STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–10.
Published online in International Academic Press (www.IAPress.org)

A novel CNN architecture for breast cancer detection

Faouzi-Ayoub El Ghanaoui [1,*], Elmehdi Aniq [1,2], Mohamed Chakraoui [1], Youness Khourdifi [1]

[1] *LS2ME, Polydisciplinary Faculty of khouribga, Sultan Moulay Slimane University, Beni mellal, Morocco*
[2] *LAMIGEP, EMSI Marrakech, Marrakech, Morocco*

Abstract Breast cancer is a major cause of death in women around the world, and early detection is essential for enhancing survival rates. While mammography is a key screening tool, its accuracy can be impacted by human interpretation. Convolutional Neural Networks (CNNs) offer advanced image analysis capabilities to enhance early detection and support healthcare professionals with higher accuracy and reliability. The aim of this study is to introduce a novel CNN architecture, developed from scratch, that would improve breast cancer detection and efficiency in diagnostics.The model was trained on a merged dataset consisting of publicly-available mammographic images obtained from MIAS, INbreast, and DDSM with data augmentation to introduce some variability. The model was then tested without the datasets combined, in order to test generalization.The proposed method demonstrates effective results on different datasets.The model achieved its best performance on the DDSM dataset, with an accuracy of 0.9996, a recall of 0.9996, a precision of 0.9996, an F1-score of 0.9996, and an AUC of 1.0000. On the combined dataset, the model obtained an accuracy of 0.9669, a recall of 0.9669, a precision of 0.9669, an F1-score of 0.9669, and an AUC of 0.9956. These results prove the potential of deep learning approach to improve early breast cancer detection, reduce errors and enhance patient outcomes.

Keywords Deep learning , Mammography, Convolutional neural network (CNNs) , Data Augmentation, Breast cancer

DOI: 10.19139/soic-2310-5070-2559

1. Introduction

Breast cancer is a major illness and a significant cause of mortality among women worldwide. With 2.3 million new cases diagnosed each year, it underlines the necessity of urgent strategy to improve survival rates [1].
Although mammograms are the most common tools of breast cancer screening, their effectiveness is susceptible to human error, depending on the ability of radiologists [2].Early detection increases the survival rates for breast cancer, but such a process, indeed, is still challenging and and time-consuming process [3].
Deep learning can boost breast cancer diagnosis, but its adoption in healthcare is constrained by the need for reliable and compelling proof to validate its efficacy [4]. Convolutional Neural Networks (CNNs) have transformed medical image analysis, improving the efficiency of interpreting medical images [5]. They enable the automated detection of abnormalities that might be missed by human observers [6]. Also, they assist radiologists in making more accurate decisions while optimizing the analysis process, thus improving the accuracy and specificity of breast cancer diagnoses [7].However, integrating such technology into real-world clinical screening poses a number of challenges, such as the need for high-quality annotated datasets, evaluating the model's performance via diverse samples, and improving result interpretability. Addressing these obstacles requires a precise and multidisciplinary approach to ensure the effectiveness and adoption of AI-based technologies in clinical settings [8].
In this context, this study explores the application of deep learning to improve the early detection of breast cancer

*Correspondence to: Faouzi-Ayoub El Ghanaoui(Email: faouziayoub.elghanaoui@usms.ac.ma). LS2ME, Polydisciplinary Faculty of khouribga, Sultan Moulay Slimane University, Beni mellal, Morocco.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

using mammographic images. Our primary goal is to provide an efficient solution, serving as a ”second opinion,” that reinforces, supports, and assists doctors to make optimal decisions and reduce diagnostic errors.

2. RELATED WORK

Convolutional Neural Networks (CNNs) have transformed breast cancer diagnostics [9]. However, challenges persist, including the rarity of large datasets and the substantial computational resources required [3]. To address these limitations, such as data augmentation [10], ensemble learning [11] and transfer learning [12] have been widely employed. Chougrad et al. ([13], 2018) developed a Computer-aided Diagnosis (CAD) system based on deep Convolutional Neural Networks (CNN) to classify mammography mass lesions. The authors analyze the role of transfer learning and how fine-tuning methods can be used to boost the performance of the models.In the study ([14], 2021)a new system using deep learning was developed from scratch to classify breast lesions in mammographic images into malignant and nonmalignant based on two approaches: the first uses patches of region of interest (ROI) and the second uses the full images. This research also includes a preprocessing step, including several data augmentation techniques. The proposed system showed great performance across MIAS, DDSM, and INbreast data sets, using 5-fold cross-validation. Moreover, Muduli et al. ([15], 2022) proposed a simplified deep convolutional neural network (CNN) model to classify breast cancer from different type of images, including mammograms and ultrasound. Data augmentation technique has been employed to reduce overfitting and provide good generalization. Thirumalaisamy et al. ([16], 2023) introduced a hybrid technique to identify breast cancer, utilizing ResNet101 architecture with a metaheuristic optimization algorithm, and transfer learning. Experimental results were conducted on the MIAS and DDSM (CBIS-DDSM) mammographic datasets.
More recent studies have leveraged pretrained models and ensemble strategies to optimize performance. Sannasi Chakravarthy et al. ([17], 2023) present the use of pretrained CNN models such as AlexNet, GoogleNet, ResNet50, and DenseNet121. These models were combined with feature fusion techniques and Principal Component Analysis (PCA) to reduce computational costs and improve the classification of mammographic images from the MIAS and INBreast datasets.Additionally, In ([18], 2024), a deep-learning ensemble classifier is proposed for breast cancer diagnosis in multimodal datasets. The approach used combines three effective transfer learning models (AlexNet, ResNet, and MobileNetV2) with LoG to boost performance.Study ([19], 2024) Underlined the role of deep learning on the diagnostic performance of radiologists as a second opinion in detecting breast cancers, reducing diagnostic errors and assisting in lesion detection, with validated generalizability across diverse equipment.Alhsnony and Sellami ([20], 2024) presented CNN model to aid in the detection and classification of breast cancer as normal, benign, or malignant from mammographic images. The efficiency of the model was validated using MIAS and DDSM datasets, where it achieved impressive results.

