Deep Learning for Mammography Classification: A CNN Approach on CBIS-DDSM

BOCHRA TRIQUI[1], AMANI BELKACEM[1], HICHAM KAID-SLIMANE[2], ABDESSAMAD AMIR[3]
[1]Computer Science and New Technologies Laboratory (CSTL),
University Abdel Hamid Ibn Badis, Mostaganem,
ALGERIA

[2]Medical Imaging Center 'ERAHMA',
Mostaganem,
ALGERIA

[3]Laboratory of Mathematics for Artificial Intelligence and Life Sciences (LMAI-LS),
University Abdel Hamid Ibn Badis, Mostaganem,
ALGERIA

Abstract: - It has widely been reported that breast cancer is one of the main causes of death among women throughout the entire world. However, its premature detection coupled with an appropriate treatment at the right time can certainly increase the chances of recovery and survival. The present work aims to propose an approach that is based on a convolutional neural network (CNN) for the automatic classification of mammograms from the open-source mammography database CBIS-DDSM. This CNN model was next trained on small size images (50×50×3) for the purpose of optimizing the computation time while ensuring high degree of accuracy. Then, once the model was trained for 50 epochs, it was able to classify benign and malignant tumors, with a test accuracy of 85.04%.

Key-Words: - Convolutional neural network; Grad-CAM; Computer-aided diagnosis; Mass classification; Deep learning; Early detection; Mammography; Breast cancer

Received: May 22, 2025. Revised: July 27, 2025. Accepted: September 11, 2025. Published: May 22, 2026.

1 Introduction

Nowadays, breast cancer is among the most widespread causes of death of women throughout the whole world. For this reason, it is highly recommended to undergo systematic early diagnoses in order to detect any potential signs of cancer and avoid invasive treatments. This would certainly allow for timely intervention to prevent any possible complications. In addition, the inter- and intra-reader variability is actually viewed as a substantial challenge that clearly underlines the need, necessity, and desire to develop a structured approach that provides automated diagnostic assistance. Over the past few years, the field of medical imaging has undergone a profound metamorphosis. It has indeed been shown that convolutional neural networks (CNNs) are capable of automatically learning and extracting discriminating images from spatial images. It is interesting to emphasize that numerous studies have previously been carried out to show the advantage of using convolutional neural networks (CNNs) and to confirm their capacities for classifying, segmenting, and detecting breast lesions, [1], [2], [3], [4], [5].

In order to bring to fruition this work, the present study proposes a CNN model trained on the Curated Breast Imaging Subset (CBIS) of Digital Database for Screening Mammography (DDSM) database which includes mammograms elucidated by experts. The primary purpose is to develop an automated system that can help radiologists interpret and analyze breast images and improve the accuracy of breast cancer screening.

2 Related Work

Convolutional neural networks (CNNs), which have proven to be effective tools for the detection, classification, and segmentation of breast lesions, have significantly changed the field of breast image analysis over the last ten years. To increase the precision and resilience of diagnostic support systems, a number of recent studies have taken advantage of several CNN designs, including ResNet, DenseNet, and YOLO. In [6], the authors performed simultaneous detection and classification of breast masses based on the YOLO (You Only Look Once) architecture. This method demonstrated the effectiveness of end-to-end approaches capable of locating and identifying lesions in a single pass. This represented a significant advancement in the design of real-time computer-aided diagnostic systems.

A previous study, [7], worked on breast cancer classification by introducing a dual-channel network (dual-channel nebula), which combines local and global features to improve the representation of complex tissue structures. Meanwhile, [8], demonstrated, in a study published in Nature Communications, that an artificial intelligence system could significantly reduce false positives in breast cancer screening, confirming the added value of AI as a clinical support tool.

The work of [9], highlighted the importance of integrating different imaging modalities, including 2D mammography and tomosynthesis, in order to enrich the information available for automatic detection. Deep architectures such as ResNet and DenseNet have also been widely adopted for mammogram image classification, thanks to their ability to learn complex hierarchical representations, as demonstrated in [10], the effectiveness of large-scale deep learning for automatic lesion detection, surpassing traditional image processing methods. More recently, attention has turned to the interpretability of models, an essential dimension for promoting their clinical adoption, [11], integrated artificial intelligence into mammography screening. This highlights the remarkable improvement in detection and reduction of false positives achieved through AI models.

