International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Breast Cancer Detection Using Mammography: From Image Processing to Deep Learning

Mr. M. Venkatarathnam[1], K. Naga Harika[2], R. Vyshnavi[3], B. Vamsi[4], P. Sunil Reddy[5], S. Sreenivasulu[6]

[1]Associate Professor, Dept of ECE, PBR VITS, Kavali, Andhra Pradesh, India.

[23456] Students of Department of Electronics and Communication Engineering, PBR Visvodaya Institute of Technology & Science, Kavali (Autonomous), SPSR Nellore (Dt.), Andhra Pradesh – 524201, India

***

Abstract - This paper presents an advanced methodology for breast cancer detection and stage classification from mammography images by integrating image processing, deep feature extraction, and ensemble learning techniques. Initially, raw mammograms undergo preprocessing to enhance diagnostic quality. Images are converted to grayscale to reduce computational complexity, followed by Contrast Limited Adaptive Histogram Equalization (CLAHE) to improve local contrast and highlight subtle tissue abnormalities. The enhanced images are resized to a uniform dimension and converted back to RGB format to ensure compatibility with deep learning architectures. Deep features are extracted using a pre-trained ResNet50 network, which effectively captures high-level spatial, structural, and texture information essential for differentiating normal and abnormal breast tissues. The extracted features are then classified using multiple base classifiers, including Decision Tree, Naïve Bayes, k-Nearest Neighbors (k-NN), and Support Vector Machine (SVM). To enhance robustness and minimize individual classifier bias, a meta-learner ensemble strategy integrates the base model predictions for final classification into Benign or Malignant categories. Furthermore, malignant cases are analyzed using a Composite CNN model to determine the cancer stage, enabling more detailed clinical assessment. The proposed framework is evaluated using accuracy, precision, recall, F1-score, and confusion matrix metrics, demonstrating reliable and effective performance for computer-aided breast cancer diagnosis.

Keywords: Automated breast cancer detection, Mammography preprocessing, Deep feature extraction, Meta-learning classification, Ensemble diagnostic framework.

***

1. INTRODUCTION

Breast cancer remains a leading cause of mortality among women worldwide, highlighting the need for accurate and early detection. Although mammography is the most widely used screening method, its manual interpretation is challenging due to low contrast, dense tissues, and subtle abnormalities. To address these limitations, this study proposes a deep learning–based framework for automated breast cancer detection. The system incorporates image preprocessing techniques such as grayscale conversion, CLAHE-based contrast enhancement, and resizing to improve image quality. Deep features are extracted using the pre-trained ResNet-50 model, and multiple machine learning classifiers—including Decision Tree, Naive Bayes, k-Nearest Neighbours, and Support Vector Machine—are applied for classification. A meta-learning approach using an XGBoost-based model combines the outputs of these classifiers to enhance prediction accuracy and robustness. The proposed hybrid framework provides reliable, real-time predictions and demonstrates strong potential as a clinical decision support tool for early breast cancer diagnosis.

2. LITERATURE SURVEY

Previous studies emphasize the vital role of mammography in early breast cancer detection while also highlighting significant diagnostic challenges. The World Health Organization recommends population-based screening programs, noting their effectiveness in reducing mortality, but also identifies limitations such as false positives, overdiagnosis, and disparities in healthcare access. Epidemiological research by Jacques Ferlay demonstrates notable regional variations in cancer incidence and mortality, largely influenced by screening availability, awareness, and treatment quality. Broader global frameworks like the United Nations Sustainable Development Goals further support improvements in healthcare systems, early diagnosis, and technological innovation. In clinical practice, the evaluation of microcalcifications plays a crucial role in early detection, with studies by Stephen A. Feig and Muttarak. Muttarak identifying specific morphological and distribution patterns that help distinguish benign from malignant lesions. Additionally, research by Daniel P. Winchester provides important insights into the diagnosis and management of ductal carcinoma in situ (DCIS), a key precursor to invasive breast cancer. Collectively, these works highlight the importance of accurate imaging interpretation, standardized screening practices, and the integration of advanced computational techniques to improve diagnostic accuracy and patient outcomes.

3. EXISTING SYSTEM

Existing methods for breast cancer detection primarily rely on traditional machine learning and standalone deep learning approaches. Earlier techniques used manual feature extraction methods, where texture, shape, and intensity-based features were derived from mammography images using image processing techniques, followed by classifiers such as Support Vector Machine (SVM), k-Nearest Neighbours (k-NN), and Decision Trees. However, these methods are highly dependent on handcrafted features and often fail to capture complex patterns in medical images. With the advancement of deep learning, Convolutional Neural Networks (CNNs) have been widely adopted to automatically learn hierarchical features directly from images, improving detection accuracy. Popular architectures such as ResNet, VGG, and DenseNet have shown promising results in mammogram classification. Despite these improvements, existing deep learning models still face challenges such as high computational cost, requirement of large annotated datasets, lack of interpretability, and reduced performance in cases of dense breast tissues or subtle abnormalities. Additionally, most approaches rely on a single model, which may limit generalization capability and robustness, highlighting the need for hybrid and ensemble-based frameworks.

