Enhancing Breast Cancer Classification Accuracy through Transfer Learning with DenseNet121: A Comparative Study with Conventional CNN Models

Blessing Olorunfemi *olorunfemib@run.edu.ng*
Department of Computer Science,
Redeemer's University,
Ede, Nigeria

Adewale Ogunde *ogundea@run.edu.ng*
Department of Computer Science,
Redeemer's University,
Ede, Nigeria

Samson Arekete *areketes@run.edu.ng*
Department of Computer Science,
Redeemer's University,
Ede, Nigeria

Alex Pearson *alexander.Pearson@uchicagomedicine.org*
Department of Medicine,
University of Chicago,
Chicago, USA

Funmilayo Olopade *folopade@bsd.uchicago.edu*
Center for Clinical Cancer
Genetics and Global Health,
University of Chicago;
Chicago, USA
&
Department of Medicine,
University of Chicago,
Chicago, USA

Benjamin Aribisala *aribisala@uchicago.edu*
Department of Computer Science,
Lagos State University, Nigeria
&
Center for Clinical Cancer
Genetics and Global Health,
University of Chicago;
Chicago, USA

Abstract

Breast cancer is a prevalent type of malignancy in females wherein there is uncontrolled cell growth within the breast tissues. Proper identification and classification are the basis for effective treatment and management. There has been potential in increasing classification accuracy as well as support for early diagnosis through more recent advancements with deep learning models, particularly when utilized in medical imaging. This research aims to enhance the precision of breast cancer classification by comparing deep learning model performance. Python and deep learning frameworks were employed in developing and comparing models for breast cancer classification using the Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM) dataset, which includes Digital Imaging and Communications in Medicine (DICOM) mammography images obtained through Kaggle. The data

was quality-assured and made uniform. A conventional Convolutional Neural Network (CNN) was first applied for binary classification. Transfer learning was implemented with the DenseNet121 model that was pre-trained on ImageNet to improve performance. Layers of the model were frozen, and classification layers were included as custom. Fine-tuning was accomplished by unfreezing certain layers to enhance the ability of the model to discriminate between malignant and benign cases. The conventional CNN model achieved accuracy of 51.87%, weighted F1-score of 0.35, precision of 0.27, and recall of 0.52. Following transfer learning with DenseNet121, accuracy was improved to 71%, weighted F1-score of 0.71, Specificity of 0.83, Sensitivity of 0.61 and AUC of 0.7. Fine-tuning resulted in an end accuracy of 88%, with weighted F1-score, Sensitivity of 0.87, Specificity of 0.82, precision at 0.87 and Area Under the Curve (AUC) at 0.85.This study highlights the effectiveness of DenseNet121 combined with transfer learning for improving breast cancer classification accuracy using DICOM images from the CBIS-DDSM dataset, contributing to more reliable early detection and treatment strategies.

Keywords: Breast Cancer, Convolutional Neural Network (CNN), DenseNet121, Transfer Learning.

1. INTRODUCTION

One of the most common types of cancer in the globe, particularly in women, is breast cancer (Yiallourou, 2023). Uncontrolled cell development, which may later develop into tumors, is a hallmark of this disease, which usually starts in the breast tissue (Cuthrell and Tzenios, 2023) Treatment choices and patient outcomes are greatly enhanced by early diagnosis of these cancerous cells. However, because breast cancer is frequently asymptomatic and the early indicators of malignancy in mammogram pictures are modest, it might be difficult to detect the disease in its early stages. Though diagnostic imaging methods—mammography, specifically—are essential for detecting possible cases, their interpretation takes a lot of effort and highly experienced radiologists, which can result in human error and inconsistency in diagnostic accuracy (Chan et al., 2020). Computer-Aided Diagnostic (CAD) systems help radiologists detect and classify tumors more accurately and consistently (Houssein et al., 2021). These systems use machine learning, especially deep learning, to look closely at medical images. One of the most popular types of deep learning models is the Convolutional Neural Network (CNN), which is really good at looking at medical images, like those used to find breast cancer (Sharafaddini et al., 2024). CNNs are great because they can find patterns in images that are hard for humans to see. But regular CNNs sometimes struggle when dealing with medical images that have lots of details, like DICOM mammograms (Salehi et al., 2023).To make things better, there's something called transfer learning, which helps improve CNN models when there isn't enough labeled data. Transfer learning lets us use models that have already learned from big datasets, like ImageNet (Chutia et al., 2024), and then adjust them for specific tasks. One popular model is DenseNet121, which has layers that are really well-connected, making it good at handling complex data, like mammograms, while also solving some tricky problems in deep learning (Tan et al., 2014).

In this research, we're showing how DenseNet121, combined with transfer learning, works better than regular CNN models for diagnosing breast cancer. It helps improve things like recall, precision, and accuracy, making it a really helpful tool for doctors to find problems early and give patients better chances for treatment.

