International Journal of Engineering Applied Science and Management
ISSN (Online): 2582-6948
Vol. 7 Issue 9, September 2026

Machine Learning Algorithm for Mammography-Based Breast Cancer Detection

Ankit Mehto [1], Pritaj Yadav [2]
[1,2] Department of Computer Science and Engineering, RNTU, Bhopal, 464993, India
[1] ankitmehto461@gmail.com, [2] yadavpritaj@gmail.com

Abstract

Breast cancer remains one of the most prevalent malignancies affecting women worldwide, and early detection through mammographic screening continues to be the most effective strategy for reducing mortality. Manual interpretation of mammograms, however, is time-consuming, subject to inter-observer variability, and constrained by radiologist availability, particularly in resource-limited settings. This paper presents a machine learning-based framework for automated detection and classification of breast abnormalities (masses, calcifications, and architectural distortions) from digital mammographic images. The proposed pipeline integrates image pre-processing (noise removal, contrast-limited adaptive histogram equalisation, and breast-region segmentation), handcrafted and deep feature extraction using a convolutional neural network (CNN) backbone, and a classification stage that distinguishes benign from malignant findings. The framework is evaluated conceptually against benchmark mammography datasets and compared with traditional machine learning classifiers such as Support Vector Machines (SVM), Random Forest, and k-Nearest Neighbours (k-NN). Results reported in the reviewed literature indicate that deep learning-based approaches, particularly transfer-learning models built on architectures such as VGG16, ResNet50, and DenseNet, consistently outperform handcrafted-feature classifiers in accuracy, sensitivity, and specificity. The paper concludes with a discussion of clinical translation challenges, dataset limitations, and directions for future research, including explainable AI and multi-modal fusion with ultrasound and MRI data.

Keywords: Breast cancer detection, mammography, machine learning, deep learning, convolutional neural network, computer-aided diagnosis (CAD), image classification

1. Introduction

Breast cancer is the most commonly diagnosed cancer among women globally and a leading cause of cancer-related mortality. According to global cancer statistics, breast cancer accounts for a substantial proportion of new cancer cases and deaths in women each year, with incidence rates continuing to rise in both developed and developing countries. Early detection significantly improves prognosis and survival rates, making screening programmes centred on mammography a cornerstone of public health policy in many countries.

Mammography is a low-dose X-ray imaging technique used to visualise the internal structure of the breast and detect abnormalities such as masses, microcalcifications, asymmetries, and architectural distortions that may indicate malignancy. Despite its widespread use, mammogram interpretation is inherently challenging: abnormalities can be subtle, overlapping with dense fibro-glandular tissue, and radiologists' diagnostic accuracy varies with experience, fatigue, and workload. Studies have shown that a meaningful percentage of cancers are missed on initial screening (false negatives), while a considerable number of biopsies performed on suspicious findings ultimately prove benign (false positives), leading to patient anxiety and unnecessary healthcare costs.

Computer-Aided Diagnosis (CAD) systems have therefore been an active area of research for over two decades, evolving from rule-based and handcrafted-feature systems to modern deep learning architectures. The advent of large-scale annotated mammography datasets (such as the Digital Database for Screening Mammography, INbreast, and the Curated Breast Imaging Subset of DDSM) combined with advances in convolutional neural networks has enabled a new generation of automated detection systems that approach, and in some studies match, expert radiologist performance.

This paper reviews and synthesises the machine learning approaches applied to mammography-based breast cancer detection and proposes a structured methodological framework. The specific objectives of this study are:

- To review the evolution of machine learning techniques applied to mammographic image analysis, from classical classifiers to deep convolutional architectures.
- To propose an end-to-end pipeline covering pre-processing, segmentation, feature extraction, and classification for benign–malignant discrimination.

- To compare the relative strengths and limitations of traditional machine learning and deep learning approaches based on evidence from the literature.
- To identify open challenges and propose future research directions for clinical deployment of ML-based mammography screening tools.

2. Literature Review

2.1 Traditional Machine Learning Approaches

Early CAD systems for mammography relied on handcrafted feature extraction techniques such as the Grey-Level Co-occurrence Matrix (GLCM) for texture analysis, Histogram of Oriented Gradients (HOG) for shape description, and wavelet transforms for multi-resolution analysis. These features were subsequently fed into classical classifiers including Support Vector Machines, Decision Trees, Random Forest, k-Nearest Neighbours, and Artificial Neural Networks. While these approaches demonstrated reasonable performance on curated datasets, they were highly dependent on the quality of feature engineering and often generalised poorly to unseen data due to variability in imaging equipment, patient anatomy, and tissue density.

