Advanced Multi-Mammographic Classification Using an Efficient Class-Balanced Transfer Learning Approach

[1, 2]Adedayo-Ajayi Victoria Oluwaseyi, [1]Marion Olubunmi Adebiyi, [1, 3, 4, 5, 6, *]Roseline Oluwaseun Ogundokun, [7]Ibidun Obagbuwa, [3, *]Rotimi-Williams Bello, [3]Pius Adewale Owolawi

[1]Department of Computer Science, Landmark University, Omu-Aran, Kwara State, Nigeria
[2]Department of Cybersecurity, Ladoke Akintola University of Technology, Ogbomoso, Nigeria.
[3]Department of Computer Systems Engineering, Tshwane University of Technology, Pretoria, South Africa.
[4]Department of Computer Science, Redeemer's University Ede, Osun State, Nigeria.
[5]Department of Mathematical and Computing Sciences, Thomas Adewumi University, Oko, Nigeria
[6]Department of Data Science, Miva Open University, Abuja, Nigeria.
[7]Department of Computer Science, Sol-Plaatje University, Kimberley 8301, South Africa

*Corresponding authors: Roseline Oluwaseun Ogundokun and Rotimi-Williams Bello

Abstract

Breast cancer ranks as the second most impactful cancer globally, and timely detection with suitable treatment can mitigate the risk of fatalities. Despite advances in mammographic disease discrimination methods, advanced multi-mammographic classification remains challenging in resource-constrained healthcare environments, especially in developing countries where GPUs are limited. This work presents a deep learning approach tailored to the CPU with an efficient class-balanced transfer learning strategy for mammographic image classification. The development and validation of this model were conducted using a set of mammographic images categorised into four diagnostic classes: benign, bilateral, malignant, and normal images, all originating from a Nigerian health facility. A modified MobileNetV2 architecture with a custom preprocessing pipeline was used, which includes Contrast Limited Adaptive Histogram Equalisation (CLAHE) and dynamic class weighting to tackle dataset imbalance. The accuracy of the model across all classes was found to be 80.87%, with a strong performance in detecting malignant cases (recall 93.94%, AUC 0.98). On standard 8GB RAM hardware, the model exhibits efficient CPU-based performance, taking 47 milliseconds for preprocessing and 150 milliseconds for inference per image. Results of cross-validation analysis showed consistent performance within folds (mean validation accuracy 84.2%, SD: 1.8), which suggested robust generalisation. Finally, the results suggest that sophisticated mammographic classification systems can be effectively deployed in resource-constrained clinical settings without compromising diagnostic accuracy. This work presents a practical approach for implementing advanced medical image analysis in developing healthcare environments where computational resources can be scarce.

Keywords: Breast cancer, Mammographic, Transfer learning, MobileNetV2.

1. Introduction

Cancer remains a major global health concern, responsible for approximately 10 million deaths in 2020 alone. Among all cancer types, breast cancer stands as the most diagnosed and the leading cause of cancer-related deaths among women (Ferlay et al., 2020; Ma et al., 2020; Wang et al., 2024; Ogundokun et al., 2024). In 2020, an estimated 2.3 million new breast cancer cases and 0.7 million deaths were recorded globally, with projections indicating a further rise by 2030. This burden is particularly acute in low- and middle-income countries (LMICs), where factors such as changing lifestyles, inadequate early detection systems, and limited access to care have contributed to rising incidence and mortality rates (Patil & Biradar, 2021). Mammography remains the standard imaging modality for early breast cancer detection, offering the potential to reduce mortality through timely intervention (Nicosia et al., 2023). However, mammography is not without limitations—particularly the occurrence of false positives and false negatives, which are more frequent among younger women and those with dense breast tissue (Zhao et al., 2015). These inaccuracies can lead to delayed diagnoses, unnecessary interventions, and significant psychological distress. Advances in medical imaging and artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL) models like Convolutional Neural Networks (CNNs), have shown significant promise in enhancing diagnostic accuracy (Zeng et al., 2023; Alom et al., 2019; Jedy-Agba et al., 2016). These methods can efficiently analyse imaging data, automate detection processes, and improve classification performance. Nevertheless, challenges persist, especially when dealing with imbalanced datasets and heterogeneous patient populations.

Breast cancer diagnosis in African populations poses additional challenges. Most existing AI-based diagnostic models are trained on mammographic datasets from Western countries, which may not reflect the unique epidemiological and clinical characteristics of African women, such as younger onset ages and more aggressive tumor subtypes (Ragab et al., 2019). These differences, compounded by variations in breast density, imaging protocols, and healthcare infrastructure, limit the generalizability and reliability of imported AI models. Furthermore, the persistent issue of class imbalance, where malignant cases are underrepresented, significantly affects the sensitivity and robustness of existing classification systems. Addressing these challenges requires developing region-specific models that are tailored to local datasets and capable of generalising effectively across diverse populations. This paper proposes to achieve the following objectives

1. Design an efficient deep learning model based on MobileNetV2 for transfer learning
2. Implement this model on computationally constrained devices while integrating class-balancing techniques
3. Evaluate its performance using standard metrics such as accuracy, precision, recall, F1-score, and AUC
4. Validate the model's generalisation capabilities through cross-validation across varied patient groups.

This research seeks to contribute a robust, scalable, and context-aware solution for improving breast cancer diagnosis in Nigeria and other underserved settings.

The paper is organised as follows: Section 2 presents a comprehensive review of the related literature and empirical findings. In Section 3, the methodology is discussed, outlining the research design, model specification, material used, and data processing techniques. Section 4 focuses on the evaluation metrics and presents the experimental findings. Finally, Section 5 offers a detailed discussion of the results, conclusions, and recommendations for future research.

2. Related Works

Medical imaging, specifically the automatic classification of breast cancer given mammographic images, is one of the areas where deep learning has been used with increasing frequency. Such a paradigm shift has led to the emergence of various CNN-based models, which can extract complex features and improve diagnostic performance compared to traditional machine learning methods. Muduli et al. (2022) developed an attention-based CNN for mammographic classification, which enhanced feature extraction efficiency and reported an AUC of 94.1%, F1-score of 90.2%, and recall of 91.5%. However, the study faced challenges related to dataset imbalance. To avoid overfitting and unbalanced class problems, Srikantamurthy et al. (2023) performed multi-scale data augmentation. Deep convolutional features were extracted using ResNet50 CNN, pre-trained on ImageNet, to classify breast cancer as benign or malignant through a hybrid CNN-LSTM-based transfer learning approach. Similarly, Garg & Singh (2022) proposed an ensemble deep learning model to detect breast cancer. The method demonstrated high precision; however, the model was overfitted, and the minority classes were not well detected due to dataset imbalance.