Limitations: Existing methods have several limitations, including reliance on binary classification (benign vs. malignant) without stage-level details, which limits clinical usefulness. They may also struggle with subtle patterns in dense tissues and often depend on large datasets and high computational resources. Additionally, lack of interpretability and reliance on single models can reduce accuracy and generalization in real-world applications.

4. PROPOSED SYSTEM

The proposed methodology presents a comprehensive hybrid framework that integrates image preprocessing, deep learning, and ensemble-based classification for accurate breast cancer detection from mammography images. Initially, input images undergo preprocessing steps such as grayscale

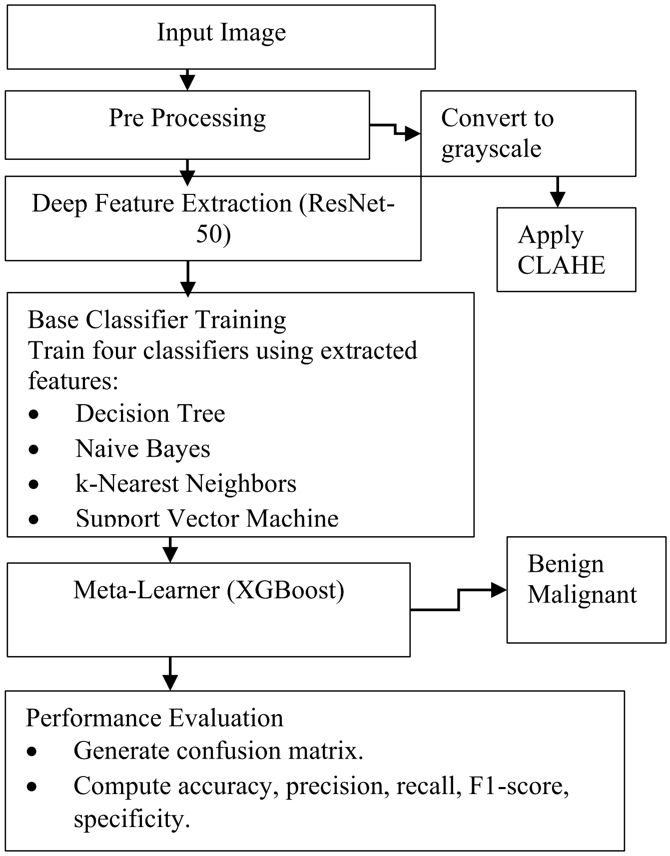

Fig1: Block Diagram of Proposed Method

conversion, Contrast Limited Adaptive Histogram Equalization (CLAHE), resizing to 224×224 pixels, and RGB channel adaptation to enhance contrast, reduce noise, and ensure uniform input representation. These steps improve the visibility of subtle features such as microcalcifications and tumor boundaries. Deep feature extraction is then performed using the pre-trained ResNet-50 Convolutional Neural Network (CNN), which leverages transfer learning to capture rich spatial, structural, and textural patterns from mammograms. The extracted high-dimensional feature vectors are subsequently classified using multiple machine learning algorithms, including Decision Tree, Naive Bayes, k-Nearest Neighbors (k-NN), and Support Vector Machine (SVM), each contributing diverse and complementary decision perspectives. The prediction outputs from these base classifiers are fused to form meta-features, which are then processed using an ensemble-based meta-learner

(XGBoost/TreeBagger) to enhance classification accuracy, robustness, and generalization while reducing individual model bias. The system ultimately classifies images as benign or malignant, and for malignant cases, a Composite CNN model is further employed to determine cancer stages, providing more detailed diagnostic insights. The overall performance of the system is evaluated using standard metrics such as accuracy, precision, recall, specificity, F1-score, and confusion matrix, demonstrating its effectiveness, reliability, and suitability for real-time clinical decision support in breast cancer diagnosis.

5. IMPLEMENTATION AND RESULTS

Fig. 2 shows the illustrates the original mammography image used as input to the proposed breast cancer detection system. Mammography employs low-dose X-rays to visualize internal breast structures, enabling early identification of abnormalities such as tumours, masses, and microcalcifications. Variations in grayscale intensity represent differences in tissue density, where dense tissues appear brighter and fatty tissues appear darker, while the black region corresponds to the background outside the breast area. However, raw mammograms may contain noise and low contrast, which can hinder accurate analysis. Therefore, preprocessing steps such as cropping, contrast enhancement, and normalization are applied to improve image quality before performing feature extraction and classification.

