A Novel CNN-Based Feature Fusion Framework for Breast Cancer Ultrasound Image Classification

Mobarak Zourhri[1], Bouchaib Cherradi[2], Mohamed El Khaili[3]
EEIS Laboratory-ENSET of Mohammedia, Hassan II University of Casablanca, Mohammedia 28830, Morocco[1, 2, 3]
STIE Team-CRMEF Casablanca-Settat, Provincial Section of El Jadida, El Jadida 24000, Morocco[2]

***Abstract*—Breast cancer remains a major global health concern and is among the leading causes of cancer-related deaths in women. Timely and precise diagnosis significantly improves treatment outcomes and patient survival rates. This paper presents a novel deep learning-based framework for breast cancer classification using ultrasound imagery, built upon the concatenation of two pre-trained Convolutional Neural Network (CNN) models: VGG19 and EfficientNetB0. By leveraging transfer learning and combining heterogeneous feature representations, the proposed method enhances the discriminative power of the extracted features. The model is evaluated on a publicly available benchmark ultrasound dataset and assessed through standard performance indicators, including accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). In addition, Gradient-weighted Class Activation Mapping (Grad-CAM) is employed to generate interpretability heatmaps, visually highlighting regions that contribute most to classification outcomes. The experimental findings reveal that the integrated architecture outperforms several existing approaches as well as individual CNN baselines. This study contributes to the growing field of AI-assisted medical diagnostics and demonstrates the effectiveness of model fusion in ultrasound-based breast cancer detection.**

***Keywords*—Breast cancer classification; ultrasound imaging; Convolutional Neural Networks (CNN); transfer learning; model fusion; Grad-CAM; deep learning**

I. Introduction

Breast cancer continues to be one of the most frequently diagnosed and life-threatening diseases among women globally. As reported by the World Health Organization (WHO), over 2.3 million women were diagnosed with breast cancer in 2020, leading to approximately 685,000 deaths globally[1]. Early diagnosis remains a cornerstone in the fight against breast cancer, contribute significantly to lower mortality rates and enhancing the effectiveness of therapeutic interventions [1].

Ultrasound remains a commonly employed adjunct technique for breast cancer screening, particularly beneficial for women with dense breast tissue where mammography may fall short in sensitivity. Its widespread use is attributed to its non-invasive nature, cost-effectiveness, and ability to provide real-time imaging [2]. Nevertheless, the diagnostic reliability of ultrasound can be compromised by its operator-dependent interpretation, often resulting in variations between observers and inconsistent assessments [3].

In recent years, deep learning, especially Convolutional Neural Networks (CNNs), has gained considerable traction for automating the interpretation of medical images [4]. These models have demonstrated strong performance in classification tasks by learning discriminative features directly from data [5], thereby eliminating the need for manual feature engineering [6]. Nevertheless, constructing deep learning models specifically for medical applications poses considerable challenges, largely because of the limited availability of large-scale, high-quality annotated datasets in this domain.

To address the challenge of limited medical data, transfer learning has become an effective and widely utilized approach. By leveraging models pre-trained on large-scale datasets such as ImageNet, researchers can fine-tune neural networks for domain-specific medical imaging tasks [7]. In addition, combining different CNN architectures through ensemble methods or feature-level concatenation has been investigated as a means to enhance classification accuracy and generalization by capturing complementary feature representations [8]. In this context, several research areas have benefited from this development, such as: Cardiovascular disease classification [9], [10], [11], diabetes disease prediction [12], [13], [14], Parkinson's disease detection [15], [16], [17], handwritten recognition [18], [19], [20], sentiment analysis [21], [22], etc.

In this study, we introduce a dual-CNN fusion framework designed for the classification of breast cancer from ultrasound images. Our proposed approach combines the feature extraction capabilities of VGG19 and EfficientNetB0 through a feature-level concatenation strategy, while keeping the pre-trained convolutional blocks frozen to reduce overfitting and accelerate convergence. The integration of Gradient-weighted Class Activation Mapping (Grad-CAM) further enhances model interpretability by providing visual explanations of the regions that most influenced the classification outcome.

The effectiveness of the proposed model was evaluated using a publicly available breast ultrasound image dataset. The model achieved highly competitive performance with an accuracy of 98.44%, precision of 98.55%, recall of 99.66%, F1-score of 99.11%, and an AUC of ≥ 0.99. These results demonstrate the potential of our fusion-based approach to support early and accurate breast cancer diagnosis, offering a valuable contribution to computer-aided diagnostic systems in clinical settings.

[1]https://www.who.int/news-room/fact-sheets/detail/breast-cancer

The rest of this article is structured as follows: Section II reviews the most relevant contributions and previous research efforts related to deep learning applications for breast cancer diagnosis. In Section III, the proposed deep convolutional framework based on the integration of VGG19 and EfficientNetB0 is described in detail, enhancement the architectural choices and fusion strategy. Section IV outlines the dataset characteristics, preprocessing techniques, and evaluation metrics adopted in this study. Section V reports the experimental findings and performance metrics achieved by our model, including the aforementioned high scores confirming the strength of the proposed method. Finally, Section VI concludes the study and outlines potential directions for future work.

II. Related Works

Recent advances in artificial intelligence, particularly in deep learning, have substantially improved the classification and diagnosis of breast cancer using medical imaging. In past studies, traditional feature extraction and dimensionality reduction techniques were most commonly applied to image classification problems, such as the use of homogeneity features in combination with mutual information for hyperspectral image analysis [23]. Among these techniques, Convolutional Neural Networks (CNNs) have demonstrated remarkable success in extracting relevant and discriminative features from complex image data, especially in ultrasound imaging, which remains a widely used modality for breast cancer screening due to its non-invasive nature and cost-effectiveness.

Several studies have explored the integration of CNN architectures and transfer learning techniques for enhancing breast lesion classification. For instance, the authors in [24] developed a computer-aided diagnosis (CAD) system based on an ensemble of VGG19 and ResNet152 to distinguish between benign and malignant lesions in breast ultrasound images. Their model achieved a sensitivity of 90.9% and an Area Under the Curve (AUC) of 0.951, showcasing the effectiveness of combining deep architectures.

Similarly, in [25], the authors employed an ensemble learning strategy involving VGG, ResNet, and DenseNet networks to classify ultrasound images of breast tumors. Their evaluation on both public and private datasets resulted in an accuracy of 94.62%, a recall of 92.31%, and an F1-score of 91.14% on the BUSI dataset, indicating the benefits of leveraging multiple pre-trained networks to improve generalization and robustness.

