) and Normal (0.83). These findings demonstrate the model's robustness across datasets. Figure 5 shows that the model correctly identified 36 benign, 45 malignant, and 31 normal cases. The model misclassified 9 normal cases as benign and 4 benign cases as normal. There were zero malignant cases misclassified.

TABLE VII. CLASSIFICATION REPORT FOR THE ENSEMBLE MODEL ON THE MIAS DATASET

	Prec.	Recall	F1-score	Support
Benign	0.80	0.90	0.85	40
Malignant	1.00	1.00	1.00	45
Normal	0.89	0.78	0.83	40
Accuracy			0.90	125
Macro Avg	0.90	0.89	0.89	125
Weighted Avg	0.90	0.90	0.90	125

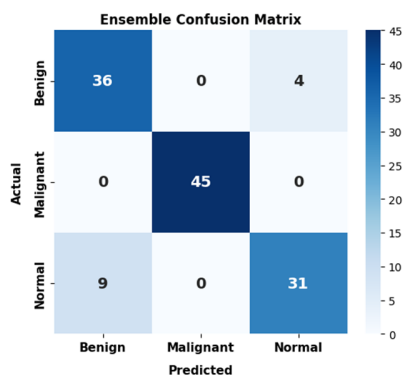


Fig. 5. Ensemble model's confusion matrix on the MIAS dataset.

The model has high sensitivity for malignant diagnosis and overall reliability, with the most misclassifications occurring between the benign and normal classes, most likely due to overlapping visual features in the MIAS dataset. The ensemble model shows outstanding generalization across both the INbreast and MIAS datasets. INbreast achieved perfect specificity (0 false positives) and great sensitivity, incorrectly recognizing only 9 cancer patients as normal. On MIAS, it obtained perfect malignant detection (0 false negatives) and good accuracy for benign and normal classifications, despite modest overlap. These findings demonstrate the pipeline's stability and adaptability to datasets with different quality and class distributions, demonstrating its usefulness in real-world clinical contexts.

Table VIII shows 13 misclassified instances. On the INbreast dataset, 9 malignant patients were misclassified as normal, most likely due to modest radiographic characteristics that reduced sensitivity. In the MIAS dataset, no malignant cases were misclassified, demonstrating great sensitivity. However, 9 normal instances were categorized as benign, 4 benign cases as normal, indicating that visual patterns overlapped between non-malignant classifications. These results highlight the model's strong cancer detection capability.

TABLE VIII. MISCLASSIFICATION REPORT ON THE MIAS DATASET

Index	True label	Predicted label
1,28,51,59,62,69,82,90,95	Normal	Benign
34,36,78,81	Benign	Normal

Table IX shows that the retrained model achieved 97% accuracy on the three-class MIAS dataset, with the malignant lesions having 100% precision, recall, and F1-score. In comparison, the model had 99% accuracy on the InBreast dataset. The slightly lower accuracy on the MIAS dataset reflects its greater complexity when compared to INbreast, but consistently high F1-scores show strong and generalizable performance across both datasets.

TABLE IX. CLASSIFICATION REPORT FOR THE RETRAINED MODEL ON THE MIAS DATASET

	Precision	Recall	F1-Score	Support
Benign	0.93	0.97	0.95	40
Malignant	1.00	1.00	1.00	45
Normal	0.97	0.93	0.95	40
Accuracy			0.97	0.97
Macro Avg.	0.97	0.97	0.97	0.97
Weighted Avg.	0.97	0.97	0.97	0.97

Figure 6 shows that the model accurately detected 39 benign, 45 malignant, and 37 normal instances. Only one benign instance was misclassified as normal, whereas three normal cases were classified as benign. In comparison, the model performed even better than the INbreast dataset, correctly diagnosing 84 of the 85 cancer cases. Overall, the model performs well on both datasets.

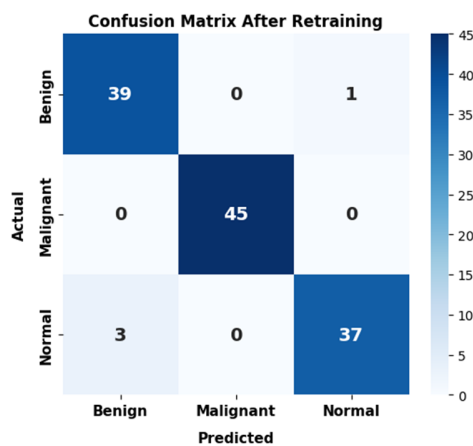


Fig. 6. Confusion matrix after retraining on misclassified samples on the MIAS dataset.

