**International Journal of Electrical and Computer Engineering (IJECE)**
Vol. 15, No. 3, June 2025, pp. 2809~2819
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i3.pp2809-2819 

# Improving breast cancer classification with a novel VGG19-based ensemble learning approach

**Chaymae Taib, Adnan El Ahmadi, Otman Abdoun, El Khatir Haimoudi**
Information Security, Intelligent Systems and Application, Faculty of science Tetouan, Abdelmalek Essaadi University, Tetouan, Morocco

**Article Info**

***Article history:***

Received Oct 14, 2024
Revised Jan 27, 2025
Accepted Mar 3, 2025

***Keywords:***

Bagging
Breast cancer
Classification
Convolutional neural network
Deep learning

**ABSTRACT**

Breast cancer is one of the most life-threatening diseases, particularly affecting women, highlighting the importance of early detection for improving survival rates. In this study, we propose a novel diagnostic framework that combines a modified VGG19 architecture with Bagging ensemble learning, using three base classifiers: decision tree (DT), logistic regression (LR), and support vector machine (SVM). We also compare this approach with twenty-four hybrid models, integrating various convolutional neural network (CNN) architectures (ResNet50, VGG19, ConvNextBase, DenseNet121, EfficientNetV2B0, EfficientNetB0, MobileNet, and NasNetMobile) with Bagging ensemble methods. Our results show that the proposed model outperforms all other architectures, especially when combined with SVM, achieving accuracy of 97% on the fine needle aspiration cytology (FNAC) dataset and 90% on the International Conference on Image Analysis and Recognition (ICIAR) dataset. This framework demonstrates strong potential for improving early breast cancer diagnosis.



***Corresponding Author:***

Chaymae Taib
Information Security, Intelligent Systems and Application, Faculty of Science Tetouan, Abdelmalek Essaadi University
Tetouan, Morocco
Email: taib.chaymae@etu.uae.ac.ma

## 1. INTRODUCTION

Breast cancer is one of the most common and life-threatening cancers affecting women worldwide. According to the World Health Organization (WHO) [1], it accounts for approximately 25% of all cancer cases among women, making it the most prevalent malignancy in this population. The incidence of breast cancer varies across regions, with higher rates observed in developed countries. This variation is likely influenced by factors such as lifestyle choices, reproductive behaviors, and more advanced screening programs. Accurate diagnosis is essential for determining the most effective treatment plan [2]. Advances in imaging techniques like mammography, ultrasound, and magnetic resonance imaging (MRI) have significantly improved the precision of detecting breast cancer lesions. Additionally, molecular and genetic testing of tumor samples provides critical insights into the characteristics of the cancer, enabling personalized treatments that can enhance patient outcomes [3].

Machine learning (ML) techniques have brought a revolution in medical diagnosis, offering powerful tools for analyzing complex datasets and identifying patterns that might be missed by traditional statistical methods [4]. ML encompasses various algorithms and models, including supervised [5]–[7], unsupervised [8], and reinforcement learning [9], [10], as well as deep learning techniques [11]. For breast cancer diagnosis, early and accurate identification is crucial for improving survival rates [12], [13]. While a variety of imaging techniques, such as MRI, mammography, ultrasound, and thermography, are used for detection, each method has its own advantages and limitations.

Recently, hybrid models combining machine learning and deep learning have gained attention as a way to improve breast cancer diagnosis. For example, Khan *et al*. developed a deep learning framework using transfer learning for breast cancer classification in cytology images, achieving 97.52% accuracy surpassing existing convolutional neural network (CNN) models like GoogLeNet, VGGNet, and ResNet [14]. Atban *et al*. proposed a fused model combining transfer learning and metaheuristic algorithms for histopathological image classification, achieving 97.73% accuracy and a 97.75% F1-score [15]. Similarly, Qasrawi *et al*. designed a hybrid deep learning ensemble model for mammogram analysis, reaching 99.7% accuracy for benign cases and 99.8% for malignant ones [16]. Other approaches, such as Raaj's hybrid CNN architecture and Aslan's end-to-end CNN and CNN-based bi-directional long short-term memory (CNN+BiLSTM) model, demonstrated impressive accuracy rates for breast cancer detection in the mammographic image analysis society (MIAS) dataset [17], [18]. Furthermore, Sreeprada *et al*. proposed a hybrid convolutional neural network-support vector machine (CNN-SVM) approach that outperformed traditional CNN techniques with a 98.70% accuracy in lung CT image classification, demonstrating the broader applicability of hybrid models in medical imaging [19].

In summary, combining CNNs with traditional machine learning techniques and ensemble methods like bagging has led to significant improvements in breast cancer diagnosis. These advancements result in more accurate, reliable, and robust predictive models, paving the way for better clinical outcomes.

The primary goal of this study is to evaluate the effectiveness of combining CNNs and bagging techniques to enhance the accuracy and reliability of breast cancer diagnosis. By leveraging CNNs' feature extraction capabilities and the ensemble stability of bagging, the study aims to develop a robust diagnostic model that outperforms traditional methods. It seeks to assess the performance of CNNs as feature extractors, evaluate the impact of bagging on reducing variance and improving accuracy in CNN-based models, and compare the performance of the combined CNN-bagging approach with standalone CNNs and traditional diagnostic techniques. Additionally, the study explores the potential of the combined method in minimizing false positives and negatives in breast cancer diagnosis and evaluates the modified VGG19 combined with bagging approach against other traditional CNN techniques.

The paper is organized into five sections to ensure clarity and coherence. Section 2 reviews related work, section 3 explains the methodology, and section 4 presents the results. Section 5 concludes with key insights and suggestions for future research.

## 2. LITERATURE REVIEW

Recent advancements in deep learning and artificial intelligence have significantly enhanced cancer diagnostics, particularly in breast cancer classification. Numerous studies have investigated the use of ensemble learning methods and deep CNN architectures to boost diagnostic accuracy and model reliability. These approaches have demonstrated strong potential in handling complex histopathological images and improving clinical decision-making.