3. METHODES

3.1. DataSet

This study uses three popular and publicly accessible mammography datasets: MIAS [21], INbreast [22], and CBIS-DDSM [23]. MIAS a smaller dataset with detailed annotations (161 cases, 322 images, MLO views) perfect for initial model development and testing. INbreast a larger dataset with high-resolution images, BI-RADS classifications, and complete annotations (115 cases, 410 images, MLO and CC views), great for advanced deep learning applications. DDSM is a large dataset with heterogeneous cases, providing a strong foundation for training robust models with good generalization capabilities (1597 cases, 3061 images, MLO and CC views).Together, these datasets offer a strong foundation for enhancing automated breast cancer detection and diagnosis.

3.2. DATA augmentation

Data augmentation creates new samples by applying random transformations to the existing dataset, thereby increasing the training data and improving model performance. This approach is especially valuable in minimizing

overfitting when working with limited datasets. For smaller datasets, common transformations such as translation, zooming, flipping, mirroring, and rotation are frequently used. These techniques not only expand the dataset size but also improve training efficiency by accelerating convergence [10].

3.3. CNN Architecture

AlexNet [24] revolutionized image recognition by increasing CNN depth and introducing Local Response Normalization (LRN) and overlapping subsampling to address overfitting. It served as a foundation for modern CNNs, inspiring further advancements in deep learning.
VGG16 [25], an extension of AlexNet, features 13 convolutional layers and 5 pooling layers, using 3×3 filters and batch normalization to optimize performance while reducing errors. Though computationally intensive, its flexibility and effectiveness have made it widely adopted in healthcare for predictive models.
MobileNetV2 [26], designed for resource-limited environments, employs inverted residuals with linear bottlenecks and depthwise separable convolutions to minimize computational demands. Its lightweight and efficient design make it well-suited for medical image analysis.
ResNet50 [27], a deep residual network, overcomes vanishing gradient issues with residual connections that allow layer skipping. Comprising 50 layers, it efficiently supports complex feature learning and has demonstrated high accuracy in breast cancer classification tasks. Table 1 provides an overview of the most recent CNN architectures and their characteristics

Table 1. Overview of CNN architectures and their descriptions.

CNN Architecture	**Description**
LeNet-5	7 layers: 2 Conv2D, 2 pooling, and 3 Fully Connected (FC) layers, with 60,000 parameters [28]
AlexNet	8 layers: 5 Conv2D and 3 FC layers, with 60 million parameters [24]
VGGNet-16	13 Conv2D and 3 FC layers, with 140 million parameters [25]
GoogLeNet	22 layers, reduced parameters from 60 million (AlexNet) to 5 million [29]
ResNet-50	250 layers: 48 Conv2D and 2 FC layers, with 23.51 million parameters [27]
Inception-v3	48 layers with less than 25 million parameters, containing multiple convolutional layers [30]
U-Net	27 layers, including approximately 23 convolutional layers, with less than 1 million parameters [27]
MobileNetV2	Approx. 53 layers, with less than 3.4 million parameters [26]

3.4. Key metrics

The confusion matrix is an important tool for evaluating model training and identifying the optimal solution in a binary classification problem. The rows represent the predicted classes, while the columns correspond to the real classes. Table 2 provides an illustration of this matrix.In this context, Tp (true positive) and Tn (true negative) represent the number of correctly classified cases with and without cancer, respectively. In contrast, Fp (false positive) and Fn (false negative) refer to instances misclassified as positive and negative, respectively.

Table 2. Confusion Matrix

	Real Positive Class	**Real Negative Class**
Predicted Positive Class	True positive (Tp)	False negative (Fn)
Predicted Negative Class	False positive (Fp)	True negative (Tn)

From the confusion matrix, various metrics can be calculated, as shown in Table 3. These metrics offer a detailed evaluation of the model's performance [31].