The authors in [26] proposed a CNN-based approach utilizing GoogLeNet for binary classification of breast lesions. Their model achieved an accuracy of approximately 92.5% and a recall of 95.8%, highlighting the capability of deep learning to match or even surpass the diagnostic performance of experienced radiologists in certain clinical scenarios.

In another comparative study [27], the authors analyzed several CNN architectures, including VGG16 and InceptionV3, for breast tumor classification using ultrasound images. Using a dataset of 947 training images and 269 testing images, their best-performing model (fine-tuned VGG16) reached an accuracy of 91.9% and an AUC of 0.934, emphasizing the critical role of fine-tuning and architecture selection in CNN-based medical applications.

In our prior work [28], we conducted a comparative analysis of individual CNN models VGG16, VGG19, MobileNetV2, and ResNet50V2 using the same breast ultrasound dataset employed in this current study. Among them, the VGG19 model achieved the highest performance with an accuracy of 98.44%. Despite these encouraging results, single-network architectures are limited in capturing diverse feature representations and often lack interpretability, which is essential for clinical adoption.

To address these limitations, the present study proposes a dual CNN concatenation framework combining VGG19 and EfficientNetB0, selected for their complementary architectural properties and feature extraction capabilities. This model enhances performance with feature-level fusion without compromising generalizability by freezing pre-trained convolutional heads. Moreover, incorporating Grad-CAM visualization advances transparency by uncovering the most discriminative region in each input image, rendering the system more interpretable and reliable for clinical use.

III. Proposed CNN-Based Model

This section describes the architecture of the proposed deep learning model, which would classify breast ultrasound images as benign or malignant. The main goal is to improve diagnostic accuracy using transfer learning and feature combination via model concatenation.

A. System Overview

Fig. 1 illustrates the whole pipeline of the breast cancer diagnosis system. The system includes several steps starting from image acquisition and preprocessing of the ultrasound images, then feature extraction through pre-trained convolutional neural networks (CNNs). The features extracted from both models are combined and fed to a series of fully connected layers to generate the final binary prediction. This framework aims to be strong, understandable, and amenable to embedding within clinical workflow.

B. Model Architecture

The major contribution of this paper is the two-model concatenation method that combines two powerful pre-trained CNNs, VGG19 and EfficientNetB0, at the feature level. Both models are used as feature extractors without their classification heads, and only the last layers are fine-tuned to the ultrasound domain.

Each backbone processes the input image in parallel, generating high-level feature representations. The feature maps are concatenated through a concatenation layer. The fused feature vector is fed through a sequence of fully connected layers with dropout regularization to avoid over-fitting. Finally, a sigmoid activation function is used to calculate the probability of malignancy. Fig. 2 illustrates the architecture of the proposed convolutional neural network.

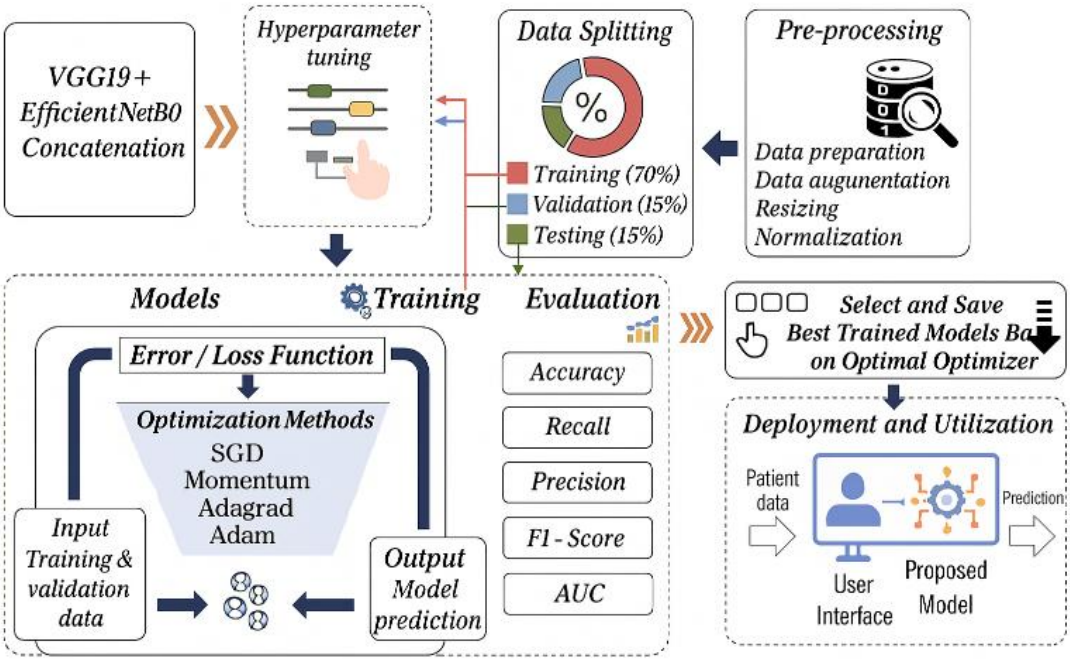


Fig. 1. Flowchart of the system for detecting breast cancer.

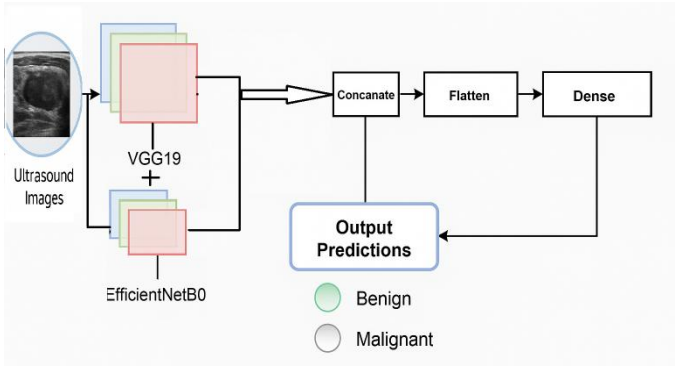


Fig. 2. The proposed CNN model architecture.

C. Motivation for Concatenation

The rationale behind using model concatenation is to harness the complementary strengths of both CNNs. VGG19 is known for its depth and ability to capture fine-grained spatial features, while EfficientNetB0 is optimized for efficiency and generalization. By combining their representations, the model can better distinguish subtle patterns present in benign and malignant lesions, which is particularly beneficial given the complexity of ultrasound data.