These studies reveal two major trends: the first is the continuous improvement of classification performance through increasingly sophisticated CNN architectures; and the second is the growing integration of explainability approaches (XAI), such as Grad-CAM (Gradient-weighted Class Activation Mapping), which allow visualization of the regions of interest activated by the model and increase clinicians' confidence in diagnostic support systems.

3 Methodology

It is worth highlighting that, during the last ten years, convolutional neural networks (CNNs), which have proven to be highly efficient tools for the detection, classification and segmentation of breast lesions, have profoundly transformed the field of breast image analysis. Recently, several studies have been carried out for the purpose of improving the accuracy and robustness of diagnostic support systems, by capitalizing on a number of convolutional neural network (CNN) architectures such as ResNet, DenseNet and YOLO. In this context, [6], achieved the simultaneous detection and classification of breast masses according to YOLO (You Only Look Once) which consists of a sequence of real-time object detection systems that are based on convolutional neural networks (CNNs). The method adopted explicitly demonstrated that the integrated approaches used allow for the localization and identification of lesions in a single step. This novel strategy represents a significant advancement in the development and design of real-time computer-aided diagnostic systems.

With regard to [7], they particularly focused on breast cancer classification by introducing a dual-channel network (dual-channel nebula) which combines local and global features in order to enhance the representation of complex tissue structures. Similarly, [8], carried out a study, which was published in the scientific journal Nature Communications, to show that an artificial intelligence system is capable of significantly reducing false positives in breast cancer screening. This further confirms the added value of artificial intelligence (AI) as a clinical decision support tool.

Furthermore, [9], revealed that it is highly important to integrate different imaging modalities, including 2D mammography and tomosynthesis, to enrich the data required for improved automatic detection.

Furthermore, it should be noted that deep architectures, such as ResNet and DenseNet, are also widely employed for mammographic image classification due to their ability to learn complex hierarchical representations. In this regard, [10], successfully demonstrated that large-scale deep learning is highly effective for automatic lesion detection, thus outperforming conventional image processing techniques.

In recent times, many researchers have primarily focused on the interpretability of models as a crucial dimension for the promotion of their clinical adoption. In this respect, the variability in the performance of CNN models was investigated among various mammography manufacturers.

Two major trends are revealed by the above-mentioned studies. The first one is the ongoing enhancement of classification performance through the increasing number of more sophisticated CNN architectures, while the second is the growing incorporation of explainability approaches (Explainable Artificial Intelligence - XAI) such as Grad-CAM (Gradient-weighted Class Activation Mapping). Explainability approaches allow visualizing the regions of interest that are activated by the model and increasing the clinicians' confidence while using diagnostic support systems as well.

3.1 Mathematical Formulation of the CNN Architecture

Let X ∈ $R^{H \times W \times C}$ be an input mammographic image. A convolutional layer applies a set of learnable kernels $K^k \in R^{m \times n}$in order to extract the feature maps A^k. It defined as:

$$A^k(i,j) = \sum_{u=0}^{m-1} \sum_{v=0}^{n-1} X_{\text{pad}}(i+u, j+v)\, K^k(u,v) + b^k \quad (1)$$

Here X_{pad} represents the zero-padded input and b^k is a bias term. The unit stride is used. Each convolution is followed by a ReLU activation:

$$\text{ReLU}(x) = \max(0, x) \quad (2)$$

This introduces non-linearity and facilitates gradient propagation during training.

3.1.1 Loss Function and Probabilistic Interpretation

The proposed model addresses a binary classification problem (benign vs. malignant). Let $y \in \{0,1\}^2$ denote the one-hot encoded ground-truth label, and $\hat{y} = \text{softmax}(z)$ is the predicted probability vector. Here z_c is the network logit for class c. Hence:

$$\hat{y}_c = \frac{e^{z_c}}{\sum_{k=1}^{2} e^{z_k}}, \ c=1,2 \quad (3)$$

The model is trained using the categorical cross-entropy loss, as follows:

$$\boldsymbol{L}_{CCE} = -\sum_{c=1}^{2} y_c \log(\hat{y}_c) \quad (4)$$

This corresponds to the negative log-likelihood of a categorical distribution and is consistent with the maximum likelihood estimation, assuming independent samples. Although no explicit class weighting was applied, data augmentation and random shuffling were used to partially mitigate class imbalance.

3.2 Exploratory Data Analysis, Dataset Overview and Clinical Annotation

The images used in this study come from the public CBIS-DDSM (Curated Breast Imaging Subset of DDSM) database, which includes mammograms annotated with benign and malignant lesions. To ensure the clinical reliability of the labels, all images were reviewed and validated by a board-certified radiologist at the ERAHMA Medical Imaging Center. The radiologist confirmed or adjusted the existing annotations, classified each lesion according to the BI-RADS system, and validated the clinical relevance of the images before they were used to train and validate the CNN model. This combination of the public database and clinical expertise guarantees high-quality data for machine learning.