Table 3. Metrics and their mathematical formulations.

Metrics	**Formula**	**Descriptions**
Accuracy	$Acc = \frac{TP+TN}{TP+TN+FP+FN}$	Measures the proportion of correct predictions, offering a general overview of model performance.
Recall	$S_v = \frac{TP}{TP+FN}$	Evaluates the model's ability to correctly identify positive cases while minimizing false negatives.
Specificity	$S_p = \frac{TN}{TN+FP}$	Evaluates the model's ability to correctly identify actual malignant cases while minimizing false positives.
Precision	$P = \frac{TP}{TP+FP}$	Indicates how many predictions labeled as malignant are truly malignant, reflecting the model's ability to identify true positives.
F1 score	$F1 = 2 \times \frac{P \times S_v}{P+S_v}$	Combines Precision and Sensitivity into a single metric, balancing their trade-off, especially useful for imbalanced datasets.
AUC		Assesses the model's ability to distinguish between classes across all classification thresholds, providing a comprehensive evaluation.

3.5. Contributions

CNN architectures, such as MobileNet, VGG, and ResNet, aim to preserve essential information during convolutions, each offering distinct advantages and limitations. MobileNet is known for its portability and efficiency, making it ideal for resource-constrained devices, though it provides lower precision for more complex tasks. VGG, with its considerable depth, provides excellent precision, but its high resource demands and computational inefficiency limit its practical use. ResNet, by utilizing residual connections, facilitates the training of very deep networks while effectively addressing gradient-related issues, but it remains more resource-intensive compared to MobileNet. By integrating the advantages of these three architectures; the lightweight design of MobileNet, the precision of VGG, and the depth and stability of ResNet; Our architecture is designed to provide an optimized solution specifically for breast cancer diagnosis, combining efficiency with robust generalization capabilities. Designed for resource-constrained environments, it ensures high performance in complex tasks, addressing the current needs for computational efficiency and clinical applicability.

It starts with an input layer designed to handle preprocessed mammographic images resized to (224,224,3), ensuring compatibility with subsequent layers.

The architecture is divided into five functional blocks, each designed to progressively extract and refine features for precise and comprehensive analysis.

The first block was inspired by the architecture of the MobileNet model, well-known for its balance between computational efficiency and compact architecture. At its input, a 1×1 convolution was applied to the images to increase the number of channels while preserving spatial dimensions, projecting the data into a richer feature space. Then a depthwise convolution with 3×3 filters was used to process each channel independently,

which considerably reduces the parameters while effectively capturing local features. Finally, a 1×1 projection convolution adjusts the channel dimensions to ensure a fluid transition to the next layer. A residual connection links the input to the output after the convolutional transformations, stabilizing the gradient flow, reducing performance degradation and improving convergence and robustness.
Let X be the input tensor, and $F(X)$ the result of the convolutional transformations applied to X. The residual output is computed as:

$$Y = P(X) + F(X) \tag{1}$$

where $P(X)$ denotes either an identity mapping (when the shapes of X and $F(X)$ match), or a 1×1 projection convolution when shape alignment is required:

$$P(X) = \text{Conv}_{1\times1}(X) \tag{2}$$

This architectural block combines computational efficiency along the channel dimension—through depthwise separable convolutions—with training stability ensured by residual connections, while maintaining a compact representation suitable for deployment in resource-constrained environments.
Next blocks include convolutions with 3×3 filters, to extract hierarchical features and capture subtle details such as edges, textures, and complex structures associated with malignant tumors. Each convolution is followed by batch normalization in order to stabilize the learning process, speed up convergence, and reduce sensitivity to weight initialization. The normalized output $\hat{F}(X)$ is given by:

$$\hat{F}(X) = \frac{F(X) - \mu}{\sqrt{\sigma^2 + \epsilon}} \tag{3}$$

where μ and σ^2 are the batch mean and variance, and ϵ is a small constant added for numerical stability. A learnable affine transformation is then applied:

$$\tilde{F}(X) = \gamma\hat{F}(X) + \beta \tag{4}$$

where γ and β are learnable scaling and shifting parameters, respectively, specific to each channel.
The ReLU activation introduces non-linearity, enabling the network to learn complex models efficiently.It is defined as:

$$A(X) = \max(0, \tilde{F}(X)) \tag{5}$$

Finally, a 2×2 Max-pooling layer subsamples the feature maps, preserving essential information while minimizing computational costs:

$$Y_{i,j} = \max_{(m,n)\in\Omega_{i,j}} A(X)_{m,n} \tag{6}$$

where $\Omega_{i,j}$ denotes the local pooling region corresponding to output location (i, j).
The architecture progressively reduces spatial dimensions with convolutional layers and subsampling operations (MaxPooling), while simultaneously increasing the number of filters to extract increasingly abstract features. When the spatial size is reduced to 7×7, a Global Average Pooling layer condenses each channel into a single averaged value, simplifying the model while retaining critical global image information. Fully connected (Dense) layers, enhanced with regularization techniques such as Dropout and L2 regularization, generate outputs adapted to the target task, such as binary classification. A detailed depiction of our architecture is presented in figure 1.

