Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 20 No. 2 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v20i02.43795

PAPER

An Integrated Multimodal Deep Learning Framework for Accurate Skin Disease Classification

Soufiane Hamida[1,2], Driss Lamrani[3], Mohammed Amine Bouqentar[3], Oussama El Gannour[3,4], Bouchaib Cherradi[1,3,5](✉)

[1]2IACS Laboratory, ENSET of Mohammedia, Hassan II University of Casablanca, Mohammedia, Morocco

[2]GENIUS Laboratory, SupMTI of Rabat, Rabat, Morocco

[3]EEIS Laboratory, ENSET of Mohammedia, Hassan II University of Casablanca, Mohammedia, Morocco

[4]LIASSE Laboratory, ENSA of Fez, Sidi Mohammed Ben Abdellah University, Fez, Morocco

[5]STIE Team, CRMEF Casablanca-Settat, Provincial Section of El Jadida, El Jadida, Morocco

bouchaib.cherradi@enset-media.ac.ma

ABSTRACT

In order to effectively treat skin diseases, an accurate and prompt diagnosis is required. In this article, a novel method for classifying skin disorders using a multimodal classifier is presented. The proposed classifier utilizes multiple information sources to enhance the accuracy of disease classification. It incorporates images of skin lesions and patient-specific data. The multimodal classifier simultaneously classifies diseases by combining image and structured data inputs. The effectiveness of the proposed classifier was evaluated using the ISIC 2018 dataset, which includes images and clinical data for seven categories of skin diseases. The results indicate that the proposed model outperforms conventional single-modal and single-task classifiers, achieving an accuracy of 98.66% for image classification and 94.40% for clinical data classification. In addition, we compare the performance of the proposed model with that of other methodologies, demonstrating its superiority. Despite yielding promising results, the proposed method has limitations in terms of data requirements and generalizability. Future research directions include incorporating additional information sources, investigating genetic data integration, and applying the method to various medical conditions. This study illustrates the potential of integrating multimodal techniques with transfer learning in deep neural networks to enhance the classification accuracy of cutaneous diseases.

KEYWORDS

multimodal classifier, cutaneous diseases, skin lesions, transfer learning, image classification

1 INTRODUCTION

Conditions affecting the skin, hair, nails, and mucous membranes are classified as dermatological diseases [1]. The clinical manifestations of these diseases range from benign and self-limiting conditions to severe and incapacitating disorders. Accurate classification of skin conditions is essential for prompt and precise diagnosis, leading to the administration of appropriate treatment and care [2], [3].

Hamida, S., Lamrani, D., Bouqentar, M.A., El Gannour, O., Cherradi, B. (2024). An Integrated Multimodal Deep Learning Framework for Accurate Skin Disease Classification. *International Journal of Online and Biomedical Engineering (iJOE)*, 20(2), pp. 78–94. https://doi.org/10.3991/ijoe.v20i02.43795

Article submitted 2023-08-06. Revision uploaded 2023-11-14. Final acceptance 2023-11-14.

A prompt diagnosis enables the initiation of the most suitable treatment, thus preventing disease progression and minimizing the risk of complications. In order to classify skin diseases, one must have a comprehensive understanding of clinical presentation, histopathological characteristics, and relevant laboratory findings. Misdiagnosis or delayed diagnosis can lead to inadequate treatment or increased healthcare costs [4].

In recent years, there has been significant interest in the potential of deep learning (DL) algorithms to improve diagnostic accuracy, decrease diagnosis times, and assist in the identification of underlying systemic diseases [5]. DL algorithms are well-suited for analyzing large datasets and identifying patterns that may be difficult for human experts to discern. Leveraging skin lesion images, clinical data analysis, and interpretation of laboratory results, Machine Learning (ML) and DL techniques have been crucial in the classification of skin diseases. These algorithms hold significant promise for improving diagnostic accuracy and speeding up the diagnostic process [6].

This paper aims to provide a comprehensive analysis of proposed DL techniques for the classification of cutaneous diseases, with a particular focus on Transfer Learning (TL), Deep Neural Networks (DNN), and the innovative concept of multimodal classifiers. TL involves refining previously trained models for specific tasks using pre-trained models, while DNNs are multilayered architectures capable of learning complex data features and patterns [7]. Multimodal classifiers are a category of DL models that are skilled at processing various data modalities, including images, text, and numeric data, while simultaneously performing multiple tasks, such as classification and regression. We compare the effectiveness of these DL techniques with traditional AI methods. The main goal of this study is to propose a new classification method for skin conditions that integrates multimodal data learning. The proposed model utilizes a multimodal classifier to improve the accuracy of disease classification by combining image data of cutaneous lesions with patient-specific clinical data. We assess the model's performance using the ISIC 2018 dataset, which is a comprehensive collection of images and clinical data from patients diagnosed with various skin diseases [8], [9]. The evaluation results unequivocally demonstrate the superiority of the proposed model over conventional single-modal classifiers. It achieved an accuracy rate of 98.66% for image classification and 94.40% for clinical data classification.

The subsequent sections of this paper are organized as follows: Section 2 outlines the design, implementation, and methodology of the proposed multimodal classifier. Section 3 provides insights into the evaluation dataset, DL algorithms, and performance metrics of the study. Section 4 presents and discusses the experimental results and findings. Finally, Section 5 provides concluding remarks and suggests potential avenues for future research.

2 LITERATURE REVIEW

Due to its ability to identify complex patterns in medical imaging data, DL has emerged as a promising tool for medical diagnosis, especially in the classification of skin diseases [10]. Utilizing various network architectures, such as Convolutional Neural Networks (CNN) [11], [12], an increasing body of research has been dedicated to applying DL techniques to classify skin diseases [13]. This literature review offers a comprehensive overview of recent studies that have utilized DL techniques to classify skin diseases and highlights their contributions to the field [14], [15].