2.2 Deep Learning and Convolutional Neural Networks

The introduction of deep learning shifted the paradigm from manual feature engineering to automatic hierarchical feature learning. Convolutional Neural Networks (CNNs) can learn low-level features (edges, textures) in early layers and progressively more abstract, task-specific representations (mass shapes, calcification clusters) in deeper layers. Architectures such as AlexNet, VGG16/VGG19, ResNet, InceptionNet, and DenseNet have been adapted for mammography classification, typically through transfer learning, where a network pre-trained on large natural-image datasets (e.g., ImageNet) is fine-tuned on mammographic images to compensate for the relatively limited size of medical imaging datasets.

Object detection frameworks, including Faster R-CNN, YOLO variants, and RetinaNet, have also been employed to localise suspicious regions within a full mammogram prior to classification, effectively combining detection and diagnosis into a single pipeline. More recent work has explored attention mechanisms and Vision Transformers (ViTs) to capture long-range spatial dependencies across the breast tissue that convolutional filters may miss.

2.3 Hybrid and Ensemble Approaches

To leverage the complementary strengths of different models, several studies propose ensemble or hybrid architectures that combine multiple CNN backbones, or fuse deep features with handcrafted radiomic features before final classification. Ensemble strategies such as majority voting, stacking, and weighted averaging have been shown to improve robustness and reduce the variance associated with any single model.

2.4 Datasets Commonly Used in the Field

Publicly available datasets play a central role in benchmarking mammography classification algorithms. Commonly used datasets include the Mammographic Image Analysis Society (MIAS) database, the Digital Database for Screening Mammography (DDSM) and its curated version CBIS-DDSM, and INbreast, which provides high-resolution full-field digital mammograms with detailed radiologist annotations. Each dataset differs in imaging modality (screen-film versus full-field digital mammography), annotation granularity, and class balance, which affects cross-study comparability.

3. Proposed Methodology

The proposed framework follows a five-stage pipeline: data acquisition, pre-processing, segmentation, feature extraction and classification, and performance evaluation. Figure 1 (conceptual) summarises this pipeline.

3.1 Data Acquisition

Mammographic images are sourced from publicly available, ethically approved repositories (e.g., CBIS-DDSM, INbreast, MIAS), each labelled as normal, benign, or malignant based on radiologist assessment and, where available, biopsy-confirmed ground truth.

3.2 Pre-processing

- Noise reduction using median filtering or non-local means denoising to suppress acquisition artefacts.
- Contrast-Limited Adaptive Histogram Equalisation (CLAHE) to enhance the visibility of dense tissue and subtle abnormalities.
- Breast-region segmentation and pectoral muscle removal using thresholding and morphological operations to isolate the region of interest.
- Image resizing and normalisation to match the input dimensions required by the chosen CNN backbone (e.g., 224×224 pixels).

3.3 Feature Extraction and Model Architecture

The core classification module is built around a convolutional neural network. Two configurations are proposed for comparison: (i) a custom CNN trained from scratch with four convolutional blocks, batch normalisation, and dropout regularisation; and (ii) a transfer-learning configuration using a pre-trained backbone (e.g., ResNet50 or DenseNet121) with the final fully connected layers replaced and fine-tuned on the mammography dataset. Data augmentation techniques, including rotation, horizontal flipping, zoom, and elastic deformation, are applied during training to mitigate overfitting given the limited size of medical imaging datasets.

3.4 Classification and Training

The final classification layer uses a softmax (multi-class) or sigmoid (binary) activation to output the probability of benign versus malignant classification. The categorical cross-entropy loss function is minimised using the Adam optimiser with a learning rate scheduler. Class-imbalance is addressed through weighted loss functions or synthetic minority oversampling (SMOTE) applied to extracted feature vectors.

3.5 Evaluation Metrics

Model performance is assessed using accuracy, sensitivity (recall), specificity, precision, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Given the clinical context, sensitivity (the ability to correctly identify malignant cases) is prioritised, since false negatives carry a substantially higher clinical cost than false positives.

4. Results and Discussion

Table 1 summarises representative performance figures reported across multiple published studies employing different machine learning and deep learning approaches on mammography datasets. These values are drawn from patterns consistently observed in the reviewed literature and are presented here for comparative illustration rather than as results of a single controlled experiment; researchers implementing this framework should substitute their own experimentally derived figures.

Table 1. Representative performance ranges reported in the literature for different modelling approaches on benchmark mammography datasets.

Model / Approach	**Typical Accuracy (%)**	**Typical Sensitivity (%)**	**Typical AUC**
SVM (handcrafted features)	82 – 88	78 – 85	0.83 – 0.88
Random Forest	80 – 86	76 – 83	0.81 – 0.87
Custom CNN (trained from scratch)	88 – 92	85 – 90	0.90 – 0.93
Transfer Learning (VGG16 / ResNet50)	92 – 96	90 – 95	0.94 – 0.97
Ensemble / Hybrid Deep Models	94 – 97	92 – 96	0.95 – 0.98

The comparative pattern indicates that deep transfer-learning models consistently outperform classical machine learning classifiers built on handcrafted features. This can be attributed to the ability of pre-trained CNNs to leverage generalised low-level visual representations learned from large natural-image corpora, which transfer effectively to texture- and shape-based patterns relevant to mammographic abnormalities. Ensemble approaches provide a further, generally modest, improvement by reducing variance and compensating for individual model weaknesses, at the cost of increased computational complexity and reduced interpretability.

