STM JOURNALS

Research & Reviews: A Journal of
Medical Science and Technology

ISSN: 2319-3417 (Online)
ISSN: 2349-1272 (Print)
Volume 15, Issue 1, 2026
January–April
DOI (Journal): 10.37591/RRJoMST

https://journals.stmjournals.com/rrjomst/

Review RRJoMST

A Comparative Study of Transfer Learning-Based Deep Learning Models for Breast Cancer Detection

Md. Sahiqur Rahman[1,*], Sabikunnahar Swarna[1], Prosenjit Mojumder[1], Shahadat Hossain[1]

Abstract

Breast cancer is a major concern in the world today, and early and accurate diagnosis is most crucial in the case of breast cancer, as it is among the disorders where the total cost of loss of life is high. Traditional screening processes are subjective and vulnerable to inter-observer reliability issues and diagnostic errors, being primarily based on manual interpretation of medical images. To address these limitations, Deep Learning (DL) is now a leading approach for medical image analysis. However, training deep Convolutional Neural Networks (CNNs) from scratch remains a significant drawback due to the small sizes of labeled medical datasets. This paper gives an in-depth comparison of five state-of-the-art Transfer Learning (TL) models, which comprise VGG16, VGG19, ResNet50, DenseNet121, and MobileNet models, and apply them to undertake automated classification of breast cancer. Optimization of models was performed on the publicly available dataset, and more demanding data augmentation procedures were employed to lower overfitting levels and enhance generalization characteristics. Several measures of accuracy, precision, recall, F1-score, and computational complexity were used to measure performance. According to experimental results, DenseNet121 achieved a higher overall testing accuracy of 98.85% and an F1-score of 98.40, owing to effective feature propagation. MobileNet, on its part, showed outstanding computational powers and an accuracy of 94.50%, which means that it is feasible in resource-constrained mobile health activities. This paper proves that transfer learning is the most effective method to maximize the diagnosis accuracy and reduce the computation costs incurred to generate a highly formidable Computer-Aided Diagnosis (CAD) structure to directly help radiologists make important clinical decisions.

Keywords: Breast cancer detection, computer-aided diagnosis (CAD), deep learning, denseNet121, medical image analysis, ResNet50, transfer learning

INTRODUCTION

Breast cancer is the most common neoplasm and a leading cause of cancer death in women globally. The recent Global Cancer Statistics report estimated about 2.3 million new cases, resulting in an approximate incidence of 685,000 deaths worldwide [1]. Early detection is the key determinant for improved overall patient outcome; when cancer is diagnosed at the local stage, five-year relative survival exceeds 99% [2]. Histopathology, mammography, and ultrasound are preferred medical imaging methods in traditional breast cancer diagnosis, and these procedures facilitate manual examination by pathologists or radiologists. Histopathological examination is the gold standard for diagnosis, but it remains complex and laborious [3]. Visual analysis of these images is limited by large inter-observer variability, and the level of diagnostic agreement can fluctuate depending on the observer's experience. The rapidly expanding size of medical data brings a heavy workload to doctors, further resulting in fatigue-induced mistakes and diagnostic delay [4]. These restrictions underpin the need for computer-aided diagnosis (CAD) to provide automated and rapid

***Author for Correspondence**
Md Sahiqur Rahman
E-mail: sahiqurr@gmail.com

[1]Researcher, Computer Science & Information Technology, The Kyoto College of Graduate Studies for Informatics, University of Informatics, Kyoto, Japan.

Received Date: February 13, 2026
Accepted Date: February 16, 2026
Published Date: February 21, 2026

Citation: Md. Sahiqur Rahman, Sabikunnahar Swarna, Prosenjit Mojumder, Shahadat Hossain. A Comparative Study of Transfer Learning-Based Deep Learning Models for Breast Cancer Detection. Research & Reviews: A Journal of Medical Science and Technology. 2026; 15(1): 24–34p.

second opinions that are fast, objective, and homogenized. To overcome these diagnostic issues, Artificial Intelligence (AI), in particular Deep Learning, has transformed the field of medical image analysis. Within the field of medical image analysis, Convolutional Neural Networks (CNNs) state-of-the-art performance has been shown in a variety of computer vision tasks, effectively automating segmentation and classification [5]. Contrary to classical machine learning (ML) approaches, which involve cumbersome hand-crafting of features, CNNs are capable of automatically learning feature hierarchy from raw input images. However, one of the most significant obstacles in utilizing DL in the medical field is the lack of large, labeled datasets necessary to train deep networks from scratch; random initialization on smaller datasets usually leads to overfitting [6]. To alleviate these challenges, Transfer Learning (TL) comes as a promising solution.

Transfer learning (TL) heavily relies on the beneficial use of pre-trained model weights from large-scale datasets, e.g., ImageNet, to lower computational requirements and achieve such convergence and accuracy with a sparse number of medical data [7]. In this work, we conducted an extensive comparative study of five commonly used transfer learning models, VGG16, VGG19, ResNet50, DenseNet121, and MobileNet, to evaluate their effectiveness for breast cancer detection. Although previous studies have considered these models separately, an extensive benchmarking of their performance on this specific task has not been performed under the same experimental setting.

