FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 12, June, 2026, pp 362 – 367
DOI: https://doi.org/10.33003/fjs-2026-1012-4866

Transfer Learning Model for Breast Cancer Detection Using Mammograms from Low-Resource Settings

***[1]Zinat Abdulkadiri, [2]Muhammad A. Suleiman and [2]Joshua Abah**

[1]Department of Information Technology, Faculty of Computing, Nile University of Nigeria, FCT Abuja, Nigeria.
[2]Department of Computer Science, Faculty of Computing, Nile University of Nigeria, FCT Abuja, Nigeria.

*Corresponding authors' email: abdulkadirizinat@gmail.com

Received: 04th May 2026
Accepted: 15th July 2026
Published: 08th August 2026

ABSTRACT
Breast cancer remains a leading cause of cancer death that affects women worldwide, and the burden is felt most acutely in low resource settings where mammography access is scarce, and radiologist coverage is thin. Transfer learning-based deep learning models were evaluated for practical utility in breast cancer screening where imaging resources are restricted. The CBIS-DDSM dataset (Kaggle JPG version) was used for this study, with pathology labels mapped into a binary benign-versus-malignant classification scheme. The dataset consisted of 3,568 full mammograms and 3,461 ROI images from over 1,500 patients after quality control. 5-fold StratifiedGroupKFold cross-validation was used to prevent patient-level data leakage, which causes performance inflation in published studies, and ensured that no patient data combined training and validation datasets. The research evaluated two types of input data which included cropped region-of-interest lesion patches and complete mammogram images. InceptionV3 achieved the highest performance among ROI models by reaching ROC-AUC 0.821 and PR-AUC 0.775 and sensitivity 0.785. The study used full mammograms to evaluate ResNet50 performance which achieved ROC-AUC 0.838 and sensitivity 0.790 results. The study shows that transfer learning acts as a strong base for AI-assisted mammography triage systems in low-resource environments when patient level assessment continues throughout system development.

Keywords: Transfer Learning, Breast Cancer Detection, Mammography, Low-Resource Settings, Convolutional Neural Networks, CBIS-DDSM

INTRODUCTION

Breast cancer remains one of the most prevalent causes of cancer related deaths among women worldwide. The International Agency for Research on Cancer estimated approximately 2.3 million new cases in 2020 alone and Sung et al., (2021) projected that annual global cases would reach 28.4 million by 2040. These numbers highlight the importance of early detection because detecting the disease before it advances to later stages results in better survival rates. Mammography serves as the primary screening method but its ability to detect breast cancer decreases for women who possess dense breast tissue which leads to both missed cancer cases and extra testing requirements (Vourtsis & Berg, 2019).

The problem affects low- and middle-income countries (LMICs) in an especially severe way. The healthcare system in Nigeria depends on outdated X-ray technology which exists in public hospitals while trained radiologists work in limited numbers and the system fails to establish effective patient referral procedures (Adeleye, 2023; Omisore et al., 2023). Health systems face infrastructure gaps while women face three barriers to screening which include screening costs and their low health knowledge and the traditional cultural beliefs that have persisted for long (Elewonibi & BeLue, 2019; Nja, 2021). The result for the system leads to advanced stage diagnoses which decrease treatment options and increase mortality rates.

Researchers have significantly explored artificial intelligence and machine learning to improve the early detection and diagnosis of breast cancer by providing decision support tools for clinicians (Olofintuyi, 2023). Convolutional neural networks have achieved specialist-level accuracy for mammography tasks in multiple research facilities that possess advanced resources according to the findings of (Carriero et al., 2024; Luo et al., 2025). The healthcare system gains better understanding about ethical and equity challenges which emerge when artificial intelligence systems get used in areas that face resource shortages because of the current trend to implement AI technologies in various healthcare fields which include mental health monitoring and radiology. Transfer learning is a promising starting point for this area as it allows leveraging feature representations from large-scale general-purpose data sets for new tasks that demand only a small amount of labeled data (Morid et al., 2021). Domain shift constitutes a real challenge because models that learn from clean high-resolution digital mammography images may fail to function properly with the lower contrast and noisier images that are typical in LMIC facilities (Matta et al., 2024). This paper is a report of an attempt to directly fill that void. Three CNN transfer learning models were evaluated to classify mammograms under conditions that simulated actual low-resource settings, with patient-level cross-validation tested to assess how well the models would perform in real-world situations. The work aimed to achieve three goals which included: (i) adopt CNN transfer learning models for mammography classification, (ii) develop and fine tune a computationally efficient transfer learning framework suitable for low-resource settings and (iii) evaluate model performance using standard metrics. Deep learning models such as VGG16, ResNet50 and InceptionV3 have shown

good performance in mammography (Alzubaidi et al., 2021; Carriero et al., 2024).