A significant study by [16] presents a set of lightweight DL networks for the detection of skin cancer. By combining several lightweight models, including DC-MobileNetV1, DC-DenseNet121, and Fusion Models, the researchers achieved accuracy rates of 86.5%, 85.5%, and 87.6% respectively, on the binary ISBI 2016 dataset. The utilization of ensemble techniques has enhanced accuracy and robustness, showcasing the potential of such methods in medical image analysis for diagnosing skin cancer. Another study [17] highlights the use of deep CNN-based methods for classifying cutaneous diseases through discriminative feature learning. On the HFSD benchmark dataset, the researchers evaluated their proposed models, ResNet152 and Inception_ResNet-V2, using the triplet method. They achieved 84.91% and 87.42% accuracy, respectively. These findings demonstrate the effectiveness of deep CNNs in extracting meaningful features from skin images, which contributes to improved classification performance for a wide range of skin conditions. Furthermore, [18] explores the utilization of ML techniques with dynamic training and testing enhancements to enhance the accuracy of skin condition prediction. Through the improvement of training and testing data, the proposed method achieved an impressive 95% accuracy on the eight-class ISEC 2019 dataset. This study illuminates the potential of data augmentation to enhance the predictive accuracy of ML models for skin disease prediction, thereby enabling early diagnosis and effective treatment.

In an attempt to enhance the accuracy of skin disease classification, [19] introduces a method that integrates a tailored loss function, balanced mini-batch logic, and real-time image enhancement. The evaluation of the ISIC 2018 dataset resulted in accuracy values of 88.46% for DenseNet169 and 89.97% for EfficientNetB4, showcasing the effectiveness of this approach in improving the classification performance of skin diseases. Another important study [20] focuses on developing a DL model for accurately classifying cutaneous lesions based on their characteristics. Using the comprehensive ISIC 2019 dataset, which contains eight classes, the researchers were able to propose two models based on DenseNet201 that achieved remarkable accuracy rates of 91.71% and 92.33%. This study highlights the potential of DL techniques for automating the classification of skin lesions, thereby contributing to the early detection and treatment of skin conditions.

In a study by researchers [6], a novel approach was explored using a combination of hybrid deep feature selection and Extreme Learning Machines (ELM) for the classification of various types of skin lesions. They developed a method to extract deep features from skin lesion images using a pre-trained CNN, followed by classification using ELM. The proposed method was rigorously evaluated using the ISIC 2018 dataset, resulting in remarkable classification accuracy for multiclass skin lesions. The accuracy rates of the NasNet Large, HWOA + ELM, EMI + ELM, and NasNet Large + ELM models were 85.96%, 87.86%, 87.12%, and 94.36%, respectively. These results demonstrate the promising potential of combining hybrid deep feature selection and ELM techniques for the diagnosis and treatment of cutaneous diseases. Recent studies summarizing the application of DL models for skin disease classification are presented in Table 1.

Table 1. Summary of relevant literature studies

Ref.	Year	Datasets	Classes	Models	Result
[16]	2020	ISBI 2016 Dataset	2	DC-MobileNetV1	Ac = 86.5%
				DC-DenseNet121	Ac = 85.5%
				Fusion_Models	Ac = 87.6%
[17]	2020	HFSD Dataset	4	ResNet152 + Triplet	Ac = 84.91%
				Inception_ResNet-V2 + Triplet	Ac = 87.42%
[18]	2020	ISIC 2019 Dataset	8	PA DPI	Ac = 95 %
[19]	2020	ISIC 2018 Dataset	7	DenseNet169	Ac = 88.46%
				EfficientNetB4	Ac = 89.97%
[20]	2021	ISIC 2019 Dataset	8	DenseNet201, Cubic SVM	Ac = 91.71%
				DenseNet201, Fine KNN	Ac = 92.34%
[21]	2022	ISIC 2018 Dataset	7	NasNet Large	Ac = 85.96%
				HWOA + ELM	Ac = 87.86%
				EMI + ELM	Ac = 87.12%
				NasNet Large + ELM	Ac = 94.36%

These studies demonstrate the rapid progress of DL techniques in classifying skin diseases. The utilization of ensemble models, deep CNNs, data augmentation, and customized loss functions has demonstrated potential for improving accuracy and resilience. The findings pave the way for the development of effective diagnostic tools in dermatology and clinical medicine, thereby facilitating the early detection and treatment of skin diseases.

3 METHODOLOGY

3.1 Design and implementation of the multimodal classifier

In the development of our multimodal classifier for skin diseases, we utilized a combination of the Efficient-Net V2L network for image data and a DNN for clinical data. To ensure clarity and transparency, we will explicitly outline the responsible classifiers and their impact on the final results. The methodology starts with gathering a large dataset that includes images of skin lesions and the corresponding clinical data for each patient. For this study, we selected the ISEC-2018 dataset, a well-established resource widely used in skin lesion research. This dataset provides valuable information to support our multimodal skin disease system [22]. The collected data undergoes rigorous preprocessing to ensure its suitability for the model. This preprocessing phase includes data cleansing, which involves eliminating irrelevant or missing information to ensure that the data is properly formatted and ready for use by the classifiers.

For feature extraction, we employ specific methodologies tailored to the characteristics of the data. The Efficient-Net V2L network is used for processing image data, enabling the extraction of meaningful visual patterns and features from skin lesion images [23]. Simultaneously, clinical data of a different modality undergoes feature extraction using a DNN architecture. This process is essential for capturing relevant clinical insights and characteristics. To leverage the strengths of both image and clinical data, we have adopted a multi-task learning framework. This framework incorporates a shared encoder architecture, enabling the model to learn simultaneously from both types of data. In this approach, the outputs of the Efficient-Net V2L network and the DNN are effectively combined and conveyed to a final classifier layer. This integration ensures that the model benefits from information derived from both modalities, thereby enhancing precision and diagnostic capabilities. To evaluate the performance of the proposed classifier, we use a separate test dataset. This evaluation process aims to assess the classifier's ability to accurately diagnose skin diseases. By utilizing both image and clinical data, the classifier can extract relevant and complementary features, leading to enhanced diagnostic accuracy and overall performance. Figure 1 illustrates the phases involved in the design of the proposed multimodal classifier.