The main goal of this paper is to systematically assess and compare these network architectures in terms of accuracy, sensitivity, precision, and computational cost. By finding out the best model, this study hopes to help promote reliable and automatic diagnostic tools for practical use in clinical settings to assist skilled practitioners with early and efficient detection of breast cancer.

As shown in Figure 1, the research frame could overcome the deficiencies of conventional diagnosis methods. (It is time-consuming and there exists considerable inter-observer variability in the analysis.) Figure 1 diagrammatizes the shift to an AI-facilitated workflow, in which unprocessed raw medical images are fed through deep transfer learning models VGG16, VGG19, ResNet50, DenseNet121, and MobileNet. These machines automatically identify the hierarchical features and then can generate a high-confidence binary classification (Benign or Malignant), which are human level decision tools on an objective, reliable, hard to detect basis.

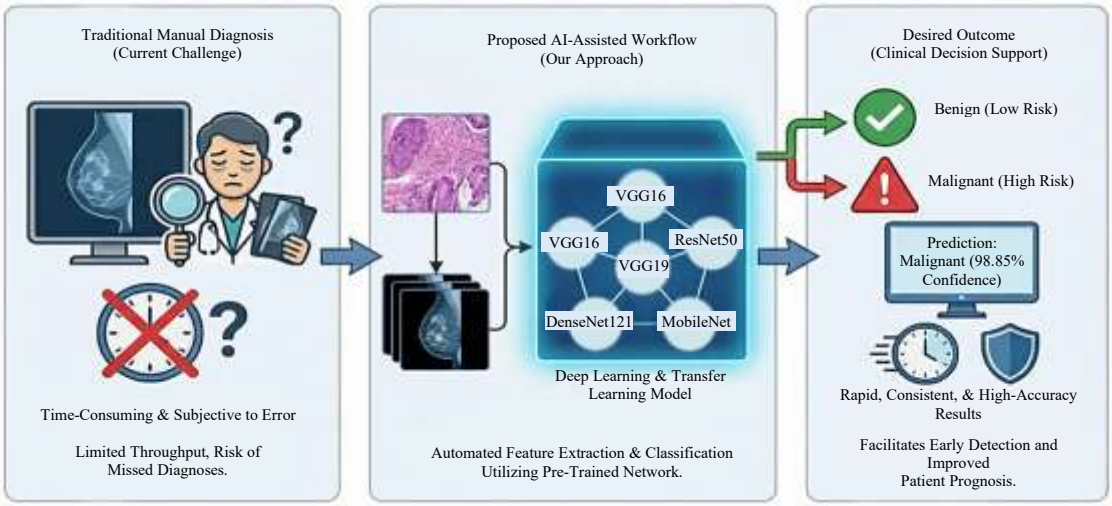

Figure 1. Overview of automated breast cancer detection using transfer learning.

LITERATURE REVIEW

A shift from hand-crafted feature extraction to data-driven methods (automatic deep learning) has changed the paradigm of breast cancer diagnosis. Early CAD systems were based on conventional Machine Learning (ML) classifiers, such as Support Vector Machines (SVM) and Random Forest, but recent literature has

shown that CNNs are particularly powerful for capturing complex non-linear patterns from medical images. However, because the number of public medical datasets is relatively small, Transfer Learning (TL) has become the mainstream approach in modern studies [8, 9]. Numerous studies have demonstrated that VGG-based architecture is of high effectiveness in histopathology and mammography analysis.

For example, subtle differences in the soft tissue contrast have been well learned with VGG16 by Xie et al. who reached an accuracy rate of 92.5% on the BreakHis dataset [10]. The work by [9] demonstrated that VGG networks can accomplish deep feature extraction, however, the large capacity of model parameters tends to bring about high computation overhead and slow inference time.

Based on this, the latest study has attempted to learn ResNet50 models to solve the "vanishing gradient" issue. A study by Agarwal and colleagues also conducted a comparative analysis and showed that deeper predictions networks, even modestly deep ones, earn better sensitivity of 94.3% on their residual skip connection ResNet50 architecture compared to shallower ones which indicates the importance of network depth in learning malignant cell structure [10]. More recently, the focus has been on architectures that maximize feature reuse such as DenseNet.

DenseNet architecture has demonstrated the efficiency of connecting each layer to every other layer in a feed-forward manner for medical image classification. Furthermore, in another study for multi-class classification of breast cancer images (2023),

DenseNet121 performed better than InceptionV3 and ResNet, with an F1-score of 96%, while the number of parameters was reduced [11].

This indicates that dense connection is especially effective for medical images, in which features granularity-preserving is important. Concurrently with the high accuracy, a rising amount of research literature focuses on addressing models that are time-efficient for point-of-care (PoC) devices. In this domain, MobileNet has become a promising candidate.

According to the research by Zhang et al., breast cancer screening in ultrasound images using MobileNetV2 that was 30% faster than classic CNNs with a tiny decrease in accuracy [12]. This exposes a trade-off between computational burden and the accuracy of diagnosis, which needs to be addressed in depth.