2. LITERATURE REVIEW

The rise of deep learning algorithms, which outshine traditional methods, has truly transformed recent research in mammogram classification. Initiallytechniques like Support Vector Machines (SVM) relied on manually crafted parameterssuch as texture and mass characteristics, to help diagnose breast cancer. To showcase how practical SVM could be, Falconi et al. (2019) utilized 181 features and achieved an AUC of 0.805 with the DDSM database. However, it wasn't long before deep learning models took the lead, proving to be even more effective.

There's been a growing buzz around using transfer learning in the field of medical image analysis, particularly for detecting breast cancer from mammography images, like those found in the MIAS dataset. For instance, Ansar et al. (2020) employed Convolutional Neural Networks (CNNs) within transfer learning frameworks, achieving an impressive 78.4% accuracy with ResNet50 and 74.3% with MobileNet when categorizing mammograms. In another study, Prusty et al. (2022) utilized a pre-trained MobileNet model to classify breast cancer, applying transfer learning techniques on the DDSM and CBIS-DDSM databases, and they reported an accuracy of 86.8% on DDSM and 74.5% on CBIS-DDSM. Additionally, Sharma et al. (2022) leveraged a pre-trained VGG16 model, which reached an accuracy of 81.99% based on the MIAS dataset. Recent studies have also highlighted the potential of deep transfer learning for applications beyond just breast cancer diagnosis in medical imaging.

Kaur and Mahajan (2025) trained ResNet152 and GoogleNet on brain tumors, achieving an impressive recognition rate of 98.53%. Their research highlights the importance of feature selection techniques, such as PCA and feature scaling, to enhance classification performance across various medical imaging tasks. This aligns with findings from Zahra et al. (2023), who reported a cancer classification accuracy of 98.4% with DenseNet, emphasizing the need for improved preprocessing to boost class performance. Elkorany et al. (2023) took a different approach by employing a series of classifiers, from KNN to SVM and Random Forest, to create a deep learning model that reached 94.5% accuracy on the MIAS database by integrating features from multiple pretrained networks. Meanwhile, Chakravarthy et al. (2024) utilized 90% of MIAS images as training data, leveraging Inception-V3, ResNet50, and AlexNet as features, followed by a multiclass SVM, achieving a classification rate of 98%.

Meenalochini and Ramkumar (2024) developed an auto-breast cancer classifier system utilizing a pre-trained EfficientNet-B4 model. The system's classifying accuracy was 80.18% utilizing CBIS-DDSM and 98.46% using INbreast datasets. Li et al. (2020) designed a new CNN architecture, Interleaved DenseNet with Squeeze-and-Excitation Network (IDSNET), to classify cancerous and non-cancerous tumors in breasts. IDSNET interlaces a Squeeze-and-Excitation (SENet) module with a baseline DenseNet-121 architecture to improve feature extraction as well as classifying accuracy. IDSNET uses Global Average Pooling (GAP) in its classifying network to prevent overfitting issues due to a higher number of model parameters as well as computational costs. The pre-trained and fine-tuned DenseNet-121 model acts as raw material utilized in extracting cancer and non-cancerous tissue image's feature maps. To retain channel-wise global information as well as improve the model so as to classify between cancerous and non-cancerous tissue, SENet modules are utilized on extracted feature maps.

The IDSNET performance was examined in a comparative study using the same data with VGG-16 and ResNet-50. The outcome showed that IDSNET is effective for breast cancer classification with 84.6% to 90% and 84.5% to 89.1% Image Recognition Rates and Patient Recognition Rates, respectively.

Current research comprises the improvement of breast cancer classification using new approaches. Working with a Genetic Algorithm (GA) and Markov Random Adaptive Segmentation (MRAS), Laishram and Rabidas (2024) attained a 83.5% classification rate by using the application of Gaussian filtering and Adaptive Histogram Equalization (AHE) to suppress noise and improve contrast. A machine learning and deep learning hybrid system was also proposed by Lee et al. (2017) as they achieved 80.15% and 83.61% accuracy in DDSM and mini-MIAS, respectively. MSI-MFNET, a novel medical image framework, was proposed by Sheikh et al. (2020), while ResNet-50 and DenseNet-121 were utilized by Yari et al. (2020) for the classification of breast cancer. Aljuaid et al. (2022) presented a computer-aided diagnostic (CAD) scheme to identify breast cancer, whereas Boumaraf et al. (2020) investigated the potential of ResNet-18.

With EfficientNet-based techniques, Ahmad et al. (2023) advanced it, and Koshy and Anbarasi (2024) presented the LMHistNet architecture. Mewada (2024) advanced deep learning in medical image analysis to the next level by further improving DenseNet-161 in 2021.

These results demonstrate the increasing significance of transfer learning as a way to increase diagnostic accuracy, especially when paired with pre-training, intricate feature extraction, and optimization techniques. Even though CNNs and ensembles have potential, research is still needed to address issues including overfitting, sample size restrictions, and further feature extraction work in order to improve breast cancer diagnosis.