Before training the model, it was deemed necessary to perform an exploratory data analysis (EDA) in order to examine the distribution of pathology types and identify the correlations existing between the numerical characteristics.

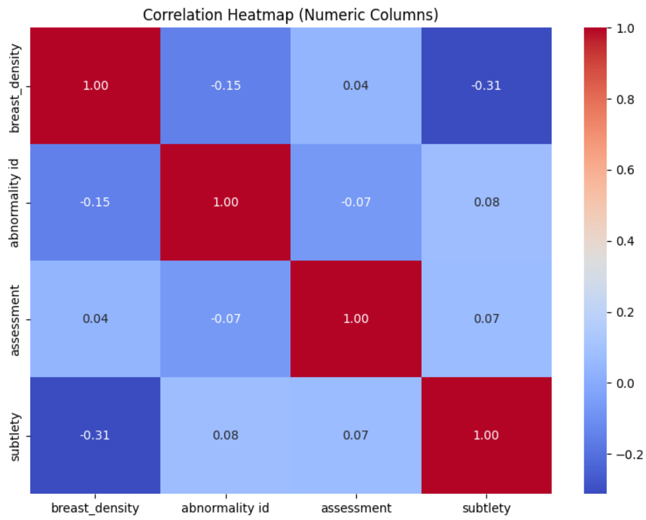

Fig. 1. Heatmap displaying the relationship between essential numerical characteristics in the mammography database.
Source: created by the authors.

The heatmap representing the correlation matrix between the numerical variables of the dataset is clearly represented in (see Figure 1) which explicitly highlights the relationships between breast density, abnormalities, evaluation score, and lesion subtlety level, thus offering insight into potential patterns before model training.

3.3 Image preprocessing and CNN training protocol

The CBIS-DDSM images were resized (50×50×3), normalized and augmented before training a CNN that was built by means of build_cnn_model, using the Adam optimizer (lr = 1e-4), a batch size of 64, over 50 epochs, with Early Stopping and ReduceLROnPlateau in order to optimize training and avoid overfitting.

3.3.1 Optimization and Regularization Strategy: Adam Optimizer

Network parameters are optimized using the Adaptive Moment Estimation (Adam) algorithm. Given the gradient $g_t = \nabla\theta\ \square_t$ at iteration t , Adam computes first and second moment estimates as:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (5)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (6)$$

Followed by bias correction:

$$\widehat{m}_t = \frac{m_t}{(1 - \beta_1^t)} \quad (7)$$

$$\hat{v}_t = \frac{v_t}{(1 - \beta_2^t)} \quad (8)$$

The parameter update rule is expressed as:

$$\theta_{t+1} = \theta_t - \eta \frac{\widehat{m}_t}{(\sqrt{\hat{v}_t} + \epsilon)} \quad (9)$$

Where η denotes the learning rate. Adam is particularly suitable for medical imaging tasks due to its robustness to noisy gradients and its ability to adapt learning rates individually for each parameter.

Early Stopping is then applied by monitoring the validation loss and terminating training when no improvement is observed for a fixed number of epochs. This may be interpreted as an implicit regularization strategy for limiting the model complexity. Learning-rate scheduling (ReduceLROnPlateau) progressively reduces the learning rate during training, hence constraining parameter updates and improving convergence stability. Using these strategies together help prevent overfitting and ensure stable and efficient optimization of the CNN parameters.

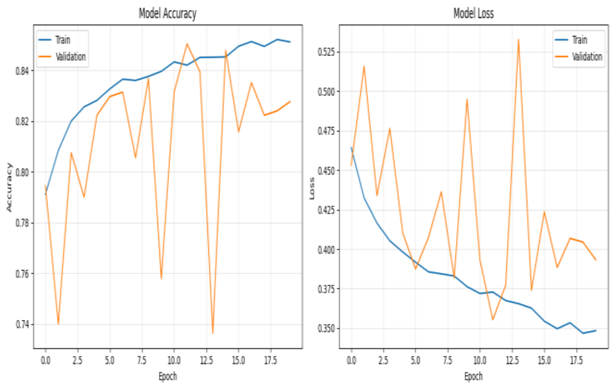

Fig. 2. Development of model performance throughout the training phase.
Source: created by the authors.