Likewise, Alshamrani & Alshomran (2024) employed a transfer learning-based lightweight ensemble model on mammogram images for multi-class classification using an imbalanced IDC dataset, an imbalanced BreakHis dataset, and a balanced BACH dataset. A lightweight, shallow CNN model with batch normalisation technology to accelerate convergence was combined with MobileNetV2 to enhance learning and adaptability. Related work by Xiong et al. (2025) proposed a multichannel feature fusion network for heart sound classification, demonstrating the effectiveness of channel-wise fusion and hierarchical feature extraction across diverse biomedical signals. Their architecture supports the value of integrating multimodal and multifeature learning strategies, which align with this study's transfer learning and feature-enhanced mammographic classification. Similarly, Xiang et al. (2025) developed a multimodal masked autoencoder with adaptive masking for vitiligo stage classification, emphasising the advantages of self-supervised representation learning and adaptive masking—concepts relevant for robust mammogram feature extraction under limited labelled data.

In addition, deep generative models and transformer architecture have gained significant traction in medical imaging for improving feature localisation and image quality. Yang et al. (2025) utilised a deep generative network for disparity estimation of stereo-endoscopic images, which provides insights into leveraging generative models for precise medical image reconstruction. Luan et al. (2023) extended this notion with a deep-learning-based super-resolution framework for ultrasound microvessel imaging, achieving high spatial fidelity and faster processing—paralleling our work's focus on computational efficiency in constrained environments. Song et al. (2024)

introduced CenterFormer, a transformer-based segmentation model that leverages cluster centre enhancement to achieve high-precision dental plaque segmentation. The attention-driven design in CenterFormer underlines how transformer paradigms can enhance spatial representation, a concept that may further improve mammographic lesion detection.

Beyond imaging, several studies have explored the biological and clinical dimensions of disease classification and modelling, showing the interdisciplinary reach of AI. Tang et al. (2024) presented a deep-learning approach for tendinopathy subtype classification using multimodal integration, advancing precision therapeutics through discriminative representation learning. Zhou et al. (2021) investigated RHPN1-AS1's oncogenic role in ovarian carcinogenesis via a microRNA-mediated pathway, showing how molecular-level insights can complement image-based AI models for cancer detection. Gao et al. (2022) analysed the molecular mechanisms of heat stress on spermatogenesis, illustrating the value of biological context in understanding pathological manifestations seen in imaging. Similarly, Yang et al. (2025) evaluated the safety and efficacy of inactivated COVID-19 vaccination in couples undergoing assisted reproductive technology, highlighting the significance of AI-assisted clinical data analysis in improving healthcare outcomes.

From a computational perspective, scalable and efficient architectures have also been explored to optimise deep learning systems. Gan (2025) proposed GraphService, a topology-aware constructor for large-scale graph applications that improves performance through communication-efficient graph partitioning. The emphasis on computational scalability and topology awareness resonates with the CPU-optimised transfer learning approach in this study, particularly in resource-constrained clinical environments.

A major gap in the literature is the scarcity of studies using African or Nigerian mammogram datasets. This lack of local data hinders the development of diagnostic models that generalise well in these populations. Furthermore, few studies have explored class-balanced transfer learning tailored to heterogeneous and imbalanced clinical datasets. Given these gaps, this study proposes an efficient multi-class classification framework that integrates class-balanced data handling with transfer learning and deep CNN architectures, leveraging mammographic images obtained from two Nigeria Hospitals. This approach aims to enhance diagnostic accuracy across all breast cancer classes, particularly in underrepresented populations.

3. Materials and Methods

3.1 Materials

3.1.1 Proposed Architecture

Figure 1 illustrates the overall architectural framework for mammographic classification using MobileNetv2 and transfer learning to enhance computational efficiency. The proposed framework shows the various components of the model, from the dataset acquisition, data pre-processing, model design and training, and model validation.

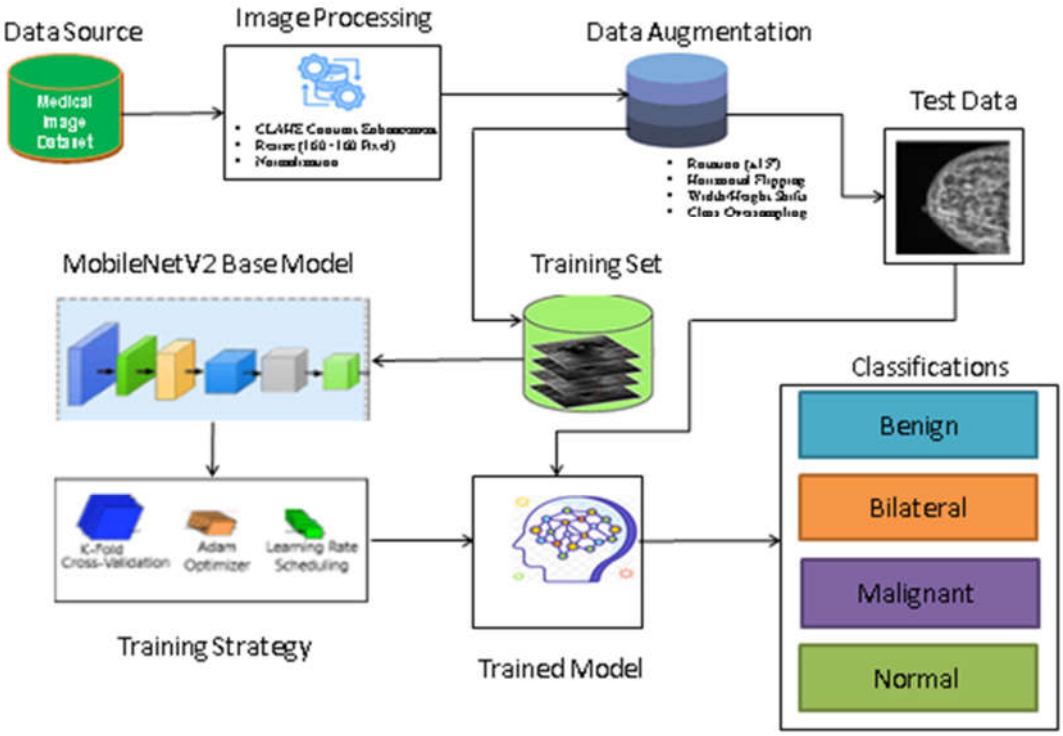

Figure 1. The Proposed Framework

3.1.2 Dataset Description

The dataset used in this study consisted of 2,310 mammographic images obtained from ME CURE Hospital and Ladoke Akintola University of Technology, Oyo State, Nigeria. The images in all cases were captured on digital mammography systems using standard imaging protocols, resulting in consistent image quality. The data collection process fully followed ethical guidelines and data anonymisation throughout the collection phase. The dataset contains a mixed population of Nigerian women aged 25 to 70, with varying breast tissue density and pathological conditions. Rigorous review and annotation of each mammographic image was performed by experienced radiologists with at least ten years of experience in breast imaging. Table 1 describes the different categories of mammographic images, categorised into four distinct pathological characteristics.