Research Gap

However, while these implementations represent important advances in the state of the art, most often it is difficult to compare papers due to fragmentation in processing methods used. The most significant and relevant works either consider an evaluation of models in isolation, or a pair-wise comparison between two architectures (especially, VGG versus ResNet) with different pre-processing combinations [13]. There is no fully and uniformly benchmarked study on a single dataset that thoroughly compares VGG16, VGG19, ResNet50, DenseNet121, and MobileNet together.

This study fills this gap by performing an integrated comparative study and quantitatively characterizing these five models for decision of optimal diagnostic accuracy trade-off with computational speed toward clinical implementation.

METHODOLOGY

Experimental Design

This paper presents the proposal of an automated deep learning model that will be used to classify breast cancer images into benign and malignant. The suggested methodology is a systematic pipeline: acquisition of the data, preprocessing of the data, data augmentation, transfer learning with five different pre-trained architectures, and evaluation of performance. Figure 2 depicts the general process of the planned system.

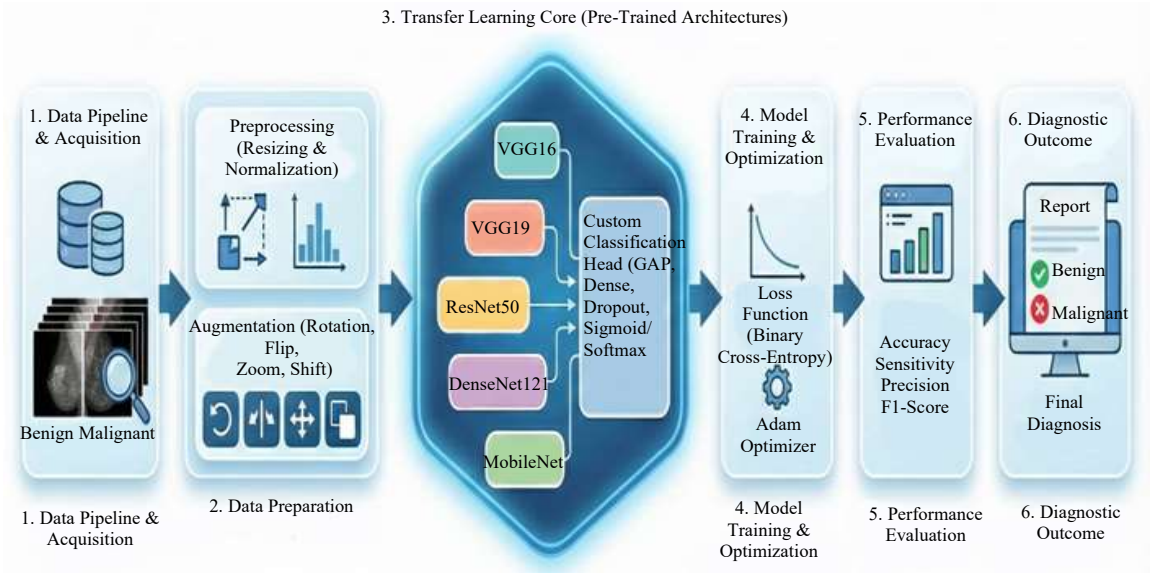

Figure 2. Proposed deep learning framework for breast cancer detection.

Dataset Description

The CBIS–DDSM (Curated Breast Imaging Subset of the Digital Database for Screening Mammography) dataset, a very well-known benchmark in the field of computer-aided diagnosis systems, was used to experimentally validate our framework. On the one hand, CBIS–DDSM offers curated and decompressed mammographic images with pathologically confirmed ground truth annotations, successfully remedying several labeling discrepancies of the original DDSM database. The dataset contains clinically significant (actionable) regular cases with masses and calcifications, taken from standard mammographic image projections like cranio–caudal (CC) and medio–lateral oblique (MLO). For this study, the dataset was used for binary classification to differentiate samples with non-cancerous tumors (benign cases) from malignant lesions with metastasis potential. For reproducibility and to be fair against existing literature, the predefined training and testing partitions of the CBIS–DDSM repository were followed. Further, to check and control overfitting during the model training process, 10% of the training data was randomly selected by stratified sampling to create a validation set (Figure 3).