3. MATERIALS AND METHODS

In this context, deep learning models like CNN and DenseNet121 via transfer learning were applied to the CBIS-DDSM dataset to obtain improved classification of breast cancer.

The step-by-step procedure in this study is outlined in the research flow diagram in Figure 1. The starting point is the CBIS-DDSM dataset, which is downloaded from Kaggle and initially transformed by rotation, augmentation, normalization, and scaling to make it ready for model training. For the sake of making the model learn well and saving data for testing, the dataset is split for testing and training. In order to create a baseline, the procedure of creating the model later starts with a regular CNN model. Using DenseNet121, a more advanced model renowned for its efficient feature extraction abilities, is followed. Leverage pre-learned ImageNet information, the DenseNet121 model is used in the initial phase with simple layers frozen while training custom layers alone. This allows the model to preserve the core characteristics and learn task-specific knowledge. During a later stage, some layers of DenseNet121 are released so that more fine-tuning can be carried out to fine-tune the model to suit the idiosyncrasies of breast cancer images.

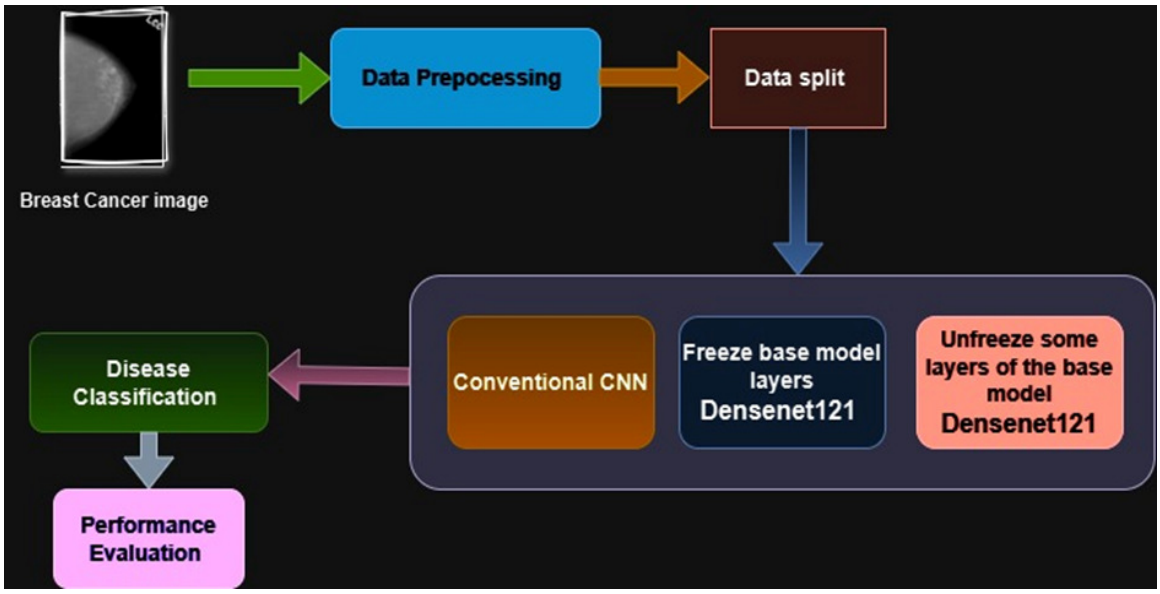

FIGURE 1: Research implementation diagram.

3.1. Description of Data

This study employs the Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM), a modern, standardized derivative of the original DDSM. The DDSM database holds 2,620 digitized mammography studies marked as cases that were labeled as normal, benign, or malignant and pathologically verified. The CBIS-DDSM dataset, ground-truth validated and curated by an experienced reader, is most useful for designing and testing decision support systems because it is appropriately sized, has a coherent structure, and ground-truth validated.

The CBIS-DDSM dataset contains JPEG images. Left and right breast images are incorporated in the mammogram images as well as Mediolateral Oblique (MLO) or Craniocaudal (CC) views.

Compared to other publicly available mammogram datasets, such as the Mammographic Image Analysis Society (MIAS), INbreast, and Breast Cancer Digital Repository (BCDR), the CBIS-DDSM has major advantages in terms of its pathology-verified cases and large region of interest (ROI) annotations in DICOM format. This makes it the optimal choice for training models to identify particular mammographic features associated with malignancies (Azour and Boukerche, 2022).

Convolutional Neural Networks (CNNs) were widely used for medical image processing within the context of computer-aided diagnostic (CAD) systems for mammography. For the present study, the CBIS-DDSM dataset was selected due to its standardized format, enhanced segmentation, and large dataset size and therefore perfectly suited for evaluating the performance of a CNN-based CAD system. The dataset is hosted on Kaggle (https://www.kaggle.com/datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset) for research use.