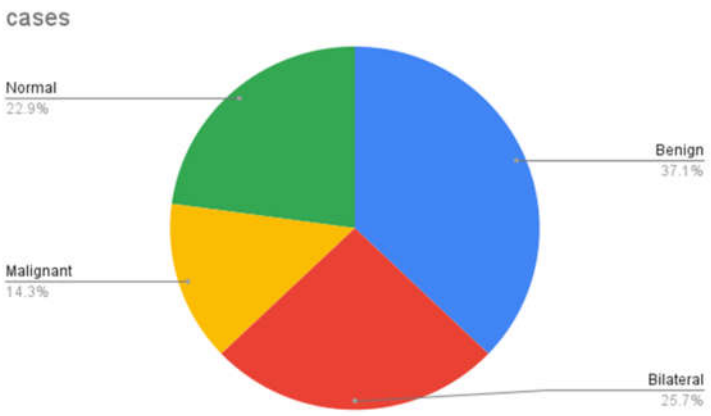

Figure 2. Distribution of Mammographic Images Across Diagnostic Categories

A notable characteristic of the dataset is its inherent class imbalance. To quantify this imbalance, we calculated the class imbalance ratio (IR) using Equation 1:

$$IR = \frac{N_{maj}}{N_{min}} = \frac{858}{330} \approx 2.6 \qquad (1)$$

where N_{maj} represents the majority class (benign) and N_{min}epresents the minority class (malignant). To address this imbalance, class weights were computed using Equation 2:

$$W_c = \frac{N_{total}}{N_c \times N_{classes}} \qquad (2)$$

where:

W_c is the weight for class c

N_{total} is the total number of samples (2,310)

N_c is the number of samples in class c

$N_{classes}$lasses are the number of classes (4)

This formula yielded the following class weights: benign cases served as the reference class and had a weight of 1.0, bilateral cases had a weight of 1.44, malignant cases were weighed at 2.60, and normal cases had a weight of 1.62. These weights are important in training the model because they ensure the process is balanced across all classes. The dataset organisation is structured, with images placed in a hierarchical directory system. Each image is at its original high resolution, originally from 2994×2294 through 3584×2816 pixels with 16-bit depth to capture all subtle tissue characteristics needed for correct diagnosis. Table 1 shows the distribution of mammographic images and the class weight across diagnostic categories.

Table 1: Distribution of Mammographic Images Across Diagnostic Categories.

Category	Number of Cases	Class weight	Percentage
Benign	858	1.0	37.14%
Bilateral	594	1.44	25.71%
Malignant	330	2.60	14.29%
Normal	528	1.62	22.86%
Total	2310	6.66	100%

3.1.3 Data Preprocessing

The preprocessing pipeline developed for this research implements a comprehensive approach to standardise and enhance mammographic images before they are fed into the neural network. The pipeline consists of three primary stages: Contrast Limited Adaptive Histogram Equalisation (CLAHE), image normalisation, and dimensional standardisation, as shown in Algorithm 3.1.

Our preprocessing pipeline deploys CLAHE as its first step because this advanced enhancement technique demonstrates special effectiveness in medical image processing. The analysis of different image sections through multiple histograms enables CLAHE to distribute lightness values throughout the image. The mammographic images are then normalised to guarantee uniform input scaling throughout the data set. This normalisation process takes the range of pixel values from the usual 0-255 scale to a float between 0 and 1. The last step in the preprocessing pipeline is to spatial standardise the images, where all images are resized to a fixed dimension of 160×160 pixels with three channels of colour (160×160×3).

Algorithm 3.1: Data preprocessing Algorithm

Step 1 for each image x_i in D do

2: Apply CLAHE enhancement: $x_i' = CLAHE(x_i, clip\_limit=2.0, grid\_size=8\times8)$