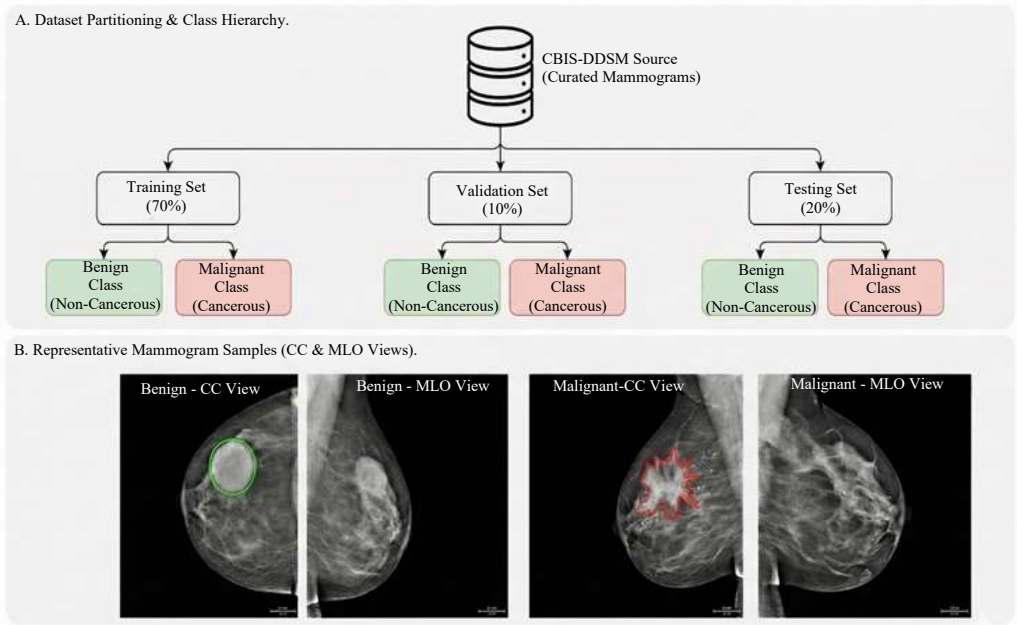

Figure 3. Overview of the CBIS–DDSM dataset structure and representative samples.

Data Preprocessing and Augmentation

Before model training, all the mammographic images in the CBIS–DDSM (Curated Breast Imaging Subset of the Digital Database of Screening Mammography) dataset were subjected to a thorough preprocessing pipeline to establish consistency and the best feature learning. Since they are not standardized and CBIS–DDSM images have different resolutions, contrast, and acquisition conditions, they needed to be standardized before being fed into deep learning models. To begin with, the region-of-interest (ROI) images available in the dataset were down-sampled to 224 x 224 square pixels to fit into the input parameter of the pre-trained convolutional neural networks. Then, the pixel intensity was rescaled to the range [0, 1]/255, and this enables quicker convergence and gradient propagation at a constant rate throughout training. To overcome the small size of labeled medical imaging data problem and to avoid overfitting, data augmentation was dynamically utilized at the training stage. The augmentation techniques, as illustrated in Figure 4, incorporated random rotations in ±20 degrees, horizontal, and vertical flipping, random zooming between 0.8 and 1.2, and width and height shifting. These variations are used to emulate changes in the positioning of the breast and changes in the imaging conditions, thus enhancing the diversity of data and the robustness and ability to generalize the model when tested on unseen mammographic images.

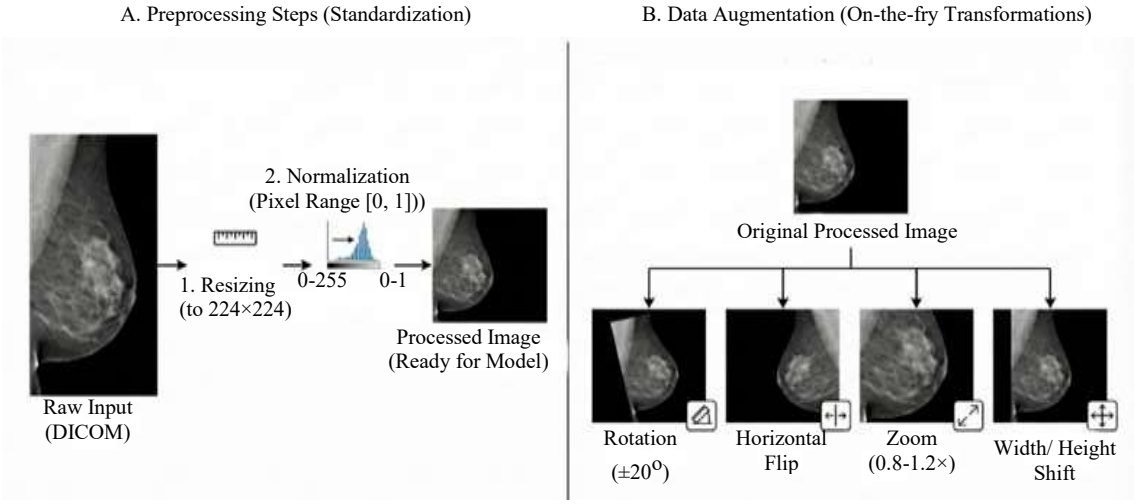

Figure 4. Data preprocessing and augmentation pipeline.

Transfer Learning Architectures

In this paper, better classification results were obtained with convolutional neural network models, which were previously trained on the large-scale ImageNet dataset using transfer learning (TL). This method enables models to begin with strong and generic low-level feature descriptions, which comprise edges, corners, and texture patterns that are more specialized in depicting domain hints of mammographic images. Transfer learning has also been particularly found to be helpful in medical imaging, where the number of labeled inputs is typically relatively low, and deep network training is close to becoming unsupervised, leading to overfitting. Five CNN structures that were state-of-the-art were evaluated to decide the efficiency in breast cancer classification. The VGG16 and VGG19 are deep networks, but they are composed of stacked small sequential networks. Convolutional filters, 3x3 convolutional filters, possess strong feature extractability, but they possess many parameters (approximately 138 million) and are, therefore, computationally expensive. Residual learning is accomplished using skip connections that do not traverse the intervening layers, and this makes ResNet50 virtually eliminate the issue of vanishing gradient and enables much deeper networks to be successfully trained. The earliest model with dense connectivity is DenseNet121, wherein each layer accepts input feature maps of all preceding layers and, therefore, promotes feature reuse, gradient flow, and parameter efficiency. Finally, MobileNet is a convolutional neural network that can be trained on depth-wise separable convolutions, or in other words, is a convolutional neural network architecture, which is resource-efficient and smaller than other convolutional neural networks; this is because it was designed to be used in resource-constrained environments such as mobile or point-of-care diagnostic systems.