Balasubramanian *et al.* utilized ensemble deep learning techniques combining VGG16 and ResNet50 architectures to classify breast cancer subtypes and invasiveness using the breast cancer histolog (BACH) dataset. They introduced an image patching technique for high-resolution preprocessing, achieving a patch classification accuracy of 95.31% and a whole-slide image classification accuracy of 98.43% [20]. While the study demonstrates the efficacy of ensemble methods, it relies heavily on existing architectures without significant innovation. Additionally, its focus on patch-level classification raises concerns about generalizability across datasets and real-world applications. This study underscores the need for more robust and adaptable models, which this research addresses through the deeper architecture and enhanced ensemble framework of VGG19. Nadkarni and Noronha proposed an ensemble of CNNs for classifying mammography images using the MIAS and igital database for screening mammography (DDSM) datasets [21]. Their approach merged these datasets to address issues of dataset imbalance and overfitting, achieving an accuracy of 95.7%. While their ensemble strategy improved robustness, it lacked architectural novelty, and their focus on mammograms limits its direct applicability to histopathological datasets like BACH and fine needle aspiration cytology (FNAC). Nevertheless, their efforts to mitigate overfitting and handle diverse datasets provide valuable insights for this study. Mamdy and Petli [22] developed a computationally efficient framework for MRI-based breast cancer detection using a modified fuzzy rough set technique and stacked autoencoders. This method achieved a remarkable 99% classification accuracy while minimizing computational complexity. Although focused on MRI imaging rather than histopathology, their dimensionality reduction techniques offer insights into optimizing feature extraction, which could complement this study's ensemble approach by enhancing the efficiency of features derived from VGG19. Mudavadkar *et al.* [23] applied ensemble learning to gastric cancer detection using the GasHisSDB dataset. By combining ResNet50, VGGNet, and ResNet34, they achieved over 99% accuracy across various resolutions. Although unrelated to breast cancer, this study highlights the potential of ensemble models to capture critical features across resolutions. However, it lacks specificity regarding breast cancer datasets, limiting its applicability to this research's focus on histopathological data. Kondejkar *et al.* [24]. demonstrated the effectiveness of ResNet models in prostate cancer grading using multi-scale patch-level pathology images, achieving near-perfect accuracy of 0.999 in identifying clinically significant cancer. While this study emphasizes the scalability and precision of CNNs, its focus on prostate cancer and grading diverges from the classification objectives of this research. Nonetheless, its success in patch-level analysis reinforces the potential of CNNs like VGG19 for histopathological tasks. Singh *et al.* [25]. introduced an AI-based web application integrating EfficientNet-B1 for prostate cancer diagnosis. While their focus was on usability and clinical workflow integration, the study highlights the importance of translating AI models into practical tools for pathologists.

In summary, previous studies highlight the promise of deep learning and ensemble techniques in cancer diagnostics but also expose key limitations. These include dependence on outdated architectures, poor generalizability across datasets, and limited innovation in ensemble strategies. To overcome these challenges, this study introduces a novel ensemble approach based on the adaptable and deeper VGG19 architecture, specifically designed for breast cancer classification.

## 3. METHOD

The breast cancer classification process, illustrated in Figure 1, follows several key steps, starting with data collection and progressing through to the classification stage. The data used consists of images of normal and cancerous breast tissue, which undergo various techniques to enhance model performance, including rotation, shear, zoom, horizontal flip, fill mode, rescaling, and width and height shifting. Feature extraction is performed using deep learning models such as ResNet50, VGG19, ConvNextBase, DenseNet121, EfficientNetV2B0, EfficientNetB0, MobileNet, and NasNetMobile.

As shown in Figure 2, the architecture utilizes the VGG19 model with additional custom layers that include a fully connected dense layer with 512 units, followed by dropout layers (with a dropout rate of 0.5) to prevent overfitting. Afterward, another dense layer with 256 units and another dropout layer are applied. The convolutional base of VGG19 is frozen to preserve pre-trained features, and the modified VGG19 extracts deep features from the dataset. These extracted features are then normalized before being fed into a bagging ensemble model. The base estimators for bagging include support vector classifiers, DT, and LR. To further optimize performance, hyperparameter tuning is performed using RandomizedSearchCV, ensuring the best combination of parameters for the classifiers.

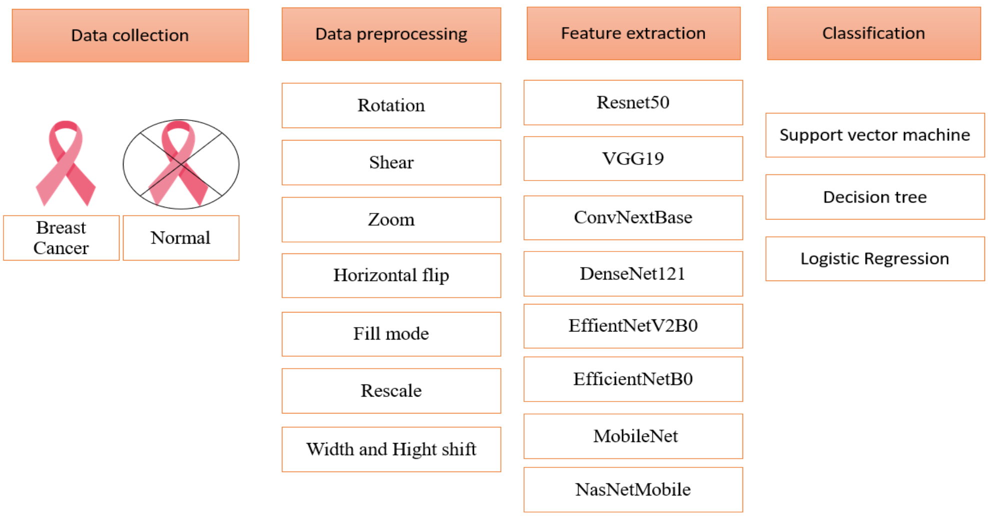



Figure 1. Overview of the combination between CNN-techniques and bagging.

In the classification stage, a bagging ensemble method is employed with base estimators like support vector machines (SVM), decision trees (DT), and logistic regression (LR) to improve model performance and robustness. This ensemble approach combines multiple predictions from different base classifiers to generate more reliable and accurate classification results. This pipeline provides a comprehensive framework for detecting and classifying breast cancer images.

This diagram represents a proposed architecture for breast cancer classification using a modified VGG19 model combined with machine learning techniques. The process begins with the dataset, which includes breast cancer images. The data is split into training and testing sets and reshaped for compatibility with the VGG19 model.