3: Normalize pixel values: $x_i'' = x_i' / 255.0$

4: Resize to standard dimensions: $x_i''' = Resize(x_i'', (160, 160))$

5: end for

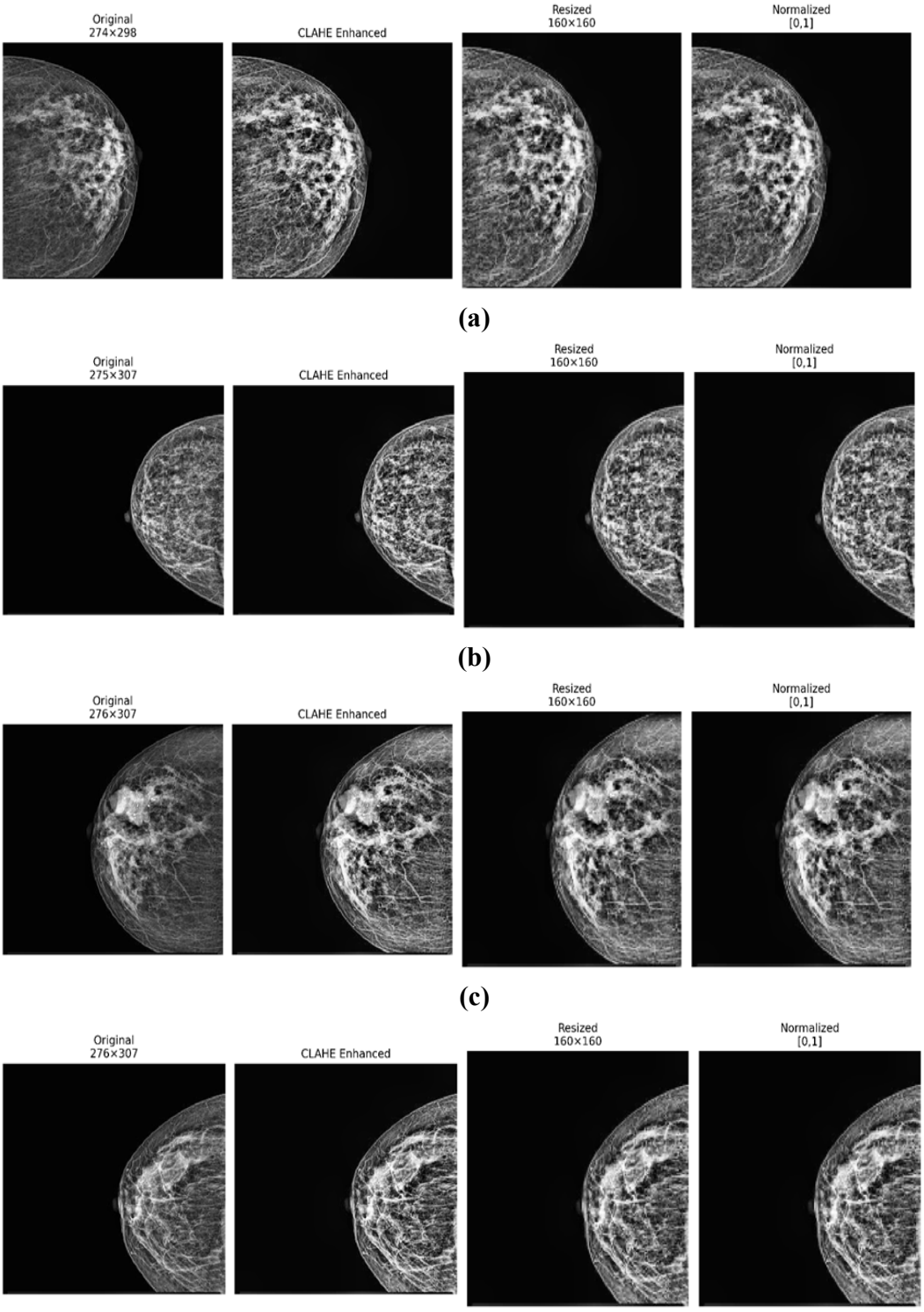

Figure 3 (a-d). Benign, Bilateral, Malignant and Normal Preprocessing, respectively.

Figure 3 above shows the Mammogram preprocessing pipeline visualisation from each class in the dataset. From left to right: (1) Original mammogram image [dimensions: WxH]; (2) After CLAHE

enhancement showing improved contrast; (3) Resized image [160×160]; (4) Normalised image scaled to [0,1] range. The enhancement in tissue detail and structural clarity is evident through each preprocessing stage. Table 2 presents the preprocessing pipeline performance metrics used for the experimentation.

Table 2. Preprocessing Pipeline Performance Metrics

Preprocessing Stage	Average Processing Time (ms)	Memory Usage (MB)
CLAHE	28	12.4
Normalization	4	2.8
Resizing	15	3.6
Total Pipeline	47	18.8

3.2 Method

3.2.1 The model

The model utilises MobileNetV2 convolutional neural network as its foundation for mammographic image classification, while having computational efficiency for CPU-based deployment. The purpose of this architectural decision was driven by the need for models to have a certain level of complexity due to performance limitations in resource-constrained clinical settings like Nigeria.

3.2.2 The MobileNetV2 layer

The base MobileNetV2 architecture was modified (see figure 4), while retaining the core inverted residual structure, by adapting the input layer to the mammographic images of size 160×160×3. With this modification, the computational overhead is reduced to be less than that of the standard 224×224×3 inputs, while still preserving enough spatial information for accurate classification. The network utilises depth-wise separable convolutions, which factorise standard convolutions into depth-wise and pointwise convolutions, reducing the computational complexity from O($h \times w \times c_{in} \times c_{out} \times k \times k$) to O($h \times w \times c_{in} \times (k \times k + c_{out})$), where h and w represent spatial dimensions, c_{in} and c_{out} The input and output channels are represented, and k represents the kernel size. Mathematically, the inverted residual block, or core component of MobileNetV2, can be expressed as:

$$X_{l+1} = F(X_l) + X_l \quad \text{If stride} = 1 \tag{3}$$

$$X_{l+1} = F(X_l) \quad \text{If stride} = 2 \tag{4}$$

where F represents the residual function.

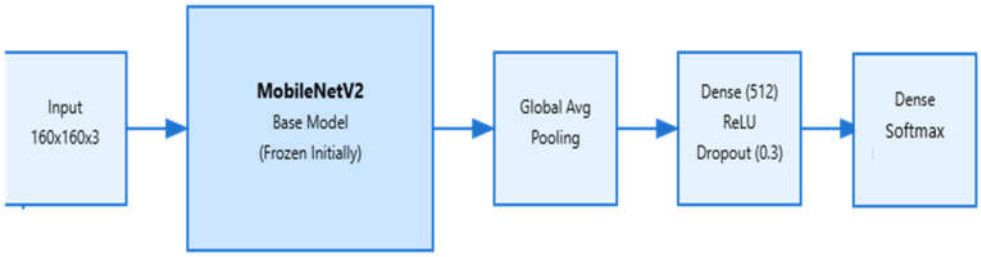

Figure 4. MobileNetV2-based Mammographic Classification Architecture

3.2.3 Transfer Learning

To address the challenges of mammographic classification in a resource-constrained environment, we employed a transfer learning approach using the MobileNetV2 architecture, pre-trained on the ImageNet dataset. This architecture, known for its lightweight and efficient structure featuring inverted residual blocks and linear bottlenecks, was selected for its suitability in CPU-based systems. The model development followed a two-phase strategy. In the initial phase, all layers of the MobileNetV2 base were frozen to preserve pre-trained weights, while only the custom classification head, which comprises a global average pooling layer, a dense layer with 512 units (ReLU activation), and a dropout layer (rate = 0.3), was trained. The final layer used softmax activation for multi-class classification.

In the second phase, selective fine-tuning of the base layers was conducted to enhance task-specific feature learning. A learning rate of 5×10^{-4} was applied initially, with dynamic adjustment via the ReduceLROnPlateau mechanism, reducing the rate by a factor of 0.5 upon performance plateau, down to a minimum of 1×10^{-6}. Early stopping was implemented to prevent overfitting by monitoring validation loss. To counter the dataset's class imbalance, class weighting was incorporated during training, allowing the model to learn from all classes effectively. This comprehensive transfer learning pipeline ensures robust feature extraction and classification performance across the four mammographic categories.

3.2.4 Cross-Validation

Our mammographic classification model underwent extensive evaluation through a k-fold cross-validation approach, which utilised a 5-fold cross-validation protocol to assess its robustness and generalisation capabilities. The dataset was partitioned into five stratified folds, ensuring class balance and allowing each sample to serve in both training and validation phases. The final cross-validation performance is assessed through both the mean and the standard deviation of the fold scores:

$$Mean\ Score = \frac{1}{5} \times \sum_{i=1}^{5}(Fold\ Score_i) \quad (5)$$

$$Standard\ Deviation = \sqrt{\left(\frac{1}{4} \times \sum_{i=1}^{5}(Fold\ Score_i - Mean\ Scoree)^2\right)} \quad (6)$$

The steps for the K-fold Cross-Validation Training are as depicted in Algorithm 3.2.