Fig.2: Input Image

Fig.3 illustrates the preprocessing stages applied to the mammography image prior to analysis. The original grayscale image is first taken as input, followed by contrast normalization to enhance tissue visibility and highlight important structures. The image is then resized to 224 × 224 pixels to meet the input requirements of the deep learning model. Finally, it is converted into RGB format to ensure compatibility with the convolutional neural network for effective feature extraction and classification.

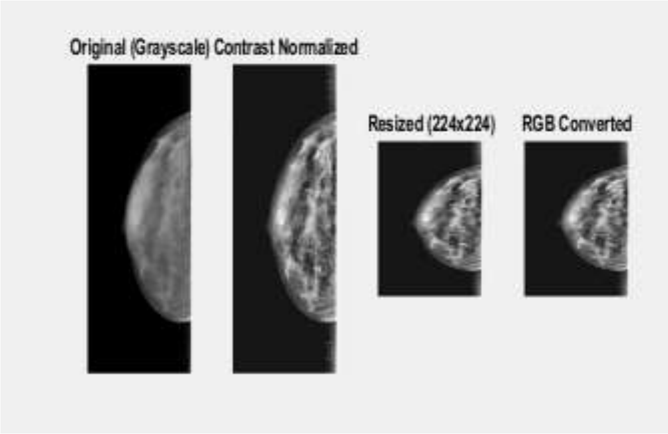

Fig3: Pre-Processing Image

Fig. 4 shows the final classification result produced by the proposed breast cancer detection system. After preprocessing and feature extraction using the deep learning model, the classifier analyzes the extracted features to determine the nature of the abnormality. The system displays the prediction "Benign," indicating a non-cancerous condition. A graphical user interface (GUI) presents the result in a simple dialog box for user interpretation.

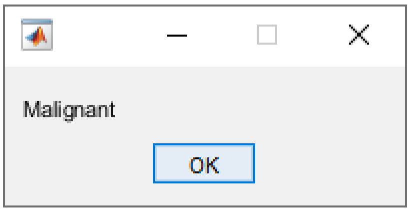

Fig4: Classification Result

Fig. 5 shows the confusion matrix of the proposed breast cancer classification model on the test dataset, comparing actual and predicted class labels. The model correctly classified 10 benign and 8 malignant cases, while 2 malignant cases were misclassified as benign. This matrix is used to evaluate the model's performance in terms of prediction accuracy.

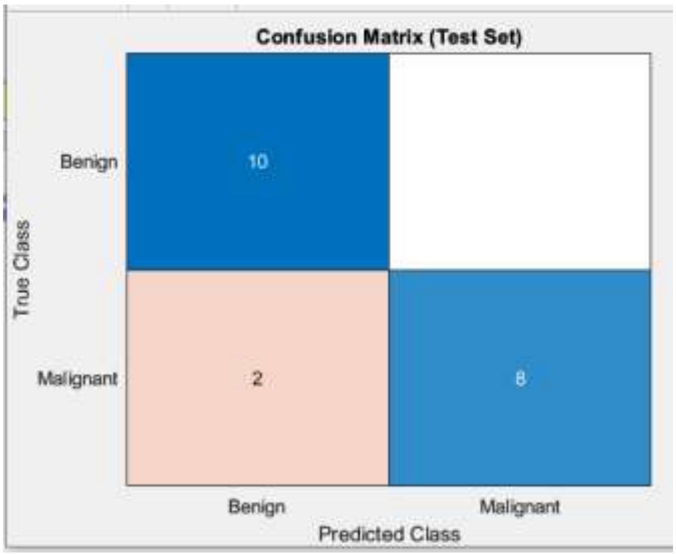

Fig5: Confusion Matrix

Fig. 6 presents the performance metrics of the proposed breast cancer classification model. The system achieved an accuracy of 90.00%, precision of 91.67%, recall (sensitivity) of 90.00%, F1-score of 90.83%, and specificity of 90.00%, demonstrating the reliability and effectiveness of the proposed system.

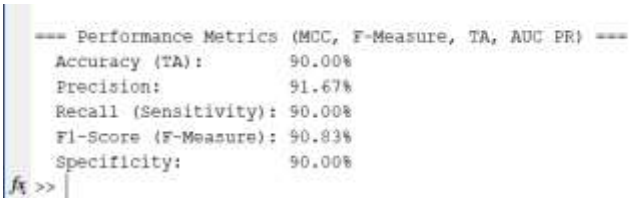

Fig6: Matrices

Fig. 7 shows the classification result where the trained model predicts the input sample as Stage 1, indicating an early stage of breast cancer.