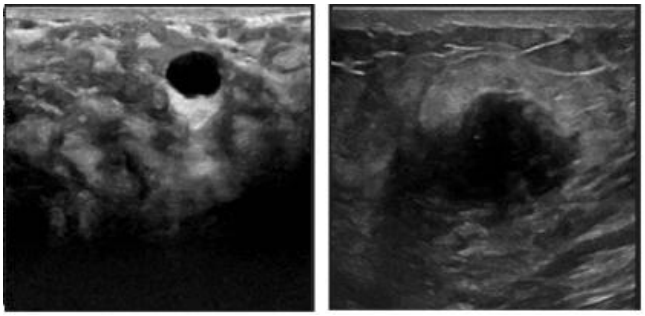
D. Output and Prediction

The output layer is a single neuron with a sigmoid activation function, producing a probability score between 0 and 1. A threshold of 0.5 is applied to categorize the image as benign or malignant. The performance of this architecture was thoroughly evaluated using various metrics, including accuracy, precision, recall, F1-score, and AUC.

IV. Materials and Methods

A. Dataset Collection

The dataset used in this study consists of breast ultrasound images categorized into two classes: benign and malignant. This dataset was previously employed in our earlier study [28] and is publicly available on Kaggle under the title "Ultrasound Breast Images for Breast Cancer"[2]. It comprises 9016 images in total, including 4574 benign and 4442 malignant samples. The images were preprocessed by resizing them to 224×224 pixels and normalizing pixel intensities to the [0,1] range. The choice of this dataset is motivated by its widespread use in prior research and its applicability to benchmarking breast cancer classification with ultrasound imaging, which remains one of the hardest modalities to handle since it is noisy and low-contrast Fig. 3 provides a visual representation of the sample data included in the database used in this paper.



(a) (b)

Fig. 3. Some samples of US images from the dataset are used: (a) Benign. (b) Malignant.

[2]https://www.kaggle.com/datasets/vuppalaadithyasairam/ultrasound-breast-images-for-breast-cancer

To ensure fair evaluation, the dataset was split into training (70%), validation (15%), and test (15%) subsets using stratified sampling to maintain class distribution. Table I summarizes the number of samples per class in each subset.

TABLE I. SAMPLE DISTRIBUTION ACROSS DATASET SPLITS

Subset	Benign	Malignant	Total
Training	3203	3109	6311
Validation	686	667	1353
Test	686	666	1352

B. Background on CNN

Convolutional Neural Networks (CNNs) are a class of deep learning models particularly effective in analyzing image data [29]. CNNs apply a series of convolutional and pooling layers to extract hierarchical features from input images. In this study, we employed a dual-branch CNN architecture combining VGG19 [30] and EfficientNetB0 [31], both pre-trained on the ImageNet dataset. The extracted features from both networks were concatenated before being passed through fully connected layers, enabling more comprehensive representation learning.

C. Confusion Matrix and Evaluation Metrics

In the context of binary classification for breast cancer diagnosis, a confusion matrix is an essential tool to assess the performance of the proposed model. It provides a detailed breakdown of the model's predictions compared to the actual ground truth labels by summarizing the number of correct and incorrect classifications.

The binary confusion matrix [32] is structured as a 2×2 table comprising the following components:

- True Positives (TP): Malignant cases correctly identified.
- True Negatives (TN): Benign cases correctly identified.
- False Positives (FP): Benign cases misclassified as malignant.
- False Negatives (FN): Malignant cases misclassified as benign.

These components form the foundation for calculating several widely adopted evaluation metrics that reflect different aspects of classification performance [33]:

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

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F1-Score} = 2 \times \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \quad (4)$$

These are particularly critical in medical imaging use cases, where false negatives, specifically misclassification, can have profound implications on patient outcomes [34], [35]. They present a holistic perspective of the classification capacity of a deep learning algorithm and are critical indicators in determining whether it is ready for clinical adoption [36].

D. Gradient-Weighted Class Activation Mapping (Grad-CAM)

Gradient-weighted Class Activation Mapping (Grad-CAM) is a visualization technique used for the improved interpretability of convolutional neural networks (CNNs), especially in high-risk applications such as medical image processing. Grad-CAM generates class-specific localization heatmaps which display the most significant areas in an input image with respect to the prediction by the model, and thereby makes an insightful diagnosis of the decision-making process of the neural network.

This is particularly helpful in medical image analysis, where it is not only important to know the decision that a model made but also why a model made that decision. The technique allows localization of discriminative regions within an image by utilizing the gradient information flowing into the last convolutional layer.

The Grad-CAM method, proposed by [37], calculates the gradients of the class score with respect to the feature maps of the final convolutional layer. They are globally averaged to obtain importance weights, which are used to compute a weighted sum of the feature maps. The resulting localization map highlights the discriminative regions that considerably influence the model's prediction.

$$\alpha_k{}^c = \frac{1}{Z} \times \Sigma_i \sum_j \left(\frac{\partial y^c}{\partial A_{ij}{}^k} \right) \quad (5)$$

Where $\alpha_k{}^c$ represents the importance weight for feature map k with respect to class c, y^c is the score for class c, $A_{ij}{}^k$ denotes the activation at position (i, j) in the k-th feature map, and Z is the total number of pixels in A^k.

$$L^c\_Grad - CAM = ReLU\ (\sum_k \alpha_k{}^c \times A^k) \quad (6)$$

This class-discriminative localization map, computed through the ReLU operation, specifically brings into view solely those features that have a positive contribution to the class of interest. This enables clinicians to better understand the model's decision-making process and ensures its alignment with medically relevant areas.

Grad-CAM was employed in this work on the presented concatenated CNN model to identify the discriminative regions responsible for classifying breast ultrasound images into benign and malignant classes. Grad-CAM identifies the spatial locations influencing the classification outcome by computing gradients of the target class with respect to the last convolutional layer. This proves useful in the medical context, where the explainability and transparency are critical towards clinical adoption [38].

The Grad-CAM technique enhances the transparency of the model to an extent that clinicians and radiologists are able to check the region of interest of the model and trust its predictions. Adding Grad-CAM to the framework not only increases diagnostic confidence but also facilitates the explanation of complex medical images.

Recent studies have emphasized the importance of explainable AI (XAI) techniques, such as Grad-CAM for

improving clinical applicability and regulatory compliance in healthcare AI solutions [39].