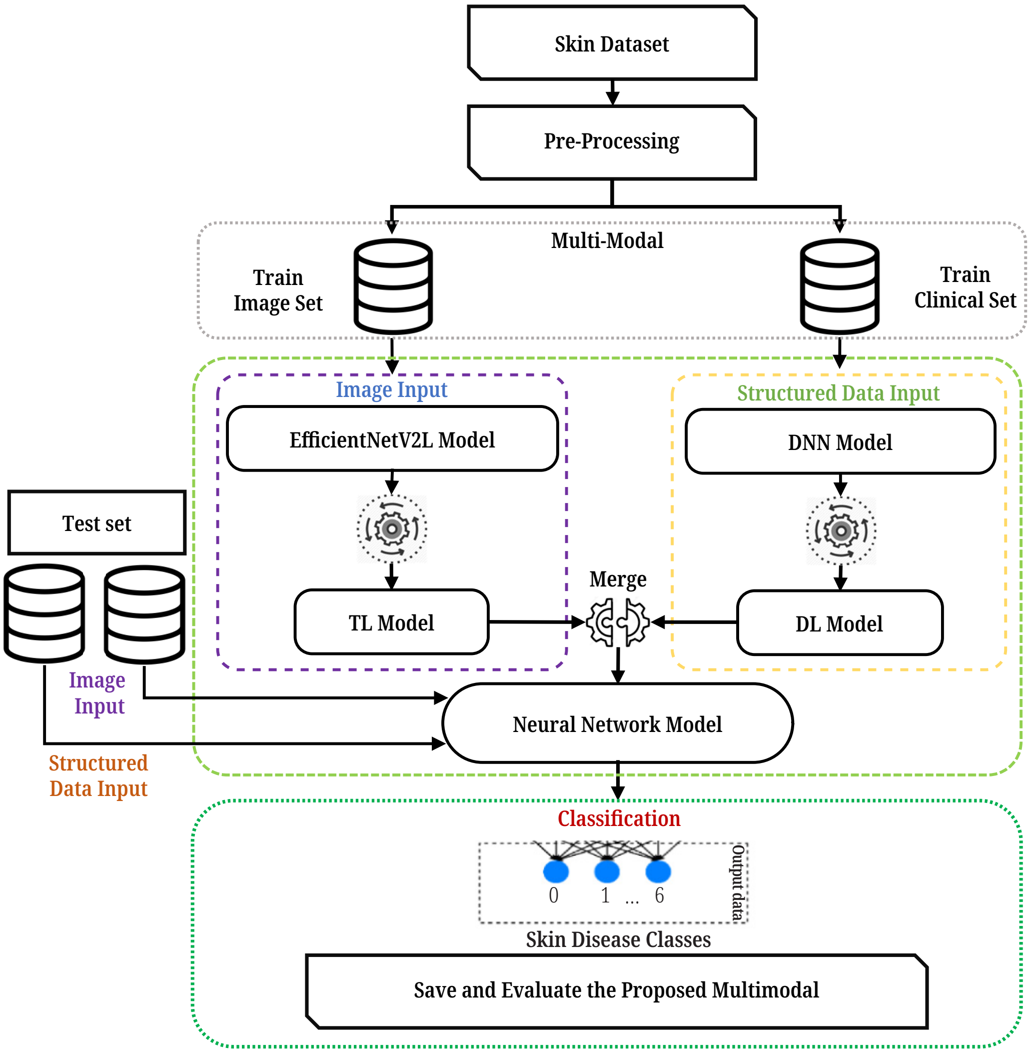

Fig. 1. The architecture of the proposed model

The flowchart shown in Figure 1 serves as both a visual representation and a comprehensive guide, illustrating the sequential progression of the design process. It outlines key stages such as data preprocessing, model construction, and model evaluation. This visual aid improves comprehension and navigation through the different stages of the design process, guaranteeing a well-organized and uniform implementation of the proposed classifier.

3.2 Data collection and preprocessing

The HAM10000 dataset is a crucial collection of dermatoscopic images of skin lesions compiled by researchers at the University of Edinburgh. It is one of the most comprehensive resources dedicated specifically to skin cancer, featuring an extensive image database of over 10,000 images. Each image in the dataset has been meticulously categorized into one of seven distinct forms of skin cancer, making it a diverse and extensive dataset suitable for in-depth examination [24].

The HAM10000 dataset offers researchers studying skin cancer a comprehensive and diverse array of images, which is perfect for training and assessing ML models for precise diagnosis. The training set consists of 7500 images, while the test set comprises 2500 images, providing ample data for robust model development and evaluation. In addition, the HAM10000 dataset is a valuable resource for researchers studying the effects of different skin cancer treatments. The availability of pre- and post-treatment images enables investigations into treatment efficacy, thereby contributing to the development of more effective treatment strategies and the advancement of medical knowledge [25]. The HAM10000 dataset is invaluable for medical professionals in aiding the diagnosis of skin cancer. Expertly labeled images with corresponding types of skin cancer assist physicians in making quick and accurate diagnoses, enabling earlier intervention and improved patient outcomes [14].

3.3 Deep learning algorithms

In this study, DL algorithms are utilized to build and train a multimodal and multitask classifier for the categorization of cutaneous diseases. DL is a subfield of ML that focuses on constructing and training neural networks with multiple layers. This enables the networks to autonomously learn hierarchical data representations. These algorithms have shown remarkable success in various applications, particularly in image recognition and natural language processing. Furthermore, a DNN is used to handle the clinical dataset. DNNs are neural networks with multiple layers that are specifically designed to learn complex patterns and relationships from structured data. The DNN accepts clinical data as input and utilizes a series of hidden layers with nonlinear activation functions to convert the clinical data into valuable, useful feature representations [26]. The DNN formulation can be represented as:

$$h = f(W_x + b) \tag{1}$$

$$y = g(V_h + c) \tag{2}$$

Where x represents the clinical data input, W and V are the weight matrices, b and c are the bias terms, and f and g are non-linear activation functions applied element-wise. The DNN's output, y, represents the extracted feature representations from the clinical data.

Efficient-Net V2L network. The Efficient-Net V2L network is one of the main DL algorithms used in this study. Efficient-Net is a series of neural network architectures designed to balance model size and performance through systematic scaling of the network's depth, width, and resolution. The Efficient-Net V2L network is a fundamental component of the proposed multimodal classifier for skin disease classification in this study [27]. The Efficient-Net V2L network is an extension of the original Efficient-Net architecture, designed specifically for image recognition tasks. It is renowned for its ability to learn complex and distinctive features from input images and is designed to efficiently handle large-scale image datasets [28].

The Efficient-Net V2L architecture features a unique combination of depth-wise separable convolutions, inverted residual blocks, and squeeze-and-excitation blocks, resulting in a highly efficient and accurate model. The following diagram illustrates the Efficient-Net V2L network's fundamental architecture:

$$\text{EfficientNet\_V2L}(x) = \text{SE} - \text{Block}(\textit{IR-Block}(x)) \tag{3}$$

Where x represents the input image data, IR-Block denotes the inverted residual block, and SE-Block represents the squeeze-and-excitation block. The inverted residual block is a unique building block that efficiently captures feature representations with fewer parameters, making the network highly scalable and computationally efficient. The Efficient-Net V2L network was chosen for its effectiveness in handling large-scale image datasets and its ability to extract informative features, which are crucial for accurate classification of skin diseases [29].