Transfer Learning for Low-Resource Settings

Several research groups have looked into transfer learning for breast cancer detection using pretrained CNN models on mammography datasets (Mutar et al., 2023). However, existing studies test their models using methods that do not adequately prevent patient level data leakage allowing images from a patient to end up in both training and validation sets. This can make the performance look better than they should, and lose the ability to generalize unseen data. Also, several papers lean mostly on accuracy and ROC-AUC, even when the classes are imbalanced, which feels a bit one dimensional. Therefore, this study uses patient-level group splits and evaluates results with both ROC-AUC and PR-AUC to provide a trustworthy assessment for low-resource healthcare environments.

MATERIALS AND METHODS

Designed Framework

Figure 1 displays the entire processing system. The system functions through five separate stages which start with the loading and cleaning process of the CBIS-DDSM dataset and continue through the creation of binary labels and their validation and image processing through preprocessing and augmentation and the implementation of 5-fold patient-level cross-validation to develop three CNN models and final performance assessment through established clinical metric standards. The subsequent sections provide details about each step of the process.

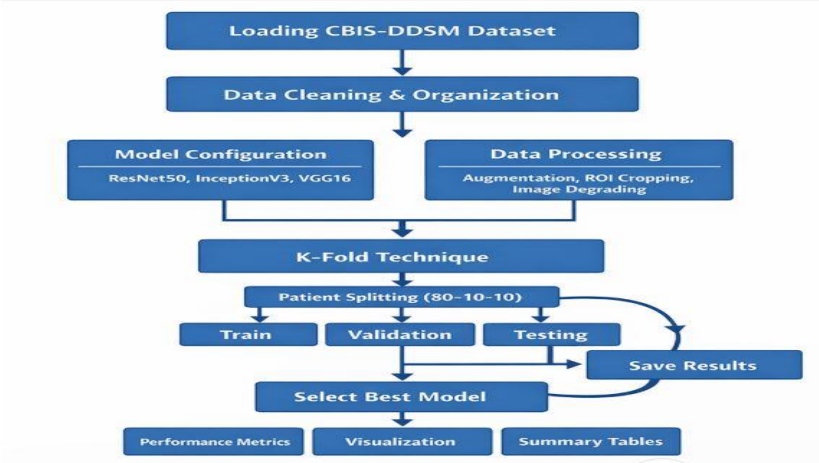

Figure 1: Design Framework for Breast Cancer Detection from Mammography

Dataset

The study used the CBIS-DDSM Kaggle JPG release which supplied mammography images that had been annotated with pathology labels for both mass and calcification cases. The label set was reduced to a binary task: cases marked MALIGNANT were assigned label 1, while BENIGN and BENIGN_WITHOUT_CALLBACK cases were both mapped to label 0. After applying the inclusion criteria, two cohorts emerged, as shown in Table 1. The two groups exhibited a class imbalance which resulted in about 41 percent of cases being classified as malignant and this imbalance influenced both metric selection and training procedures.

Table 1: Cohort Characteristics after Quality-Control Filtering and Label Consolidation

Cohort	Input Type	Patients	Samples	Benign (n, %)	Malignant (n, %)
ROI	ROI crop	1,514	3,461	2,028 (58.60%)	1,433 (41.40%)
Full	Full mammogram	1,566	3,568	2,111 (59.16%)	1,457 (40.84%)

Data Partitioning

Data leakage was avoided by splitting the dataset at the patient level, so that images from the same patient were not included in more than one subset. A stratified split was used to maintain the benign and malignant class proportions in all subsets. The last partition was 80% training, 10% validation, and 10% testing data, as illustrated in Figure 1. The training set was used to learn the model, the validation set was used to tune the hyperparameters and to stop training early, and the independent test set was used only to evaluate the final performance of the model.