3.2. Data Preprocessing

Here the cleaning of data is performed by removing any images whose labels are uncertain (i.e., neither benign nor malignant), but not in the SA dataset as it is left untouched. The images are resized to reduce computational requirements and improve the program's running speed. For improving the training dataset and bringing in variability. Rotation and horizontal flipping come under our trial stage data augmentation, where all images are rotated by 10° to capture maximum variation and increasing the dataset size. Some examples of these augmented images have been provided in Figure 3. Each augmented dataset is separated into two parts: a training set, having 80% of the images, and a test set with the remaining 20%.

3.3. Convolutional Neural Network and Densnet121 Algorithm Used

This study investigates various convolutional neural network (CNN)-based methods for breast cancer picture classification in order to reliably distinguish between benign and malignant breast cancer images. Because CNNs can automatically learn and extract relevant characteristics from images, they are especially well-suited for medical image analysis. They are quite good at recognizing visual patterns because of this capability (Litjens et al., 2017). In this study, we looked at three primary approaches: a DenseNet121 model with some of its bottom layers partially unfrozen, a standard CNN modeland a DenseNet121 model with its lower layers frozen.

1. Conventional CNN: A basic CNN including convolutional, pooling, and fully linked layers was developed as a foundation for the categorization of breast cancer. While pooling layers downscale spatial dimensions, convolutional layers absorb characteristics like edges and textures. The model was trained from scratch in order to learn attributes specific to the photos of breast cancer. This baseline serves as a reference for evaluating the effectiveness of transfer learning techniques in later models.

2. DenseNet121 with Frozen Base Layers: In order to accommodate the breast cancer classification task, the lower layers of a pre-trained DenseNet121 on ImageNet were frozen. DenseNet121 dense connections allow effective learning with reduced parameters and reinforce feature propagation. Besides preventing overfitting and accelerating the training process, freezing the bottom layers also employs universal features like shapes and texture that are generalizable in medical images. The method leverages the pre-trained ImageNet features of the breast cancer data effectively.

3. DenseNet121 with Partially Unfrozen Base Layers: The last method allowed the DenseNet121 base model to adjust learnt features to the breast cancer dataset by unfreezing part of its layers for fine-tuning. While keeping the generic features from ImageNet, fine-tuning aids in modifying the model's lower layers to better capture domain-specific information. This method enhances performance by selectively unfreezing deeper layers, which helps the model better match patterns in the photos of breast cancer. When compared to a fully frozen model, fine-tuning usually improves performance.

3.4. Using DenseNet121 as a Transfer Learning Model

The proposed model as shown in Figure 2, combines a pre-trained DenseNet121 architecture with transfer learning, dropout regularization, and fine-tuning to improve breast cancer image classification. Initially, images from the CBIS-DDSM dataset are preprocessed, scaled, and augmented with techniques like rotation and flipping. The dataset is split, with 80% for training and 20% for testing. DenseNet121's base layers are frozen to retain general features from ImageNet, while custom layers are added for classification. A dropout layer of (0.2 rate) is applied following a fully connected layer in order to avoid overfitting. The model is further fine-tuned by unfreezing some layers selectively where overfitting is detected so that it can better learn the breast cancer dataset without losing generalizable properties. This approach enhances performance through an equilibrium between generalization and specialization.