In (see Figure 2), the curve on the left depicts the variation in model accuracy across the training and validation sets, while the curve on the right illustrates the loss function. It should be emphasized that, at the end of the training process, the model's accuracy increases steadily to approximately 85%, hence indicating strong learning potential. The performance variability, which is due to the complexity of the validation set, can clearly be noticed in the variations observed on the validation curve. In addition, the steady loss function decline across the training set validates the model's convergence, while the loss function oscillations over the validation set indicate slight overfitting. Therefore, the model under consideration shows good generalization capacities with acceptable steadiness on the test data.

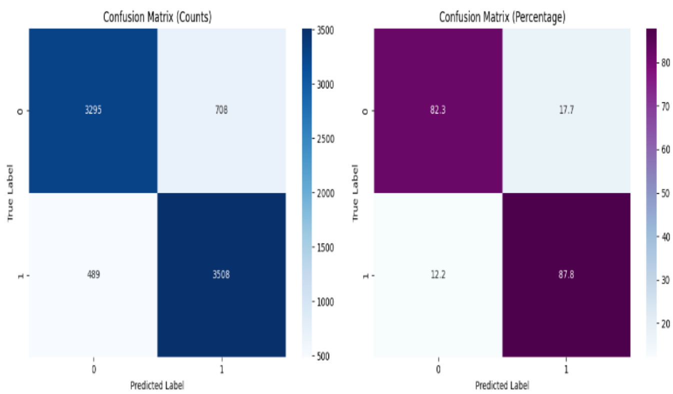

Fig. 3. Confusion matrices depicting the performance of the breast mass classification model.
Source: created by the authors.

The above figure illustrates the performance of the breast cancer mass image classification model. The results are presented in the form of percentage of correct predictions for each class in the right-hand matrix, while the predictions are displayed as absolute values in the left-hand matrix, as clearly shown in (see Figure 3).

The results suggest that the model performs well in distinguishing between benign (0) and malignant (1) masses. Indeed, this model was able to correctly identify 82.3% of benign masses and 87.8% of malignant masses. The CNN model exhibited high robustness against misclassification, with only 17.7% of false positives, i.e. benign masses classified as malignant, and 12.2% of false negatives. Although the results explicitly indicate that the model is capable of automatically detecting and classifying breast abnormalities, they nevertheless suggest that its overall sensitivity could be enhanced by adjusting the decision threshold or by providing additional training to avoid the risk of false negatives.

3.3.2 Explainability Method: Grad-CAM

The Gradient- weighted Class Activation Mapping (Grad-CAM) is employed to provide visual explanations of the CNN decisions. Let y^c denote the score of class c before the softmax layer, and A^k is the k-th feature map of the last convolutional later. The importance weights α_k^c are then computed as:

$$\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \qquad (10)$$

Where Z is a normalization factor. The Grad-CAM localization map is then obtained using:

$$L_{Grad-CAM}^c = \mathrm{ReLU}\left(\sum_k \alpha_k^c\, A^k\right) \qquad (11)$$

This contributes to highlighting the spatial regions that most strongly influence the network's prediction for class c. The resulting heatmaps are upsampled to the input image resolution and overlaid on the original mammograms for visual interpretation, as illustrated in Figure 4.

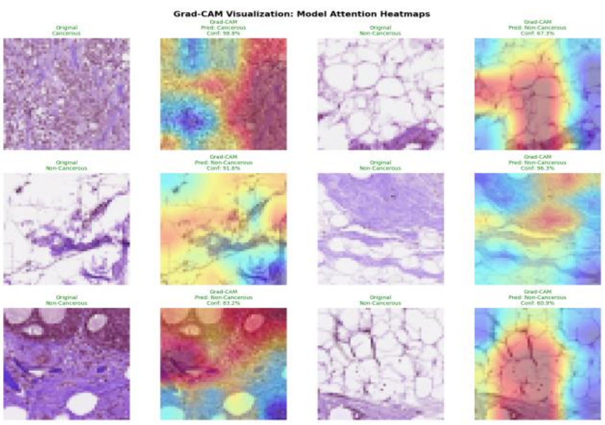

Fig. 4. Grad-CAM visualization displaying areas of interest of the model on a breast mass image.
Source: created by the authors.

Furthermore, when predicting a malignant mass, the model focuses primarily on the red or orange areas highlighted in Figure 4. The blue or neutral areas correspond to zones considered less significant and the model mainly concentrates on the tumor area itself, which confirms that the convolutional neural network (CNN) architecture learns consistent discriminatory features from both anatomical and radiological perspectives.