In addition, the Efficient-Net V2L network incorporates a technique called compound scaling, which systematically scales the network's depth, width, and resolution. This strategy allows the model to achieve a better balance between model size and performance, leading to enhanced accuracy for challenging image recognition tasks. By integrating the Efficient-Net V2L network into the proposed multimodal classifier, the gains acquire the ability to effectively analyze and process image data, which is crucial for accurately diagnosing skin diseases. The Efficient-Net V2L network serves as a powerful feature extractor, capturing informative representations from the input images. These representations are then integrated with clinical data features to produce strong and precise predictions.

Multimodal neural network. A multimodal neural network is a powerful architecture for DL that can process and analyze multiple categories of input data simultaneously. Multiple modalities, such as images, text, and audio, or any combination thereof, can be efficiently processed by this network [30]. A multimodal neural network can handle various tasks, including classification and segmentation, by integrating a single network structure. This offers a unified and efficient approach to data processing. One of the main advantages of a multimodal neural network is its capability to use shared representations and features across different tasks. Through weight sharing, the network is able to acquire knowledge from one task and apply it to improve the performance of other related tasks. This not only enhances the overall performance of the model but also enables it to generalize more effectively, reducing the requirement for extensive training data for each specific task [31].

In medical applications such as medical imaging and clinical data analysis, where multiple categories of data (e.g., images and patient records) must be jointly analyzed to make accurate diagnoses, a multimodal neural network is particularly advantageous. By integrating and simultaneously learning from multiple modalities, the network can make more informed decisions, leading to improved diagnosis and treatment planning [32]. Figure 2 depicts the architectural components and connections of a multimodal neural network model.

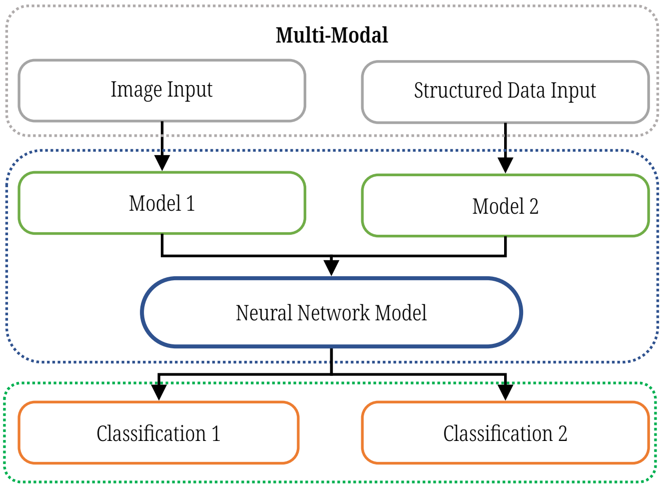

Fig. 2. An example of a multi-modal neural network

This model graphically illustrates the integration and processing of various modalities within the network, demonstrating its ability to manage multiple information sources and perform complex tasks using shared representations. The multimodal neural network is a valuable tool in various fields, such as healthcare, multimedia analysis, and natural language processing, because of its adaptability and versatility.

3.4 Performance metrics

The Confusion Matrix is an essential tool for evaluating the performance of classification algorithms, particularly in the field of medical diagnosis, where precision and reliability are of the utmost importance [33]. For the classification of skin diseases, the confusion matrix provides a concise representation of the results of a diagnostic test, including the counts of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). These values provide crucial information for calculating various performance measures that help evaluate the diagnostic test's ability to accurately identify skin diseases.

Several performance metrics are commonly used to quantitatively evaluate diagnostic tests, and four important measures are widely utilized in the classification of skin diseases. Accuracy is a fundamental measure that indicates the percentage of correctly classified cases out of the total number of instances. The formula for precision is as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{4}$$

Second, precision refers to the proportion of true positive predictions made by the model out of all positive predictions. It is essential for assessing the accuracy of positive predictions. The formula for precision is defined as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{5}$$

Third, the recall, also referred to as sensitivity, is the proportion of true positive cases correctly identified by the model. It highlights the classifier's capability to identify positive cases. This is the formula for recall:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{6}$$

Fourth, specificity is a measure of the classifier's ability to accurately identify negative cases. It quantifies the proportion of true negative cases that the model correctly identifies. The specificity formula is expressed as follows:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \tag{7}$$

By calculating these metrics using the values from the confusion matrix, a thorough performance evaluation of the diagnostic test can be obtained. These performance measures provide valuable insight into the strengths and limitations of the classification model, enabling informed decisions and potential improvements to enhance the accuracy and effectiveness of skin disease diagnosis.

4 RESULTS AND DISCUSSION

4.1 Configuration of the experimental

The experimental setup for this study was implemented in the Jupyter Lab environment. We used the Google Colab platform, a cloud-based service optimized for research and AI algorithm training through Jupyter Notebooks. Our computational resources included a Tesla K-80 GPU, a 2.20 GHz Intel Xeon processor with dual CPUs, and 13 GB of RAM. These resources were carefully selected to ensure sufficient computing capacity for training and evaluating DL models.

For the design of our DNN, we utilized Keras and Autokeras, which are widely recognized for their flexibility and usability in constructing DL architectures. Keras served as the primary framework for constructing models, offering a strong foundation for designing and implementing DNN. Autokeras played a pivotal role in automating hyperparameter optimization, which expedited the model development process.

In the training process, we used standard settings appropriate for multiclass classification tasks. We specifically utilized the Adam optimizer and the categorical cross-entropy loss function. The Adam optimizer was selected for its capability to offer adaptive learning rates, which accelerate convergence during the training phase. The categorical cross-entropy loss function is well-suited for multi-class classification because it measures the difference between predicted class probabilities and actual class labels, enabling accurate learning. To strike an effective balance between model performance and the prevention of overfitting, the proposed DNN model underwent fine-tuning for 50 epochs with a batch size of 16. Each epoch represents a full iteration through the entire training dataset, while the batch size determines the number of samples processed during each forward and backward pass. The learning rate for the Adam optimizer was set to 1e-4, which is a critical hyperparameter that regulates the step size during gradient descent and influences the model's convergence rate. This parameter was carefully selected to ensure optimal learning and accurate training outcomes. Table 2 provides a concise summary of the key hyperparameters and functions used during the training process of the proposed DNN model. These parameters were carefully selected to optimize the model's performance and efficient convergence during training.