Image Preprocessing and Augmentation

PIL was used to read all images, which were converted to three-channel RGB format. The images were zero-padding to square dimensions before resizing. The ROI crops were resized to 384×384 pixels while full mammograms were resized to 512×512 pixels to preserve structural details. The pixel values were transformed to standard ImageNet channel means and standard deviations through the normalization process. The training process employed system training through two forms of horizontal flips and rotation operations within 10 degrees and minor brightness and contrast changes.

Model Training Protocol

The research tested three different architectures, which included ResNet50, VGG16, and InceptionV3, that all used ImageNet-pretrained weights for their initial setup. The original classification head was replaced with a single binary output for each test case. The training process started with a two-phase system that required a warm-up period. The first phase required the training team to freeze the convolutional backbone while updating only the new head, which allowed them to settle before testing deeper weights. The second phase required the complete unfreezing of all training layers, which needed to run through an entire training process at a

reduced learning speed. The training used a cosine annealing schedule for learning rate adjustments, while early training termination occurred when validation ROC-AUC reached its maximum. The loss function used inverse-frequency weights to address the problem of class imbalance.

Model Implementation Details

The three transfer learning models were implemented using the PyTorch deep learning framework with pretrained ImageNet weights obtained from the torchvision model repository. For each architecture, the original ImageNet classification layer was removed and replaced with a single fully connected output neuron followed by a sigmoid activation function for binary classification.

For ResNet50, the final fully connected layer consisting of 2048 input features was replaced with a binary classifier.

For VGG16, the final classifier layer containing 4096 input features was replaced with a single-neuron binary output layer.

For InceptionV3, the auxiliary classifier was disabled during inference, while the final fully connected layer with 2048 input features was replaced with a binary output layer.

The same implementation protocol was applied across all three architectures to ensure a fair comparison. The only differences among the models were their inherent network architectures and computational complexities.

Training Hyperparameters

Unless otherwise stated, identical hyperparameter settings were employed for all three pretrained models to ensure consistency during evaluation. The training configuration is summarized in Table 2.

Table 2: Training Hyperparameters

Hyperparameter	Value
Optimizer	AdamW
Initial learning rate (Warm-up)	1×10^{-3}
Fine-tuning learning rate	1×10^{-4}
Weight decay	1×10^{-4}
Batch size (ROI)	32
Batch size (Full Mammogram)	16
Maximum epochs	50
Warm-up epochs	5
Learning rate scheduler	Cosine Annealing
Early stopping patience	10 epochs
Loss function	Weighted Binary Cross Entropy
Random seed	42
Image normalization	ImageNet Mean and Standard Deviation

The batch size for full mammograms was reduced because of their larger image resolution and corresponding increase in GPU memory requirements.

Evaluation in Resource-Constrained Environments

i. To evaluate the suitability of the proposed models for deployment in resource-constrained environments, computational efficiency was assessed in addition to predictive performance. The evaluation considered model complexity, inference speed, and memory requirements, since these factors directly influence deployment on low-cost clinical computing devices. Each pretrained model was evaluated using the following computational metrics:
ii. Total number of trainable parameters.
iii. Model size on disk (MB).
iv. Peak GPU memory consumption during inference.
v. Average inference time per image.
vi. Floating Point Operations (FLOPs).
vii. Classification throughput measured as images processed per second.

Inference time was measured on the independent test set using a batch size of one image to simulate real-world deployment. The average inference latency was computed over the complete test set after excluding the initial warm-up iterations. To further assess deployment efficiency, the models were compared using an overall trade-off between diagnostic performance and computational cost. Models achieving comparable ROC-AUC with fewer parameters, lower memory consumption, and faster inference were considered more suitable for resource-constrained healthcare environments.

All experiments were conducted using the same hardware and software environment to ensure a fair comparison among the three pretrained architectures.

RESULTS AND DISCUSSION

ROI-Based Model Performance

Table 3 shows the performance of the three pre-trained models using ROI lesion crops. Inception achieved the highest ROC-AUC (0.821), PR-AUC (0.775) and Sensitivity (0.785) while ResNet achieved the highest Specificity (0.774) and VGG16 showed the lowest overall performance among the three models.