Algorithm 3.2: K-fold Cross-Validation Training

1nitialize 5-fold cross-validation: KFold(n_splits=5, shuffle=True, random_state=42)
for each fold i from 1 to 5 do
Split D into training set Dtrain and validation set Dvalid
Apply data augmentation on Dtrain: rotation(±15°), width/height shifts(±10%), horizontal flip
Apply class balancing on Dtrain using weights W
Initialise Adam optimiser with learning rate lr = 5 × 10⁻⁴
Train model with batch size = 16 using class weights W
Monitor validation loss and implement early stopping with patience = 8
Implement learning rate reduction: LRt+1 = max(0.5 × LRt, 1 × 10⁻⁶) when plateau
Save the best model weights based on validation accuracy
Evaluate the model on the validation set and record metrics
Clear memory and reset model state
end for

4. Results and Discussions

This section presents the results and findings of the research focused on developing an Advanced Multi-Mammographic Classification Using an Efficient Class-Balanced Transfer Learning Approach.

4.1 Cross-validation performance

Our transfer learning method shows robust performance across all the evaluation metrics considered in this study. Achieving an overall accuracy of 80.87% on the test set, the model displayed variations across the different classes due to the intrinsic difficulty of mammographic classification. The model performance was consistent across different data folds in the cross-validation analysis, with a validity accuracy of 82-87%. In Figure 5, we observe that the peak validation accuracy (87%) was achieved in Fold 3, and the lowest performance (82%) was in Fold 4. This stability across folds implies the model's ability to learn generalizable features without overfitting specific data subsets. The Learning was stable, as the mean validation accuracy across all folds was 84.2% with a Standard deviation of 1.8.

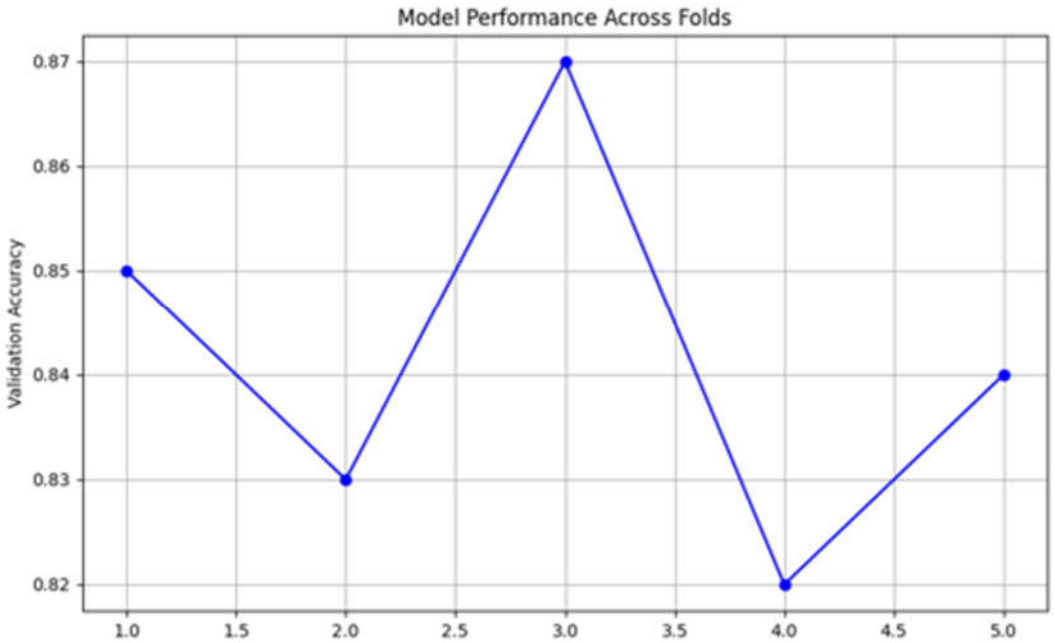

Figure 5. Model performance across the five folds

4.1.1 Efficient Analysis

The implementation of our CPU-based mammographic classification system demonstrates very good performance in terms of resource utilisation and training time optimisation, as shown in Figure 6. We demonstrate the practical viability for deployment in resource-constrained environments by successfully executing on a system with 8GB of RAM. An interesting pattern is uncovered in the training time analysis across the fivefold cross-validation process. The time to complete the initial fold was 2.9 minutes, while subsequent folds varied, ranging from 2.0 minutes for fold 1, to 1.0 minutes for fold 2, to 1.1 minutes for fold 3, and 1.5 minutes for fold 4. With this architecture, the training time on all folds was 8.5 minutes, and the average training time per fold was 1.7 minutes. This variation in the training times stems from the dynamic nature of our learning rate scheduling and early stopping criteria, which optimise the training duration based on model convergence patterns. Table 3 illustrates the temporal distribution of training times across folds.

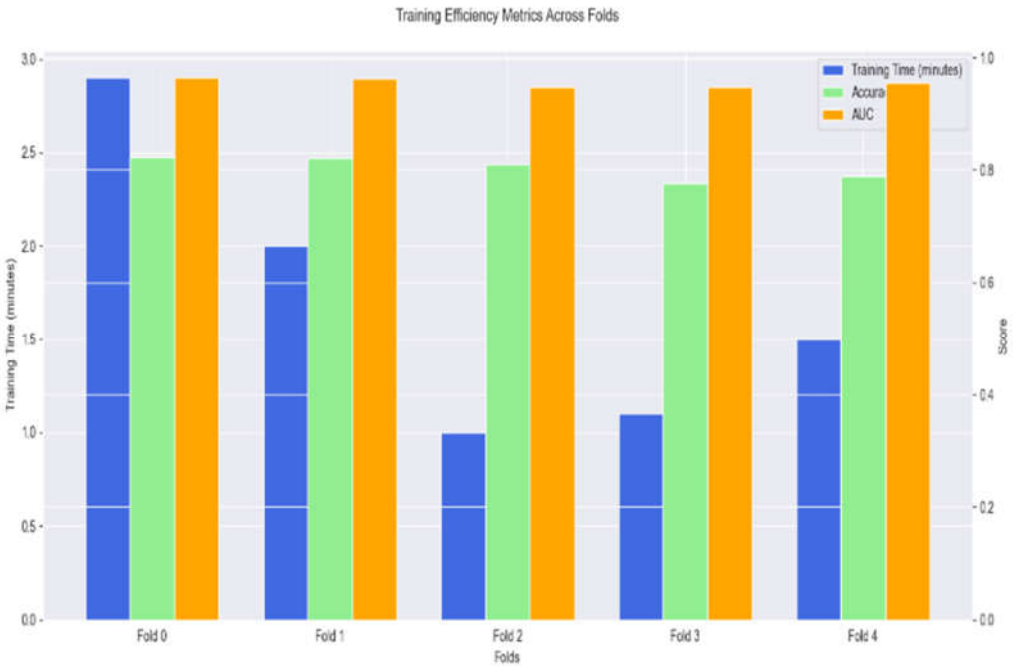

Figure 6. Training Efficiency metrics across folds

Table 3. Training Time and Reduction across Folds

Fold	Time (minutes)	Reduction from Initial Fold
Fold 0	2.9	-
Fold 1	2.0	31%
Fold 2	1.0	66%
Fold 3	1.1	62%
Fold 4	1.5	48%

4.2 Performance Evaluation Metrics

The performance of the proposed model is assessed using a range of classification metrics, including accuracy, precision, recall, F1-score, and area under the ROC curve (AUC) across the four different categories: benign, bilateral, malignant, and normal. These metrics provide a comprehensive evaluation of the model's effectiveness in accurately detecting and classifying mammographic lesions across multiple classes. The model's achieved overall accuracy of 80.87%

across all classes suggests good. The classification report obtained from the experiment conducted is shown in Table 4.