Table 2. Utilized training procedure parameters and functions

Model	Iterations	Batch	Loss	Optimizer	Learning Rate
Proposed Multimodel	50	16	Categorical Cross entropy	Adam	$1e^{-4}$

4.2 Performance of the proposed multimodal classifier

Results of model training. The training of the proposed multimodal with TL was evaluated using a variety of performance metrics, including precision, recall, and specificity. These metrics provide valuable insight into the accuracy of the classifier in diagnosing cutaneous diseases. In addition, learning curves were used to monitor the model's performance during the training process. The learning curves show that as the number of training epochs increased, the model's performance steadily improved. This observation suggests that the proposed method can effectively utilize the additional information from multimodal learning, resulting in improved classification performance.

During the training phase, the loss trajectory was also analyzed, revealing a gradual decrease over time. The convergence of the accuracy and loss curves indicates the successful integration of multimodal information, resulting in a robust classifier for classification. Figure 3 illustrates the accuracy and loss curves during the training phase, providing a visual representation of the model's performance and its capacity to learn from multimodal data.

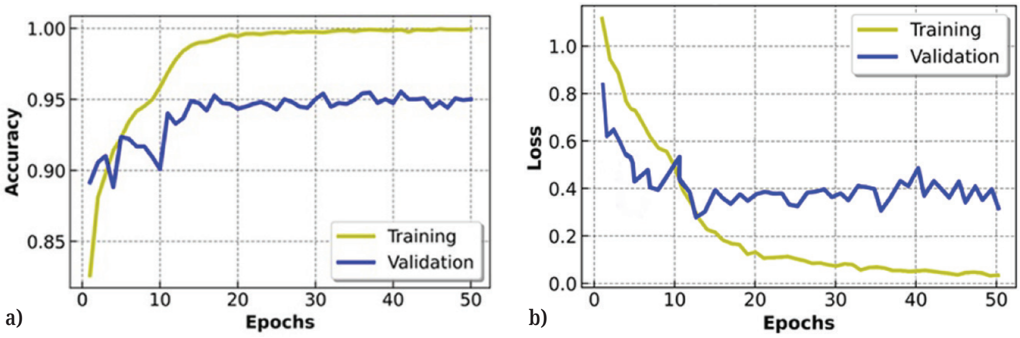

Fig. 3. Classifier performance of the multimodal DNN with TL (a) The model's accuracy variation; (b) The model's loss variation

The results obtained clearly demonstrate that the proposed method is effective in improving the classification of cutaneous diseases. By integrating both image data and patient-specific clinical data using a multimodal architecture, the model's effectiveness in diagnosing various skin conditions was significantly enhanced. The consistent improvement in accuracy and decreasing loss over training epochs further validate the reliability and effectiveness of the proposed method for classifying skin diseases.

Results of model testing. A separate set of image and clinical inputs was utilized as test data to assess the effectiveness of the proposed multimodal DNN with transfer learning. The model's classification accuracy was assessed by creating a confusion matrix for each class of cutaneous disease. Figure 4 illustrates the model's performance for each skin disease class using a multi-class classification confusion matrix.

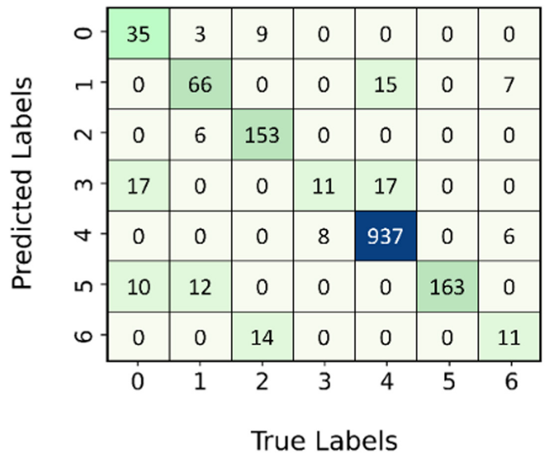

Fig. 4. Confusion matrix for multi-class skin disease classification

The matrix indicates that the model performed well in identifying the majority of classes, with a high proportion of TP and low percentages of FP and FN. However, there were several classes that exhibited misclassifications, which could be investigated to improve model performance. In addition to the confusion matrix, Table 3 presents a quantitative assessment of the classification performance across multiple tasks.

Table 3. Performance metrics of multi-modal DNN with transfer learning for skin disease classification

Class	Acc.	Prec.	Sens.	Spec.
Actinic keratoses	97.23%	58.45%	76.47%	99.16%
Basal cell carcinoma	98.16%	78.86%	73.00%	97.53%
Benign keratosis-like lesions	97.09%	83.93%	97.23%	96.31%
Dermatofibroma	98.23%	54.89%	27.44%	98.46%
Melanoma	97.97%	97.75%	97.55%	95.17%
Melanocytic nevi	97.55%	100.00%	89.11%	100.00%
Vascular lesions	98.62%	67.83%	51.00%	98.13%
Avg.	98.66%	77.25%	79.54%	97.25%

4.3 Analysis of experimental findings

The experimental results of this study illuminate the performance and effectiveness of the proposed multimodal DNN with transfer learning for skin disease classification. The analysis focuses on the accuracy, precision, recall, and specificity of the model, as well as the overall performance metrics obtained during the training and assessment phases.

Training evaluation. The performance of the proposed multi-modal DNN with transfer learning improved as the number of training epochs increased during the training phase. The model effectively utilized the additional data provided by the multimodal method, as demonstrated by the learning curves. The consistent decrease in the loss curve indicates that the model learned and generalized effectively during training. The convergence of both curves during training indicates that the proposed technique efficiently leveraged multi-modality information from skin lesion visual representations and patient-specific clinical data. The use of multi-modality information enabled our model to learn and utilize valuable features from both sources, resulting in a robust classifier for accurate categorization of cutaneous illnesses. The training results indicate that our model outperforms standard techniques in categorizing cutaneous illnesses. By integrating and learning from various data modalities, our method overcomes the limitations of single-modal classifiers and enhances the accuracy of illness classification. This improvement in accuracy is essential for precise diagnosis and early intervention, leading to better patient outcomes and disease management. The data strongly suggest that the proposed approach is capable of classifying complex skin diseases. Our technique captures subtle skin condition patterns and fluctuations by integrating data from multiple sources. It provides doctors with a valuable tool to make accurate and rapid diagnoses, thereby enhancing dermatological patient care and treatment.