Table 3: ROI-Based Model Performance (5-fold Stratified Group K Fold Cross-Validation)

Model (ROI)	ROC-AUC	PR-AUC	Sensitivity	Specificity	F1-Score	Accuracy
ResNet50	0.782	0.727	0.657	0.774	0.664	0.726
VGG16	0.740	0.665	0.703	0.663	0.643	0.678
InceptionV3	0.821	0.775	0.785	0.710	0.715	0.741

The ROC and PR curves in Figure 2 give a fuller picture of how these differences play out across thresholds. InceptionV3 demonstrates superior performance to the other two models for most of the ROC curve while the PR panel displays a

larger performance difference because it shows results under class imbalance conditions.

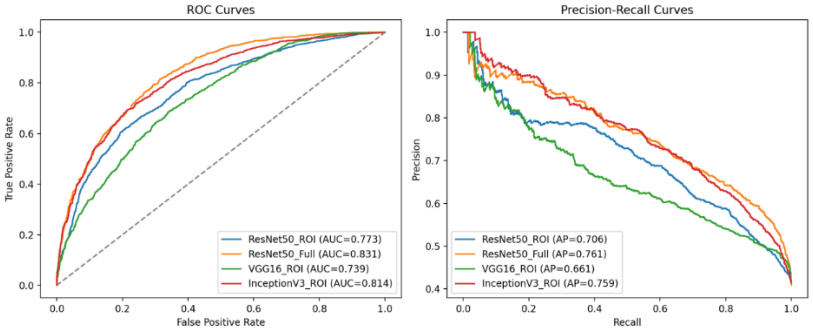

Figure 2: Pooled ROC curves (left) and PR curves (right) for ROI-based models, ResNet50, VGG16, and InceptionV3

Confusion Matrices

Figure 3 shows the confusion matrices at the Youden-selected threshold for each model. The visualisations make it easy to see where the trade-offs fall. InceptionV3 achieves accurate detection of most cancer cases however this success leads to increased incorrect detection of noncancerous cases when compared to ResNet50. VGG16 displays a different mistake pattern which causes it to demonstrate equal strength across all classes but weaker overall performance. The plots provide additional information to the AUC measurements because they establish performance metrics at specific operational points.

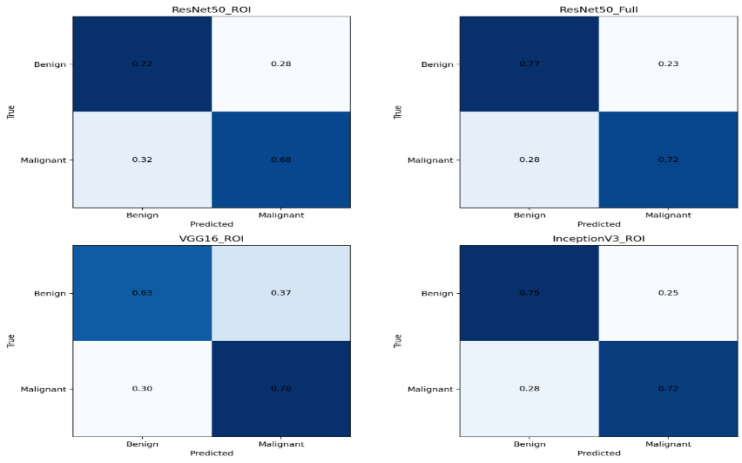

Figure 3: Confusion matrices for ROI-based models (ResNet50, VGG16, Inception V3) at Youden-selected operating thresholds

Full-Mammogram Model Performance

Table 4 presents the results for ResNet50 trained on full 512×512 mammogram images. The study achieved an ROC-AUC score of 0.838 and a sensitivity level of 0.790. Showing the benefits of preserving global breast tissue information.

Table 4: Full-Mammogram ResNet50 Performance (5-fold Stratified Group KFold cross-validation)

Model (Full)	ROC-AUC	PR-AUC	Sensitivity	Specificity	F1-Score	Accuracy
ResNet50 (512×512)	0.838	0.775	0.790	0.727	0.722	0.753

Comparative Analysis: ROI Crops vs. Full Mammograms

Table 5 presents the comparison between the performance of ROI crops and full mammograms. The full mammogram analysis achieved higher ROC-AUC, and sensitivity while the ROI analysis achieved higher specificity. This shows that the full mammograms improve cancer detection where ROI crops may be suitable when computational resources are limited