V. Experimental Results

A. Algorithm Best Parameters

In order to optimize the performance of the proposed deep learning model for breast cancer diagnosis from ultrasound images, various training settings and hyperparameters were chosen based on extensive experimentation. The training approach used was carried out within the TensorFlow Keras environment of a Kaggle notebook setup, which gave the computational capacity necessary for handling deep learning operations.

The input images were resized to a fixed resolution of 224 × 224 pixels, a size that is widely accepted to ensure compatibility with the input layer of most pre-trained CNN models, including VGG19 and EfficientNetB0. The data was divided into three sets: 70% for training, 15% for validation, and 15% for testing. This split allowed confident model estimation while still having sufficient data for training.

The training procedure was established with a batch size of 16 and a total of 24 epochs, which were empirically discovered to achieve an optimal trade-off between convergence rate and model generalization. The Adam optimizer [40] was utilized with learning rate set as 1e-5, enabling adaptive learning as well as stable training, particularly when fine-tuning the pre-trained networks.

To prevent overfitting and ensure generalizability, early stopping was employed with patience of 10 epochs, monitoring the validation loss as the primary stopping point. This ensured that training would be stopped once the performance of the model stabilized, preventing unnecessary computation and overfitting.

Besides, transfer learning was leveraged by loading the ImageNet pre-trained weights for VGG19 and EfficientNetB0. To train, the classification heads of both models were removed and only the shared feature maps were retained. Importantly, all convolutional layers in both networks were frozen to preserve the learned representations, except for the last classification block consisting of concatenated features and fully connected layers. This architecture, dubbed the "frozen heads" setup, takes advantage of powerful feature extractors while concentrating learning capacity on classification layers specific to the task.

To ensure maximum performance of the proposed concatenated model, a group of experiments was conducted to optimize the training configuration. Certain combinations of hyperparameters were investigated, and the choice was based on validation performance, model generalizability, and computational efficiency. The set of optimal hyperparameters utilized in the training of the model is depicted in Table II.

TABLE II. Optimal Hyperparameters Used for Training the Proposed Model

Network	Learning Rate	Batch Size	Optimizer	Loss Function	Epochs
The Proposed Model	1e-5	16	Adam	Binary Cross entropy	24

These parameters were set based on empirical results of repeated training iterations in the Kaggle Notebooks platform. The selected parameters gave a balance between achieving high accuracy and minimizing overfitting, especially when working with a relatively limited medical imaging dataset.

B. Training Results

During the training phase, the performance of the proposed model was monitored closely with standard parameters, including training loss, training accuracy, validation loss, and validation accuracy. All these parameters were recorded over 24 epochs to observe how the model learns and to detect evidence of overfitting or underfitting.

Fig. 4 and Fig. 5 present the accuracy and loss curves of training and validation datasets. In Fig. 4, training and validation accuracy grew with every epoch, which reflects the model to learn discriminative features from input ultrasound images. Fig. 5 exhibits the training and validation loss curves, which gradually came down over the period, reflecting convergence.

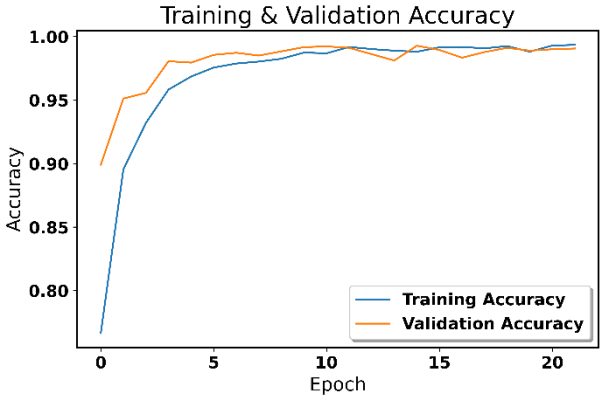


Fig. 4. Accuracy of the proposed model.

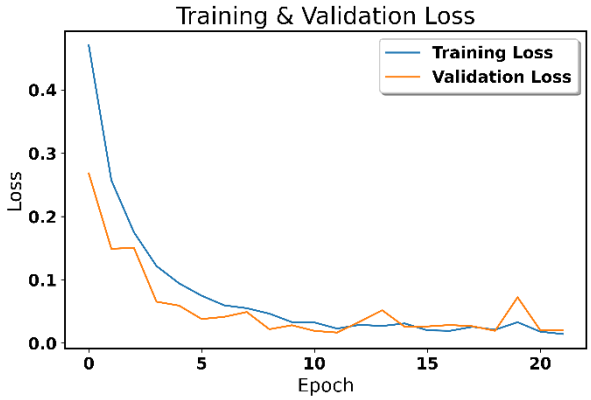


Fig. 5. Loss of the proposed model.

The training results indicate that the model generalizes well and is not overfitting, as the gap between the training and validation metrics remained minimal throughout the training process. Early stopping was employed to prevent overfitting, halting the training if no improvement was observed in the validation loss after 10 consecutive epochs.

These findings confirm the effectiveness of the proposed concatenated architecture in learning meaningful

representations from breast ultrasound images, contributing to robust classification performance.

C. Testing Results

To assess the generalization performance of the proposed model, the testing phase was conducted using the reserved 15% of the dataset. The model demonstrated strong predictive capabilities, as evidenced by the metrics summarized in Table III. These results highlight the model's effectiveness in correctly identifying benign and malignant breast tumors from ultrasound images.

TABLE III. TESTING PERFORMANCE METRICS OF THE PROPOSED MODEL

Metric	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
Proposed Model (VGG19 + EfficientNetB0)	98.44	98.55	99.66	99.11	0.99

The evaluation metrics were computed from the confusion matrix and the ROC curve.

1) Confusion matrix: The confusion matrix shown in Fig. 6 summarizes the classification outcomes:

- True Positives (TP) = 886 malignant cases correctly predicted as malignant.
- True Negatives (TN) = 902 benign cases correctly predicted as benign.
- False Positives (FP) = 13 benign cases incorrectly classified as malignant.
- False Negatives (FN) = 3 malignant cases incorrectly classified as benign.

These results indicate a strong classification capability, especially with a very low false-negative rate, which is critical in clinical settings to avoid missing malignant cases.

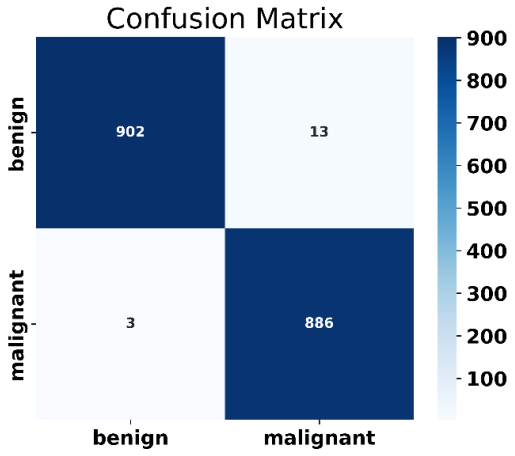


Fig. 6. The confusion matrix of the proposed model.