Testing evaluation. We assessed our model using different sets of image data to fully comprehend its ability to classify skin diseases. The confusion matrix describes the model's predictions for each skin disease class. The findings revealed that our model accurately identified most skin disease classifications with a high frequency of true positive predictions. The model also had few false positive and false negative predictions, demonstrating its ability to accurately identify healthy samples and minimize incorrect diagnoses. As with any classification methodology, some classes were misclassified. These occurrences highlight opportunities for optimization. By conducting a thorough study of these misclassifications, we may identify issues and trends that require further investigation. This misclassification analysis is essential for refining the model and enhancing its performance across all skin conditions. The confusion matrix showed that our strategy for identifying cutaneous disorders worked well. It demonstrated its ability to accurately predict a range of skin issues and pinpointed areas for enhancement to improve its precision and diagnostic capabilities. Continuous progress is essential to developing a reliable and efficient tool for diagnosing and managing skin diseases.

Performance measurement. The performance metrics for the proposed multimodal DNN with transfer learning are presented in Tables 2 and 3. These metrics include accuracy, precision, recall, and specificity, which allow for a quantitative evaluation of the model's classification performance. The numerical values in the tables offer specific performance measures for each category of skin disease, enabling a comprehensive analysis of the model's performance across disease categories.

The analysis of experimental results demonstrates that the proposed multimodal DNN with transfer learning is effective for classifying cutaneous diseases. The model's ability to use both image and clinical data and to simultaneously perform multiple tasks contributes to its improved accuracy and robustness in diagnosing skin diseases.

4.4 Comparison with traditional AI methods

In this section, we present a comprehensive comparative analysis of our proposed multimodal DNN classifier with traditional AI methods. Our main focus is on evaluating the classifier's performance in categorizing cutaneous disorders using various data modalities, thus emphasizing the strengths and benefits of our approach. The proposed classifier is designed to utilize multiple information sources, combining visual representations of skin lesions with patient-specific features. This unique approach enables the classifier to multitask by categorizing illnesses using both image-based and structured data inputs, ultimately leading to a more comprehensive and accurate diagnosis. To rigorously evaluate the effectiveness of our methodology, we conducted experiments using the ISIC 2018 dataset, which comprises imaging and clinical data for individuals with seven distinct skin disorders. The results obtained from these experiments are compelling, with our proposed model achieving an impressive accuracy rate of 98.66% for image categorization. These results demonstrate the clear superiority of our proposed methodology over single-modal classifiers. Despite the promising results, it is essential to acknowledge the limitations of our approach. Our methodology is currently only applicable to the specific dataset of skin illnesses used in this study. Moreover, the process of training the model requires a significant amount of labeled data, which can be difficult and time-consuming to obtain in real-world situations. Therefore, we acknowledge the necessity for further research endeavors to address these limitations in order to make the model more flexible and applicable to a wider range of skin-related concerns. To enable a comprehensive comparative analysis, we present Table 4, which offers a detailed comparison of our multimodal DNN classifier with transfer learning with other approaches documented in the existing literature. This comparative assessment reveals that our system consistently outperforms alternative methods, highlighting its potential as a practical and effective tool for categorizing dermatological conditions. Table 4 provides a comprehensive comparison of our multimodal DNN classifier with transfer learning with other approaches in the literature. The comparison demonstrates that our system outperforms the others, indicating its potential as a reliable and practical tool for categorizing dermatological illnesses.

Table 4. A comparison of the proposed method's results to those of recent published studies

Ref.	Datasets	Number of Class	Models	Result
[34]	ISIC 2018 Dataset	7	RegNetY-3.2G-Drop	Acc = 85.8%
[35]	ISIC 2019 Dataset	8	SSD-KD	Acc = 84.6%
[36]	ISIC 2018 Dataset	2	2-HDCNN	Acc = 92.15%
[37]		2	CLCM-net	Acc = 94.42%
[38]		7	E2EDT	Acc = 87%
Proposed Model		7	M_DNN_TL	Acc = 98.66%

5 CONCLUSION AND PERSPECTIVES

In this study, we proposed a new approach for classifying skin diseases using a multimodal deep neural network. The inclusion of visual representations of skin lesions and patient-specific clinical data enabled us to create a strong classifier with remarkable precision. The experimental results demonstrated the effectiveness of the proposed model, achieving an accuracy of 98.66% for image classification, surpassing the performance of traditional single-modal classifiers. Nevertheless, despite the encouraging results, our method has some limitations.

Numerous opportunities for future research exist in the classification of skin diseases using multimodal and multitask deep neural network approaches. First, we aim to incorporate genetic data into our model, which could offer further insight into the fundamental mechanisms of skin diseases and enhance classification accuracy. Second, exploring the potential of our model to be applied to medical conditions other than skin disorders could be a promising strategy for expanding the scope and impact of this research.

6 REFERENCES

[1] A. Wells, S. Patel, J. B. Lee, and K. Motaparthi, "Artificial intelligence in dermatopathology: Diagnosis, education, and research," *J. Cutan. Pathol.*, vol. 48, no. 8, pp. 1061–1068, 2021. https://doi.org/10.1111/cup.13954

[2] E. W. M. Verhoeven *et al.*, "Prevalence of physical symptoms of itch, pain and fatigue in patients with skin diseases in general practice," *Br. J. Dermatol.*, vol. 156, no. 6, pp. 1346–1349, 2007. https://doi.org/10.1111/j.1365-2133.2007.07916.x

[3] A. Moreira *et al.*, "Skin symptoms as diagnostic clue for autoinflammatory diseases," *An. Bras. Dermatol.*, vol. 92, no. 1, pp. 72–80, 2017. https://doi.org/10.1590/abd1806-4841.20175208

[4] F. Suryani, I. Muhimmah, and S. Kusumadewi, "Preferred model of dialog style in expert system of physical examination of skin disease," in *2015 International Conference on Science in Information Technology (ICSITech)*, Yogyakarta: IEEE, 2015, pp. 247–252. https://doi.org/10.1109/ICSITech.2015.7407812