Table 5: Direct Comparison of ResNet50 on ROI Crops Versus Full Mammograms

Input Type	ROC-AUC	PR-AUC	Sensitivity	Specificity	F1-Score	Accuracy
ROI crop (384×384)	0.782	0.727	0.657	0.774	0.664	0.726
Full mammogram (512×512)	0.838	0.775	0.790	0.727	0.722	0.753

Discussion

The primary objective of this study was not only to compare the diagnostic performance of pretrained convolutional neural networks but also to identify architectures that are suitable for deployment in resource-constrained healthcare environments, where computational resources, memory capacity, storage,

and processing power are often limited. Consequently, model selection should consider both predictive performance and computational efficiency rather than diagnostic accuracy alone. Among the ROI- based models, Inception V3 achieved the highest diagnostic performance but required greater computational resources, making deployment more challenging in resource- constrained environment. ResNet50 provided a better balance between diagnostic performance and computational efficiency making it more practical choice for low resource healthcare settings. VGG16 required more parameters and memory while providing lower classification performance, making it less suitable for deployment in low resource health care settings.

The comparison between ROI-based and full mammogram revealed that full mammograms had better ROC-AUC and sensitivity, but needed more computational resources due to the larger size. In terms of deployment, this is a vital trade-off. For healthcare facilities with limited computational resources, ROI-based analysis might be a better choice due to the significant reduction in input size, memory usage and inference time when using cropped images. ROI-based inference can also be used to process larger numbers of images on less expensive computing systems, while the centers with more powerful computational resources can afford full-mammogram analysis for the improved diagnostic performance. Thus, the choice of a deployment strategy will depend on the accuracy of the diagnosis and the computational resources available.

This is also in line with previous studies that have shown the great potential of deep learning in breast cancer screening (McKinney et al., 2020). The complexity of the network and the representation of the input should be taken into account at the same time. Models that offer competitive diagnostic performance with less computational resources are more likely to be used in routine clinical practice, especially in developing countries where access to high performance computing infrastructure is limited.

The results are similar to those of earlier research. A similar finding was reported by Wu et al. (2019), who noted that adding global breast context to the classification task of mammograms improves classification accuracy, but at the cost of higher computational costs. Their results confirm that better diagnostic accuracy tends to be associated with greater resource usage, which highlights the need to consider performance as well as computational efficiency.

Lastly, it is important to note that the CBIS-DDSM dataset was obtained from digitized film mammograms from the United States, and thus may not fully reflect imaging conditions in Nigerian or other West African healthcare facilities. Therefore, it is still necessary to validate the model with local mammography data before it can be used in clinical practice. In addition, computational optimization methods like model pruning, quantization, knowledge distillation, and lightweight network architectures should be explored in future research to enhance the deployment suitability in resource-constrained healthcare settings.

Table 6 compares the computational efficiency of the pretrained models for deployment in a low resource setting.

Table 6: Comparison of the Pretrained Models in Resource Constrained Environments

Model	Parameters(M)	Model Size(MB)	FLOPs(G)	GPU Memory	Inference Time(ms)	FPS
ResNet50(ROI ,384px)	23.5	94	24.0	348 MB	7.6	131
VGG16(ROI, 384px)	134.3	537	90.4	1.2 GB	11.9	84
InceptionV3 (ROI, 299px)	25.1	101	11.4	249MB	14.3	70
ResNet50(Full, 512px)	23.5	94	42.7	370 MB	7.1	140

Measured on an NVIDIA GeForce RTX 4060 Laptop GPU

CONCLUSION

In this study, three CNN transfer learning models were tested for breast cancer detection using mammograms in a simulated low - resource healthcare environment. Inception V3 had the highest diagnostic performance among the ROI-based models, and ResNet50 trained on full mammograms had the highest overall ROC-AUC and sensitivity. The results show that transfer can be used to enable accurate AI-based breast cancer screening and reveal the balance between diagnostic performance and computational efficiency for choosing models in low resource settings. Patient level cross-validation also helps to avoid data leakage and gives a more accurate estimate of real-world performance. The results are encouraging, but external validation with locally collected mammography data is crucial prior to clinical use. Lightweight models, optimization techniques and multi-view mammographic analysis are recommended for future studies to further enhance performance and deployment in low resource healthcare environments.