P-NET has approximately 1.05M parameters with a memory footprint of 4.0 MB, and a computational cost of 1.69 GFLOPs per inference performed on 224×224×3 images. P-NET provides advantages of efficiency in terms of computation and size making it useable in embedded systems and real-time applications. P-NET is nearly as fast as VGG16 and it is 18 times smaller. VGG16 achieves high miters/sec despite its power in FLOPs due to a simple pipeline. ResNet-50 suffers from long inference latency due to the sequential operations through the network. MobileNetV2 is performing efficient FLOPs but is slowed by the execution of depthwise convolutions. Table 4 presents a comparison of our model's parameter against those of widely used CNN architectures.

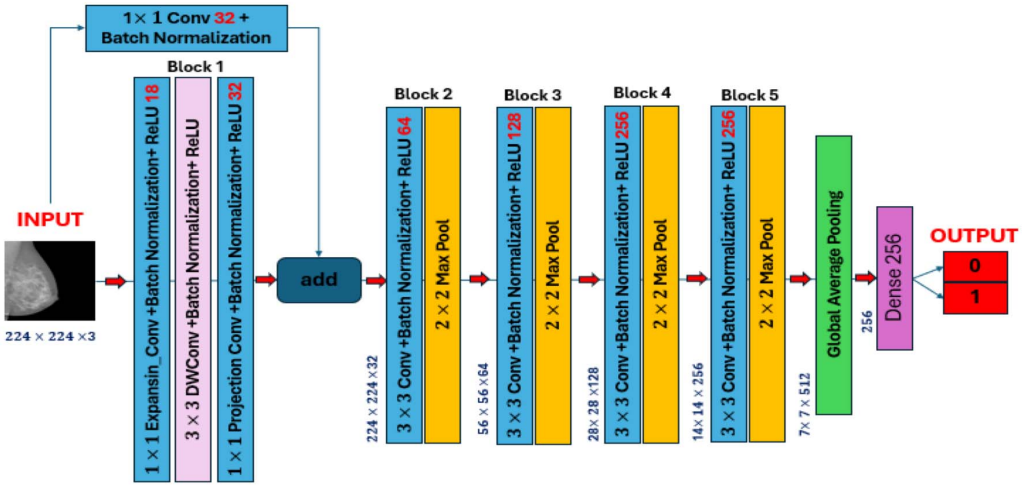

Figure 1. Our CNN Architecture.

The Binary Cross-Entropy loss function and the Adam optimizer were applied to allow for accurate and efficient updates to parameters in the course of training. In addition to these functions, a Dropout layer was incorporated after the dense layer of 256 units, randomly dropping out a number of neurons during the training iteration to avoid reliance on specific neurons. Additionally, L2 regularization was imposed to place a penalty on the large weight magnitudes, favoring simpler and more generalizable models. These two regularization measures can improve the robustness and performance of the model and lower the risk of overfitting which is a common issue for deep learning-based classification tasks.

Table 4. Comparison of CNN architectures in terms of parameter count, computational cost, model size, and inference time.

Model	**Params (M)**	**FLOPs (G)**	**Model Size (MB)**	**Inference (ms)**
VGG16	138.36	30.693261	∼527.8	76.89
ResNet-50	25.64	7.711850	∼97.8	797.19
MobileNetV2	3.54	0.557556	∼13.5	455.18
Proposed Model (P_NET)	**1.05**	**1.691333**	**∼4.0**	**83.69**

4. EXPERIMENTAL RESULTS

4.1. Data preprocessing

To ensure accurate results, we first split the data into 70/20/10 for training, validation, and testing. Then, we applied data augmentation techniques to each image dataset and throughout the model training process, augmenting each image using random transformations, including shifts, rotations of ±90°, shearing, zooming (from –0.2 to +0.2), and horizontal flipping. These augmentations were performed in real time using Keras's ImageDataGenerator, ensuring that each iteration produced a different image. All datasets were subsequently balanced to maintain an equal number of images per class, thereby reducing the potential for bias. We then combined all these datasets into a single dataset, referred to as Merged Data. Table 5 provides detailed information on the number of images

allocated to each phase for each dataset.

Table 5. Number of images allocated for training, validation, and testing across all datasets.

DATA	Total Images		Total Images Allocated To		
	Before Augmentation	After Augmentation	Train	Val	Test
MIAS	330	2880	2014	576	290
INBREAST	410	5040	3528	1008	504
DDSM	3061	11940	8358	2388	1194
Merged	3781	21732	15212	4346	2174

We additionally included an intensity inversion to improve the appearance of masses and relevant characteristics present in the breast images. An example of the applied augmentation approach with intensity inversion is demonstrated in Figure 2. These transformations were employed to create more variety and thus enhance the generalized ability of the model.