[5] O. M. Al-hazaimeh, A. Abu-Ein, N. Tahat, M. Al-Smadi, and M. Al-Nawashi, "Combining artificial intelligence and image processing for diagnosing diabetic retinopathy in retinal fundus images," *Int. J. Online Biomed. Eng.*, vol. 18, no. 13, pp. 131–151, 2022. https://doi.org/10.3991/ijoe.v18i13.33985

[6] A. Taleb, C. Lippert, T. Klein, and M. Nabi, "Multimodal self-supervised learning for medical image analysis," in *Information Processing in Medical Imaging*, A. Feragen, S. Sommer, J. Schnabel, and M. Nielsen, Eds., in Lecture Notes in Computer Science, vol. 12729. Cham: Springer International Publishing, 2021, pp. 661–673. https://doi.org/10.1007/978-3-030-78191-051

[7] O. E. Gannour, S. Hamida, S. Saleh, Y. Lamalem, B. Cherradi, and A. Raihani, "COVID-19 detection on X-ray images using a combining mechanism of pre-trained CNNs," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 6, 2022. https://doi.org/10.14569/IJACSA.2022.0130668

[8] Y. Lamalem, S. Hamida, Y. Tazouti, O. E. Gannour, K. Housni, and B. Cherradi, "Evaluating multi-state systems reliability with a new improved method," *Bull. Electr. Eng. Inform.*, vol. 11, no. 3, pp. 1568–1576, 2022. https://doi.org/10.11591/eei.v11i3.3509

[9] Y. Lamalem, S. Hamida, K. Housni, A. Ouhmida, and B. Cherradi, "Evaluating systems reliability with a new method based on node cutset," in *2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Meknes, Morocco: IEEE, 2022, pp. 1–4. https://doi.org/10.1109/IRASET52964.2022.9737956

[10] Y. Nie, P. Sommella, M. Carratù, M. O'Nils, and J. Lundgren, "A deep CNN transformer hybrid model for skin lesion classification of dermoscopic images using focal loss," *Diagnostics*, vol. 13, no. 1, p. 72, 2022. https://doi.org/10.3390/diagnostics13010072

[11] 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.*, 2022. https://doi.org/10.1007/s11042-022-14105-2

[12] 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*, 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. https://doi.org/10.1007/978-3-030-36778-7_41

[13] M. M. Ahsan, S. A. Luna, and Z. Siddique, "Machine-learning-based disease diagnosis: A comprehensive review," *Healthcare*, vol. 10, no. 3, p. 541, 2022. https://doi.org/10.3390/healthcare10030541

[14] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Sci. Data*, vol. 5, no. 1, p. 180161, 2018. https://doi.org/10.1038/sdata.2018.161

[15] M. A. Mahjoubi, S. Hamida, O. E. Gannour, B. Cherradi, A. E. Abbassi, and A. Raihani, "Improved multiclass brain tumor detection using convolutional neural networks and magnetic resonance imaging," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 3, 2023. https://doi.org/10.14569/IJACSA.2023.0140346

[16] L. Wei, K. Ding, and H. Hu, "Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network," *IEEE Access*, vol. 8, pp. 99633–99647, 2020. https://doi.org/10.1109/ACCESS.2020.2997710

[17] B. Ahmad, M. Usama, C.-M. Huang, K. Hwang, M. S. Hossain, and G. Muhammad, "Discriminative feature learning for skin disease classification using deep convolutional neural network," *IEEE Access*, vol. 8, pp. 39025–39033, 2020. https://doi.org/10.1109/ACCESS.2020.2975198

[18] F. Afza, M. Sharif, M. Mittal, M. A. Khan, and D. Jude Hemanth, "A hierarchical three-step superpixels and deep learning framework for skin lesion classification," *Methods*, vol. 202, pp. 88–102, 2022. https://doi.org/10.1016/j.ymeth.2021.02.013

[19] T.-C. Pham, A. Doucet, C.-M. Luong, C.-T. Tran, and V.-D. Hoang, "Improving skin-disease classification based on customized loss function combined with balanced mini-batch logic and real-time image augmentation," *IEEE Access*, vol. 8, pp. 150725–150737, 2020. https://doi.org/10.1109/ACCESS.2020.3016653

[20] S. Benyahia, B. Meftah, and O. Lézoray, "Multi-features extraction based on deep learning for skin lesion classification," *Tissue Cell*, vol. 74, p. 101701, 2022. https://doi.org/10.1016/j.tice.2021.101701

[21] F. Afza, M. Sharif, M. A. Khan, U. Tariq, H.-S. Yong, and J. Cha, "Multiclass skin lesion classification using hybrid deep features selection and extreme learning machine," *Sensors*, vol. 22, nos. 3, p. 799, 2022. https://doi.org/10.3390/s22030799

[22] P. Cerda, G. Varoquaux, and B. Kégl, "Similarity encoding for learning with dirty categorical variables," *Mach. Learn.*, vol. 107, nos. 8–10, pp. 1477–1494, 2018. https://doi.org/10.1007/s10994-018-5724-2

[23] A. S. Ashour *et al.*, "Ensemble-based bag of features for automated classification of normal and COVID-19 CXR images," *Biomed. Signal Process. Control*, vol. 68, p. 102656, 2021. https://doi.org/10.1016/j.bspc.2021.102656

[24] T. Diwan, R. Shukla, E. Ghuse, and J. V. Tembhurne, "Model hybridization & learning rate annealing for skin cancer detection," *Multimed. Tools Appl.*, vol. 82, no. 2, pp. 2369–2392, 2023. https://doi.org/10.1007/s11042-022-12633-5

[25] N. C. F. Codella *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Washington, DC: IEEE, 2018, pp. 168–172. https://doi.org/10.1109/ISBI.2018.8363547

[26] H. Alaskar, A. Hussain, B. Almaslukh, T. Vaiyapuri, Z. Sbai, and A. K. Dubey, "Deep learning approaches for automatic localization in medical images," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–17, 2022. https://doi.org/10.1155/2022/6347307

[27] J. Wang, Q. Liu, H. Xie, Z. Yang, and H. Zhou, "Boosted EfficientNet: Detection of lymph node metastases in breast cancer using convolutional neural networks," *Cancers*, vol. 13, no. 4, p. 661, 2021. https://doi.org/10.3390/cancers13040661