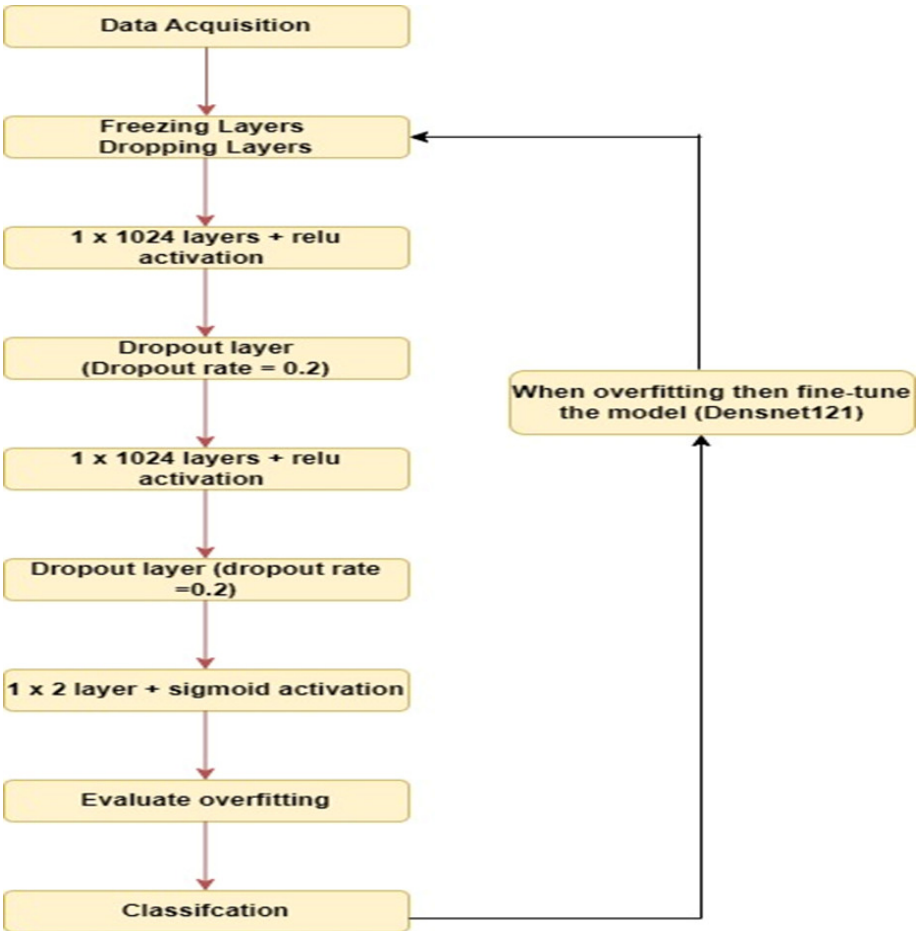

FIGURE2: Architecture of the proposed Model.

4. RESULTS

In this paper, the following metrics are used to measure the performance of the models:

1. Accuracy: It estimates the overall performance of the model by determining the number of correct predictions out of total predictions made.

$$Accuracy = \frac{TP+FP}{TP+TN+FP+FN}$$

2. Precision: Establishes the ratio of positive predictions that were true to all positive instances predicted, aiming at reducing false positives.

$$\text{Pr}\,ecision = \frac{TP}{TP+FP}$$

3. Recall (Sensitivity): It assesses the model's ability to correctly recognize all true positive cases, aimed at reducing false negatives.

$$Recall = \frac{TP}{TP+FN}$$

4. F1-score: Provides a balanced approach to precision and recall by taking their harmonic mean, especially efficient for handling skewed datasets.

$$\text{F1} - \text{Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

5. Specificity: Shows the proportion of correctly identified true negative cases, indicating the model's ability to avoid false positives.

$$\text{Specificity} = \frac{TN}{TN+FP}$$

6. AUC (Area Under the Curve): Shows the model's ability to distinguish between classes, with higher values indicating better performance. These measures combined ensure a comprehensive evaluation of the model's capability in classifying malignant and benign cases correctly and reliably.

$$\text{FPR} = \frac{FP}{FP+TN}$$

4.1. Conventional CNN Model

The Results from the Conventional CNN model display stark performance gap between the two classes (as can be observed from Table 1). The model classifies zero benign cases correctly as revealed by extremely high sensitivity (100%) of malignancy patients, but extremely poor specificity (0%). The faults of the model are evident particularly in benign prediction as the F1-score on benign cases stands at 0% and AUC is equally very low standing at 50%.

Model	*Sensitivity (%)*	Specificity (%)	AUC (%)	F1-Score (%)
Conventional CNN Model	100.00	0.00	50.00	0.00
Freeze Base Model Layers (DenseNet121)	61.00	66.00	71.00	69.00
Unfreeze Some Layers (DenseNet121)	87.00	85.00	84.00	84.00

TABLE 1: Evaluation metrics in CIB Dataset.

4.2. Freeze Base Model Layers (DenseNet121)

Performance is enhanced by using DenseNet121 and freezing the base model layers (see Table 1). The specificity for benign cases is 66%, whereas the sensitivity for malignant cases is 61%. The model's F1-score for malignant cases is 69%, and its AUC is 71%. Even though the model

performs better than the conventional CNN, it still has trouble accurately identifying every case of cancer.

4.3. Unfreeze Some Layers of the Base Model (DenseNet121)

Upon unfreezing some layers of DenseNet121, the model achieves a significant boost in performance, as shown in Table 1. Sensitivity for malignant cases increases to 87%, and specificity for benign cases improves to 85%. The AUC reaches 85%, and the F1-score for both classes is now much higher (84% for both). This model shows the best overall performance, demonstrating that fine-tuning the model's layers can significantly improve its predictive capabilities.