This architecture is designed to enhance both accuracy and robustness in classifying breast cancer images. It leverages the powerful feature extraction capabilities of the deep learning model VGG19, combined with the ensemble stability of a bagging approach using machine learning base estimators. This integration aims to capitalize on the strengths of both methods for improved diagnostic performance.

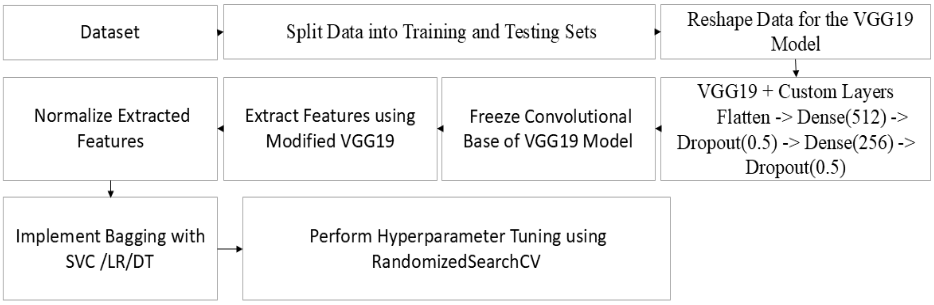



Figure 2. Overview of the modified VGG19

### 3.1. The dataset description and preprocessing

International Conference on Image Analysis and Recognition (ICIAR) [26]: Hematoxylin and eosin (H&E) stained breast histology microscopy and whole-slide images comprise the dataset. The dataset comprises 400 microscope pictures, which are distributed as follows: 100 is considered normal, Positive: 100, 100 cases of in situ carcinoma,100 cases of invasive carcinoma Microscopy pictures are in.tiff format and meet the following requirements: Red Green Blue is the color model, Dimensions: 2048×1536 pixels, 0.42×0.42 m pixel scale, 10-20 MB of memory space, Type of label: image-wise. In this study we used the two classes normal and benign that contained 100 images each.

FNAC dataset: Database of fine needle aspiration cytology (FNAC) [27]: We acquired images of the FNAC dataset using a Leica ICC50 HD microscope with 400 resolution and 24-bit color depth, as well as a 5-megapixel camera attached to the microscope. The digitized images were then evaluated by competent certified cytopathologists, who picked 212 images in total 113 Malignant and 99 Benign.

The ICIAR dataset was chosen for its high-quality H&E-stained breast histology images, offering balanced and well-labeled categories ideal for early-stage cancer detection. The normal and benign classes were specifically selected to align with the study's focus on early-stage abnormalities. The FNAC dataset complements this with cytological-level imaging, providing fine-grained diagnostic details validated by expert cytopathologists. Together, these datasets offer diverse and reliable data to support robust model development and evaluation.

### 3.2. Data pre-processing

The initial and most important step in building a reliable predictive model is preprocessing the input images using various preprocessing techniques. To augment and balance the data, geometric transformations were applied. Initially, the images underwent Histogram Equalization to enhance contrast and normalize intensity distribution. This was followed by Data Augmentation to expand the dataset and improve model generalization. The augmentation process included transformations such as rotation up to 30 degrees, zooming with a range of 0.2, shearing with a range of 0.2, width shifting with a range of 0.3, height shifting with a range of 0.3, horizontal flipping, and rescaling by dividing pixel values by 255. Additionally, a fill mode set to “nearest” was used to handle gaps during transformations, ensuring robustness in the preprocessing pipeline.

The dataset was then split into 80% for training and 20% for testing to ensure a comprehensive evaluation of the model's performance. This split provided a sufficient amount of data for training while reserving an appropriate portion for assessing the model's ability to generalize to unseen data. A consistent random seed was used to ensure the reproducibility of the split. Furthermore, class labels were preprocessed into a one-hot encoded format to facilitate multi-class classification, ensuring compatibility with neural network models and enhancing the efficiency of the learning process. This standardized preprocessing approach ensured consistency across the training and testing phases for reliable evaluation.

### 3.3. Performance measures

The performance of the hybrid techniques was assessed using the following metrics [28]: accuracy, precision, recall and F1-score, defined in the (1)-(4).

$$Accuracy = \frac{TP+TN}{TN+TP+FP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1 = 2 \times \frac{Re\,call \times precision}{Re\,call + precision} \tag{4}$$

where: TP (True Positive): The number of cases correctly predicted as positive.
TN (True Negative): The number of cases correctly predicted as negative.
FP (False Positive): The number of cases incorrectly predicted as positive.
FN (False Negative): The number of cases incorrectly predicted as negative

### 3.4. Performance setup

All experiments were conducted on Google Colab Pro with 50 GB RAM and 107.7 GB disk storage. The codes were implemented in TensorFlow using Python 3 as the programming language. The computational backend was powered by Google Compute Engine with CPU support. The hyperparameters used in the BaggingClassifier with LR were carefully selected to optimize the model's performance and ensure reproducibility. The ensemble was configured with a varying number of estimators, specifically 50, 100, and 200, to assess the impact of the ensemble size on classification performance. For LR, the regularization strength (C) was tuned over a range of values, including 0.01, 0.1, 1, 10, 100, and 1000, allowing for control over the trade-off between bias and variance. A random state of 42 was used throughout the process to guarantee consistent results. These choices reflect a structured and methodical approach to identifying the optimal model configuration for robust and reliable classification.

## 4. RESULTS AND DISCUSSION

The evaluation metrics accuracy, F1-score, recall, and precision for eight CNN architectures VGG19, ResNet50, NasNetMobile, MobileNet, EfficientNetB0, EfficientNetV2B0, DenseNet121, and ConvNeXtTiny were combined with three classifiers DT, LR, and SVM and assessed over two datasets, FNAC and ICIAR. The results, presented in Figures 3 to 9 and Tables 1 and 2, provide a comprehensive analysis of model performance across various configurations. The confusion matrices, particularly for the FNAC dataset as shown in Figures 3, 4, 5, and 9, demonstrate that models integrated with SVM consistently achieve the highest accuracy. These models produce significantly fewer false negatives and false positives, highlighting their robustness in correctly classifying cancerous and non-cancerous samples. For the ICIAR dataset, Figure 6 reveals a similar trend, where CNNs combined with SVM outperform DT and LR, demonstrating fewer misclassifications and superior classification capability. This indicates the potential of SVM to enhance the diagnostic accuracy of CNN-based models across varying datasets.