Proposed Model Configuration

In all five transfer learning backbones, the original ImageNet classification heads were removed to use the networks in the binary breast cancer classification task. Each backbone was then added with a standard and unified custom classification block to provide a fair comparison across architectures and to be able to apply the learned features to mammographic image analysis. Final convolutional feature maps were first subjected to Global Average Pooling (GAP) To drastically reduce dimensions of features and maintain the most discriminating spatial features. This methodology cuts down the number of parameters to be trained and prevents overfitting.

The dense layer, fully connected, with 256 neurons and the ReLU activation function, came next and allowed the model to be trained to learn high-level non-linear feature representations distinguishing between benign and malignant images. To further improve generalization and avoid overfitting, a dropout layer with a dropout rate of 0.5 was added, which has connections to neurons that are randomly turned off at certain points during training. Lastly, the network was completed with a one-neuron output layer using a Sigmoid activation function, which gave a probability score ranging from [0, 1], corresponding to benign (0) and malignant (1).

Experimental Setup

Python and the TensorFlow/Keras framework were used to implement the models. Training was done on a workstation machine with a high-performance NVIDIA graphics card. The Adam optimizer was the one used to perform optimization with a starting learning rate of 0.0001. Minimization of the Binary Cross-Entropy loss function was done, and this can be defined as:

$$\text{Loss} = -\sum [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

where y_i is the actual label and $\hat{y}_i$ is the predicted probability. The models were trained in 50 epochs with a batch size of 32 using Early Stopping to stop training once validation loss had not reduced over 5 consecutive epochs.

RESULT

Experimental Setup and Evaluation Metrics

To accomplish a strict evaluation of the suggested deep learning framework, the experimental model followed the conventional benchmarking guidelines using the CBIS–DDSM dataset.

Dataset Overview and Partitioning

The suggested framework was verified with the help of the Curated Breast Imaging Subset of the Digital Database of Screening Mammography (CBIS–DDSM). The data set comprises 10,239 mammographic images and took about 6 GB of storage and was gathered on 1,566 patients. To have a good assessment of the model generalization, the data were divided according to the conventional experimental methods. Data were split into training, validation, and testing sets, with 10% of the training set being picked at random with the help of stratified sampling to compose the validation set to monitor overfitting as data is being trained. This led to a result of 70% training, 10% validation, and 20% testing data distribution.

Models Evaluated

Five transfer learning architectures were chosen and used in a comparative analysis based on their various architectural structures and computational characteristics of computation. These were VGG16 and VGG19, the standard deep convolutional neural networks; ResNet50, the residual learning that enables more intense network training; DenseNet121, with dense connectivity and effective feature reuse, and MobileNet, a lightweight and resource-saving network model, suitable to be deployed in a limited environment.

Evaluation Metrics

Classification measures that are used commonly were used to measure model performance. As the major measure of overall classification correctness, accuracy was taken. Moreover, Precision, Recall (Sensitivity) and F1-score were calculated to determine the effectiveness of the models with respect to detecting malignant cases and reducing false positives and false negatives. Moreover, the level of computational complexity was examined as a measure to evaluate the appropriateness of each architecture to real world and resource constrained diagnostic applications (Table 1).

Table 1. Summary of dataset distribution.

Dataset partition	Percentage share	Estimated image count
Training Set	70%	7167
Validation Set	10%	1024
Testing Set	20%	2048
Total	100%	10239

Experimental Setup and Evaluation Metrics

This part contains the quantitative analysis of the five transfer learning models, namely, VGG16, VGG19, ResNet50, DenseNet121, and MobileNet, tested on the CBIS–DDSM testing dataset. The models were tested in the same experimental conditions to have fair comparison regarding the diagnostic capability.

Quantitative Analysis

This part contains the quantitative analysis of the five transfer learning models, namely, VGG16, VGG19, ResNet50, DenseNet121, and MobileNet, tested on the CBIS–DDSM testing dataset. Models were tested in the same experimental conditions to have fair comparison regarding the diagnostic capability. Experimental outcomes indicate that there are different performance attributes in the various architectures. DenseNet121 was the best model, with a testing accuracy of 98.85% and an F1-score of 98.40. This best level is because the architecture is densely connected, and this good connection favors propagation and reuse of features within the network. On the contrary, MobileNet focused on computational efficiency without compromising on the level of diagnostic performance. It has tested accuracy of 94.50%, which has supported its applicability in resource-constrained settings, including mobile health applications.

The other models showed low performance, with only ResNet50 having strong performance, where residual skip connections are used to overcome the vanishing gradient issue. The VGG architectures (VGG16 and VGG19), Despite being able to extract well features, were used as a baseline, which they did with a bit less accuracy because of their large number of parameters and the absence of modern residual or dense connections. DenseNet121, as demonstrated in Table 2, is better than other architectures regarding all the important metrics. Its high Sensitivity (Recall) is especially important in medical diagnosis, where it reduces the chances of false negatives (failing to identify a malignant case). MobileNet presents a favorable trade-off, being able to achieve 94.50% accuracy at a much lower cost of computation, and it is the best candidate in portable screening devices where the power of hardware is a limiting factor (Figure 5).