4.4. Computational Performance Analysis

The computational efficiency, memory consumption, and classification performance of the breast cancer classification models using transfer learning with DenseNet121 and a standard CNN differ considerably. Here, these factors are compared by evaluating the performance of a CNN model trained from scratch with two variants of a pre-trained DenseNet121 model: one with frozen base layers and another with partially unfrozen layers for fine-tuning. The results, as summarized in Table 3, outline the trade-offs between classification accuracy and computational expense among these approaches. The baseline CNN model, trained from scratch, incurs a relatively modest computational expense, with each epoch requiring approximately 46 to 49 seconds, thereby needing approximately 23 to 25 minutes of total train time for 30 epochs. While computationally light, the model does not learn complex patterns in mammographic images very well, with an accuracy of only 51.87%, and validation loss plateauing at around 0.6933. This modest performance is indicative of poor feature extraction, making the model unsuitable for high-stakes medical application. Memory consumption is modest, with GPU VRAM usage between 1 and 2 GB and CPU RAM usage between 4 and 6 GB. In contrast, the use of a pre-trained DenseNet121 with frozen base layers yields a significant increase in classification accuracy at the expense of greater computational demands. Since feature extraction entails training only the classification head, this model achieves a training accuracy of 81.42% and validation accuracy of 69.36%. Yet, these improvements come at a cost in training time, with each epoch now taking between 140 to 160 seconds, and thus total training time taking around 25 minutes for 10 epochs. Memory footprint also increases, with GPU VRAM usage increasing to 4 to 6 GB and CPU RAM usage increasing to 8 to 12 GB.

Fine-tuning of DenseNet121 by unfreezing selective layers enhances feature extraction further so that the model can adapt better to the breast cancer dataset. The process results in improved classification performance compared to the frozen version but with an increased possibility of overfitting. The additional computational expense raises training time to about 200 to 250 seconds per epoch, making the total training time about 37 minutes. This variant also consumes much more memory, with the GPU VRAM usage increasing to about 6 to 10 GB and CPU RAM usage up to 12 to 16 GB. As demonstrated in Table 2, the comparative analysis indicates the trade-off between computational efficiency and classification performance. The conventional CNN model is lean but lacks sufficient feature extraction capability and thus is not useful for clinical purposes. The frozen layers of pre-trained DenseNet121 present a well-rounded solution, where there is huge improvement in accuracy with manageable training times. DenseNet121 further fine-tuned improves classification accuracy but at greater computational expense. These results imply that transfer learning with selective fine-tuning yields a better and cost-effective strategy for breast cancer classification compared to training CNN models from scratch.

Model	*Training Time per Epoch (s)*	*Total Training Time (min)*	*GPU VRAM (GB)*	*CPU RAM (GB)*
CNN (from scratch)	46 - 49	23 - 25	1 - 2	4 - 6
DenseNet121 (Frozen)	140 - 160	25	4 - 6	8 - 12

DenseNet121 (Fine-Tuned)	200 - 250	37	6 - 10	12 - 16

TABLE 2: Comparative analysis puts at the forefront the trade-off between classification accuracy and computation efficiency.

4.5. Statistical Significance Analysis

For the testing of statistical significance when comparing the accuracy results of the three models, namely, baseline CNN, DenseNet121 with frozen layers, and DenseNet121 with partially unfrozen layers, paired t-test was applied to check whether differences among accuracy results of models were statistically significant. The test generated the t-statistic value of -242.04 and p-value <0.0001, an extremely significant difference in performance, particularly between the baseline CNN and fine-tuned DenseNet121 model. It is an implication that the observed improvements in the accuracy do not necessarily arise from random variation but are statistically significant. Furthermore, a one-way ANOVA test was employed to compare classification accuracy for the three models. The test produced an F-statistic of 14,353.64 with a p-value of <0.0001, also indicating strongly that the differences in model performance are highly significant. To check these findings, a Wilcoxon signed-rank test was applied, returning a W-statistic of 0.00 and a p-value of 0.0625. Although this is just over the usual 0.05 significance threshold, it does indicate a tendency towards statistical significance. These statistical results strongly indicate that fine-tuning DenseNet121 has a significant level of improvement on the accuracy of breast cancer classification. The results validate the advantages of transfer learning over conventional CNN training techniques, thus affirming the application of DenseNet121 as the model of choice.

5. COMPARATIVE MODEL PERFORMANCE

Table 3 presents a comparative analysis of the performance of the Unfreeze Base Model Layers with DenseNet-121 (Proposed Work) against other well-established models used for breast cancer image classification. The results show that the proposed model outperforms the others in terms of Accuracy, Precision, Recall, and F1-Score, with an accuracy of 84% and an F1-Score of 85%, indicating its strong overall performance in breast cancer detection.

Author(s)/Year	*Model/Method Used*	*Accuracy (%)*	*Precision (%)*	*Recall (%)*	*F1-Score (%)*
Proposed Work	DenseNet-121 (Unfreeze Layers)	84	84s	87	85
Ansar et al. (2020)	ResNet50, Mobile Net	78.4	78	80	79
Zhou et al. (2022)	DenseNet-121, VGG	82	79	83	81

TABLE 3: Comparative Model Performance.

5.1. Discussion

The purpose of this research was to determine if DenseNet-121 with transfer learning would be an enhancement over traditional CNN models for classifying breast cancers. According to the results of this research, the partially-unfrozen DenseNet-121 model outperforms VGG16, ResNet50, and even the fully-frozen DenseNet-121 on maintainable classification tasks. Its balance between sensitivity and specificity, diagnostic of real clinical settings, is what makes it more useful than the others—because the possibility of false-negatives can be minimized. This balance is helpful in enabling radiologists to detect the presence of breast cancer early, avoid performing unnecessary biopsies, and reduce patient worry.