On the FNAC dataset, Table 1 highlights that the proposed model achieved the highest accuracy of 95%, outperforming NasNetMobile and MobileNet, which scored around 80%, and DenseNet121, which also showed competitive performance. For the ICIAR dataset shows in Table 2, accuracy declined slightly for all models, with VGG19 and MobileNet achieving between 70% and 77%, while the proposed model still outperformed others with an accuracy of 82%. These results suggest that DT, though effective, are outperformed by other classifiers in terms of robustness and precision. LR exhibited strong performance on the FNAC dataset, with VGG19 and ResNet50 achieving accuracies of 92% and 80%, respectively, as shown in Table 1. However, the proposed model surpassed all others with a remarkable accuracy of 95%, indicating enhanced robustness when using the integrated method. On the ICIAR dataset, similar trends were observed, with VGG19 and MobileNet emerging as the top-performing models after the proposed model, which maintained superior accuracy and reliability.

SVM consistently delivered the best performance across both datasets. On FNAC, models such as VGG19, MobileNet, and the proposed architecture achieved accuracies ranging from 90% to 97%, as indicated in Table 1. For ICIAR, the performance was slightly lower, with accuracies ranging from 83% to 90%, as shown in Table 2. The proposed model demonstrated exceptional effectiveness, achieving the highest accuracy of 90% on the ICIAR dataset. This emphasizes the reliability of the combined CNN-SVM approach in minimizing false positives and negatives, critical for applications in breast cancer diagnosis. The results underscore the advantages of integrating CNN architectures with SVM classifiers, particularly in medical image classification tasks where precision and reliability are paramount. The proposed model consistently outperformed traditional standalone approaches and alternative classifier combinations across both datasets. These findings highlight the robustness of the combined CNN-SVM approach, particularly for models like VGG19 and MobileNet, which exhibit superior feature extraction capabilities. The reduction in false negatives and positives is especially significant for diagnostic applications, where errors can have critical implications. Overall, the integration of CNNs with ensemble techniques such as SVM offers a promising path toward developing accurate and reliable diagnostic systems for breast cancer.

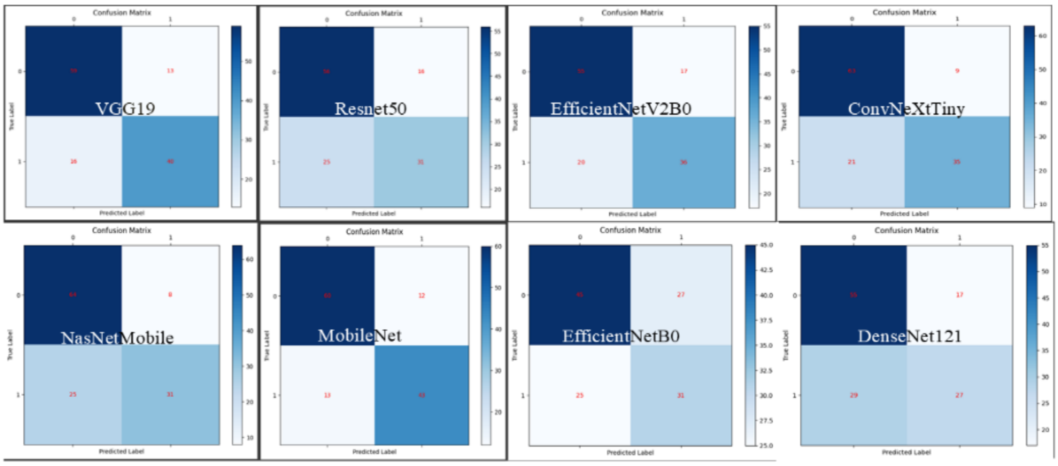



Figure 3. Confusion matrix of the eight CNN techniques combined with decision tree Over FNAC

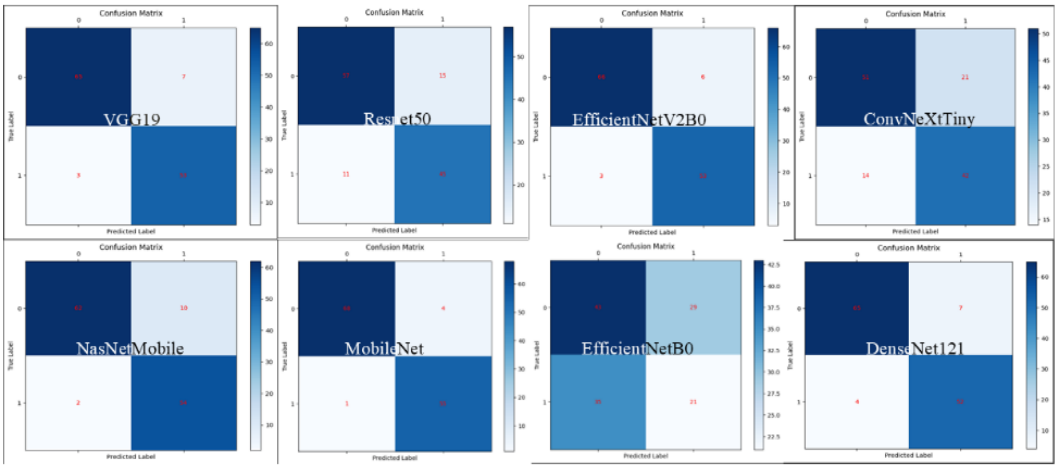



Figure 4.Confusion matrix of the eight CNN techniques combined with logistic regression Over FNAC

The comparison of CNN architectures reveals varying levels of performance across different configurations. VGG19 demonstrates consistently high results, achieving a strong balance of accuracy, F1-score, recall, and precision, particularly when combined with SVM and bagging, as highlighted in Figure 8. MobileNet and NasNetMobile also perform well across both datasets and classifiers, showcasing that lightweight architectures can be highly effective for image classification tasks. In contrast, EfficientNet (B0 and V2B0) underperforms compared to VGG19 and MobileNet, indicating that increased model complexity does not always lead to better performance in this specific context. Similarly, DenseNet121 and ConvNeXtTiny exhibit lower performance, especially when paired with DT and LR.