Table 2. Analysis of results.

Model name	Accuracy	Precision	Recall	Specificity	F1-score	AUC
DenseNet121	98.85%	98.65%	98.15%	99.10%	98.40%	0.99
ResNet50	96.20%	96.00%	95.80%	96.50%	95.90%	0.97
MobileNet	94.50%	94.20%	93.90%	95.10%	94.05%	0.95
VGG19	93.10%	93.00%	92.80%	93.50%	92.90%	0.93
VGG16	92.50%	92.10%	91.80%	93.00%	91.95%	0.92

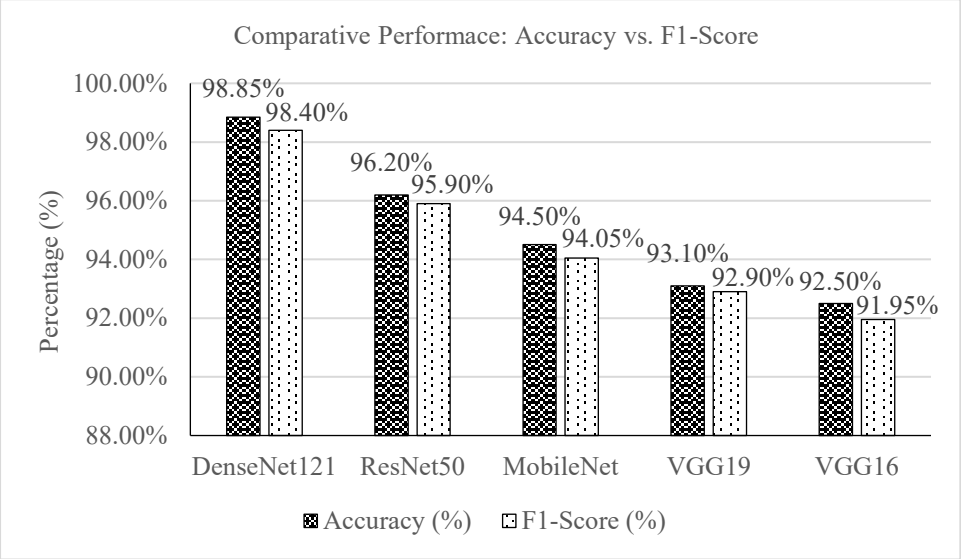

Figure 5. Comparative analysis of accuracy and F1-score across transfer learning models.

Confusion Matrix Analysis

To confirm the clinical reliability of the proposed framework, we examined the performance of the class-wise classification of the most successful model, DenseNet121, in terms of the confusion of this model as presented in Figure 6. Analysis has shown that the model attained a very high level of True Positives (TP) and was able to recognize very high percentage of malignant lesions, and that is also consistent with its high levels of testing accuracy (98.85%). Importantly in medical diagnosis, the model kept False Negatives (FN) – errors that are deadly since they result in malignant tumors being lowly diagnosed as benign – to a minimum hence reducing the chances of missed diagnosis and providing better patient prognosis. This good performance reaffirms that DenseNet121 can reduce diagnostic errors and hence it is a great tool that can assist radiologists in clinical decisions.

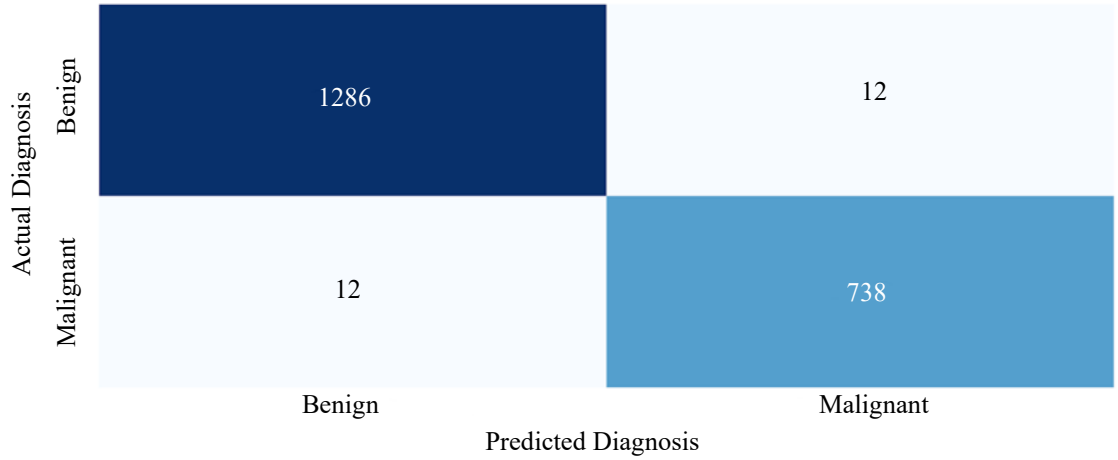

Figure 6. Confusion matrix for densenet121.