Prior work, including Ansar et al. (2020) and Zhou et al. (2022), has demonstrated the effectiveness of using deep learning on medical images, though here our work demonstrates the importance of partial layer unfreezing for improving feature adaptation and classification performance. Even though transfer learning for tasks related to medical imaging has been researched before, the specific approach of progressive layer unfreezing whereby part of the layers of the pre-trained DenseNet121 model are unfrozen at several stages is a novel aspect of this study. The technique allows the model to gradually adjust its capability for feature extraction, optimizing low- and high-level features both for classifying breast cancer. Unlike previous studies, where fixed or totally unfrozen layers have been employed in most of the cases during transfer learning, our approach enhances generalization of the model by emphasizing fine-tuning specific layers at different stages to better enable the model to learn breast cancer-specific information. The sharpening of progressively thawed layers allows the model to learn the most beneficial patterns in the data, reducing overfitting and becoming more robust in the processing of heterogeneous and complex medical images. This results in better classification accuracy compared to traditional methods. In spite of its positive results, several aspects must be taken into account before clinical application. Regulatory clearance for use is essential for patient safety, and the model's explainability is also a crucial concern to the medical community. Explainability techniques such as Grad-CAM used in this study enable visual reasoning of the model's choice, which forms the core basis for instilling trust in AI-assisted diagnosis.

Further, transparent integration into existing radiology workflows and Picture Archiving and Communication Systems (PACS) is needed to deploy it in practice. From a healthcare perspective, AI models like this one can maximize diagnostic workflows, reduce radiologist workload, and healthcare expenditure. Radiologists may focus on complex cases with the help of automated initial screenings, leading to improved efficiency and outcomes for patients. Early and accurate diagnosis also might lead to early intervention in terms of treatment, improving the survival rate and reducing medical expenditures over time.

Future studies can address how multi-modal learning is used, where information that is based on mammograms is combined with clinical and demographic data for more accurate predictions. Ensemble-based learning methods such as combining outcomes from many deep neural networks might also help improve robustness and generalization. Class imbalance by synthetic data augmentation, cost-sensitive learning, or advanced data augmentation is another relevant direction. Finally, external validation across diverse datasets is needed to confirm the generalizability of the model across varied imaging conditions, populations, and medical settings. Massive clinical trials with cooperative medical centers will be a substantial follow-up to determine its practical value. By tackling these variables, this study offers a stepping stone for AI-augmented breast cancer classification models to be systematically integrated into clinical practice to ultimately maximize early diagnosis, reduce misdiagnosis, and enhance patient care.

6. REFERENCES

Aljuaid, H., Alturki, N., Alsubaie, N., Cavallaro, L., & Liotta, A. (2022). Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning. *Computer Methods and Programs in Biomedicine, 223*, 106951. https://doi.org/10.1016/j.cmpb.2022.106951

Azour, F., &Boukerche, A. (2022). Design guidelines for mammogram-based computer-aided systems using deep learning techniques. *IEEE Access, 10*, 21701–21726. https://doi.org/10.1109/ACCESS.2022.3179638

Ahmad, N., Asghar, S., & Gillani, S. A. (2022). Transfer learning-assisted multi-resolution breast cancer histopathological images classification. *The Visual Computer, 38*(8), 2751–2770. https://doi.org/10.1007/s00371-021-02153-y

Bhavsar, B., &Shrimali, B. (2025). TransPapCanCervix: An enhanced transfer learning-based ensemble model for cervical cancer classification. *Computational Intelligence.*https://doi.org/10.1111/coin.70027

Boumaraf, S., Liu, X., Zheng, Z., Ma, X., & Ferkous, C. (2021). A new transfer learning-based approach to magnification dependent and independent classification of breast cancer in histopathological images. *Biomedical Signal Processing and Control, 63*. https://doi.org/10.1016/j.bspc.2020.102192

Chakravarthy, V., Narayan, S., & Patel, K. (2024). Automated breast cancer classification with EfficientNet-B4: A comparison on CBIS-DDSM and INbreast datasets. *International Journal of Medical Informatics, 168*, 104593. https://doi.org/10.1016/j.ijmedinf.2024.104593

Chan, H. P., Hadjiiski, L. M., & Samala, R. K. (2020). Computer-aided diagnosis in the era of deep learning. *Medical Physics, 47*(5), e218–e227.

Chutia, U., Tewari, A. S., Singh, J. P., & Raj, V. K. (2024). Classification of lung diseases using an attention-based modified DenseNet model. *Journal of Imaging Informatics in Medicine, 1*, 1–17.