IV. COMPARATIVE ANALYSIS OF RESULTS

Table X contrasts ensemble deep learning techniques on the INbreast and MIAS datasets. The proposed ensemble with misclassification-based retraining achieved 99.0% and 97.0% accuracy when trained on INbreast and deployed to MIAS without modifications, in contrast to previous works assessed on a single dataset. In contrast, the studies in [27-29] reported accuracies of 99.94%, 97.6%, and 95.1% on individual datasets. These findings show the cross-dataset generalizability and robustness of the proposed model.

TABLE X. COMPARATIVE EVALUATION OF ENSEMBLE DEEP LEARNING MODELS ON THE INBREAST AND MIAS DATASETS

Study (Year)	Backbones	Ensemble Strategy	Datasets	Acc. (%)
[27] (2024)	SE-ResNet152, MobileNetV2, VGG19, ResNet18, InceptionV3, DenseNet121	Hard voting over TL models	INbreast	99.94
[28] (2022)	ResNet101, ResNet152	Majority voting	INbreast	97.6
[29] (2023)	VGG16, VGG19	Soft voting	MIAS	95.1
[30] (2020)	FCRN (single DL model, non-ensemble)	—	MIAS	94.7
Proposed model	DenseNet121, InceptionV3, EfficientNetB0	Soft voting retraining on misclassified samples	INbreast and MIAS	99.0/97.0

V. CONCLUSION

In order to classify breast cancer accurately and early, this study used mammograms from the INbreast and MIAS datasets to develop a strong ensemble-based deep learning framework. The ensemble model successfully used transfer learning to overcome the limited availability of data by optimizing pre-trained DenseNet121, InceptionV3, and EfficientNetB0 models, applying the same pipeline first on the INbreast dataset and subsequently on the MIAS dataset.

CLAHE for contrast enhancement, focal loss to reduce class imbalance, and a two-stage misclassification refinement strategy greatly improved the model's robustness and classification accuracy. The proposed ensemble model with misclassification refinement achieved excellent performance with 99% accuracy on the INbreast dataset and 97% accuracy on the MIAS dataset. Furthermore, cross-dataset validation confirmed the generalizability of the proposed framework across heterogeneous image distributions. The model's potential for practical clinical use was demonstrated through evaluation using a variety of performance metrics.

This study provides a solid basis for further developments in deep learning-based breast cancer detection, especially when considering small, imbalanced, and heterogeneous medical datasets. Although the proposed model performs well, there is still a chance of improvement. Using larger and more diverse datasets from several hospitals can improve the model's performance for more patients. To further minimize false negatives and resolve diagnostic ambiguity, future research may further investigate more complex loss functions, such as dynamic task adaptive loss or focal Tversky loss.