2) ROC curve and AUC: The ROC curve, depicted in Fig. 7, evaluates the model's discriminative power by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR). The Area Under the Curve (AUC) for the proposed model reached an impressive 0.9999, which signifies excellent reparability between the benign and malignant classes.

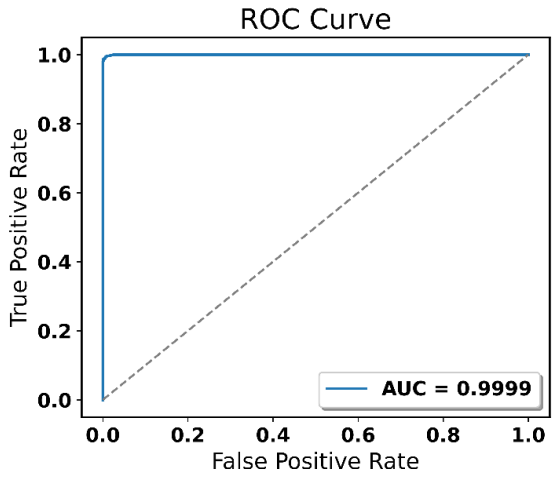


Fig. 7. ROC curve finding for the proposed model.

These findings confirm that the model not only achieves high accuracy (98.44%), but also maintains excellent precision (98.55%), recall (99.66%), and F1-score (99.11%), making it highly reliable for real-world breast cancer diagnosis using ultrasound imaging.

D. Discussion

The proposed dual-model CNN framework, based on the concatenation of VGG19 and EfficientNetB0, demonstrated notable improvements in breast cancer ultrasound image classification. This section discusses the significance of the obtained results, the visual interpretability provided by Grad-CAM, comparative performance with prior studies, limitations, and clinical implications. While the model was specially designed and optimized for breast ultrasound images, it is specifically well-suited for medical imaging modalities with similar issues of low contrast, noise, and high intra-class variability. However, the architecture can easily be generalized to other imaging applications, provided sufficient annotated data, which proves its potential generalizability to breast ultrasound.

1) Performance superiority of the proposed model: Our proposed model achieved an accuracy of 98.44%, a precision of 98.55%, a recall of 99.66%, an F1-score of 99.11%, and an AUC value of 0.9999. These accuracies beat those of individual backbone networks and most prior models reported in the literature. The complementarity of feature representations learned by EfficientNetB0 and VGG19 enabled the model to learn nuanced spatial and contextual patterns, resulting in a greatly improved classification performance.

The improved performance of the proposed model owes to the complementary capability of the backbone networks. VGG19 captures deep spatial features, and EfficientNetB0 offers parameter efficiency and enhanced generalizability.

Concatenation of both at the feature level increases the richness in the extracted features, enabling more precise classification.

2) Grad-CAM visualizations for model interpretability

a) Interpretation of the proposed model: To improve the interpretability of the proposed concatenated model (VGG19 + EfficientNetB0), Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to visualize the discriminative regions that influence the model's classification decisions. Grad-CAM enables a qualitative assessment by projecting class-specific activation maps onto the original ultrasound images, thereby highlighting the areas the model focuses on when distinguishing between benign and malignant cases.

Fig. 8 shows an original breast ultrasound image of a benign lesion. The grayscale image serves as a reference, showing the anatomical structures without any model-generated overlay.

Fig. 8. Original breast ultrasound image showing a benign lesion.

Fig. 9 shows the Grad-CAM visualization generated by the new model. The heatmap overlays reveal that the model is paying attention most of the time to the lesion itself, with the most intense activations marked in red and yellow. This pattern of activation is in line with clinical expectations for benign lesions, where the model's attention is contained within the lesion pattern, aligning with clinical expectations for benign lesions, as the model's attention is confined within the lesion boundaries, avoiding unnecessary activation in adjacent healthy tissues.

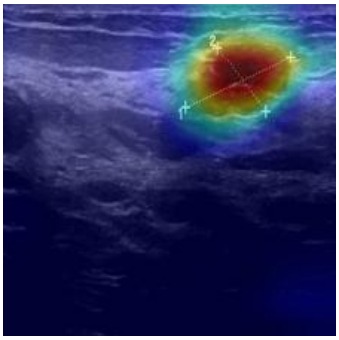


Fig. 9. Grad-CAM visualization from the proposed VGG19 + EfficientNetB0 model, emphasizing the lesion boundary influencing the benign classification decision.

The model's accuracy in targeting the lesion area means that the concatenated structure is capable of encoding the salient aspects of benign cases. This contributes significantly towards encouraging both the interpretability and clinical validity of its predictions. Moreover, the Grad-CAM visualizations positively establish that the model arrives at decisions on the basis of anatomically relevant areas, rather than being confused by noise or noise-like image artifacts.

This interpretability analysis places the model's clinical value into perspective by showing how accurately it can identify important diagnostic areas on ultrasound scans. Providing such visual feedback is crucial when integrating AI-based diagnostic systems into medical procedures, as it fosters transparency and credibility with clinicians who rely on these systems for well-educated choices.

b) Comparative Analysis with other CNN architectures: To further compare the interpretability analysis of the proposed dual-model architecture (VGG19 + EfficientNetB0), there was qualitative comparison against three top single standalone convolutional neural networks: VGG16, EfficientNetB0, and GoogleNet. Comparison was done for all the models using Grad-CAM heatmaps that were implemented on a corresponding benign breast ultrasound image, hence preserving consistency in comparative visual interpretation.

Fig. 10 shows the corresponding Grad-CAM heatmaps in 2×2 formation: top-left shows VGG16, top-right shows EfficientNetB0, bottom-left shows GoogleNet, and bottom-right shows the proposed model (VGG19 + EfficientNetB0).

