



Ensemble Learning for Breast Cancer Detection in Mammography

Faouzi-Ayoub El Ghanaoui^{1*}, Elmehdi Aniq^{1,2}, Mohamed Chakraoui¹, Youness Khourdifi¹

¹LS2ME, Polydisciplinary Faculty of Khouribga, Sultan Moulay Slimane University, Beni Mellal 23020, Morocco

²LAMIGEP, EMSI Marrakech, Marrakech 40000, Morocco

Corresponding Author Email: ayoub.elghanaoui@gmail.com

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.300724>

ABSTRACT

Received: 24 March 2025

Revised: 20 June 2025

Accepted: 12 July 2025

Available online: 31 July 2025

Keywords:

deep learning, mammography, convolutional neural network (CNNs), transfer learning, ensemble learning, breast cancer, computer-aided diagnosis (CAD), medical imaging

Breast cancer is among the most common and lethal malignancies affecting women worldwide, with its rising incidence constituting a major public health concern. Early detection remains essential to reducing mortality, as it continues to be the second leading cause of cancer-related deaths among women globally. This study explored deep learning-based computer-aided diagnosis (CAD) systems using convolutional neural networks (CNNs), integrating data augmentation, transfer learning, and ensemble learning. Experiments were performed on publicly available mammography datasets (MIAS, CBIS-DDSM, and INbreast). Across evaluation metrics, ensemble learning achieved the best performance on the INbreast database, with an accuracy of 0.9976, AUC of 1.0, recall of 0.9959, precision of 0.9983, and F1-score of 0.9971. These findings highlight the potential of deep learning-driven CAD systems to improve early breast cancer detection, enhance diagnostic precision, and reduce the overall burden on healthcare systems.

1. INTRODUCTION

Abnormal cell division in the mammary gland leads to the formation of either malignant or benign tumors, which is the cause of breast cancer (BC) [1]. Breast cancer is one of the most dangerous cancers worldwide. The World Health Organization reported 2.3 million new cases and 666,000 related deaths in 2022. These figures are projected to reach 3.3 million cases and 1.2 million deaths by 2050 [2]. Figure 1 illustrates a normal breast image and a suspicious breast from the INbreast database.

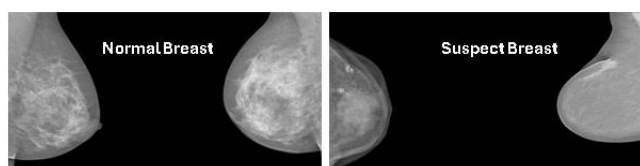


Figure 1. Normal breast image and a suspicious breast

Early diagnosis is critical for improving survival rates, but it remains complex and time-consuming [3]. Emerging technologies such as Deep Learning (DL) offer more efficient, accurate methods for diagnosing breast cancer, particularly when integrated into Computer-Aided Detection (CAD) systems, improving diagnostic accuracy and reducing false results [4, 5]. However, DL adoption is still limited as healthcare professionals require proof of its effectiveness [6, 7]. Deep learning and Deep Convolutional Neural Networks rely on analyzing patient data to identify patterns and make predictions, requiring large datasets and accurate parameters for better results [8]. The research also explores the potential

of deep learning algorithms to enhance mammography-based breast cancer classification and detection, despite digital mammography being the most widely used method [9].

Convolutional Neural Networks (CNNs), inspired by the human visual cortex, have emerged as powerful tools for medical image analysis [10]. Many studies have applied CNNs to breast cancer diagnosis in mammography [11]. Ensemble strategies remain relatively underexplored, despite their potential to deliver more robust and reliable results by integrating complementary features from pre-trained architectures such as VGG16, ResNet50, and MobileNetV2. In this study, ensemble learning is combined with transfer learning and data augmentation, and the framework is evaluated on three benchmark datasets. This integrated approach enables a more comprehensive evaluation of CAD systems for mammographic classification, providing valuable insights into their clinical applicability. The ultimate objective is to develop a breast cancer classification model that can act as a reliable “second opinion” for radiologists, thereby reducing the risk of diagnostic errors.