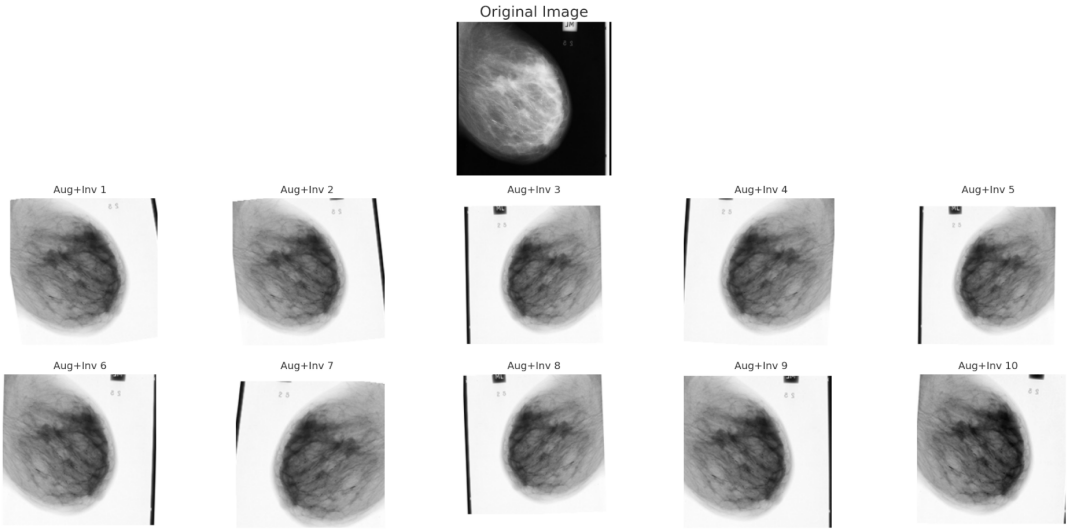

Figure 2. Example of data augmentation.

4.2. Ablation studies

We conducted ablation studies to explore how important hyperparameters affected model performance: dropout, L2 regularization, learning rate, and residual connections. The dropout rate was sampled uniformly from the range of 0.2 to 0.7, while both the L2 penalty and the learning rate were sampled within the range of 10^{-2} to 10^{-4}. The batch size was set to 32 to keep training processes identical. We conducted hyperparameter optimization with Keras Tuner's GridSearch by evaluating a total of 36 configurations. Each model was trained on the images resized to 224×224×3 for up to 30 epochs, and we used validation accuracy to evaluate performance. Analysis of the best performing configurations revealed that the moderate dropout values (0.2 to 0.4) provided the best regularization balance, while higher rates (0.5) negatively affected accuracy. Similarly, the lowest L2 penalty (10^{-4}) led to better generalization, whereas stronger regularization impaired performance. A learning rate of 0.001 consistently outperformed 0.01, indicating more stable convergence with smaller updates. Beyond hyperparameter tuning,

architectural ablation experiments showed that removing the residual block, skip connections, or regularization layers resulted in significant drops in accuracy—highlighting the synergistic contribution of both the architecture and fine-tuned hyperparameters. The three best-performing configurations among the 36 evaluated are summarized in the table 6.

Table 6. Performance obtained with different hyperparameter configurations.

Dropout	L2 Regularization	Learning Rate	Batch Size	Accuracy
0.2	0.0001	0.001	32	0.8944
0.3	0.0001	0.001	32	0.8790
0.4	0.0001	0.001	32	0.8787

The best-performing configuration—with a dropout rate of 0.2, L2 regularization of 10^{-4}, a learning rate of 10^{-3}, and a batch size of 32—achieved a validation accuracy of 0.8944, confirming the effectiveness of the proposed deep learning architecture in breast lesion detection. These results further validate that combining lightweight convolutional blocks with carefully tuned regularization and optimization strategies can significantly improve generalization and overall performance in medical imaging tasks.
The model was then trained on the merged dataset using the best performing configuration for a total of 100 epochs. On the merged dataset the model had an accuracy of 0.9669, a recall of 0.9669, a precision of 0.9669, an AUC of 0.9956 and an F1 score of 0.9669.