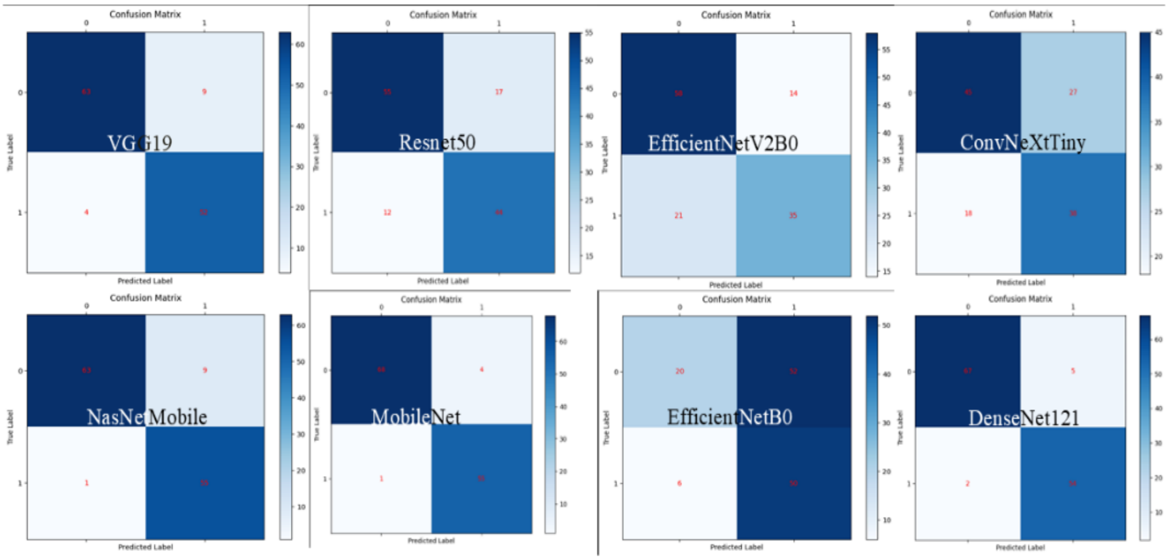



Figure 5. Confusion matrix of eight CNN techniques combined with SVM over FNAC

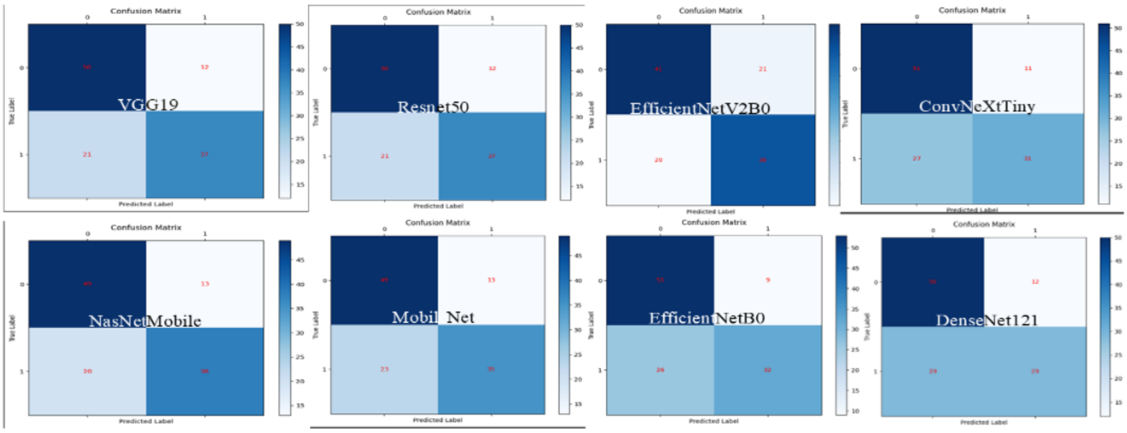



Figure 6. Confusion matrix of eight CNN techniques combined with DT over ICIAR

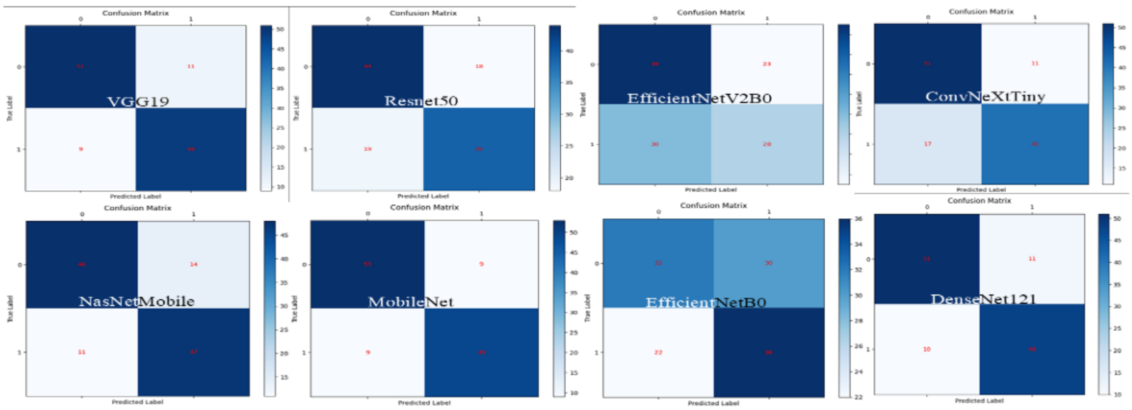



Figure 7.Confusion matrix of the eight CNN techniques combined with LR over ICIAR