This paper is structured as follows: Section 2 provides an overview of recent work in this field. In Section 3, the method for classifying breast anomalies into normal and suspicious is described. Section 4 presents and discusses the results of the experiments carried out. Finally, section 5 concludes and highlights directions for future work.

2. RELATED WORK

The introduction of convolutional neural networks (CNNs) for breast cancer diagnosis has led to significant advances in

architectures such as VGG16, ResNet50, and MobileNetV2, all of which have proven dominant in analyzing difficult medical images. VGG16 provides a clean structure for effective classification; ResNet50 introduces residual connections to help learn in deep networks; and MobileNetV2 is designated for the edge with low-compute resources. These models have collectively advanced breast cancer image classification, leading to more accurate detection of malignant patterns within medical datasets. But the limitation of sizes of the existing datasets and the heavy computational requirement of these models pose challenges that urgently need addressing through various methods such as data augmentation, kernel optimization, and transfer learning.

A CNN-based approach for segmenting and classifying two-dimensional imagery in improved breast cancer detection is proposed in study [12]. The proposed method includes augmentation techniques and testing on various datasets. In study [13], a comparative analysis is presented here regarding respective CNN architectures for early diagnosis of breast cancer. An additional preprocessing stage has been employed here to remove artifacts and improve image quality, while also optimizing hyperparameters for the experiments, which include the number of epochs and batch size.

In Sahu et al. [14], an ensemble classifier based on deep learning is proposed for breast cancer detection by combining three robust transfer learning models: AlexNet, ResNet and MobileNetV2: The authors use the modified amplification filter based on the Gaussian Laplacian (LoGMHBF) to further enhance image quality and increase performance. The proposed model can be employed in multimodal datasets thanks to its versatility and reliability.

The investigations carried out in Yasaka et al. [15] demonstrate the power of deep learning in breast cancer detection and how it provides an automated second opinion to radiologists and allied medical personnel. The method found to work well in a clinical setting was applied to patients in varying stages of the disease with validation data covering multiple equipment providers. In the study, Han et al. [16], ultrasound and mammogram images are combined with clinical and pathological data to predict breast cancer

recurrence. All patients were observed for more than three years. The deep learning models was used to predict which patients will experience a recurrence which ones will not. Also, Jeneffa et al. [17] presents a method that combines MobileNetV2 with Long Term Memory (LSTM) to improve tumor detection in mammographic images using the Digital Database for screening Mammography (DDSM). The finding shows that deep learning-based systems demonstrate their capacity to enhance cancer diagnosis and to predict the recurrence of this illness. In study [18], a CNN-based classification system was proposed with extensive preprocessing steps, including denoising, image enhancement, region of interest (ROI) extraction, data augmentation, and resizing. The methodology was validated across multiple datasets, emphasizing its adaptability to different imaging sources.

These studies confirm the impact of CNN architectures on breast cancer diagnosis, supported by data optimization and the integration of various learning techniques for improved performance

3. METHODS

3.1 Datasets

In this study, we employ three open-access mammography datasets that are suitable for developing, training, and testing breast cancer models based on deep learning. The Mammographic Image Analysis Society (MIAS) dataset [19], although dated, remains ideal for small-scale studies and foundational research. INbreast [20] is appropriate for advanced deep learning techniques due to its high-resolution digital images, although it has not been updated since 2017. The Curated Breast Imaging Subset of DDSM (CBIS-DDSM) [21] excels in large-scale studies, offering a diverse and extensive collection of cases. Table 1 provides a detailed overview of the specific characteristics of each dataset used in this analysis.

Table 1. The specific characteristics of each dataset

Dataset	Information
MIAS UK 1994	161 cases, 322 images, MLO views, all types of anomalies, normal, benign, and malignant categories, no BI-RADS classifications
Inbreast Portugal 2010	115 cases, 410 images, MLO and CC views, all types of anomalies, detailed lesion annotations, benign and malignant categories, BI-RADS classifications
CBIS DDSM USA 2017	1644 cases, 3103 images, MLO and CC views, all types of anomalies, normal, benign, and malignant categories, BI-RADS classifications

3.2 Data augmentation

Data augmentation techniques generate new samples of the training dataset through random transformations on the available data. This approach has multiple benefits, including accelerating the convergence process and preventing overfitting. For small datasets, the simplest approach is to perform basic transformations such as translation, zooming, flipping, mirroring, and rotation [22].