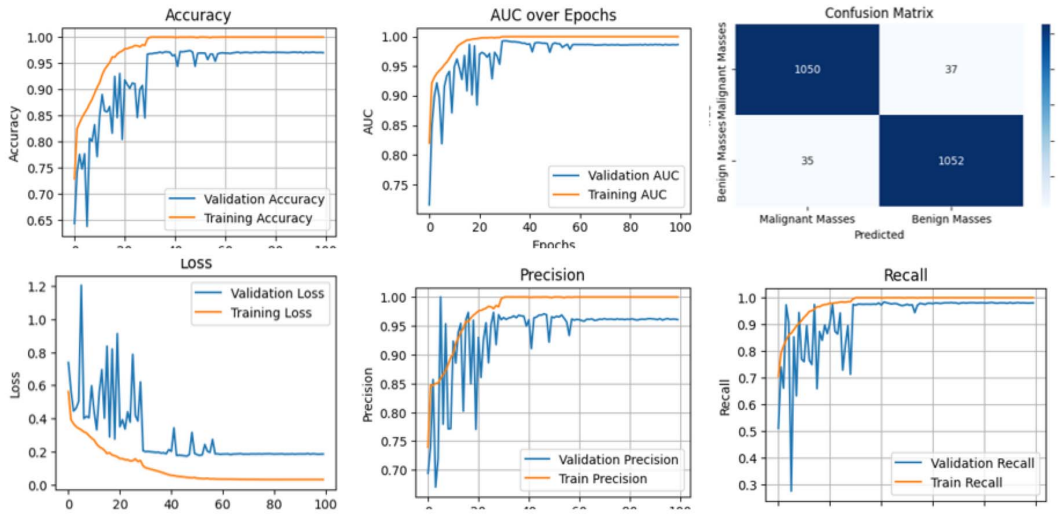

Figure 3. The model's performance on the Merged dataset.

Figure 3 shows the model training process on the merged dataset over 100 epochs for the training and validation sets. Training loss consistently decreased, approaching zero, while validation loss initially followed a similar trend before exhibiting brief fluctuations that later stabilized, suggesting mild overfitting. Despite this, the model maintained strong generalization. Accuracy and precision improved rapidly, with the training accuracy approaching 100% and the validation accuracy consistently exceeding 96% after epoch 15. Precision remained high, reflecting low false positive rates, and recall on the validation set also stabilized above 96%—a crucial factor in medical diagnostics, where missing positive cases is unacceptable. These learning dynamics highlight the robust and reliable performance of the model across all epochs, making it well suited for sensitive tasks such as breast lesion detection.

The model was evaluated separately on the MIAS, INbreast, and DDSM datasets and achieved competitive results compared to recent studies. All results are displayed in Table 7. This confirms the potential of the model and provides a good foundation for future improvements and potential implementation in clinical practice.

Table 7. Comparison of Our Model's Results with State-of-the-Art Approaches.

Ref	Dataset	Accuracy	AUC	Recall	Precision	F1 Score	Time (min)
Our work	**MIAS**	**0.9375**	**0.9906**	**0.9375**	**0.9417**	**0.9374**	**-**
	INBREAST	**0.9792**	**0.9964**	**0.9792**	**0.9792**	**0.9792**	**-**
	CBIS-DDSM	**0.9996**	**1**	**0.9996**	**0.9996**	**0.9996**	**-**
	Merged Data	**0.9669**	**0.9956**	**0.9669**	**0.9669**	**0.9669**	**60.51**
[13] - 2018	MIAS	0.9823	0.990	-	-	-	-
	INBREAST	0.9550	0.970	-	-	-	14.3
	DDSM	0.9735	0.980	-	-	-	91.6
[14] - 2021	MIAS	0.9530	0.974	0.9800	-	-	-
	INBREAST	0.9652	0.980	0.9655	-	-	-
	Merged Data	0.9017	0.914	0.8908	-	0.9078	-
[15] - 2022	MIAS	0.9655	-	0.9728	-	-	-
	INBREAST	0.9128	-	0.9943	-	-	-
	DDSM	0.9068	-	0.9272	-	-	-
[16] - 2023	MIAS	0.9915	0.9912	0.9786	0.9830	0.9760	-
	CBIS-DDSM	0.9860	-	0.9870	0.9870	0.9804	-
[17] - 2022	MIAS	0.9793	-	-	-	-	-
	INBREAST	0.9664	-	-	-	-	-
[20] - 2024	MIAS	-	-	-	0.9423	-	-
	DDSM	-	-	-	0.9553	-	-

The absence of recent and accessible mammography data was a barrier that led us to work with small and old data sets that were often inadequate for training, thereby limiting their applicability to more complicated data or clinical case analyses. As such, the consideration of data from the African population becomes imperative as the African continent possesses unique genetic diversity. The absence of such data represents a major error, making models trained on non-diverse datasets unsuitable and unreliable for use in screening centers across Africa.
In addition, despite the fact that AI has confirmed its effectiveness by proposing innovative solutions to combat this disease, it is still dropped by healthcare professionals because it is seen as a 'black box' due to its complexity and lack of transparency. Integrating Explainable Artificial Intelligence (XAI) techniques into breast cancer screening could solve this problem by highlighting the elements that have an impact on diagnostic decisions. Such approaches could increase the healthcare professionals' trust, lend help to their decisions, and provide insight into the diagnostic process.