The model achieved the highest precision of 0.9222 for the Benign class, indicating strong reliability in correctly identifying benign cases. However, the *Malignant* class recorded the lowest Precision (0.6352), suggesting some misclassifications, likely due to class imbalance or overlapping feature characteristics. Despite this, the model showed an excellent Recall of 0.9394 for Malignant cases, meaning it successfully identified most true malignant instances. This is crucial in medical diagnostics, where missing malignant cases could have severe implications.

For the Bilateral and Normal classes, both Precision and Recall were consistently strong, reflecting the model's stable performance across non-malignant categories. The F1-Score, which harmonises Precision and Recall, ranged from 0.7579 (Malignant) to 0.8332 (Benign), indicating a balanced trade-off between false positives and false negatives.

Table 4. Performance Metrics of the Mammographic Classification

Class	Precision	Recall	F1-Score	Support
Benign	0.922206506	0.75990676	0.833226837	858
Bilateral	0.829015544	0.808080808	0.818414322	594
Malignant	0.635245902	0.939393939	0.75794621	330
Normal	0.794776119	0.806818182	0.80075188	528
Accuracy	0.808658009	0.808658009	0.808658009	0.808658009
Macro Avg	0.795311018	0.828549922	0.802584812	2310
Weighted Avg	0.828121798	0.808658009	0.811240682	2310

To further evaluate the model’s performance, both the AUC-ROC were plotted for visualisation, as presented in Figure 7.

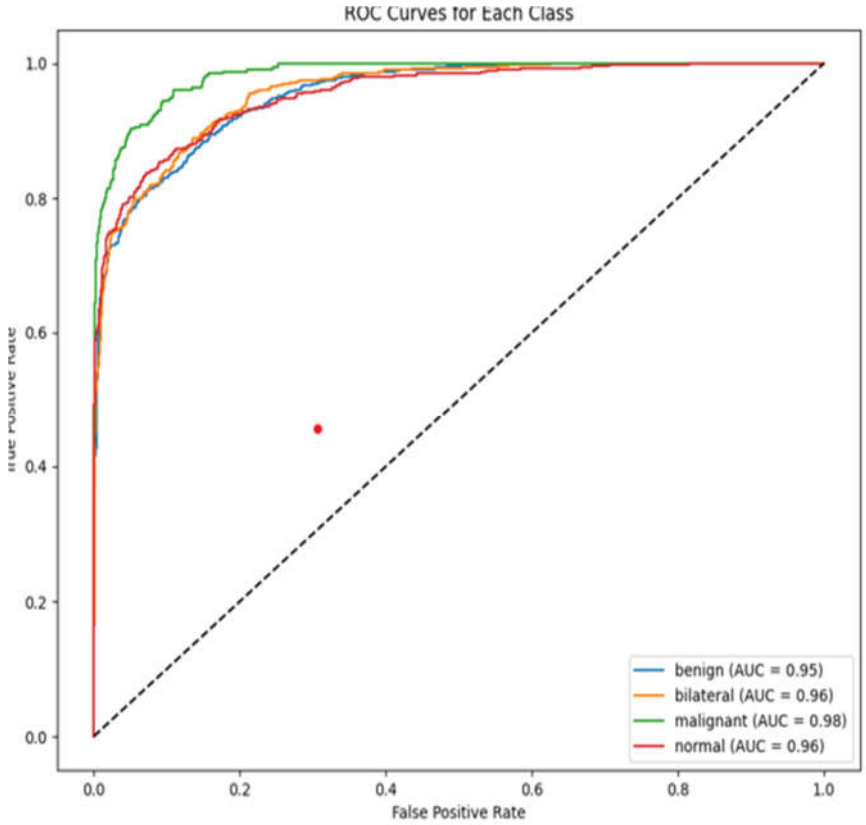

Figure 7. Area Under the Receiver Operating Characteristic curve (AUC-ROC)

4.2. Confusion Matrix

Figure 8 illustrates a confusion matrix for the multi-class classification of mammographic images into Benign, Bilateral, Malignant, and Normal categories, further demonstrating the model's performance. From this matrix, correctly classified cases are 652 as benign, 480 as Bilateral, 310 as Malignant and 426 as Normal, while misclassifications occurred in all cases as follows:

- For benign cases, 65 were misclassified as bilateral, 84 as malignant, and 57 as normal.
- For the bilateral cases, 28 were misclassified as benign, 45 as malignant and 41 as normal.
- For malignant cases, there were minimal misclassifications, with 1 benign, 7 bilateral and 12 normal.
- For normal cases, 26 were misclassified as benign, 27 as bilateral and 49 as malignant.

These results indicate a high overall recognition rate, particularly for the malignant class, which is clinically critical. However, there is observable confusion between benign and malignant cases, which can be attributed to overlapping features in mammographic images. The relatively high number of correct predictions along the diagonal of the matrix supports the model’s reliability in real-world breast cancer diagnosis scenarios.