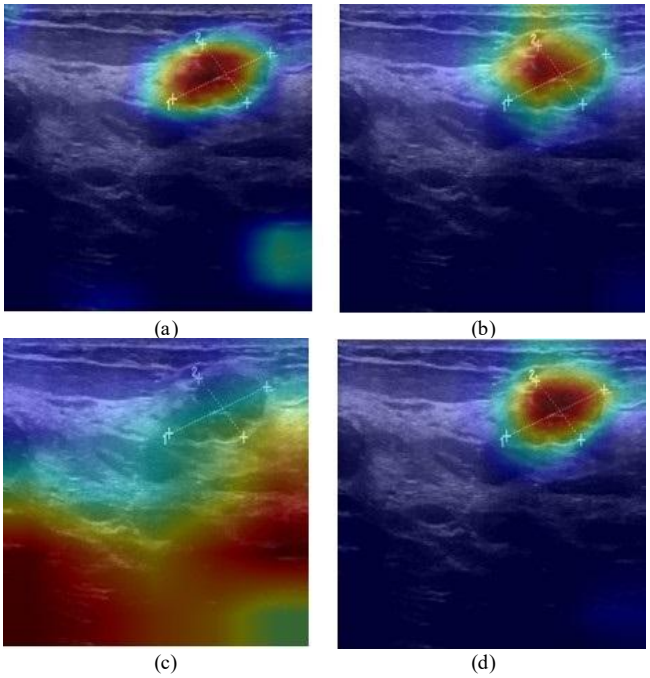


Fig. 10. Visualizations of benign lesion classification Grad-CAM comparison between different CNN architectures: (a) VGG16, (b) EfficientNetB0, (c) GoogleNet, (d) Proposed VGG19 + EfficientNetB0.

These heat maps represent the spatial attention of each model in processing the same benign case. As one may observe, the proposed concatenated model has the most concentrated activation within the diseased area, with minimal interference from the surrounding tissue, suggesting good clinical concordance.

In contrast, VGG16 and EfficientNetB0 show moderate attention to the periphery, while GoogleNet tends to produce more dispersed activation maps, sometimes extending beyond the boundaries of the lesion. These distinctions highlight the value of combining complementary feature representations in the proposed model.

To support the visual observations, a qualitative analysis was conducted based on four interpretability criteria: Activation Focus, Localization Precision, Background Noise, and Clinical Interpretability. The results are summarized in Table IV.

TABLE IV. QUALITATIVE COMPARISON OF GRAD-CAM INTERPRETABILITY ACROSS CNN MODELS

Model	Activation Focus	Localization Precision	Background Noise	Clinical Interpretability
VGG16	Moderate focus on lesion	Medium	Low to medium	Acceptable
EfficientNet B0	Peripheral focus	High at edges	Low	Good
GoogleNet	Scattered focus	Low	High	Limited
Proposed Model	**Centered on lesion**	**Very High**	**Very Low**	**Excellent**

This comparative analysis reinforces the superior interpretability and diagnostic accuracy of the proposed model. The combination of precise localization via heat maps and high classification metrics suggests that feature fusion via model concatenation can significantly improve the performance and clinical usability of AI-based diagnostic tools. This information is particularly relevant for real-world deployment, where model transparency and reliability are crucial factors for their integration into medical decision-making processes.

3) Comparative analysis with prior work: In TABLE V. Table V, we provide a comparison of our proposed model, which leverages the concatenation of VGG19 and EfficientNetB0, against several recent state-of-the-art techniques designed for breast cancer classification using ultrasound imaging and related modalities.

TABLE V. COMPARISON WITH PREVIOUS WORK

Authors	Approach	Breast cancer data	Accuracy (%)	F1-Score (%)	AUC
[9]	VGG19 + ResNet152	Breast US images	90.90	89.20	0.95
[10]	VGG + ResNet + DenseNet	Breast US images	94.62	91.14	0.97
[11]	GoogleNet	Ultrasound DICOM images	92.50	N/A	0.91
[12]	VGG16 + InceptionV3	AUTOMATED BREAST ULTRASOUND	91.90	N/A	0.93
[13]	VGG19	Ultrasound images	98.44	98.39	0.98
This Work	**VGG19 + EfficientNetB0**	**Ultrasound images**	**98.44**	**99.11**	**0.99**

As illustrated in Fig. 11, the proposed model achieved a classification accuracy of 98.44%, marking a clear improvement over earlier approach. For instance, the method presented in [25], which combined VGG19 and ResNet152, reported an accuracy of 90.90%, while the approach in [28] using VGG16 with InceptionV3 reached 91.90%. In addition to accuracy, our model attained an F1-Score of 99.11%, reflecting a well-balanced performance between precision and recall. This level of consistency in performance metrics was either not documented or was significantly lower in prior studies.

The AUC (Area Under the ROC Curve) for our approach is 0.99, demonstrating superior discriminative ability compared to other frameworks, with the closest being [10] at 0.97. This improvement confirms the benefit of our model's dual feature extraction mechanism, which leverages the strengths of both VGG19's spatial depth and EfficientNetB0's efficiency and generalization capabilities.

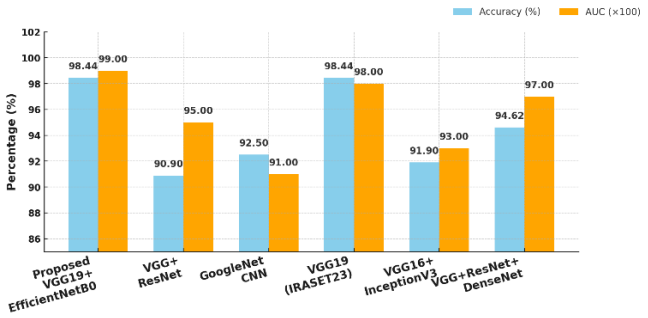


Fig. 11. Comparison of accuracy and AUC for different approaches.

In summary, these results confirm the stability and effectiveness of our proposed model. With the achievement of higher scores in all the measures of evaluation, the model is able to establish itself as suitable for valid clinical application in diagnosing breast cancer via ultrasound imaging.

4) Limitations and clinical relevance: Despite its promising results, the proposed model has its limitations. While the current dataset provides a valuable foundation for model training and evaluation, it does not fully encompass the range of variations typically observed in clinical practice. Moreover, the generalization capabilities of the model require further assessment through rigorous testing on external and more diverse datasets. Moving forward, research efforts should prioritize domain adaptation strategies and conduct large-scale validation studies to ensure robustness across different clinical scenarios. Despite these considerations, the model's demonstrated accuracy and its capacity to visually explain predictions make it a promising tool to support radiologists, especially in settings where resources are limited and diagnostic expertise may not always be readily available.