3.3 Transfer learning

Transfer learning enables the use of small datasets, such as medical images, by eliminating the need for costly training of

deep models from scratch [23].

3.4 Ensemble learning

Ensemble learning is a robust strategy that integrates predictions from multiple models to achieve superior outcomes. By harnessing the strengths of various models, this approach can significantly improve classification accuracy compared to individual models [24].

3.5 CNN architecture

VGGNet [25] overcomes dataset limitations, allowing deep model training with flexible input sizes. As an extension of

AlexNet. VGG16 include 13 convolutional layers and 5 pooling layers within five convolutional blocks. Each block employs 3×3 filters, followed by max-pooling to halve input size and double filters, optimizing with fewer parameters. Batch normalization accelerates convergence and minimizes error, preventing overfitting, while dropout reduces error rates further. The final block has three fully connected layers (4096, 4096, and 1000 units). For this study's two-class output, the last layer (FC8) is modified to two units. Despite demanding significant computational resources, managing around 140 million parameters, this architecture has been quickly adopted and applied in healthcare for developing simple and effective predictive models [13, 26, 27]. Figure 2 illustrates the VGG16 architecture.

ResNet50, also known as a deep residual network, is a pre-trained model built on ImageNet, designed to address vanishing gradient issues by allowing the network to skip one or more layers. This structure comprises 50 layers arranged in residual blocks with shortcut connections, enabling layer skipping by adding the convolutional output to the block's input tensor, thereby reducing computational complexity while supporting complex feature learning. The architecture begins with an initial convolutional layer, followed by batch normalization and two pooling layers, and includes 16 residual modules. These modules alternate between blocks with 4 convolutional layers and those with 3, each followed by batch normalization and ReLU activation for efficient connectivity. ResNet50 achieves a top-5 error rate of 7.8 on ImageNet [28]. This architecture is further validated by researchers and has demonstrated effectiveness in various predictive models. The results obtained reinforce its strong reputation in breast cancer classification tasks [13, 26, 27].

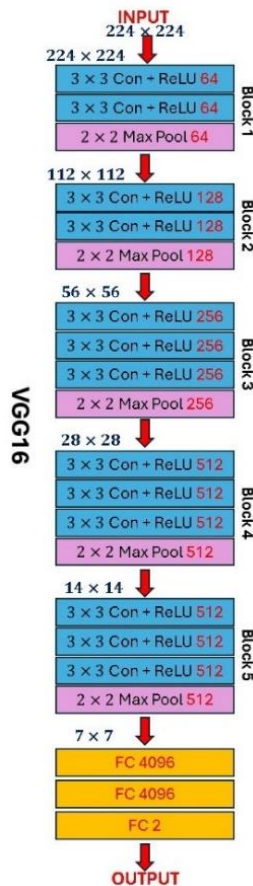


Figure 2. VGG16 architecture

MobileNetV2, an evolution of MobileNetV1 by Howard et al. [29], is designed for embedded systems, mobile devices, and resource-limited environments, minimizing computational and memory demands with minimal loss in accuracy. Its core innovation, the inverted residual with a linear bottleneck, processes low-dimensional inputs using depthwise convolution before compressing them, enhancing efficiency. This replaces standard convolution with depthwise convolution for lightweight filtering and pointwise (1×1) convolution for output merging, allowing MobileNetV2 to operate on individual channels, reducing complexity while preserving feature extraction.

MobileNetV2 contains 53 layers, fewer parameters, and an input size of 224×224 . It leverages depthwise separable convolutions with two residual block types: stride 1 for regular processing and stride 2 for downsampling, striking a balance between performance and efficiency [30]. MobileNetV2 is widely applied in medical image analysis [14, 26]. Figure 3 illustrates the MobileNetV2 building blocks [30].