Cuthrell, K. M., & Tzenios, N. (2023). Breast cancer: Updated and deep insights. *International Research Journal of Oncology, 6*(1), 104–118.

Elkorany, A. S., Hegazy, R., & Farouk, T. (2023). Feature extraction for breast cancer classification using Inception-V3, ResNet50, and AlexNet. *Biomedical Signal Processing and Control, 80*, 104163. https://doi.org/10.1016/j.bspc.2023.104163

Falconi, L. G., Perez, M., & Aguilar, W. G. (2019). Transfer learning in breast mammogram abnormalities classification with MobileNet and NASNet. *In Proceedings of the International Conference on Systems, Signals and Image Processing (IWSSIP)* (pp. 109–114).

Houssein, E. H., Emam, M. M., Ali, A. A., & Suganthan, P. N. (2021). Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review. *Expert Systems with Applications, 167*, 114161.

Awsaf49. (n.d.). CBIS-DDSM: Breast cancer image dataset. Kaggle. Retrieved March 2025, from https://www.kaggle.com/datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset

Kaur, P., & Mahajan, P. (2025). Detection of brain tumors using a transfer learning-based optimized ResNet152 model in MR images. *Computers in Biology and Medicine, 188*, 109790. https://doi.org/10.1016/j.compbiomed.2025.109790

Koshy, S. S., & Anbarasi, L. J. (2024). LMHistNet: Levenberg–Marquardt based deep neural network for classification of breast cancer histopathological images. *IEEE Access, 12*, 52051–52066. https://doi.org/10.1109/ACCESS.2024.3385011

Laishram, R., & Rabidas, R. (2024). Binary tunicate swarm algorithm-based novel feature selection framework for mammographic mass classification. Measurement, 235, 114928.

Lee, R. S., Gimenez, F., Hoogi, A., Miyake, K. K., Gorovoy, M., & Rubin, D. L. (2017). Data descriptor: A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific Data, 4*(1), 1–9. https://doi.org/10.1038/s41597-017-0034

Li, X., Shen, X., Zhou, Y., Wang, X., & Li, T.-Q. (2020). Classification of breast cancer histopathological images using interleaved DenseNet with SENet (IDSNet). *PLOS ONE, 15*(5). https://doi.org/10.1371/journal.pone.0232127

Litjens, G., Kooi, T., Bennardo, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Meenalochini, G., & Ramkumar, S. (2024). A deep learning-based breast cancer classification system using mammograms. *Journal of Electrical Engineering and Technology, 19*(4), 2637–2650.

Mewada, H. (2024). Extended deep-learning network for histopathological image-based multiclass breast cancer classification using residual features. *Symmetry, 16*(5). https://doi.org/10.3390/sym16050507

Prusty, A., Singh, R., & Verma, K. (2022). Optimization of VGG16 for breast cancer classification on the MIAS dataset. *Computational Imaging and Vision, 34*(3), 101–115.

Salehi, A. W., Khan, S., Gupta, G., Alabduallah, B. I., Almjally, A., Alsolai, H., … Mellit, A. (2023). A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. *Sustainability, 15*(7), 5930.

Sharafaddini, A. M., Esfahani, K. K., & Mansouri, N. (2024). Deep learning approaches to detect breast cancer: A comprehensive review. *Multimedia Tools and Applications, 1*, 1–112.

Sharma, P., Gupta, R., & Mishra, S. (2022). DenseNet for malignant breast anomaly classification: Insights and challenges. *Histopathology Insights, 28*(1), 23–39.

Sheikh, T. S., Lee, Y., & Cho, M. (2020). Histopathological classification of breast cancer images using a multi-scale input and multi-feature network. *Cancers, 12*(8). https://doi.org/10.3390/cancers12082031

Tan, M., Pu, J., & Zheng, B. (2014). Optimization of breast mass classification using sequential forward floating selection (SFFS) and a support vector machine (SVM) model. *International Journal of Computer Assisted Radiology and Surgery, 9*(6), 1005–1020. https://doi.org/10.1007/s11548-014-0950-4

Yari, Y., Nguyen, T. V., & Nguyen, H. T. (2020). Deep learning applied for histological diagnosis of breast cancer. *IEEE Access, 8*, 162432–162448. https://doi.org/10.1109/ACCESS.2020.3021557

Yiallourou, A. I. (2023). Hereditary breast cancer syndromes. *Breast Cancer Management for Surgeons: An Examination Guide*, 79.

Zahra, N., Ahmed, H., & Malik, S. (2023). Deep learning-based multi-network feature fusion for breast cancer detection. *Artificial Intelligence in Medicine, 120*(4), 1–12.

Zhou, Q., Zhu, W., Li, F., Yuan, M., Zheng, L., & Liu, X. (2022). Transfer learning of the ResNet-18 and DenseNet-121 model used to diagnose intracranial hemorrhage in CT scanning. *Current Pharmaceutical Design, 28*(4), 287–295.