5. Conclusion

In conclusion, this study has proved the efficient use of deep learning models to improve the classification of mammographic images for breast cancer detection. We presented a CNN architecture that includes techniques such as depth convolutions, residual connections and optimization methods such as dropout and L2 regularization. These enabled the model to effectively capture spatial and contextual relationships in the images, improving classification accuracy. The results are encouraging and outperforms state-of-the-art results, with the best performance reaching 0.9669 accuracy, 0.9656 AUC, 0.9669 recall, 0.9669 precision and 0.9669 F1-score on the Merged database.
Although this study produced promising results, several limitations must be considered. First, the performance of the model was evaluated using the MIAS, INBreast and DDSM databases, which, although widely recognized, are relatively small and may not fully reflect the diversity of mammographic images encountered in the real clinical world. Furthermore, the restriction to outdated or small datasets may limit the generalizability of the results to larger or newer datasets. Second, the high computational demands of training deep learning models with advanced architectures pose challenges in resource-constrained environments. Although techniques such as dropout and L2 regularization have been used to reduce overfitting, the lack of external validation in different datasets or healthcare contexts can compromise the robustness of the model and limit its clinical applicability.
Future research should focus on developing deep architectures adapted to complex imaging, integrating advanced imaging techniques, and creating unified diagnostic frameworks. Key priorities include building representative datasets, combining segmentation and augmentation methods, and building compact deep learning models.The exploration of new imaging modalities, such as 3D mammography, computed tomography, or histology, could further transform breast cancer diagnosis and treatment.
It is important to collaborate with healthcare professionals and validate the proposed approaches inside a real-world healthcare setting. These partnerships help address prejudices and trade-offs often associated with such models while contributing to enhanced diagnostic outcomes, patient care, and personalized medicine in breast cancer management. These initiatives illustrate our commitment to leveraging innovative deep learning-based techniques to address the complex challenges associated with breast cancer detection and treatment.

Acknowledgement

We would like to express our sincere gratitude to all individuals whose efforts and support made this research possible. We are especially grateful to our thesis directors for their help and valuable feedback.

REFERENCES

1. F. Bray, M. Laversanne, H. Sung, J. Ferlay, Rebecca L. Siegel, Isabelle Soerjomataram, Ahmedin Jemal *Global cancer statistics 2022*, Statistics World Health Day 2024, vol. 74, Issue3, pp. 229–263, May/June 2024, doi: 10.3322/caac.21834
2. S. E. Lee, H. Hong, and E.-K. Kim, *Diagnostic performance with and without artificial intelligence assistance in real-world screening mammography*, European Journal of Radiology Open, vol. 12, pp. 100545, 2024, doi: 10.1016/j.ejro.2023.100545
3. Ophira Ginsburg MD, Cheng-Har Yip MD, Ari Brooks MD..., *Breast cancer early detection: A phased approach to implementation*, Cancer, Volume 126, Issue S10, pp. 2325-2493, May 15, 2020, doi: 10.1002/cncr.32887
4. M. Khalifa and M. Albadawy, *AI in diagnostic imaging: Revolutionising accuracy and efficiency*, Computer Methods and Programs in Biomedicine Update, vol. 5, pp. 100146, 2024, doi: 10.1016/j.cmpbup.2024.100146
5. E. M. Nwanosike, B. R. Conway, H. A. Merchant, and S. S. Hasan, *Potential applications and performance of machine learning techniques and algorithms in clinical practice: A systematic review*, International Journal of Medical Informatics, vol. 159, pp. 104679, 2022, doi: 10.1016/j.ijmedinf.2021.104679
6. Aniq, E., Chakraoui, M., Mouhni, N., Aboulfalah, A. and Rais, H. *Breast Cancer Stage Determination Using Deep Learning*, Lecture Notes in Networks and SystemsVolume 799 LNNS, Pages 550–5582024, 11th World Conference on Information Systems and Technologies, WorldCIST 2023, doi: 10.1007/978-3-031-45642-8$_5$3
7. X. Wen, X. Guo, S. Wang, Z. Lu, and Y. Zhang, *Breast cancer diagnosis: A systematic review*, Biocybernetics and Biomedical Engineering, vol. 44, no. 1, pp. 119–148, 2024, doi: 10.1016/j.bbe.2024.01.002
8. O. Díaz, A. Rodríguez-Ruíz, and I. Sechopoulos, *Artificial Intelligence for breast cancer detection: Technology, challenges, and prospects*, European Journal of Radiology, vol. 175, p. 111457, Jun. 2024, doi: 10.1016/j.ejrad.2024.111457