VI. CONCLUSIONS AND PERSPECTIVES

This work suggests a deep learning architecture for enhancing the classification of breast lesions in ultrasound images. The approach leverages the complementary attributes of two pre-trained convolutional neural networks, VGG19 and EfficientNetB0, by combining their feature extraction layers.

Employing the dual-model fusion, the constructed system offered remarkable performance, with the classification accuracy being 98.44%, F1-score 99.11%, and AUC 0.99.

One of the key arguments of this paper is placing emphasis on the interpretation of the model. Using Grad-CAM visualizations, we had identified and visualized the influential areas of ultrasound images that resulted in the model's classification outputs. These heatmaps provide valuable visual insights into the model's reasoning process and hence improve its transparency and integration into clinical workflow. This interpretability is essential to gaining the trust of medical practitioners since it implies that the choices made by AI are based on medically relevant features.

The new model demonstrated enhanced accuracy and robustness when compared to the previous published methods. Such performance improvement is primarily due to careful blending of complementary CNN structures, combined with training techniques optimized for them. In addition, techniques such as dropout and early stopping were applied strategically to reduce overfitting, hence enabling the model to generalize to new data.

Despite the positive results, there are certain limitations that should be appreciated. Even though the dataset used in the current study does include a heterogeneous collection of cases, future extension of the dataset to include a wider patient population and image qualities would further strengthen the model's robustness. Additional external validation by independent datasets from different clinical scenarios is also required to critically test the model's ability to generalize beyond the training scenario.

There are some ways in which this work will be taken forward in the future. One of the main areas of focus is to expand the size of the present dataset using large-scale data collection and synthetic augmentation approaches. This will strengthen the resilience and flexibility of the model for various clinical scenarios. Also, adding other tools of explainability, such as SHAP and LIME, to work with Grad-CAM to gain a deeper insight into model behavior is also a potential area for study in the future. Lastly, the implementation of this system into a real-time clinical decision support tool and determining the effect of the system on the accuracy and efficiency of the diagnosis will be the step required toward adoption and applied practice in the clinical environment.

These advancements in the future will seek to unveil the relationship between experimental research and clinical application, thus advancing AI-based breast cancer imaging diagnostic tools.

REFERENCES

[1] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, 'Global cancer statistics, 2012', CA. Cancer J. Clin., vol. 65, no. 2, pp. 87–108, 2015, doi: 10.3322/caac.21262.

[2] W. A. Berg, 'Tailored supplemental screening for breast cancer: what now and what next?', Radiol. Clin. North Am., vol. 42, no. 5, pp. 935–947, 2004, doi: 10.1016/j.rcl.2004.04.005.

[3] A. T. Stavros, D. Thickman, C. L. Rapp, M. A. Dennis, S. H. Parker, and G. A. Sisney, 'Solid breast nodules: use of sonography to distinguish benign and malignant lesions', Radiology, vol. 196, no. 1, pp. 123–134, 1995, doi: 10.1148/radiology.196.1.7784555.

[4] G. Litjens et al., 'A survey on deep learning in medical image analysis', Med. Image Anal., vol. 42, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.

[5] S. Laghmati, S. Hamida, K. Hicham, B. Cherradi, and A. Tmiri, 'An improved breast cancer disease prediction system using ML and PCA', Multimed. Tools Appl., vol. 83, no. 11, pp. 33785–33821, Sept. 2023, doi: 10.1007/s11042-023-16874-w.

[6] D. Shen, G. Wu, and H.-I. Suk, 'Deep learning in medical image analysis', Annu. Rev. Biomed. Eng., vol. 19, pp. 221–248, 2017, doi: 10.1146/annurev-bioeng-071516-044442.

[7] H.-C. Shin et al., 'Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning', IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1285–1298, 2016, doi: 10.1109/TMI.2016.2528162.

[8] Z. Z, S. Mmr, T. N, and L. J, 'UNet++: A Nested U-Net Architecture for Medical Image Segmentation', Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support 4th Int. Workshop DLMIA 2018 8th Int. Workshop ML-CDS 2018 Held Conjunction MICCAI 2018 Granada Spain S, vol. 11045, Sept. 2018, doi: 10.1007/978-3-030-00889-5_1.

[9] O. Terrada, B. Cherradi, A. Raihani, and O. Bouattane, 'A fuzzy medical diagnostic support system for cardiovascular diseases diagnosis using risk factors', in 2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra: IEEE, Dec. 2018, pp. 1–6. doi: 10.1109/ICECOCS.2018.8610649.

[10] P. Kumar and A. Kumar, 'Heart Disease Classification and Recommendation by Optimized Features and Adaptive Boost Learning', Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 3, 2023, doi: 10.14569/IJACSA.2023.01403103.

[11] O. Terrada, A. Raihani, O. Bouattane, and B. Cherradi, 'Fuzzy cardiovascular diagnosis system using clinical data', in 2018 4th International Conference on Optimization and Applications (ICOA), Mohammedia, Morocco: IEEE, Apr. 2018, pp. 1–4. doi: 10.1109/ICOA.2018.8370549.

[12] O. Daanouni, B. Cherradi, and A. Tmiri, 'Type 2 Diabetes Mellitus Prediction Model Based on Machine Learning Approach', in Innovations in Smart Cities Applications Edition 3, M. Ben Ahmed, A. A. Boudhir, D. Santos, M. El Aroussi, and İ. R. Karas, Eds, in Lecture Notes in Intelligent Transportation and Infrastructure. , Cham: Springer International Publishing, 2020, pp. 454–469. doi: 10.1007/978-3-030-37629-1_33.

[13] E. Sabitha and M. Durgadevi, 'Improving the Diabetes Diagnosis Prediction Rate Using Data Preprocessing, Data Augmentation and Recursive Feature Elimination Method', Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 9, 2022, doi: 10.14569/IJACSA.2022.01309107.

[14] O. Daanouni, B. Cherradi, and A. Tmiri, 'Automatic Detection of Diabetic Retinopathy Using Custom CNN and Grad-CAM', in Advances on Smart and Soft Computing, vol. 1188, F. Saeed, T. Al-Hadhrami, F. Mohammed, and E. Mohammed, Eds, in Advances in Intelligent Systems and Computing, vol. 1188. , Singapore: Springer Singapore, 2021, pp. 15–26. doi: 10.1007/978-981-15-6048-4_2.

[15] A. Ouhmida, A. Raihani, B. Cherradi, and Y. Lamalem, 'Parkinson's disease classification using machine learning algorithms: performance analysis and comparison', in 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco: IEEE, Mar. 2022, pp. 1–6. doi: 10.1109/IRASET52964.2022.9738264.