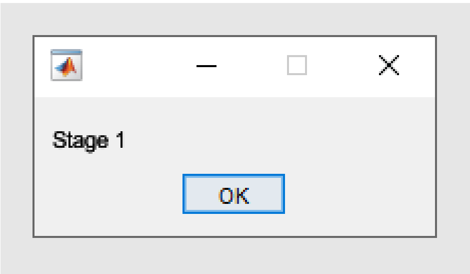

Fig7: Stage Classification

6. DISCUSSION

The proposed breast cancer detection system can be further improved by incorporating larger and more diverse datasets from different medical institutions, which would enhance accuracy, robustness, and generalization in real-world applications. The use of advanced deep learning architectures, including improved convolutional neural networks and hybrid models, can further strengthen feature extraction and classification performance, enabling better detection of subtle abnormalities.Enhancing preprocessing techniques such as image enhancement, noise reduction, and segmentation can improve image quality and highlight suspicious regions more clearly. Integration with real-time clinical systems can assist radiologists in faster diagnosis, while user-friendly interfaces can support easier interaction and interpretation of results. Additionally, combining multiple imaging modalities and adopting cloud-based platforms can improve accessibility, reliability, and overall effectiveness of the system, especially in resource-limited settings.

7. CONCLUSION

In this work, a hybrid breast cancer detection system based on mammography images has been proposed, integrating image preprocessing, deep learning, and ensemble machine learning techniques. The use of ResNet-50 for feature extraction combined with multiple classifiers and a meta-learner improves classification accuracy and robustness. Preprocessing methods enhance image quality, enabling better detection of subtle abnormalities. Additionally, the system provides both classification and stage-level prediction, supporting effective clinical decision-making. Overall, the proposed approach demonstrates reliable performance and has the potential to assist radiologists in early diagnosis, reducing errors and improving patient outcomes.

ACKNOWLEDGEMENT

The authors sincerely thank Mr. M. Venkatarathnam (Associate Professor, ECE, PBR VITS Kavali) for his guidance, Dr. R. Sravanthi (Professor & HoD, ECE) for providing facilities, and Dr. V. Anil Kumar (Principal, PBR VITS Kavali) for the academic environment that enabled this work.

REFERENCES

[1] WHO Position Paper on Mammography Screening. World Health Org., Geneva, Switzerland, 2014.

[2] J. J. Fenton et al., "Influence of computer-aided detection on performance of screening mammography," *New England Journal of Medicine*, vol. 356, no. 14, pp. 1399–1409, 2007.

[3] J. Wang et al., "Discrimination of breast cancer with microcalcifications on mammography by deep learning," *Scientific Reports*, vol. 6, 2016.

[4] V. K. Singh et al., "Breast tumor segmentation and classification using CNN and GAN," *Expert Systems with Applications*, 2020.

[5] D. Ribli et al., "Detecting and classifying lesions in mammograms with deep learning," *Scientific Reports*, 2018.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, 2015.

[7] A. Hamidinekoo et al., "Deep learning in mammography and breast histology: Overview and future trends," *Medical Image Analysis*, 2018.

[8] J. R. Burt et al., "Deep learning in breast cancer diagnosis," *British Journal of Radiology*, 2018.

[9] T. Kooi et al., "Large scale deep learning for computer-aided detection of mammographic lesions," *Medical Image Analysis*, 2017.

[10] R. Agarwal et al., "Automatic mass detection in mammograms using deep CNN," *Journal of Medical Imaging*, 2019.

[11] T. R. Mahesh et al., "Breast cancer classification using blended ensemble learning," *International Journal of System Assurance Engineering*, 2024.

[12] S. Zhou et al., "Breast cancer prediction using multiple machine learning algorithms," *Technology in Cancer Research & Treatment*, 2024.

[13] T. Islam et al., "Breast cancer classification using machine learning with explainable AI," *Scientific Reports*, 2024.

[14] S. Ünalan et al., "Comparative study on breast cancer classification using ensemble ML," *Journal of Radiation Research and Applied Sciences*, 2024.

[15] D. Abdelhafiz et al., "Deep CNNs for mammography: Advances and challenges," *BMC Bioinformatics*, 2019.

[16] S. J. S. Gardezi et al., "Breast cancer detection using mammographic data: Systematic review," *Journal of Medical Internet Research*, 2019.

[17] K. J. Geras et al., "AI for mammography: Current concepts and future perspectives," *Radiology*, 2019.

[18] S. E. Hickman et al., "Machine learning in screening mammography," *Radiology*, 2022.

[19] M. Jiang et al., "Computer-aided diagnosis of mammographic masses," *IEEE Transactions on Biomedical Engineering*, 2015.