Despite these promising results, several factors constrain direct clinical translation. First, most published models are trained and validated on curated, relatively homogeneous datasets that may not reflect the full diversity of imaging equipment, patient demographics, and breast density encountered in routine screening. Second, deep learning models are often criticised as 'black boxes', which limits clinician trust; this has motivated growing interest in explainable AI techniques such as Gradientweighted Class Activation Mapping (Grad-CAM) to visualise the image regions driving a model's prediction. Third, class imbalance, since malignant cases are typically far less frequent than benign or normal findings, continues to bias models toward the majority class unless explicitly addressed.

5. Challenges and Future Directions

- Dataset diversity: Larger, multi-institutional, and demographically diverse datasets are needed to improve model generalisability across populations and imaging equipment.
- Explainability: Integration of interpretable AI techniques (Grad-CAM, SHAP, attention visualisation) to build clinician trust and support regulatory approval.
- Multi-modal fusion: Combining mammography with ultrasound, MRI, and clinical/genetic risk factors to improve diagnostic accuracy, particularly in dense breast tissue where mammography sensitivity is reduced.
- Federated learning: Enabling collaborative model training across hospitals without sharing sensitive patient data, addressing privacy and data-governance concerns.
- Prospective clinical validation: Moving beyond retrospective dataset benchmarking toward prospective trials embedded in real screening workflows, alongside regulatory and ethical review.
- Bias and fairness auditing: Systematic evaluation of model performance across age groups, ethnicities, and breast density categories to prevent inequitable diagnostic performance.

6. Conclusion

Machine learning, and deep learning in particular, has demonstrated substantial potential to augment mammography-based breast cancer screening by improving detection consistency and reducing radiologist workload. This paper has outlined a structured pipeline encompassing pre-processing, segmentation, feature extraction, and classification, and reviewed comparative evidence indicating the superiority of transfer-learning-based CNN architectures over traditional handcrafted-feature classifiers. Nevertheless, clinical adoption requires addressing dataset diversity, model interpretability, and rigorous prospective validation. Future work integrating explainable AI, multi-modal imaging, and federated learning frameworks is likely to accelerate the safe and equitable deployment of ML-based computer-aided diagnosis systems in breast cancer screening programmes worldwide.

References

[1] American Cancer Society. (2024). Breast cancer facts & figures. American Cancer Society.

[2] Arevalo, J., González, F. A., Ramos-Pollán, R., Oliveira, J. L., & Guevara Lopez, M. A. (2016). Representation learning for mammography mass lesion classification using convolutional neural networks. Computer Methods and Programs in Biomedicine, 127, 248–257.

[3] Heath, M., Bowyer, K., Kopans, D., Moore, R., & Kegelmeyer, W. P. (2001). The Digital Database for Screening Mammography. In Proceedings of the Fifth International Workshop on Digital Mammography.

[4] Lotter, W., Diab, A. R., Haslam, B., et al. (2021). Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. Nature Medicine, 27, 244–249.

[5] McKinney, S. M., Sieniek, M., Godbole, V., et al. (2020). International evaluation of an AI system for breast cancer screening. Nature, 577, 89–94.

[6] Moreira, I. C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M. J., & Cardoso, J. S. (2012). INbreast: Toward a full-field digital mammographic database. Academic Radiology, 19(2), 236–248.

[7] Harsh Pratap Singh, Sandeep Govindrao Panchal, Nagendra Singh, Sachin Sahu, Shantilal Bhayal, Urmila Mahor “A Deep Learning Framework for Multi-Disease Diagnosis using Medical Imaging and Clinical Data”, 2nd International Conference on Big Data & Machine Learning (ICBDML), 2026

[8] Mehto, A., & Yadav, P. (2026). Machine learning algorithm for mammography-based breast cancer detection. ShodhKosh: Journal of Visual and Performing Arts.

[9] Bhayal, Shantilal, Lokendra Singh Songare, and Harsh Pratap Singh. "Multi-Modal Deep Learning Framework for Chest Radiology: Integrating Images and Textual Reports for

Enhanced Multi-Label Disease Classification." 2026 2nd International Conference on Big Data & Machine Learning (ICBDML). IEEE, 2026.

[10] Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. (2019). Deep learning to improve breast cancer detection on screening mammography. Scientific Reports, 9, 12495.

[11] Sung, H., Ferlay, J., Siegel, R. L., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 71(3), 209–249.

[12] Wu, N., Phang, J., Park, J., et al. (2020). Deep neural networks improve radiologists' performance in breast cancer screening. IEEE Transactions on Medical Imaging, 39(4), 1184–1194.