REFERENCES

Alzubaidi, L., Zhang, J., Humaidi, A.J. *et al*. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* **8**, 53 (2021). https://doi.org/10.1186/s40537-021-00444-8

Vourtsis, A., & Berg, W. A. (2019). Breast density implications and supplemental screening. *European radiology*, *29*(4), 1762–1777. https://doi.org/10.1007/s00330-018-5668-8

Carriero, A., Groenhoff, L., Vologina, E., Basile, P., & Albera, M. (2024). Deep Learning in Breast Cancer Imaging: State of the Art and Recent Advancements in Early 2024. *Diagnostics (Basel, Switzerland)*, *14*(8), 848. https://doi.org/10.3390/diagnostics14080848

Mutar, M. T., Majid, M., Ibrahim, M. J., Obaid, A. H., Alsammarraie, A. Z., Altameemi, E., & Kareem, T. F. (2023). Transfer learning with different modified convolutional neural network models for classifying digital mammograms utilizing Local Dataset. *The Gulf journal of oncology*, *1*(41), 66–71.

Adeleye, D. (2023). *Mammography in Nigeria: A rigorous retrospective analysis of one-year utilization in a tertiary hospital*. *Annals of Breast Cancer Research*, *7*(1), 1021. https://www.jscimedcentral.com/public/assets/articles/breast cancer-7-1021.pdf

McKinney, S.M., Sieniek, M., Godbole, V. *et al*. International evaluation of an AI system for breast cancer

screening. *Nature* 577, 89–94 (2020). https://doi.org/10.1038/s41586-019-1799-6

Morid, M. A., Borjali, A., & Del Fiol, G. (2021). A scoping review of transfer learning research on medical image analysis using ImageNet. *Computers in biology and medicine*, *128*, 104115. https://doi.org/10.1016/j.compbiomed.2020.104115

Elewonibi, B., & BeLue, R. (2019). The influence of socio-cultural factors on breast cancer screening behaviors in Lagos, Nigeria. *Ethnicity & health*, *24*(5), 544–559. https://doi.org/10.1080/13557858.2017.1348489

Omisore, A. D., Sutton, E. J., Akinola, R. A., Towoju, A. G., Akhigbe, A., Ebubedike, U. R., Tansley, G., Olasehinde, O., Goyal, A., Akinde, A. O., Alatise, O. I., Mango, V. L., Kingham, T. P., & Knapp, G. C. (2023). Population-Level Access to Breast Cancer Early Detection and Diagnosis in Nigeria. *JCO global oncology*, *9*, e2300093. https://doi.org/10.1200/GO.23.00093

Matta, S., Lamard, M., Zhang, P., Le Guilcher, A., Borderie, L., Cochener, B., & Quellec, G. (2024). *A systematic review of generalization research in medical image classification*. *Computers in Biology and Medicine*, *183*, 109256. https://doi.org/10.1016/j.compbiomed.2024.109256

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a cancer journal for clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Nja, G. (2021). Breast Cancer Knowledge and Mammography Uptake among Women Aged 40 Years and Above in Calabar Municipality, Nigeria. *Asian Journal of Medicine and Health*, 19(8). https://doi.org/10.9734/AJMAH/2021/v19i830351

Luo, L., Wang, X., Lin, Y., Ma, X., Tan, A., Chan, R., Vardhanabhuti, V., Chu, W. C. W., Cheng, K. T., & Chen, H. (2025). *Deep Learning in Breast Cancer Imaging: A Decade of Progress and Future Directions*. IEEE Reviews in Biomedical Engineering, 18, 130–151. https://doi.org/10.1109/RBME.2024.3357877

Wu, N., Phang, J., Park, J., Shen, Y., Huang, Z., Zorin, M., Jastrzębski, S., Févry, T., Katsnelson, J., Kim, E., Wolfson, S., Parikh, U., Gaddam, S., Lin, L. L. Y., Ho, K., Weinstein, J. D., Reig, B., Gao, Y., Toth, H., Geras, K. J. (2019). *Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening*. ArXiv:1903.08297.

Olofintuyi, . S. S. (2023). BREAST CANCER DETECTION WITH MACHINE LEARNING APPROACH. FUDMA JOURNAL OF SCIENCES, 7(2), 216-222. https://doi.org/10.33003/fjs-2023-0702-1392