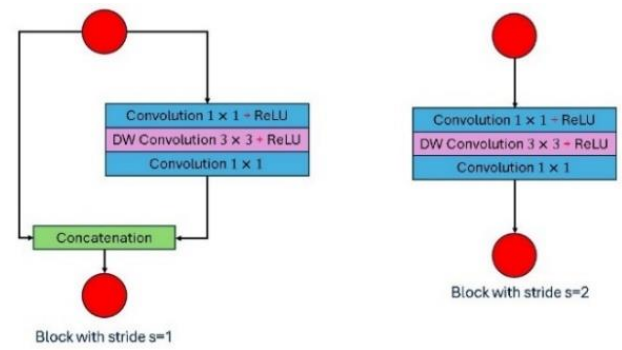


Figure 3. Structure of the building blocks of lightweight MobileNetV2 CNN model

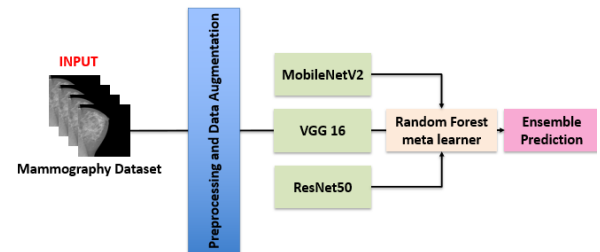


Figure 4. Flowchart of the proposed ensemble pipeline

Our pipeline begins with preprocessing and data augmentation to increase dataset variability and reduce overfitting. Using transfer learning, we employed three convolutional neural networks (CNNs): VGG16, ResNet50 and MobileNetV2, each independently estimating the probability of an image belonging to the benign or malignant class. The choice of these networks was motivated by their complementary strengths. VGG16 provides a widely used baseline with a straightforward convolutional design; ResNet50 incorporates residual connections, enabling deeper and more generalizable feature extraction; and MobileNetV2 offers a lightweight architecture well suited for resource-constrained environments. To combine their outputs, a stacking strategy was applied in which the predictions from the three CNNs were used as input features for a meta-classifier. A Random Forest was chosen as the meta-learner due to its robustness and ability to capture nonlinear

relationships. Its hyperparameters (number of estimators and tree depth) were optimized using RandomizedSearchCV with cross-validation, while the area under the ROC curve (AUC) served as the evaluation metric. Figure 4 presents an overview of the complete workflow.

3.6 Key metrics

Key metrics for model evaluation and comparison include accuracy, which measures the proportion of correct predictions (Eq. (1)). Precision, or positive predictive value, represents the percentage of predicted malignant cases that are correct (Eq. (2)). Sensitivity (Recall), or true positive rate, indicates the proportion of malignant cases detected (Eq. (3)). The F1 score, defined as the harmonic mean of Precision and Sensitivity, provides a balanced assessment (Eq. (4)). AUC, which represents the area under the ROC curve, serves as a global performance measure. Formal definitions are given below [31].

$$\textbf{Accuracy} \quad \text{Acc} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \quad (1)$$

$$\textbf{Recall} \quad \text{Sv} = \frac{\text{TP}}{\text{TP}+\text{FN}} \quad (2)$$

$$\textbf{Specificity} \quad \text{S}_p = \frac{\text{TN}}{\text{TN}+\text{FP}} \quad (3)$$

$$\textbf{Precision} \quad \text{P} = \frac{\text{TP}}{\text{TP}+\text{FP}} \quad (4)$$

$$\textbf{F1 score} \quad \text{F1} = 2 \times \frac{\text{P} \times \text{Sv}}{\text{P} + \text{Sv}} \quad (5)$$

- **TN:** True negatives (correctly identified negatives)
- **TP:** True positives (correctly identified positives)
- **FN:** False negatives (positives incorrectly identified as negatives)
- **FP:** False positives (negatives incorrectly identified as positives)

4. EXPERIMENTAL RESULTS

Three open access datasets are utilized in this study: MIAS, INBREAST and CBIS-DDSM, each providing craniocaudal (CC) and mediolateral oblique (MLO) views for both breasts, with cases categorized as benign or malignant. The MIAS, INBREAST and CBIS-DDSM datasets contain 161, 115 and 1,597 cases respectively, corresponding to 330, 410 and 10,237 images.