The explainability of the network's decisions, which represents an essential component in the field of AI-assisted diagnostics in medical imaging, is confirmed, and visual interpretation can then enhance the transparency and reliability of the model.

4 Results and Discussion

Model performance is evaluated using standard metrics derived from the confusion matrix. Let TP, TN, FP and FN, denote the true positives, true negatives, false positives and false negatives, respectively. The metrics are then defined as:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \qquad (12)$$

$$\text{Sensitivity (Recall)} = \frac{\text{TP}}{\text{TP+FN}} \qquad (13)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \qquad (14)$$

These metrics provide a comprehensive assessment of the classification performance, particularly in the context of medical diagnosis in which false negatives and false positives have distinct clinical implications.

Accuracy is assessed for each execution, and variability over several executions can be reported as mean ± standard deviation, giving an estimate of performance robustness.

In this regard, the proposed model gave quite satisfactory results throughout the test set, with an overall accuracy of 85.04% and a test loss of 0.3553.

The evolution of validation accuracy and training loss in the training and validation sets is represented by the learning curves, as illustrated in Figure 2. Furthermore, a progressive improvement in learning performance, reflecting good convergence of the model, along with a relative stability of the validation curve, confirming correct generalization without significant overfitting, are then noticed.

In addition, recognition rates of 87.8% and 82.3% were achieved respectively for malignant and benign masses by the model under consideration, according to the confusion matrix shown in Figure 3. The recorded results evidence a good balance between sensitivity and specificity, with an effective ability to distinguish between the two classes. Moreover, the residual errors primarily involve false positives which are often associated with benign masses showing an irregular morphology. These are often difficult to recognize visually.

Furthermore, (see Figure 4) depicts the Grad-CAM images that were used to conduct a qualitative analysis in order to determine the model's focus areas during categorization. Likewise, the heatmaps suggest that the network mainly focuses on the relevant tumor locations in the image. It can therefore be asserted that the model's selection is corroborated by consistent radiological and anatomical features.

It should be emphasized that the explainability phase is highly essential in the medical context because it strengthens clinical confidence in the results generated by the diagnostic support system and promotes transparency in the decision-making process.

Overall, it may be concluded that the results clearly indicate that explainable deep learning techniques for computer-assisted medical imaging and the reliability of the proposed CNN model show great potential for the automatic detection, identification and categorization of breast masses.

It should be noted that statistical significance testing and confidence interval estimation were not performed in this study. The reported results correspond to point estimates obtained on a fixed train–test split. Future work will focus on assessing statistical robustness through cross-validation and confidence interval analysis.

5 Conclusion and Perspectives

Convolutional neural networks (CNNs) were employed in this study to automatically classify breast masses based on high-quality mammographic images. In addition, an experimental analysis was performed to confirm the proposed model's ability to reliably differentiate between benign and malignant tumors. This analysis achieved an accuracy of 85.04%, which represents a significant advancement in the field of computer-aided diagnostics.

Likewise, the learning curves, which exhibited stable convergence with a good compromise between training and validation, confirmed the effectiveness of the selected CNN architecture and optimized hyperparameters. Moreover, the model's ability to identify anomalies with high accuracy, while reducing false positives, was confirmed by the high sensitivity (87.8%) and robust specificity (82.3%) of the confusion matrix. These features are highly essential for clinical applications requiring high diagnostic accuracy.

Furthermore, the integration of Grad-CAM images confirmed the interpretability and explainability of the model. It was also shown that the activated regions correspond to anatomically significant sections of the mammogram. This underscores the crucial importance of interpretable deep learning techniques in medical imaging, as they help improve transparency and enhance clinical confidence in the system.

The above-mentioned results pave the way for the adoption of intelligent systems capable of assisting radiologists in their daily work. They also confirm the potential of CNN models for computer-aided diagnosis, particularly for the premature identification of breast cancer.

In addition, future work will focus on refining the preprocessing pipeline through advanced enhancement and normalization strategies, which are expected to improve feature representation and classification performance. Furthermore, the integration of advanced regularization techniques to minimize overfitting, the use of more diverse and targeted data augmentation strategies, and the investigation of hybrid CNN-Transformer architectures are promising directions to enhance the robustness and interpretability of the proposed system.

Finally, the generalizability and clinical reliability of the model will be further validated through adaptation to multicenter datasets and evaluation on sufficiently large cohorts, paving the way for its potential integration into AI-assisted medical diagnostic platforms.