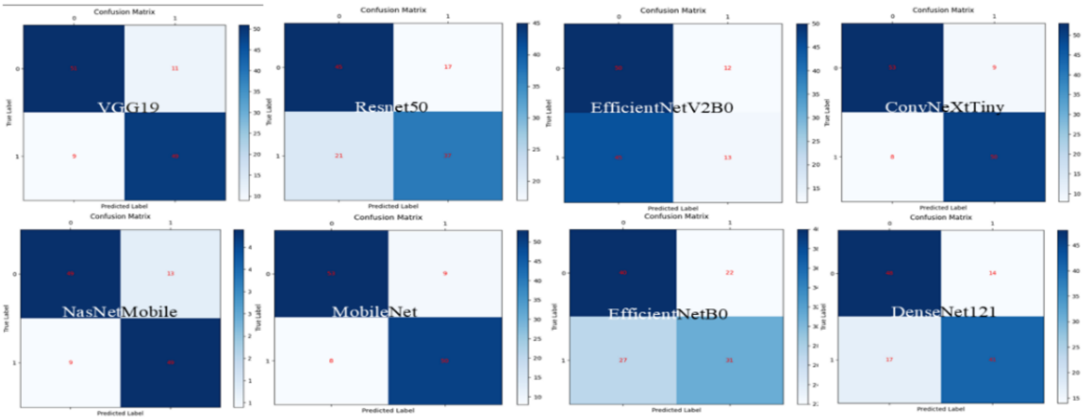



Figure 8. Confusion matrix of the eight CNN techniques combined with SVM over ICIAR

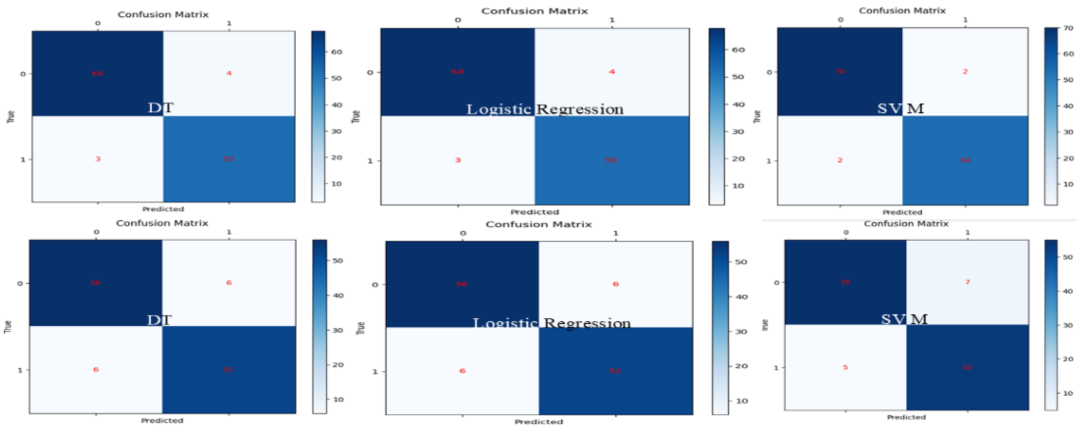



Figure 9. Confusion matrix of the modified VGG19 combined with the bagging over two datasets

Table 1. The results of the four merics of evaluation over FNAC

| Estimator | CNN | ACC | F1-SCORE | Recall | Precision |
|---|---|---|---|---|---|
| DT | VGG19 | 77% | 77% | 77% | 77% |
| | Resnet50 | 68% | 67% | 67% | 67% |
| | NasNetMobile | 74% | 72% | 72% | 76% |
| | MobileNet | 80% | 80% | 80% | 80% |
| | EfficientNetB0 | 59% | 59% | 59% | 59% |
| | EfficientNetV2B0 | 71% | 70% | 70% | 71% |
| | DenseNet121 | 77% | 75% | 75% | 77% |
| | ConvNeXtTiny | 64% | 62% | 62% | 63% |
| | Proposed | 95% | 95% | 95% | 95% |
| LR | VGG19 | 92% | 92% | 92% | 92% |
| | Resnet50 | 80% | 80% | 80% | 79% |
| | NasNetMobile | 91% | 91% | 91% | 91% |
| | MobileNet | 96% | 96% | 96% | 96% |
| | EfficientNetB0 | 50% | 48% | 49% | 49% |
| | EfficientNetV2B0 | 93% | 93% | 93% | 93% |
| | DenseNet121 | 91% | 91% | 92% | 91% |
| | ConvNeXtTiny | 73% | 73% | 73% | 73% |
| | Proposed | 95% | 95% | 95% | 95% |
| SVM | VGG19 | 90% | 90% | 90% | 90% |
| | Resnet50 | 77% | 77% | 77% | 77% |
| | NasNetMobile | 92% | 92% | 93% | 92% |
| | MobileNet | 96% | 96% | 96% | 96% |
| | EfficientNetB0 | 55% | 55% | 55% | 55% |
| | EfficientNetV2B0 | 73% | 73% | 73% | 73% |
| | DenseNet121 | 95% | 95% | 95% | 95% |
| | ConvNeXtTiny | 65% | 65% | 65% | 65% |
| | Proposed | 97% | 97% | 97% | 97% |

Regarding classifier effectiveness, SVM consistently emerges as the best-performing classifier across both datasets, particularly when combined with VGG19 and MobileNet. Its ability to effectively handle high-dimensional feature spaces and distinguish between classes likely accounts for its superior performance. LR also performs reasonably well, though it falls short of SVM, particularly when applied to more complex architectures like EfficientNet. However, it achieves competitive results when combined with VGG19, highlighting its potential as a simpler yet effective alternative in some configurations. These findings underscore the importance of selecting appropriate classifier-architecture combinations to optimize performance for specific datasets and tasks.

Table 2. The results of the four metrics of evaluation over ICIAR

| Estimator | CNN | ACC | F1-SCORE | Recall | Precision |
|---|---|---|---|---|---|
| DT | VGG19 | 72% | 72% | 72% | 72% |
| | Resnet50 | 72% | 72% | 73% | 72% |
| | NasNetMobile | 72% | 72% | 72% | 72% |
| | MobileNet | 70% | 70% | 70% | 70% |
| | EfficientNetB0 | 71% | 70% | 70% | 73% |
| | EfficientNetV2B0 | 66% | 66% | 66% | 66% |
| | DenseNet121 | 66% | 65% | 65% | 67% |
| | ConvNeXtTiny | 68% | 67% | 68% | 70% |
| | Proposed | 86% | 86% | 86% | 86% |
| LR | VGG19 | 83% | 83% | 83% | 83% |
| | Resnet50 | 69% | 69% | 69% | 69% |
| | NasNetMobile | 79% | 79% | 79% | 79% |
| | MobileNet | 85% | 85% | 85% | 85% |
| | EfficientNetB0 | 56% | 56% | 56% | 56% |
| | EfficientNetV2B0 | 57% | 57% | 57% | 57% |
| | DenseNet121 | 82% | 82% | 83% | 82% |
| | ConvNeXtTiny | 77% | 77% | 77% | 77% |
| | Proposed | 90% | 90% | 90% | 90% |
| SVM | VGG19 | 83% | 83% | 83% | 83% |
| | Resnet50 | 68% | 68% | 68% | 68% |
| | NasNetMobile | 82% | 82% | 82% | 82% |
| | MobileNet | 86% | 86% | 86% | 86% |
| | EfficientNetB0 | 53% | 53% | 53% | 53% |
| | EfficientNetV2B0 | 59% | 59% | 59% | 59% |
| | DenseNet121 | 86% | 86% | 86% | 86% |
| | ConvNeXtTiny | 74% | 74% | 74% | 74% |
| | Proposed | 90% | 90% | 90% | 90% |