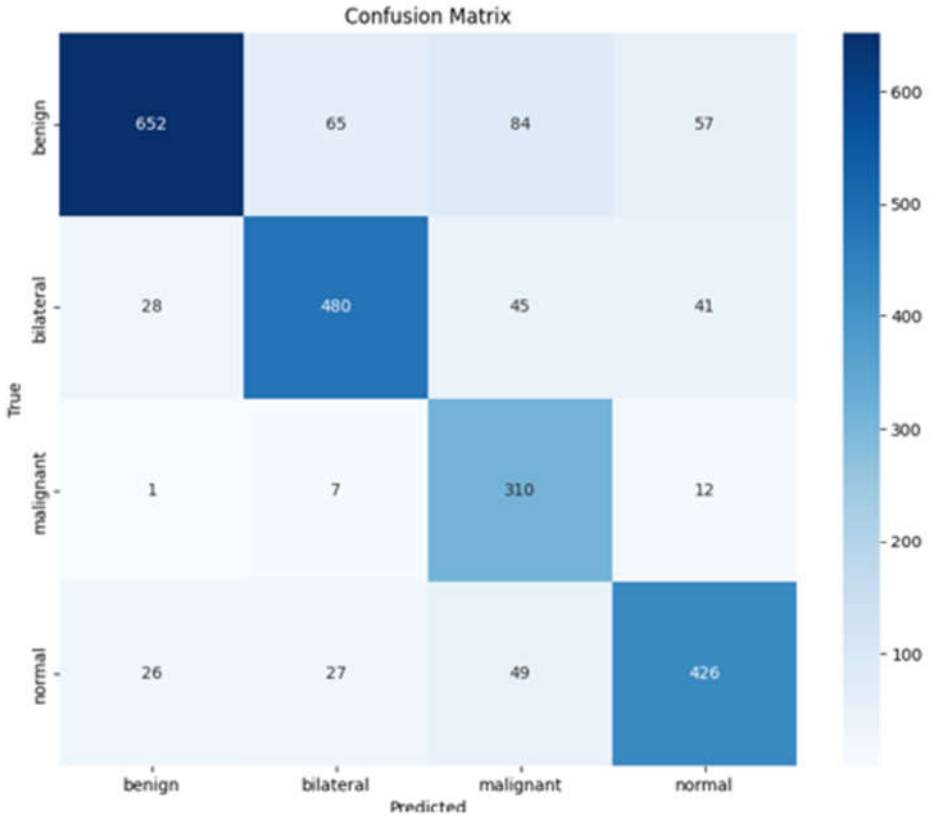

Figure 8. Model Confusion Metrics

4.3. Comparative Analysis

As illustrated in Table 5, this section presents a comparison of the performance of the proposed model with some other studies. The comparative evaluation of mammographic classification models highlights the strengths and trade-offs of various architectural and training approaches across diverse datasets and computational settings, as shown in Table 5. The proposed MobileNetV2-based model achieves an accuracy of 80.87%, a malignant recall of 93.94%, and an AUC of 0.98, outperforming all other studies regarding malignancy detection sensitivity. This is particularly significant for clinical applications where false negatives in malignant cases can have critical consequences. This result is even more remarkable because the model was optimised explicitly for CPU deployment.

Table 5: Performance Comparison.

Study	Model	Dataset	Accuracy (%)	Recall (Malignant)	AUC (Malignant)	Remarks
Proposed Study	**MobileNetV2 (Transfer Learning + Class Weights)**	**2,310 Nigerian mammograms (4 classes)**	**80.87**	**93.94**	**0.98**	**CPU-optimised with CLAHE and class-balancing; lightweight deployment**
Rakhlin *et al.* (2018).	DenseNet + Ensemble Learning	INbreast	86.7	90.2	0.96	Strong results using ensembling and a deep DenseNet backbone
Shen *et al.* (2019).	ResNet50 (Transfer Learning)	DDSM	82.3	89.5	0.94	Fine-tuned ResNet50 for 2-class mammogram diagnosis
Arevalo *et al.* (2016).	CNN + MLP (Feature fusion)	BCDR-F03	78.1	87.0	0.92	Early CNN fusion model using handcrafted and deep features
Khamparia *et al.* (2021).	VGG16 + LSTM hybrid	DDSM	84.9	91.8	0.95	Hybrid spatial-temporal modelling of mammogram sequences

While other models demonstrate solid accuracy in controlled environments, the proposed approach distinguishes itself by delivering high sensitivity to malignancy, strong generalisation, and efficient performance within a low-resource, real-world clinical deployment context.

5. Conclusion and Future Works

The results of this study indicate an efficient Mammographic classification model enabling CPU-based deployment within resource-constrained healthcare settings. Our transfer learning approach and class balancing strategies achieve an overall accuracy of 80.87%, with great performance in malignant case detection (93.94% recall, AUC 0.98), validating the efficacy of our proposed and applied ideas. Being able to run the system within standard hardware constraints (8GB RAM) and only taking 47 milliseconds for preprocessing and 150 milliseconds for inference per image makes

it particularly suitable for practical clinical deployment in developing healthcare settings. The ability to successfully integrate MobileNetV2 architecture with custom preprocessing techniques and dynamic class weighting shows that sophisticated medical image analysis can be performed without reliance on specialised hardware acceleration. The ability to generalise well across cross-validation folds (Mean validation accuracy of 84.2%, SD: 1.8%) implies robust generalisation and efficient parameter utilisation (657,924 trainable parameters), indicating successful transfer of knowledge from the pre-trained network to mammographic classification.

Future research directions for this mammographic classification system present numerous promising avenues for advancement and expansion. It is hoped that the contribution of advanced attention mechanisms and transformer architectures may improve the model's performance, particularly in feature extraction for cases involving subtle tissue variations in bilateral regions.

Acknowledgement: The authors gratefully acknowledge the Department of Computer Systems Engineering, Tshwane University of Technology (TUT), for providing financial support for the Article Processing Charge (APC) associated with the publication of this paper.

References

1. Ferlay, J.; Laversanne, M.; Ervik, M.; Lam, F.; Colombet, M.; Mery, L.; Piñeros, M.; Znaor, A.; Soerjomataram, I.; Bray, F. Global Cancer Observatory: Cancer Tomorrow. International Agency for Research on Cancer: Lyon, France, 2020; Available online:https://gco.iarc.fr/tomorrow (accessed on 9 July 2021).
2. Ma, X., Cheng, H., Hou, J., Jia, Z., Wu, G., Lü, X.,... Chen, C. (2020). Detection of breast cancer based on novel porous silicon Bragg reflector surface-enhanced Raman spectroscopy-active structure. Chinese Optics Letters, 18(5), 051701. doi: 10.3788/COL202018.051701.
3. Wang, Z. B., Zhang, X., Fang, C., Liu, X. T., Liao, Q. J., Wu, N., & Wang, J. (2024). Immunotherapy and the ovarian cancer microenvironment: Exploring potential strategies for enhanced treatment efficacy. Immunology, 173(1), 14–32. https://doi.org/10.1111/imm.13793.
4. Zeng, Q., Chen, C., Chen, C., Song, H., Li, M., Yan, J.,... Lv, X. (2023). Serum Raman spectroscopy combined with convolutional neural network for rapid diagnosis of HER2 positive and triple-negative breast cancer. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 286, 122000. doi: https://doi.org/10.1016/j.saa.2022.122000.
5. Ogundokun R. O. et al., "Hybrid Deep Learning for Breast Cancer Diagnosis: Evaluating CNN and ANN on BreakHis_v1_400X," 2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG), Omu-Aran, Nigeria, 2024, pp. 1-6, doi: 10.1109/SEB4SDG60871.2024.10629774.