To enhance the training process, each image was augmented four times with rotations of 0°, 90°, 180°, and 270°. Additional random transformations included horizontal and vertical shifts, rotations ranging from 0° to 30°, slight shearing, zooming between -0.2 and +0.2, and horizontal flipping. Newly generated pixels from these transformations were managed using a fill strategy, while test images remained unaltered. These augmentations were performed in real time using the Keras ImageDataGenerator, ensuring unique training images. Figure 5 illustrates examples of images produced through data augmentation.

To ensure balanced training, each dataset was adjusted to maintain an equal number of images per category, reducing potential bias during model training. The MIAS and CBIS-DDSM datasets were used independently for training the

models, employing a 70/20/10 split for training, validation, and testing, respectively. An ensemble learning approach was then applied, and the final model was evaluated on the INbreast database. Table 2 provides details on the number of images allocated for training, validation, and testing across all datasets, along with the data augmentation techniques used and the total images generated.

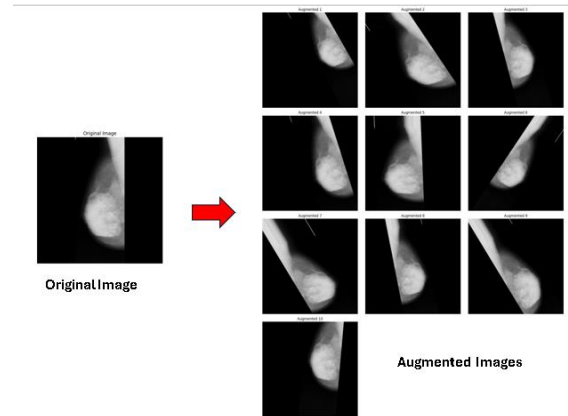


Figure 5. Example of data augmentation

Table 2. Number of images allocated for training, validation, and testing across all datasets

DATABASES	Total Images	
	Before Augmentation	After Augmentation
MIAS	330	4554
CBIS-DDSM	10237	14316
INbreast	410	410

Table 3. Classification results without data augmentation for different models

Model	Acc	AUC	Rel	P	F1	T (s)
MIAS						
VGG 16	0.59	0.57	0.54	0.54	0.54	62
ResNet50	0.55	0.45	0.42	0.31	0.34	55
MobileNetV2	0.66	0.43	0.50	0.33	0.40	70
CBIS DDSM						
VGG 16	0.75	0.84	0.75	0.75	0.75	287
ResNet50	0.63	0.70	0.61	0.61	0.61	294
MobileNetV2	0.66	0.71	0.63	0.65	0.64	225

Table 4. Classification results with data augmentation for different models

Model	Acc	AUC	Rel	P	F1	T (s)
MIAS						
VGG 16	0.90	0.97	0.90	0.91	0.90	430
ResNet50	0.73	0.80	0.73	0.74	0.72	423
MobileNetV2	0.95	0.99	0.95	0.95	0.95	317
CBIS DDSM						
VGG 16	0.99	0.99	0.99	0.99	0.99	1160
ResNet50	0.96	0.98	0.96	0.96	0.96	1147
MobileNetV2	0.95	0.99	0.95	0.95	0.95	863

To identify the optimal parameters for the model, we investigated a variety of hyperparameters, including the number of epochs (ranging from 80 to 200), learning rates 0.0001 to 0.0000001, and the implementation of techniques such as data augmentation, dropout, regularization, and callbacks. The open-source OPTUNA library played a crucial role in testing and optimizing multiple hyperparameters

concurrently, enabling the efficient discovery of the best-performing configuration. Table 3 presents the classification results without data augmentation, while Table 4 shows the classification results with data augmentation. Then, an ensemble learning approach was implemented with the models that achieved the best scores on the MIAS and CBIS-DDSM datasets. The INbreast dataset later served as the test set. Table 5 presents the classification results achieved using the ensemble learning strategy.

Table 5. The classification results using an ensemble learning strategy

Model	Accuracy	AUC	Recall	Precision	F1 Score
1	0.9976	1	0.9959	0.9983	0.9971
2	0.9927	1	0.9878	0.9948	0.9912

Table 3 demonstrates that limited data negatively affects each model's performance, indicating that training with small, unbalanced datasets and limited resources is suboptimal. Data augmentation, however, enables training on larger datasets, facilitating better parameter tuning and enhancing

generalization to real-world cases. This approach is validated in Table 4, where data augmentation leads to results comparable to or even surpassing those of current state-of-the-art models.