REFERENCES

- [1] H. Sung *et al.*, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021, <https://doi.org/10.3322/caac.21660>.
- [2] X. Jing *et al.*, "Automated Breast Density Assessment in MRI Using Deep Learning and Radiomics: Strategies for Reducing Inter-Observer Variability," *Journal of Magnetic Resonance Imaging*, vol. 60, no. 1, pp. 80–91, 2024, <https://doi.org/10.1002/jmri.29058>.
- [3] A. F. A. Alshamrani and F. Saleh Zuhair Alshomrani, "Optimizing Breast Cancer Mammogram Classification Through a Dual Approach: A Deep Learning Framework Combining ResNet50, SMOTE, and Fully Connected Layers for Balanced and Imbalanced Data," *IEEE Access*, vol. 13, pp. 4815–4826, 2025, <https://doi.org/10.1109/ACCESS.2024.3524633>.
- [4] F. Azour and A. Boukerche, "An Efficient Transfer and Ensemble Learning Based Computer Aided Breast Abnormality Diagnosis System," *IEEE Access*, vol. 11, pp. 21199–21209, 2023, <https://doi.org/10.1109/ACCESS.2022.3192857>.
- [5] S. DemiRel, A. Urfali, Ö. F. Bozkir, A. ÇeliKten, A. Budak, and H. Karataş, "Improving Mass Detection in Mammography Using Focal Loss Based RetinaNet," *Turkish Journal of Forecasting*, vol. 07, no. 1, pp. 1–9, Oct. 2023, <https://doi.org/10.34110/forecasting.1326245>.
- [6] Y. Gao, J. Lin, Y. Zhou, and R. Lin, "The application of traditional machine learning and deep learning techniques in mammography: a review," *Frontiers in Oncology*, vol. 13, Aug. 2023, Art. no. 1213045, <https://doi.org/10.3389/fonc.2023.1213045>.
- [7] M. A. Al-antari, M. A. Al-masni, and T. S. Kim, "Deep Learning Computer-Aided Diagnosis for Breast Lesion in Digital Mammogram," in *Deep Learning in Medical Image Analysis*, vol. 1213, G. Lee and H. Fujita, Eds. Springer International Publishing, 2020, pp. 59–72.
- [8] Y. Khourdifi, A. El Alami, M. Zaydi, Y. Maleh, and O. Er-Remyly, "Early Breast Cancer Detection Based on Deep Learning: An Ensemble Approach Applied to Mammograms," *BioMedInformatics*, vol. 4, no. 4, pp. 2338–2373, Dec. 2024, <https://doi.org/10.3390/biomedinformatics4040127>.
- [9] H. Avci and J. Karakaya, "A Novel Medical Image Enhancement Algorithm for Breast Cancer Detection on Mammography Images Using Machine Learning," *Diagnostics*, vol. 13, no. 3, Jan. 2023, Art. no. 348, <https://doi.org/10.3390/diagnostics13030348>.
- [10] Z. Liu, J. Peng, X. Guo, S. Chen, and L. Liu, "Breast cancer classification method based on improved VGG16 using mammography