## 5. CONCLUSION

In this study, we applied a modified VGG19 model combined with bagging, using SVM as the base classifier, to classify images from two datasets. The results demonstrated strong performance, with high accuracy and consistent results across test sets. The confusion matrices revealed that the model was effective in correctly classifying both true positives and true negatives, with minimal misclassifications. The hyperparameter tuning of the SVM using RandomizedSearchCV further optimized the model, enhancing its generalization ability. These findings suggest that the combination of deep feature extraction using VGG19 and the ensemble learning approach provided by bagging has the potential to handle complex image classification tasks effectively. Future work could explore additional enhancements to improve model performance further. One direction could involve experimenting with different ensemble methods, such as boosting, to compare its efficacy with bagging. Another avenue is to investigate hybrid deep learning approaches by incorporating other convolutional neural networks CNNs alongside VGG19, or by integrating feature fusion techniques to combine information from multiple layers. Moreover, the use of more advanced optimization techniques, such as Bayesian optimization, could improve the hyperparameter search process.

## FUNDING INFORMATION

Authors state no funding involved.

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

| Name of Author | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chaymae TAIB | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ |
| Adnan El Ahmadi | | ✓ | | | | ✓ | | ✓ | | | ✓ | | | |
| Otman Abdoun | ✓ | ✓ | | ✓ | | | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| El khatir Haimoudi | ✓ | ✓ | | ✓ | | | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ |

| | | |
|---|---|---|
| C : **C**onceptualization | I : **I**nvestigation | Vi : **Vi**sualization |
| M : **M**ethodology | R : **R**esources | Su : **Su**pervision |
| So : **So**ftware | D : **D**ata Curation | P : **P**roject administration |
| Va : **Va**lidation | O : Writing - **O**riginal Draft | Fu : **Fu**nding acquisition |
| Fo : **Fo**rmal analysis | E : Writing - Review & **E**diting | |

## CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

## DATA AVAILABILITY

Two public datasets were used in this study.

- The data that support the findings of this study are openly available in in *Medical Image Analysis* at https://doi.org/10.1016/j.media.2019.05.010, reference number [26] .
- The data that support the findings of this study are openly available in *Pathology - Research and Practice* at https://doi.org/10.1016/j.tice.2019.02.001, reference number [27].

## REFERENCES


[1] "Breast cancer," 2024. https://www.who.int/news-room/fact-sheets/detail/breast-cancer (accessed Sep. 23, 2024).

[2] B. S. Alotaibi *et al.*, "The accuracy of breast cancer diagnostic tools," *Cureus*, Jan. 2024, doi: 10.7759/cureus.51776.

[3] M. Kataoka, M. Iima, K. K. Miyake, and Y. Matsumoto, "Multiparametric imaging of breast cancer: An update of current applications," *Diagnostic and Interventional Imaging*, vol. 103, no. 12, pp. 574–583, Dec. 2022, doi: 10.1016/j.diii.2022.10.012.

[4] C. Chakraborty, M. Bhattacharya, S. Pal, and S.-S. Lee, "From machine learning to deep learning: Advances of the recent data-driven paradigm shift in medicine and healthcare," *Current Research in Biotechnology*, vol. 7, 2024, doi: 10.1016/j.crbiot.2023.100164.

[5] H. Kang *et al.*, "Assessing the performance of fully supervised and weakly supervised learning in breast cancer histopathology," *Expert Systems with Applications*, vol. 237, Mar. 2024, doi: 10.1016/j.eswa.2023.121575.

[6] C. Taib, O. Abdoun, and E. Haimoudi, "Performance evaluation of diagnostic and classification systems using deep learning on Apache spark," in *Lecture Notes in Mechanical Engineering*, Springer International Publishing, 2023, pp. 145–154. doi: 10.1007/978-3-031-23615-0_15.

[7] C. Taib, O. Abdoun, and E. K. Haimoudi, "Optimizing breast cancer diagnosis: combining hybrid architectures through Apache Spark," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 4, pp. 4261–4272, Aug. 2024, doi: 10.11591/ijece.v14i4.pp4261-4272.

[8] C. Zhang, P. Chen, and T. Lei, "Category-weight instance fusion learning for unsupervised domain adaptation on breast cancer histopathology images," *Biomedical Signal Processing and Control*, vol. 99, Jan. 2025, doi: 10.1016/j.bspc.2024.106794.

[9] S. Muksimova, S. Umirzakova, S. Kang, and Y. I. Cho, "CerviLearnNet: Advancing cervical cancer diagnosis with reinforcement learning-enhanced convolutional networks," *Heliyon*, vol. 10, no. 9, May 2024, doi: 10.1016/j.heliyon.2024.e29913.

[10] A. Aghanim, H. Chekenbah, O. Oulhaj, and R. Lasri, "Q-Learning empowered cavity filter tuning with epsilon decay strategy," *Progress In Electromagnetics Research C*, vol. 140, pp. 31–40, 2024, doi: 10.2528/PIERC23111903.

[11] H. Slimi, S. Abid, and M. Sayadi, "Advanced deep learning strategies for breast cancer image analysis," *Journal of Radiation Research and Applied Sciences*, vol. 17, no. 4, Dec. 2024, doi: 10.1016/j.jrras.2024.101136.