9. Vineela Nalla, Seyedamin Pouriyeh, Reza M. Parizi, Hari Trivedi, Quan Z. Sheng, Inchan Hwang, Laleh Seyyed-Kalantari, MinJae Woo, *Deep learning for computer-aided abnormalities classification in digital mammogram: A data-centric perspective*, Current Problems in Diagnostic Radiology, Volume 53, Issue 3, 2024, Pages 346-352, ISSN 0363-0188. doi: 10.1067/j.cpradiol.2024.01.007
10. S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, *Understanding data augmentation for classification: when to warp?*, ArXiv,Nov. 26, 2016. doi: 0.48550/arXiv.1609.08764
11. A. Mohammed and R. Kora, *A comprehensive review on ensemble deep learning: Opportunities and challenges*, Journal of King Saud University - Computer and Information Sciences, vol. 35, no. 2, pp. 757–774, Feb. 2023, doi: 10.1016/j.jksuci.2023.01.014
12. F. Zhuang et al., *A Comprehensive Survey on Transfer Learning*, Proceedings of the IEEE, vol. 109, no. 1, pp. 43-76, Jan. 2021, doi: 10.1109/JPROC.2020.3004555
13. H. Chougrad, H. Zouaki, and O. Alheyane, *Deep Convolutional Neural Networks for breast cancer screening*, Computer Methods and Programs in Biomedicine, vol. 157, pp. 19–30, Apr. 2018, doi: 10.1016/j.cmpb.2018.01.011
14. E. M. F. El Houby and N. I. R. Yassin, *Malignant and nonmalignant classification of breast lesions in mammograms using convolutional neural networks*, Biomedical Signal Processing and Control, vol. 70, p. 102954, Sep. 2021, doi: 10.1016/j.bspc.2021.102954
15. D. Muduli, R. Dash, and B. Majhi, *Automated diagnosis of breast cancer using multi-modal datasets: A deep convolution neural network based approach*, Biomedical Signal Processing and Control, vol. 71, p. 102825, Jan. 2022, doi: 10.1016/j.bspc.2021.102825
16. S. Thirumalaisamy et al., *Breast Cancer Classification Using Synthesized Deep Learning Model with Metaheuristic Optimization Algorithm*, Diagnostics, vol. 13, no. 18, Art. no. 18, Jan. 2023, doi: 10.3390/diagnostics13182925
17. Sannasi Chakravarthy, S. R., Bharanidharan, N., and Rajaguru, H, *Multi-Deep CNN based Experimentations for Early Diagnosis of Breast Cancer*, IETE Journal of Research, 69(10), 7326–7341. doi:10.1080/03772063.2022.2028584
18. A. Sahu, P. K. Das, and S. Meher, *An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images*, Biomedical Signal Processing and Control, vol. 87, p. 105377, Jan. 2024, doi: 10.1016/j.bspc.2023.105377
19. Yasaka K, Sato C, Hirakawa H, Fujita N, Kurokawa M, Watanabe Y, Kubo T, Abe O, *Impact of deep learning on radiologists and radiology residents in detecting breast cancer on CT: a cross-vendor test study*, Clinical Radiology, vol. 79, no. 1, pp. e41–e47, Jan. 2024, doi: 10.1016/j.crad.2023.09.022
20. F. H. Alhsnony and L. Sellami, *Advancing Breast Cancer Detection with Convolutional Neural Networks: A Comparative Analysis of MIAS and DDSM Datasets*, IEEE 7th International Conference on Advanced Technologies, Signal and Image Processing (ATSIP), Jul. 2024, pp. 194–199. doi: 10.1109/ATSIP62566.2024.10638886
21. Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S., Taylor, P., Betal, D., and Savage, J., *Mammographic Image Analysis Society (MIAS) database*, Apollo - University of Cambridge Repository, Aug. 2015, doi: 10.17863/CAM.105113
22. I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, *INbreast: Toward a Full-field Digital Mammographic Database*, Academic Radiology, vol. 19, no. 2, pp. 236–248, Feb. 2012, doi: 10.1016/j.acra.2011.09.014
23. Lee, R., Gimenez, F., Hoogi, A. et al, *A curated mammography data set for use in computer-aided detection and diagnosis research*, Sci Data 4, 170177 (2017), https://doi.org/10.1038/sdata.2017.177
24. A. Krizhevsky, I. Sutskever, and G. E. Hinton, *ImageNet classification with deep convolutional neural networks*, Commun. ACM, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386
25. K. Simonyan and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*, ArXiv, Apr. 10, 2015. doi: 10.48550/arXiv.1409.1556
26. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, *MobileNetV2: Inverted Residuals and Linear Bottlenecks*, ArXiv, Mar. 21, 2019. doi: 10.48550/arXiv.1801.04381
27. K. He, X. Zhang, S. Ren, and J. Sun, *Deep Residual Learning for Image Recognition*, ArXiv, Dec. 10, 2015. doi: 10.48550/arXiv.1512.03385
28. Yann Lecun, L.D. Jackel, Leon Bottou, Corinna Cortes, J. S. Denker, Harris Drucker, I. Guyon, U.A. Muller, Eduard Sackinger, Patrice Simard, V. Vapnik, *Learning algorithms for classification: A comparison on handwritten digit recognition*, World Scientific, p. 261-276, 1995
29. C. Szegedy et al., *Going deeper with convolutions*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2015, pp. 1–9. doi: 10.1109/CVPR.2015.7298594
30. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, *Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning*, roceedings of the AAAI Conference on Artificial Intelligence, vol. 31, no. 1, Art. no. 1, Feb. 2017, doi: 10.1609/aaai.v31i1.11231
31. M. Hossin and S. M.N, *A Review on Evaluation Metrics for Data Classification Evaluations*, International Journal of Data Mining and Knowledge Management Process, vol. 5, no 2, pp. 01–11, Mar. 2015, doi: 10.5121/ijdkp.2015.5201