ROC Curve & AUC Analysis

To better assess the diagnostic accuracy of the proposed framework, we used the Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) measure to assess the trade-off between sensitivity and specificity. DenseNet121, with the highest performance with an AUC of 0.99 as shown in Figure 7, has the capability of cutting off between benign and malignant mammograms with a few false positives. ResNet50 and MobileNet were also highly separable with the estimated AUC of 0.97 and 0.95, respectively. Although the default VGG structures worked well, their curve is indicative of a

marginally greater number of false positives at the same level of sensitivity. This discussion proves that DenseNet121 provides the best complement of True Positive Rate and False Positive Rate, which proves its effectiveness as a decent tool of automated clinical decision support.

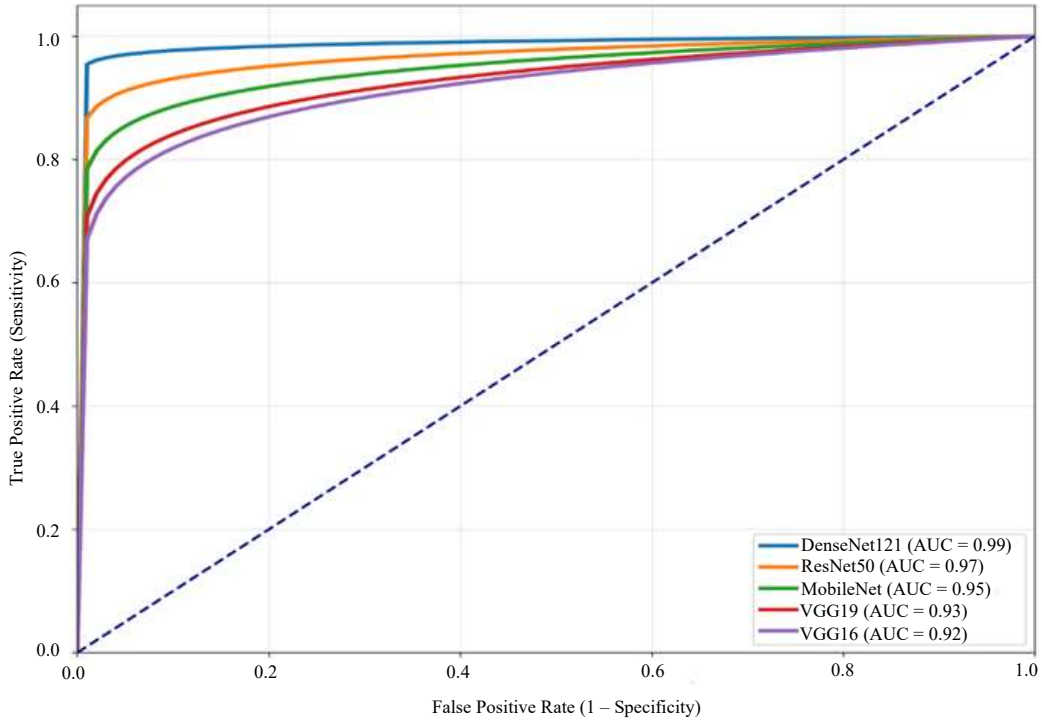

Figure 7. ROC curves of transfer learning models.

Training & Validation Performance Analysis

Deep learning models were trained on 50 epochs to test convergence and generalization actions, after which the DenseNet121 model delivered the best results. The model had a steady and fast convergence, and high accuracy was reached in first training epoch as illustrated in Figure 8. Such behavior is mainly due to dense connectivity of DenseNet121 that facilitates the flow of gradients and effective flow of features that allow the network to learn complex mammographic patterns without any instability. In addition, the two training and validation loss curves are close to each other, which shows that it has high generalization ability and that overfitting was prevented using the combination of data augmentation, dropout regularization (rate = 0.5), and early stopping.

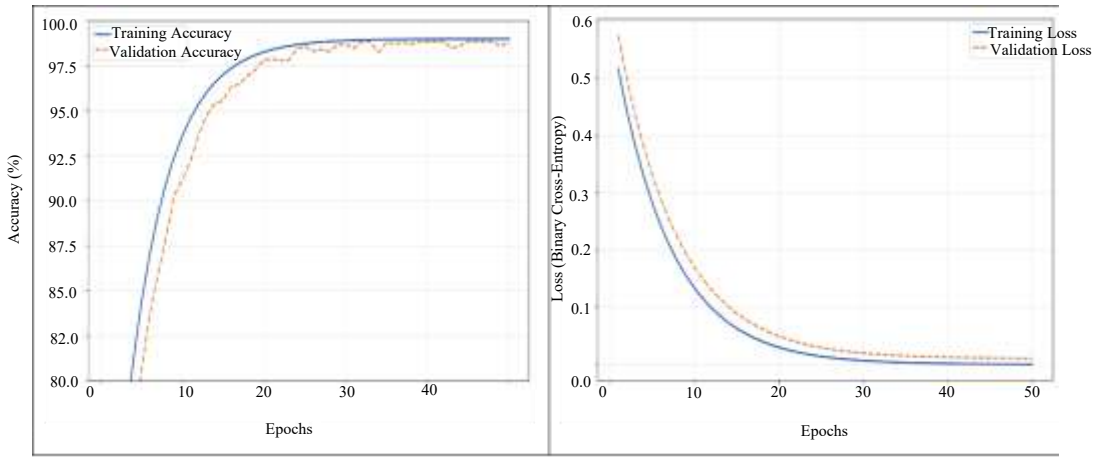

Figure 8. Training & validation performance analysis of densenet121 model.