- images," *Journal of Radiation Research and Applied Sciences*, vol. 17, no. 2, Jun. 2024, Art. no. 100885, <https://doi.org/10.1016/j.jrras.2024.100885>.
- [11] P. Aliniya, M. Nicolescu, M. Nicolescu, and G. Bebis, "Improved Loss Function for Mass Segmentation in Mammography Images Using Density and Mass Size," *Journal of Imaging*, vol. 10, no. 1, Jan. 2024, Art. no. 20, <https://doi.org/10.3390/jimaging10010020>.
- [12] M. M. Dhas and N. S. Singh, "Breast Cancer Diagnosis Using Majority Voting Ensemble Classifier Approach," *Journal of Soft Computing and Data Mining*, vol. 5, no. 1, Jun. 2024, <https://doi.org/10.30880/jscdm.2024.05.01.013>.
- [13] M. Khaled, F. Touazi, and D. Gaceb, "Improving Breast Cancer Diagnosis in Mammograms with Progressive Transfer Learning and Ensemble Deep Learning," *Arabian Journal for Science and Engineering*, vol. 50, no. 10, pp. 7697–7720, May 2025, <https://doi.org/10.1007/s13369-024-09428-1>.
- [14] "INbreast Dataset." Kaggle, [Online]. Available: <https://www.kaggle.com/datasets/ramanathansp20/inbreast-dataset>.
- [15] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep Learning to Improve Breast Cancer Detection on Screening Mammography," *Scientific Reports*, vol. 9, no. 1, Aug. 2019, Art. no. 12495, <https://doi.org/10.1038/s41598-019-48995-4>.
- [16] J. Suckling, "The mammographic images analysis society digital mammogram database," in *Excerpta Medica. International Congress Series, 1994*, vol. 1069, pp. 375–378.
- [17] S. Aymaz, "A new framework for early diagnosis of breast cancer using mammography images," *Neural Computing and Applications*, vol. 36, no. 4, pp. 1665–1680, Feb. 2024, <https://doi.org/10.1007/s00521-023-09156-x>.
- [18] P. Oza, P. Sharma, S. Patel, F. Adedoyin, and A. Bruno, "Image Augmentation Techniques for Mammogram Analysis," *Journal of Imaging*, vol. 8, no. 5, May 2022, Art. no. 141, <https://doi.org/10.3390/jimaging8050141>.
- [19] R. K. Pattanaik, S. Mishra, M. Siddique, T. Gopikrishna, and S. Satapathy, "Breast Cancer Classification from Mammogram Images Using Extreme Learning Machine-Based DenseNet121 Model," *Journal of Sensors*, vol. 2022, pp. 1–12, Dec. 2022, <https://doi.org/10.1155/2022/2731364>.
- [20] W. M. Salama and M. H. Aly, "Deep learning in mammography images segmentation and classification: Automated CNN approach," *Alexandria Engineering Journal*, vol. 60, no. 5, pp. 4701–4709, Oct. 2021, <https://doi.org/10.1016/j.aej.2021.03.048>.
- [21] S. Hu, M. Zhou, S. Jin, J. Peng, and Y. Wang, "A Comparative Study of Different Deep Learning Models Over Diverse Mammography Datasets," in *2024 IEEE International Conference on Medical Artificial Intelligence (MedAI)*, Chongqing, China, Nov. 2024, pp. 168–173, <https://doi.org/10.1109/MedAI62885.2024.00029>.
- [22] A. Darakh, A. Shah, and P. Oza, "Exploring the Benefits of Data Augmentation for Breast Cancer Classification using Transfer Learning," in *Information Systems for Intelligent Systems*, vol. 379, C. S. In, N. D. Londhe, N. Bhatt, and M. Kitsing, Eds. Springer Nature Singapore, 2024, pp. 509–520.
- [23] A. Saber, A. G. Hussien, W. A. Awad, A. Mahmoud, and A. Allakany, "Adapting the pre-trained convolutional neural networks to improve the anomaly detection and classification in mammographic images," *Scientific Reports*, vol. 13, no. 1, Sep. 2023, Art. no. 14877, <https://doi.org/10.1038/s41598-023-41633-0>.
- [24] E. Justaniah, G. Aldabbagh, A. Alhothali, and N. Abourokbah, "Classifying Breast Density from Mammogram with Pretrained CNNs and Weighted Average Ensembles," *Applied Sciences*, vol. 12, no. 11, May 2022, Art. no. 5599, <https://doi.org/10.3390/app12115599>.
- [25] J. J. Puspo, "A Novel Approach to Classify Breast Cancer using Transfer Learning," in *2024 27th International Conference on Computer and Information Technology (ICCIT)*, Cox's Bazar, Bangladesh, Dec. 2024, pp. 3354–3359, <https://doi.org/10.1109/ICCIT64611.2024.11021716>.
- [26] M. S. Darweesh *et al.*, "Early breast cancer diagnostics based on hierarchical machine learning classification for mammography images," *Cogent Engineering*, vol. 8, no. 1, Jan. 2021, Art. no. 1968324, <https://doi.org/10.1080/23311916.2021.1968324>.
- [27] M. T. Ahad, S. Mustofa, F. Ahmed, Y. R. Emon, and A. D. Anu, "A study on Deep Convolutional Neural Networks, Transfer Learning and Ensemble Model for Breast Cancer Detection." arXiv, 2024, <https://doi.org/10.48550/ARXIV.2409.06699>.
- [28] N. A. Samee, G. Attea, S. Meshoul, M. A. Al-antari, and Y. M. Kadah, "Deep Learning Cascaded Feature Selection Framework for Breast Cancer Classification: Hybrid CNN with Univariate-Based Approach," *Mathematics*, vol. 10, no. 19, Oct. 2022, Art. no. 3631, <https://doi.org/10.3390/math10193631>.
- [29] S. Khan, F. T. Zahra, A. Asif, S. Arif, and U. Fatima, "Comparative Study of Transfer Learning with VGG16 and VGG19 for Breast Cancer Classification," in *2024 26th International Multi-Topic Conference (INMIC)*, Karachi, Pakistan, Dec. 2024, pp. 1–6, <https://doi.org/10.1109/INMIC64792.2024.11004305>.
- [30] D. Saraswathi and E. Srinivasan, "Mammogram Analysis using League Championship Algorithm Optimized Ensembled FCRN Classifier," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 5, no. 2, pp. 451–461, Feb. 2017, <https://doi.org/10.11591/ijeecs.v5.i2.pp451-461>.