[28] Y. Ye *et al.*, "An improved EfficientNetV2 model based on visual attention mechanism: Application to identification of cassava disease," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–16, 2022. https://doi.org/10.1155/2022/1569911

[29] T. M. Alam *et al.*, "An efficient deep learning-based skin cancer classifier for an imbalanced dataset," *Diagnostics*, vol. 12, no. 9, p. 2115, 2022. https://doi.org/10.3390/diagnostics12092115

[30] R. Carvalho, J. Pedrosa, and T. Nedelcu, "Multimodal multi-tasking for skin lesion classification using deep neural networks," in *Advances in Visual Computing*, G. Bebis, V. Athitsos, T. Yan, M. Lau, F. Li, C. Shi, X. Yuan, C. Mousas, and G. Bruder, Eds., in Lecture Notes in Computer Science, vol. 13017. Cham: Springer International Publishing, 2021, pp. 27–38. https://doi.org/10.1007/978-3-030-90439-5_3

[31] G. Cai, Y. Zhu, Y. Wu, X. Jiang, J. Ye, and D. Yang, "A multimodal transformer to fuse images and metadata for skin disease classification," *Vis. Comput.*, vol. 39, no. 7, pp. 2781–2793, 2023. https://doi.org/10.1007/s00371-022-02492-4

[32] A. Ouahab, "Multimodal convolutional neural networks for detection of Covid-19 using chest x-ray and CT images," *Opt. Mem. Neural Netw.*, vol. 30, no. 4, pp. 276–283, 2021. https://doi.org/10.3103/S1060992X21040044

[33] H. Dalianis, "Evaluation metrics and evaluation," in *Clinical Text Mining*, Cham: Springer International Publishing, 2018, pp. 45–53. https://doi.org/10.1007/978-3-319-78503-5_6

[34] P. Yao *et al.*, "Single model deep learning on imbalanced small datasets for skin lesion classification," *IEEE Trans. Med. Imaging*, vol. 41, no. 5, pp. 1242–1254, 2022. https://doi.org/10.1109/TMI.2021.3136682

[35] Y. Wang, Y. Wang, J. Cai, T. K. Lee, C. Miao, and Z. J. Wang, "SSD-KD: A self-supervised diverse knowledge distillation method for lightweight skin lesion classification using dermoscopic images," *Med. Image Anal.*, vol. 84, p. 102693, 2023. https://doi.org/10.1016/j.media.2022.102693

[36] JaneY. Nancy, S. Charanya, M. Amsaprabhaa, P. Jayashanker, and H. K. Nehemiah, "2-HDCNN: A two-tier hybrid dual convolution neural network feature fusion approach for diagnosing malignant melanoma," *Comput. Biol. Med.*, vol. 152, p. 106333, 2023. https://doi.org/10.1016/j.compbiomed.2022.106333

[37] S. Gopikha and M. Balamurugan, "Regularised layerwise weight norm based skin lesion features extraction and classification," *Comput. Syst. Sci. Eng.*, vol. 44, no. 3, pp. 2727–2742, 2023. https://doi.org/10.32604/csse.2023.028609

[38] A. C. Foahom Gouabou, R. Iguernaissi, J.-L. Damoiseaux, A. Moudafi, and D. Merad, "End-to-end decoupled training: A robust deep learning method for long-tailed classification of dermoscopic images for skin lesion classification," *Electronics*, vol. 11, no. 20, p. 3275, 2022. https://doi.org/10.3390/electronics11203275

7 AUTHORS

Soufiane Hamida a researcher from Rabat, Morocco. His did his Ph.D. on Machine Learning Methodologies for Pattern Recognition. In 2017, he got his Master's degree in educational technology from the Higher Normal School of Tetouan, Abdelmalek Essaadi University. He has actively participated in a number of study projects in the areas of recognizing handwriting characters and words and improving machine learning models (E-mail: hamida.93s@gmail.com).

Driss Lamrani earned an engineering degree in mechatronics in 2013 from Fes' Faculty of Science and Technology. Currently, he is a senior manager in the research and engineering branch of the Moroccan vocational training and employment office. He is a Ph.D. researcher at the EEIS Laboratory at ENSET Mohammadia, University Hassan II of Casablanca (UH2C). He does research on artificial intelligence and deep learning algorithms to medical picture segmentation and recognition (E-mail: lamranidriss2013@gmail.com).

Mohammed Amine Bouqentar earned an engineering degree in the Biomedical field in 2019 from National Higher School of Arts and Crafts ENSAM Rabat. He is a PhD student in the Research Laboratory on Electrical Engineering and Intelligent Systems at ENSET Mohammadia, Hassan II University of Casablanca (UH2C). He does research on the heart diseases classification and prediction using artificial intelligence and deep learning algorithms (E-mail: m.amine.bouqentar@gmail.com).

Oussama El Gannour earned a Master's Specialized Degree in Internet of Things and Mobile Services from Rabat's National Higher School of Computer Science and Systems Analysis, Mohammed V University. He is a PhD student in the Research Laboratory on Electrical Engineering and Intelligent Systems at Hassan II University of Casablanca in 2019. His study focuses on integrating the Internet of Things and artificial intelligence with medicine. He has published research on screening patients for heart disease and COVID-19 using AI and IoT (E-mail: oussama.elgannour@gmail.com).

Bouchaib Cherradi obtained a Doctor of Philosophy degree in Electronics and Image Processing from the Faculty of Science and Technology at Mohammedia. He obtained a Bachelor of Science degree in Electronics in the year 1990 and a Master of Science degree in Applied Electronics in 1994 from the ENSET Institute of Mohammedia, Morocco. In 2004, he obtained the DESA diploma in Instrumentation of Measure and Control (IMC) from Chouaib Doukkali University located in El Jadida. His areas of focus include the use of massive parallel architectures, cluster analysis, pattern recognition, image processing, fuzzy logic systems, artificial intelligence, machine learning, and deep learning in the study of medical and educational data. Dr. Cherradi currently holds the position of Associate Professor at CRMEF-El Jadida. Furthermore, he holds the position of associate researcher at the Electrical Engineering and Intelligent Systems (EEIS) laboratory in ENSET Mohammedia, which is affiliated with Hassan II University of Casablanca (UH2C). Additionally, he is associated with the LaROSERI Laboratory while on leave from the Faculty of Science at El Jadida, Chouaib Doukali University in Morocco (E-mail: bouchaib.cherradi@enset-media.ac.ma).