References

[1] Dhungel N., Carneiro G., Bradley A. P., "Deep learning and structured prediction for segmentation of masses in mammograms," in Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015, Cham, Switzerland: Springer, pp. 605-612, 2015. https://doi.org/10.1007/978-3-319-24553-9_74

[2] McKinney S. M., Sieniek M., Godbole V., Godwin J., Antropova N., Ashrafian H., Back T., Chesus M., Corrado G. S., Darzi A., Etemadi M., Garcia-Vicente F., Gilbert F. J., Halling-Brown M., Hassabis D., Jansen S., Karthikesalingam A., Kelly C. J., King D., Ledsam J. R., Melnick D., Mostofi H., Peng L., Reicher J. J., Romera-Paredes B., Sidebottom R., Suleyman M., Tse D., Young K. C., De Fauw J., Shetty S., "International evaluation of an AI system for breast cancer screening," *Nature*, vol. 577, pp. 89–94, 2020. https://doi.org/10.1038/s41586-019-1799-6

[3] Lotter W., Diab A. R., Haslam B., Kim J. G., Grisot G., Wu E., Wu K., Onieva J. O., Boyer Y., Boxerman J. L., Wang M., Bandler M., Vijayaraghavan G. R., Sorensen A. G., "Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach," *Nature Medicine*, vol. 27, pp.244–249,2021.https://doi.org/10.1038/s41591-020-01174-9

[4] Shen L., Margolies L. R., Rothstein J. H., Fluder E., McBride R., Sieh W., "Deep learning to improve breast cancer detection on screening mammography," *Scientific Reports*, vol. 9, no. 1, Article 12495, 2019. https://doi.org/10.1038/s41598-019-48995-4

[5] Fathimathul Rajeena P. P., Tehsin S., "A framework for breast cancer classification with deep features and modified grey wolf optimization," *Mathematics*, vol. 13, no. 8, Article 1236, 2025. https://doi.org/10.3390/math13081236

[6] Al-Masni M. A., Al-Antari M. A., Choi M. T., Han S. M., Kim T. S., "Simultaneous detection and classification of breast masses via YOLO-based CAD," *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 85–94, 2018. https://doi.org/10.1016/j.cmpb.2018.01.017

[7] Zhang Q., Song S., Xiao Y., Chen S., Shi J., Zheng H., "Dual-mode artificially-intelligent diagnosis of breast tumours in shear-wave elastography and B-mode ultrasound using deep polynomial networks," Medical Engineering & Physics, vol. 64, pp. 1–6, 2019. https://doi.org/10.1016/j.medengphy.2018.12.005

[8] Shen Y., Shamout F. E., Oliver J. R., Witowski J., Kannan K., Park J., Wu N., Huddleston C., Wolfson S., Millet A., Ehrenpreis R., Awal D., Tyma C., Samreen N., Gao Y., Chhor C., Gandhi S., Lee C., Kumari-Subaiya S., Leonard C., Mohammed R., Moczulski C., Altabet J., Babb J., Lewin A., Reig B., Moy L., Heacock L., Geras K.J., "Artificial intelligence system reduces false-positive findings in the interpretation of breast ultrasound exams," *Nature Communications*, vol. 12, Article 5645, 2021. https://doi.org/10.1038/s41467-021-26023-2

[9] Geras K. J., Mann R. M., Moy L., "Artificial intelligence for mammography and digital breast tomosynthesis: Current concepts and future perspectives," *Radiology*, vol. 293, no. 2, pp. 246–259, 2019. https://doi.org/10.1148/radiol.2019182627

[10] Kooi T., Litjens G., van Ginneken B., Sánchez C. I., "Large-scale deep learning for computer-aided detection of mammographic lesions," *Medical Image Analysis*, vol. 35, pp.303–312, 2017.https://doi.org/10.1016/j.media.2016.07.007

[11] Abu Abeelh E., Abuabeileh Z., "Screening mammography and artificial intelligence: A comprehensive systematic review,"*Cureus*, vol. 17, no. 2, Article 79353, pp. e79353, 2025. https://doi.org/10.7759/cureus.79353

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Bochra Triqui and Amani Belkacem equally contributed to the computational modeling, CNN implementation, and data analysis. Hicham Kaid-Slimane contributed to the clinical interpretation and radiological validation. Abdessamad Amir contributed to the mathematical formulation and statistical analysis. All authors contributed to study design, interpretation of results, and manuscript preparation.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.