[12] M. Nasser and U. K. Yusof, "Deep learning based methods for breast cancer diagnosis: A systematic review and future direction," *Diagnostics*, vol. 13, no. 1, Jan. 2023, doi: 10.3390/diagnostics13010161.

[13] M. Amrane, S. Oukid, I. Gagaoua, and T. Ensari, "Breast cancer classification using machine learning," in *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, Apr. 2018, pp. 1–4. doi: 10.1109/EBBT.2018.8391453.

[14] S. Khan, N. Islam, Z. Jan, I. Ud Din, and J. J. P. C. Rodrigues, "A novel deep learning based framework for the detection and classification of breast cancer using transfer learning," *Pattern Recognition Letters*, vol. 125, pp. 1–6, Jul. 2019, doi: 10.1016/j.patrec.2019.03.022.

[15] F. Atban, E. Ekinci, and Z. Garip, "Traditional machine learning algorithms for breast cancer image classification with optimized deep features," *Biomedical Signal Processing and Control*, vol. 81, Mar. 2023, doi: 10.1016/j.bspc.2022.104534.

[16] R. Qasrawi *et al.*, "Hybrid ensemble deep learning model for advancing breast cancer detection and classification in clinical applications," *Heliyon*, vol. 10, no. 19, Oct. 2024, doi: 10.1016/j.heliyon.2024.e38374.

[17] R. S. Raaj, "Breast cancer detection and diagnosis using hybrid deep learning architecture," *Biomedical Signal Processing and Control*, vol. 82, Apr. 2023, doi: 10.1016/j.bspc.2022.104558.

[18] M. F. Aslan, "A hybrid end-to-end learning approach for breast cancer diagnosis: convolutional recurrent network," *Computers and Electrical Engineering*, vol. 105, Jan. 2023, doi: 10.1016/j.compeleceng.2022.108562.

[19] V. Sreeprada and D. K. Vedavathi, "Lung cancer detection from X-ray images using hybrid deep learning technique," *Procedia Computer Science*, vol. 230, pp. 467–474, 2023, doi: 10.1016/j.procs.2023.12.102.

[20] A. A. Balasubramanian *et al.*, "Ensemble deep learning-based image classification for breast cancer subtype and invasiveness diagnosis from whole slide image histopathology," *Cancers*, vol. 16, no. 12, Jun. 2024, doi: 10.3390/cancers16122222.

[21] S. Nadkarni and K. Noronha, “Breast cancer detection using ensemble of convolutional neural networks,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 1, pp. 1041–1047, Feb. 2024, doi: 10.11591/ijece.v14i1.pp1041-1047.

[22] S. Kumar Mamdy and V. Petli, “Modified fuzzy rough set technique with stacked autoencoder model for magnetic resonance imaging based breast cancer detection,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 1, pp. 294–304, Feb. 2024, doi: 10.11591/ijece.v14i1.pp294-304.

[23] G. R. Mudavadkar *et al.*, “Gastric cancer detection with ensemble learning on digital pathology: use case of gastric cancer on GasHisSDB Dataset,” *Diagnostics*, vol. 14, no. 16, Aug. 2024, doi: 10.3390/diagnostics14161746.

[24] T. Kondejkar *et al.*, “Multi-scale digital pathology patch-level prostate cancer grading using deep learning: Use case evaluation of DiagSet Dataset,” *Bioengineering*, vol. 11, no. 6, Jun. 2024, doi: 10.3390/bioengineering11060624.

[25] A. Singh *et al.*, “Visualizing decisions and analytics of artificial intelligence based cancer diagnosis and grading of specimen digitized biopsy: Case study for prostate cancer.” Dec. 2022. doi: 10.1101/2022.12.21.22283754.

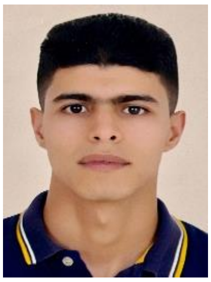
[26] G. Aresta *et al.*, “BACH: grand challenge on breast cancer histology images,” *Medical Image Analysis*, vol. 56, pp. 122–139, Aug. 2019, doi: 10.1016/j.media.2019.05.010.

[27] A. R. Saikia, K. Bora, L. B. Mahanta, and A. K. Das, “Comparative assessment of CNN architectures for classification of breast FNAC images,” *Tissue and Cell*, vol. 57, pp. 8–14, Apr. 2019, doi: 10.1016/j.tice.2019.02.001.

[28] D. Berrar, “Performance measures for binary classification,” in *Encyclopedia of Bioinformatics and Computational Biology*, Elsevier, 2019, pp. 546–560. doi: 10.1016/B978-0-12-809633-8.20351-8.


## BIOGRAPHIES OF AUTHORS

**Chaymae Taib** a Ph.D. student at the computer science Department of Abdelmalek Essadi University, has a Master's degree in big data and IOT from ENSAM Casablanca, which she earned in 2020. Her ongoing research focuses on neural networks, machine learning, data preprocessing, big data, and intelligent systems. With a collection of published works, she can be reached via email at taib.chaymae@etu.uae.ac.ma.



**Adnan El Ahmadi** completed his license degree in Mathematics and Computer Studies at the Faculty Polydisciplinary of Larache in 2015. He then obtained a master's degree in security of information systems from Natioanal School of Applied Sciences of Kenitra in 2017. Currently, he is preparing his Ph.D. degree and he is focusing on research in artificial intelligence and the internet of things, he can be reached via email at adnanelahmadi@gmail.com.

**Otman Abdoun** is currently a professor in the Department of Computer Science at the Faculty of Science, Abdelmalek Essaadi University, Tetouan, Morocco. He earned his Ph.D. degree in computer science from Ibn Tofail Kenitra University in Morocco. His primary research interests lie in multi-agent systems, optimization algorithms, and e-learning. With numerous publications to his credit, Otman Abdoun can be contacted via email at abdoun.otman@gmail.com.

**Elkhatir Haimoudi** having successfully earned his Ph.D. degree in computer science from Ukraine University, El Khatir Haimoudi now serves as a professor in the Computer Science Department of the Polydisciplinary Faculty at Larache, Abdelmalek Essadi University. His passion for research encompasses areas such as neural networks, deep learning, pattern recognition, and classification problems. With an array of publications to his credit, El Khatir Haimoudi can be contacted through email at helkhatir@gmail.com.