The ensemble learning strategy further improved performance in two cases: Model 1, which combines models trained on the MIAS dataset, and Model 2, which integrates models trained on the CBIS-DDSM dataset. In both scenarios, the INbreast dataset was used as the test set, showcasing enhanced classification accuracy and greater robustness across diverse training data sources. Both ensemble model 1 (trained on MIAS) and ensemble model 2 (trained on CBIS-DDSM) accurately classified all 287 normal images as normal. Among 123 suspicious images, model 1 misclassified 1 case, while model 2 misclassified 4 cases. Figure 6 illustrates the combined confusion matrices for Models 1 (left) and 2 (right) for the INbreast dataset.

The proposed classification system was compared with recent CAD systems. This comparison highlights the superiority of our approach regarding accuracy, AUC, precision, and F1 score, as shown in Table 6.

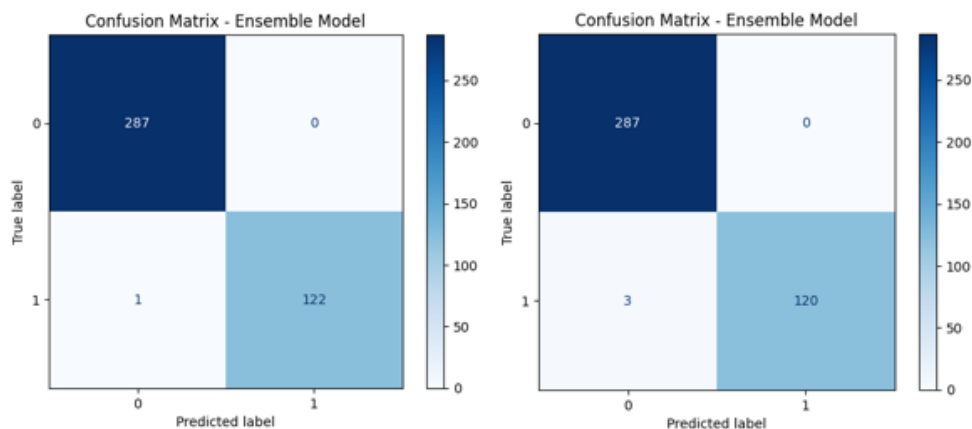


Figure 6. Combined confusion matrices for the INbreast dataset

Table 6. Comparison of our approach results with state-of-the-art approaches

Ref.	Dataset	Acc	AUC	Rel	P	F1
our work	MIAS	0.9976	1	0.9959	0.9983	0.9971
	CBIS DDSAM	0.9902	1	0.9837	0.9931	0.9883
[26]	DDSM	0.9887	0.988	0.9898	0.9879	0.9799
[27]	DDSM	0.9798	0.9846	0.9763	0.9651	0.9597
[13]	CBIS DDSM	0.9758	-	-	-	-
[18]	INBREAST	0.9652	0.98	0.9655	-	-
[32]	INBREAST	0.9550	0.97	-	-	-

5. CONCLUSION

This study investigated the potential of deep learning–based CAD systems for breast cancer diagnosis through an ensemble of VGG16, ResNet50, and MobileNetV2. The proposed approach achieved strong performance across multiple benchmark datasets, offering valuable support as a “second opinion” to radiologists. However, these results must be interpreted with caution. In particular, the small size and limited diversity of datasets such as INbreast restrict the ability to generalize findings to larger and more heterogeneous clinical populations. Moreover, while many studies, including

ours, report improved results compared to the state of the art in controlled testing setups, this does not guarantee robustness in real-world clinical deployment. AI models remain highly sensitive to variations in patient cohorts, imaging protocols, and annotation quality, which often limit their applicability outside the training environment.