6. Patil, R. S., & Biradar, N. (2021). Automated mammogram breast cancer detection using the optimized combination of convolutional and recurrent neural network. Evolutionary intelligence, 14(4), 1459-1474.
7. Nicosia L, Gnocchi G, Gorini I, Venturini M, Fontana F, Pesapane F, Abiuso I, Bozzini AC, Pizzamiglio M, Latronico A, Abbate F, Meneghetti L, Battaglia O, Pellegrino G, Cassano E. History of Mammography: Analysis of Breast Imaging Diagnostic Achievements over the Last Century. Healthcare (Basel). 2023 May 30;11(11):1596. doi: 10.3390/healthcare11111596. PMID: 37297735; PMCID: PMC10252579.
8. Zhao, H., Zou, L., Geng, X. et al. Limitations of mammography in the diagnosis of breast diseases compared with ultrasonography: a single-center retrospective analysis of 274 cases. Eur J Med Res 20, 49 (2015). https://doi.org/10.1186/s40001-015-0140-6
9. Alom, Md. Zahangir & Yakopcic, Chris & Hasan, Mahmudul & Taha, Tarek & Asari, Vijayan. (2019). Recurrent residual U-Net for medical image segmentation. Journal of Medical Imaging. 6. 10.1117/1.JMI.6.1.014006.
10. Jedy-Agba, E., Curado, M. P., Ogunbiyi, O., & et al. (2016). Cancer incidence in Nigeria: A report from population-based cancer registries. Cancer Epidemiology, 44, 9–13. https://doi.org/10.1016/j.canep.2016.07.007
11. Ragab DA, Sharkas M, Marshall S, Ren J. 2019. Breast cancer detection using deep convolutional neural networks and support vector machines. PeerJ 7:e6201 https://doi.org/10.7717/peerj.6201
12. Muduli, D., Dash, R., & Majhi, B. (2022). Automated diagnosis of breast cancer using multi-modal datasets: A deep convolution neural network based approach. Biomedical Signal Processing and Control, 71, Article 102825. https://doi.org/10.1016/j.bspc.2021.102825
13. Srikantamurthy, M.M., Rallabandi, V.P.S., Dudekula, D.B. *et al.* Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning. *BMC Med Imaging* **23**, 19 (2023). https://doi.org/10.1186/s12880-023-00964-0
14. Garg, S., & Singh, P. (2022). Transfer learning based lightweight ensemble model for imbalanced breast cancer classification. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 20(2), 1529-1539.
15. Alshamrani, A. F. A., & Alshomran, F. (2024). Optimizing Breast Cancer Mammogram Classification Through a Dual Approach: A Deep Learning Framework Combining ResNet50, SMOTE, and Fully Connected Layers for Balanced and Imbalanced Data. IEEE Access.
16. Xiong, W., Zhang, G., Yan, D., Cao, L., Huang, X.,... Li, D. (2025). Multichannel feature fusion network-based technique for heart sound signal classification and recognition. Expert Systems with Applications, 273, 126839. doi: https://doi.org/10.1016/j.eswa.2025.126839

17. Xiang, F., Li, Z., Jiang, S., Li, C., Li, S., Gao, T.,... Zhang, J. (2025). Multimodal Masked Autoencoder Based on Adaptive Masking for Vitiligo Stage Classification. Journal of Imaging Informatics in Medicine. doi: 10.1007/s10278-025-01521-7

18. Yang, B., Xu, S., Yin, L., Liu, C., & Zheng, W. (2025). Disparity estimation of stereo-endoscopic images using deep generative network. ICT Express, 11(1), 74-79. doi: https://doi.org/10.1016/j.icte.2024.09.017

19. Luan, S., Yu, X., Lei, S., Ma, C., Wang, X., Xue, X.,... Zhu, B. (2023). Deep learning for fast super-resolution ultrasound microvessel imaging. Physics in Medicine & Biology, 68(24), 245023. doi: 10.1088/1361-6560/ad0a5a

20. Song, W., Wang, X., Guo, Y., Li, S., Xia, B.,... Hao, A. (2024). CenterFormer: A Novel Cluster Center Enhanced Transformer for Unconstrained Dental Plaque Segmentation. IEEE Transactions on Multimedia, 26, 10965-10978. doi: 10.1109/TMM.2024.3428349

21. Tang, C., Wang, Z., Xie, Y., Fei, Y., Luo, J., Wang, C.,... Shen, W. (2024). Classification of distinct tendinopathy subtypes for precision therapeutics. Nature Communications, 15(1), 9460. doi: 10.1038/s41467-024-53826-w

22. Zhou, Y., Li, J., Yang, X., Song, Y., & Li, H. (2021). Rhophilin rho GTPase binding protein 1-antisense RNA 1 (RHPN1-AS1) promotes ovarian carcinogenesis by sponging microRNA-485-5p and releasing DNA topoisomerase II alpha (TOP2A). Bioengineered, 12(2), 12003–12022. https://doi.org/10.1080/21655979.2021.2002494

23. Gao, Y., Wang, C., Wang, K., He, C., Hu, K.,... Liang, M. (2022). The effects and molecular mechanism of heat stress on spermatogenesis and the mitigation measures. Systems Biology in Reproductive Medicine, 68(5-6), 331-347. doi: https://doi.org/10.1080/19396368.2022.2074325

24. Yang, J., Yao, Y. L., Lv, X. Y., Geng, L. H., Wang, Y., Adu-Gyamfi, E. A., Wang, X. J., Qian, Y., Chen, M. X., Zhong, Z. H., Li, R. Y., Wan, Q., & Ding, Y. B. (2025). The Safety and Efficacy of inactivated COVID-19 vaccination in couples undergoing assisted reproductive technology: A prospective cohort study. Vaccine, 45, 126635. https://doi.org/10.1016/j.vaccine.2024.126635

25. Gan, X. (2025). GraphService: Topology-aware Constructor for Large-scale Graph Applications. ACM Trans. Archit. Code Optim., 22(1), 2. doi: 10.1145/3689341

26. Rakhlin, A., Shvets, A., Iglovikov, V., & Kalinin, A. A. (2018). Deep convolutional neural networks for breast cancer histology image analysis. In Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27–29, 2018, Proceedings 15 (pp. 737-744). Springer International Publishing.

27. Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. (2019). Deep learning to improve breast cancer detection on mammography. Scientific Reports, 9(1), 1–12. https://doi.org/10.1038/s41598-019-48995-4
28. Arevalo, J., González, F. A., Ramos-Pollán, R., Oliveira, J. L., & Lopez, M. A. G. (2016). Representation learning for mammography mass lesion classification with convolutional neural networks. Computer methods and programs in biomedicine, 127, 248-257.
29. Khamparia, A., Gupta, D., de Albuquerque, V. H. C., Sangaiah, A. K., & Jhaveri, R. H. (2020). Internet of health things-driven deep learning system for detection and classification of cervical cells using transfer learning. The Journal of Supercomputing, 76, 8590-8608.