[16] A. M and P. Gera, 'Parkinson's Disease Identification using Deep Neural Network with RESNET50', Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 11, 2022, doi: 10.14569/IJACSA.2022.0131156.

[17] A. Ouhmida, A. Raihani, B. Cherradi, and S. Sandabad, 'Parkinson's diagnosis hybrid system based on deep learning classification with imbalanced dataset', Int. J. Electr. Comput. Eng. IJECE, vol. 13, no. 3, p. 3204, June 2023, doi: 10.11591/ijece.v13i3.pp3204-3216.

[18] S. Hamida, O. El Gannour, B. Cherradi, H. Ouajji, and A. Raihani, 'Handwritten computer science words vocabulary recognition using concatenated convolutional neural networks', Multimed. Tools Appl., vol. 82, no. 15, pp. 23091–23117, June 2023, doi: 10.1007/s11042-022-14105-2.

[19] S. Hamida, B. Cherradi, O. El Gannour, O. Terrada, A. Raihani, and H. Ouajji, 'New Database of French Computer Science Words Handwritten Vocabulary', in 2021 International Congress of Advanced Technology and Engineering (ICOTEN), Taiz, Yemen: IEEE, July 2021, pp. 1–5. doi: 10.1109/ICOTEN52080.2021.9493438.

[20] S. Hamida, B. Cherradi, H. Ouajji, and A. Raihani, 'Convolutional Neural Network Architecture for Offline Handwritten Characters Recognition', in Innovation in Information Systems and Technologies to Support Learning Research, vol. 7, M. Serrhini, C. Silva, and S. Aljahdali, Eds, in Learning and Analytics in Intelligent Systems, vol. 7. , Cham: Springer International Publishing, 2020, pp. 368–377. doi: 10.1007/978-3-030-36778-7_41.

[21] M. Errami, M. A. Ouassil, R. Rachidi, B. Cherradi, S. Hamida, and A. Raihani, 'Sentiment Analysis on Moroccan Dialect based on ML and Social Media Content Detection', Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 3, 2023, doi: 10.14569/IJACSA.2023.0140347.

[22] I. Sutedja and H. -, 'Sentiment Analysis: An Insightful Literature Review', Int. J. Adv. Comput. Sci. Appl., vol. 16, no. 3, 2025, doi: 10.14569/IJACSA.2025.0160351.

[23] H. Nhaila, M. Merzouqi, E. Sarhrouni, and A. Hammouch, 'Hyperspectral images classification and Dimensionality Reduction using Homogeneity feature and mutual information', in 2015 Intelligent Systems and Computer Vision (ISCV), Mar. 2015, pp. 1–5. doi: 10.1109/ISACV.2015.7106167.

[24] M. Tanaka and et al., 'Computer-aided diagnosis system for breast ultrasound images using deep learning', Phys. Med. Biol., vol. 64, no. 21, 2019, doi: 10.1088/1361-6560/ab5093.

[25] W. K. Moon, R. F. Chang, and others, 'Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks', Comput. Methods Programs Biomed., vol. 190, 2020, doi: 10.1016/j.cmpb.2020.105361.

[26] T. Fujioka and others, 'Distinction between benign and malignant breast masses at breast ultrasound using deep learning method with convolutional neural network', Jpn. J. Radiol., vol. 38, no. 4, 2019, doi: 10.1007/s11604-019-00831-5.

[27] J. F. Lazo, S. Moccia, E. Frontoni, and E. D. Momi, 'Comparison of different CNNs for breast tumor classification from ultrasound images', Dec. 28, 2020, arXiv: arXiv:2012.14517. doi: 10.48550/arXiv.2012.14517.

[28] M. Zourhri, S. Hamida, N. Akouz, B. Cherradi, H. Nhaila, and M. E. Khaili, 'Deep Learning Technique for Classification of Breast Cancer using Ultrasound Images', in 2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Mohammedia, Morocco: IEEE, May 2023, pp. 1–8. doi: 10.1109/IRASET57153.2023.10153069.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, 'ImageNet Classification with Deep Convolutional Neural Networks', in Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[30] K. Simonyan and A. Zisserman, 'Very Deep Convolutional Networks for Large-Scale Image Recognition', in International Conference on Learning Representations (ICLR), 2015.

[31] M. Tan and Q. V. Le, 'EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks', in International Conference on Machine Learning (ICML), 2019, pp. 6105–6114.

[32] O. EL GANNOUR, S. HAMIDA, B. CHERRADI, A. RAIHANI, and H. MOUJAHID, 'Performance Evaluation of Transfer Learning Technique for Automatic Detection of Patients with COVID-19 on X-Ray Images', in 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Dec. 2020, pp. 1–6. doi: 10.1109/ICECOCS50124.2020.9314458.

[33] H. Dalianis, 'Evaluation Metrics and Evaluation', in Clinical Text Mining: Secondary Use of Electronic Patient Records, H. Dalianis, Ed., Cham: Springer International Publishing, 2018, pp. 45–53. doi: 10.1007/978-3-319-78503-5_6.

[34] M. Sokolova and G. Lapalme, 'A systematic analysis of performance measures for classification tasks', Inf. Process. Manag., vol. 45, no. 4, pp. 427–437, 2009, doi: https://doi.org/10.1016/j.ipm.2009.03.002.

[35] D. M. W. Powers, 'Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation', Oct. 11, 2020, arXiv: arXiv:2010.16061. doi: 10.48550/arXiv.2010.16061.

[36] D. Chicco and G. Jurman, 'The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation', BMC Genomics, vol. 21, no. 1, p. 6, Jan. 2020, doi: 10.1186/s12864-019-6413-7.

[37] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, 'Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization', in 2017 IEEE International Conference on Computer Vision (ICCV), Oct. 2017, pp. 618–626. doi: 10.1109/ICCV.2017.74.

[38] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, 'Causability and explainability of artificial intelligence in medicine', WIREs Data Min. Knowl. Discov., vol. 9, no. 4, p. e1312, 2019, doi: 10.1002/widm.1312.

[39] J. Gupta and K. R. Seeja, 'A Comparative Study and Systematic Analysis of XAI Models and their Applications in Healthcare', Arch. Comput. Methods Eng., vol. 31, no. 7, pp. 3977–4002, Sept. 2024, doi: 10.1007/s11831-024-10103-9.

[40] D. P. Kingma and J. Ba, 'Adam: A Method for Stochastic Optimization', Jan. 30, 2017, arXiv: arXiv:1412.6980. doi: 10.48550/arXiv.1412.6980.