Future work should therefore prioritize validation on larger, multi-center, and demographically diverse datasets, along with systematic analyses of model robustness under varying acquisition and labeling conditions. Beyond reporting incremental performance gains, it is equally important to highlight scenarios where AI systems underperform, as such

insights are crucial to building clinically reliable tools. Shifting the notion of “novel contribution” from ever-increasing model complexity toward a deeper understanding of AI limitations and failure modes will ultimately help bridge the gap between promising experimental results and safe, trustworthy deployment in medical practice.

ACKNOWLEDGMENT

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

REFERENCES

- [1] Dar, R.A., Rasool, M., Assad, A. (2022). Breast cancer detection using deep learning: Datasets, methods, and challenges ahead. *Computers in biology and medicine*, 149: 106073. <https://doi.org/10.1016/j.compbiomed.2022.106073>
- [2] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74(3): 229-263. <https://doi.org/10.3322/caac.21834>
- [3] Aniq, E., Chakraoui, M., Mouhni, N., Aboulfalah, A., Rais, H. (2024). Breast cancer stage determination using deep learning. In *Information Systems and Technologies: WorldCIST 2023*, pp. 550-558. https://doi.org/10.1007/978-3-031-45642-8_53
- [4] Elmehdi, A.N.I.Q., El Ghanaoui, F.A., Chakraoui, M. (2025). Vision transformers for breast cancer mammographic image classification. *Statistics, Optimization & Information Computing*. <https://doi.org/10.19139/soic-2310-5070-2539>
- [5] El Ghanaoui, F.A., Aniq, E., Chakraoui, M., Khouridifi, Y. (2025). A novel CNN architecture for breast cancer detection. *Statistics, Optimization & Information Computing*. <https://doi.org/10.19139/soic-2310-5070-2559>
- [6] Khalifa, M., Albadawy, M. (2024). AI in diagnostic imaging: Revolutionising accuracy and efficiency. *Computer Methods and Programs in Biomedicine Update*, 5: 100146. <https://doi.org/10.1016/j.cmpbup.2024.100146>
- [7] Nwanosike, E.M., Conway, B.R., Merchant, H.A., Hasan, S.S. (2022). Potential applications and performance of machine learning techniques and algorithms in clinical practice: A systematic review. *International Journal of Medical Informatics*, 159: 104679. <https://doi.org/10.1016/j.ijmedinf.2021.104679>
- [8] Goecks, J., Jalili, V., Heiser, L.M., Gray, J.W. (2020). How machine learning will transform biomedicine. *Cell*, 181(1): 92-101. <https://doi.org/10.1016/j.cell.2020.03.022>
- [9] Lee, S.E., Hong, H., Kim, E.K. (2024). Diagnostic performance with and without artificial intelligence assistance in real-world screening mammography. *European Journal of Radiology Open*, 12: 100545. <https://doi.org/10.1016/j.ejro.2023.100545>
- [10] Chakraoui, M., Mouhni, N., Elkalay, A., Nemiche, M. (2022). Deep negative effects of misleading information about COVID-19 on populations through twitter. *Ingénierie des Systèmes d'Information*, 27(2): 185. <https://doi.org/10.18280/isi.270202>
- [11] Khouridifi, Y., El Alami, A., Zaydi, M., Yassine, M., Er-Remyly, O. (2024). Early breast cancer detection based on deep learning: An ensemble approach applied to mammograms. *BioMedInformatics*, 4(4): 2338-2373. <https://doi.org/10.3390/biomedinformatics4040127>
- [12] Aniq, E., Chakraoui, M., Mouhni, N. (2024). Innovative: A Novel Deep Learning-Based Semantic Segmentation Architecture for Medical Applications. *Ingénierie des Systèmes d'Information*, 29(4): 1603-1609. <https://doi.org/10.18280/isi.290433>
- [13] Leong, Y.S., Hasikin, K., Lai, K.W., Mohd Zain, N., Azizan, M.M. (2022). Microcalcification discrimination in mammography using deep convolutional neural network: Towards rapid and early breast cancer diagnosis. *Frontiers in Public Health*, 10: 875305. <https://doi.org/10.3389/fpubh.2022.875305>
- [14] Sahu, A., Das, P.K., Meher, S. (2024). An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images. *Biomedical Signal Processing and Control*, 87: 105377. <https://doi.org/10.1016/j.bspc.2023.105377>
- [15] Yasaka, K., Akai, H., Kunimatsu, A., Kiryu, S., Abe, O. (2024). Impact of deep learning on radiologists and radiology residents in detecting breast cancer on CT: A cross-vendor test study. *Clinical Radiology*, 79(1): e41-e47. <https://doi.org/10.1016/j.crad.2023.09.022>
- [16] Han, J., Kim, E.K., Kim, S.Y., Moon, H.J., Yoon, J.H. (2024). Prediction of disease-free survival in breast cancer using deep learning with ultrasound and mammography: A multicenter study. *Clinical Breast Cancer*, 24(3): 215-226. <https://doi.org/10.1016/j.clbc.2024.01.005>
- [17] Jenefa, A., Lincy, A., Edward Naveen, V. (2024). Chapter 5 - A framework for breast cancer diagnostics based on MobileNetV2 and LSTM-based deep learning. In *Computational Intelligence and Modelling Techniques for Disease Detection in Mammogram Images*, pp. 91-110. <https://doi.org/10.1016/B978-0-443-13999-4.00013-4>
- [18] El Houby, E.M.F., Yassin, N.I.R. (2021). Malignant and nonmalignant classification of breast lesions in mammograms using convolutional neural networks. *Biomedical Signal Processing and Control*, 70: 102954. <https://doi.org/10.1016/j.bspc.2021.102954>
- [19] Suckling, J., Parker, J., Dance, D., Astley, S., et al. (2015). Mammographic Image Analysis Society (MIAS) database v1.21. Apollo-University of Cambridge Repository. <https://doi.org/10.17863/CAM.105113>
- [20] Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S. (2012). INbreast: Toward a full-field digital mammographic database. *Academic Radiology*, 19(2): 236-248. <https://doi.org/10.1016/j.acra.2011.09.014>
- [21] Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L. (2017). A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific Data*, 4: 170177. <https://doi.org/10.1038/sdata.2017.177>