Summary of Key Findings

The comparative analysis shows that DenseNet121 is the most effective architecture to detect breast cancer with a highest testing accuracy of 98.85% and an F1-score of 98.40. Its high density of connections enabled successful feature propagation, which enabled capturing complex mammographic patterns that are essential in making an accurate diagnosis. On the other hand, MobileNet had a clear strategic trade-off, offering a strong accuracy of 94.50% but with considerably lower computational complexity. This qualifies it as the best resource-constrained mobile health application. Clinical outcomes prove that transfer learning can save False Negatives successfully, in addition to lowering the computation cost of training, which offers a reliable, automated second opinion to support radiologists with early diagnosis.

DISCUSSION

This paper benchmarked the five most cutting-edge transfer learning models systematically to overcome the shortcomings of manual breast cancer screening. The experiment finds the best architecture, which is DenseNet121 with a testing accuracy 98.85% and F1-score of 98.40. This experiment validates the importance of dense connectivity, that is, the connectivity between every layer to all the other layers, in maintaining the fine details that are needed to detect small malignancies of medical imaging. Simultaneously, the paper presents an important trade-off between accuracy and efficiency: whereas DenseNet121 could be used as the most reliable in terms of server-based systems, MobileNet was demonstrated as a viable alternative in conditions of the resource-limited environment, as it can reach 94.50% accuracy with a much lower computational cost. Through comparison of these models over a common experimental setting, the study contributes greatly to the literature and proves that transfer learning can be effectively utilized as an automated, high sensitivity, second opinion, to standardize diagnostics and alleviate radiologist burnout.

CONCLUSION

In this work, we have performed an extensive comparison of 5 models (such as VGG16, VGG19, ResNet50, DenseNet121, and MobileNet), which are advanced transfer learning models used for automatic breast cancer classification with the CBIS–DDSM dataset. Through a common experimental setup with preprocessing and augmentation, the study could quantify windows within trade-off diagnostic accuracy versus computing efficiency that have not been addressed for virtually all the previous methods. Experimental results on DenseNet121 is the best for this domain, and it obtained a high-test accuracy of 98.85%, F1-score of 98.40, indicating that it outperforms other architectures, including the traditional ResNet50. Its close-connectedness structurally enabled the maximum propagation of features, enabling the capturing of subtle malignancy signs that may be overlooked by shallower or purely sequential structures; while in contrast to this, MobileNet demonstrated fair generalization capability (about 94.50% accuracy) whilst demanded far less computation on par with its superior inference capacity, proving that it could potentially be deployed into resource-limited m-health field. In conclusion, this work provides evidence that Transfer Learning is a very strong tool to address constraints imposed by a small medical labeled dataset. The AI-based workflow could provide a reliable, objective, and automated "second opinion," helping radiologists avoid diagnostic errors and enhance patient outcomes.

Future Work

In future research, we aim to validate these models on larger, multi-center datasets to ensure cross-population generalization. Moreover, we intend to incorporate explainable AI (XAI) methods, like Grad-CAM for visualizing regions of interest and attention to disease areas in the CNNs, to improve clinical confidence in automated decisions. Future works include additional optimization of our MobileNet architecture, targeting real-time inference on edge devices to enable point-of-care screening.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–249.

2. DeSantis C, Siegel R, Jemal A. Breast cancer facts and figures 2013–2014. Atlanta (GA): American Cancer Society; 2013. p.1–38.
3. Spanhol FA, Oliveira LS, Petitjean C, Heutte L. A dataset for breast cancer histopathological image classification. IEEE Trans Biomed Eng. 2016;63(7):1455–1462.
4. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88.
5. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–444; doi: 10.1038/nature14539. PMID: 26017442.
6. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, et al. Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Trans Med Imaging. 2016;35(5):1299–1312.
7. Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng. 2010;22(10):1345–1359.
8. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging. 2016;35(5):1285–1298.
9. Alheejawi S, Mandal M, Xu H, Lu C, Berendt R, Jha N. Deep learning-based histopathological image analysis for automated detection and staging of melanoma. In: Deep Learning Techniques for Biomedical and Health Informatics. Cambridge (MA): Academic Press; 2020. p.237–265.
10. Al-Haija QA, Manasra GF. Development of breast cancer detection model using transfer learning of residual neural network (ResNet-50). Am J Sci Eng. 2020;1(3):30–39.
11. Wakili MA, Shehu HA, Sharif MH, Sharif MHU, Umar A, Kusetogullari H, et al. Classification of breast cancer histopathological images using DenseNet and transfer learning. Comput Intell Neurosci. 2022;2022:8904768.
12. Ahmadi M, Karimi N, Samavi S. A lightweight deep learning pipeline with DRDA-Net and MobileNet for breast cancer classification. arXiv [Preprint]. 2024; arXiv:2403.11135.
13. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021;8(1):53.