- [22] Wong, S.C., Gatt, A., Stamatescu, V., McDonnell, M.D. (2016). Understanding data augmentation for classification: When to warp? arXiv:1609.08764. <https://doi.org/10.48550/arXiv.1609.08764>
- [23] Aljuaaid, H., Alturki, N., Alsubaie, N., Cavallaro, L., Liotta, A. (2022). Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning. *Computer Methods and Programs in Biomedicine*, 223: 106951. <https://doi.org/10.1016/j.cmpb.2022.106951>
- [24] Mohammed, A., Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University - Computer and Information Sciences*, 35(2): 757-774. <https://doi.org/10.1016/j.jksuci.2023.01.014>
- [25] Simonyan, K., Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556. <https://doi.org/10.48550/arXiv.1409.1556>
- [26] Salama, W.M., Aly, M.H. (2021). Deep learning in mammography images segmentation and classification: Automated CNN approach. *Alexandria Engineering Journal*, 60(5): 4701-4709. <https://doi.org/10.1016/j.aej.2021.03.048>
- [27] Salama, W.M., Elbagoury, A.M., Aly, M.H. (2021). Novel breast cancer classification framework based on deep learning. *IET Image Processing*, 15(8): 1724-1737. <https://doi.org/10.1049/iet-ipr.2020.0122>
- [28] He, K., Zhang, X., Ren, S., Sun, J. (2015). Deep residual learning for image recognition. arXiv:1512.03385. <https://doi.org/10.48550/arXiv.1512.03385>
- [29] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861. <https://doi.org/10.48550/arXiv.1704.04861>
- [30] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C. (2019). MobileNetV2: Inverted residuals and linear bottlenecks. arXiv:1801.04381. <https://doi.org/10.48550/arXiv.1801.04381>
- [31] Hossin, M., Sulaiman, M.N. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2): 1-11. <https://doi.org/10.5121/ijdkp.2015.5201>
- [32] Chougrad, H., Zouaki, H., Alheyane, O. (2018). Deep convolutional neural networks for breast cancer screening. *Computer Methods and Programs in Biomedicine*, 157: 19-30. <https://doi.org/10.1016/j.cmpb.2018.01.011>