IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 2, April 2025, pp. 1484~1495
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i2.pp1484-1495

Adaptive kernel integration in visual geometry group 16 for enhanced classification of diabetic retinopathy stages in retinal images

Mustafa Hiri[1], Nabil Ourdani[1], Mohamed Chrayah[2], Abeer Alsadoon[3,4], Noura Aknin[1]
[1]Faculty of Sciences, Abdelmalek Essaadi University, TIMS Laboratory, Tetuan, Morocco
[2]ENSATE, Abdelmalek Essaadi University, TIMS Laboratory, Tetuan, Morocco
[3]School of Computer Data and Mathematical Sciences, Western Sydney University (WSU), Sydney, Australia
[4]Asia Pacific International College (APIC), Sydney, Australia

Article Info

Article history:

Received May 12, 2024
Revised Oct 26, 2024
Accepted Nov 14, 2024

Keywords:

Adaptive kernel techniques
Deep learning
Diabetic retinopathy
Image classification
Medical imaging
Visual geometry group 16

ABSTRACT

Diabetic retinopathy (DR) is a major cause of vision impairment globally, with early detection remaining a significant challenge. The limitations of current diagnostic methods, particularly in identifying early-stage DR, highlight a pressing need for more accurate diagnostic technologies. In response, our research introduces an innovative model that enhances the visual geometry group 16 (VGG16) architecture with adaptive kernel techniques. Traditionally, the VGG16 model deploys consistent kernel sizes throughout its convolutional layers. In this study, multiple convolutional branches with varying kernel sizes (3×3, 5×5, and 7×7) were seamlessly integrated after the 'block5_conv1' layer of VGG16. These branches were adaptively merged using a softmax-weighted combination, enabling the model to automatically prioritize kernel sizes based on the image's intricate features. To combat the challenge of imbalanced datasets, the synthetic minority over-sampling technique (SMOTE) was employed before training, harmonizing the distribution of the five DR stages. Our results are promising, showing a training accuracy above 94.17% and a validation accuracy over 90.24%, our model significantly outperforms traditional methods. This study represents a significant stride in applying adaptive kernels to deep learning for precise medical imaging tasks. The model's accuracy in classifying DR stages highlights its potential as a valuable diagnostic tool, paving the way for future enhancements in DR detection and management.

Corresponding Author:

Mustafa Hiri
Faculty of Sciences, Abdelmalek Essaadi University, TIMS Laboratory
Tetuan, Morocco
Email: mustafa.hiri@gmail.com

1. INTRODUCTION

Diabetic retinopathy (DR), a microvascular complication arising from diabetes mellitus, is the leading cause of visual impairment and blindness worldwide [1]. The global prevalence of DR among diabetes patients is significant, ranging from 20% to 30% in the United States alone [2]. The progression of DR from modest non-proliferative damage to proliferative DR, which can cause severe vision loss, poses important diagnostic issues [3]. Early identification and appropriate staging of DR are crucial for avoiding its negative consequences on vision and quality of life [4]. The increasing prevalence of diabetes, projected to impact 642 million individuals by 2040 [5], underscores the necessity for effective DR screening methods.

Fundus photography has been the prevailing method for detecting DR due to its ability to provide immediate visual screening [6]. However, the interpretation of these images requires specialized expertise, and variability in image analysis can lead to inconsistent diagnoses [2]. Additionally, a significant proportion of individuals with diabetes, particularly those residing in disadvantaged regions, lack access to screening initiatives. Consequently, there is an elevated likelihood that DR may remain undetected [7].

Advancements in artificial intelligence (AI), particularly in the field of deep learning, have offered promising solutions to these challenges [8]–[11]. Researchers have made significant progress in the automatic classification of fundus images by utilizing advanced approaches like convolutional neural networks (CNNs) [4]. Despite these advancements, existing techniques face limitations in accurately classifying the various stages of DR due to factors such as image heterogeneity, subtle differences in early-stage features, and imbalanced datasets [2].

Previous research efforts have employed preprocessing methods, like measuring image entropy, to improve image heterogeneity, addressing the variability in retinal images from existing datasets [12]. These methods enhance CNN performance in distinguishing between the various stages of DR [2]. However, challenges remain in effectively capturing multi-scale features and adapting to the diverse characteristics of retinal images.

To bridge these gaps, this study aims to improve existing techniques by integrating an adaptive kernel approach into the VGG16 framework, enhancing the classification of images across different stages of DR. Unlike the traditional VGG16 model that deploys consistent kernel sizes throughout its convolutional layers, the proposed model introduces multiple convolutional branches with varying kernel sizes (3×3, 5×5, and 7×7) after the 'block5_conv1' layer. These branches are adaptively merged using a SoftMax-weighted combination, allowing the model to dynamically prioritize kernel sizes based on the intricate features of the images. Furthermore, to address the challenge of imbalanced datasets, the synthetic minority over-sampling technique (SMOTE) is employed before training. This method harmonizes the distribution of the five DR stages, enhancing the model's predictive performance across all stages.

The primary contributions of this study are:

- Adaptive kernel integration: introducing a novel adaptive kernel approach within the VGG16 architecture to improve feature extraction and classification accuracy for different DR stages.
- Dataset balancing with SMOTE: utilizing SMOTE to address dataset imbalance, ensuring equitable representation of all DR stages during model training.
- Enhanced model performance: demonstrating significant improvements over traditional methods, achieving a training accuracy above 94.17% and a validation accuracy over 90.24%.

This study not only leverages the existing advantages of the VGG16 model but also introduces innovative modifications to enhance its applicability in medical imaging classification tasks. By improving the accuracy of DR stage classification, the proposed model has the potential to serve as a valuable diagnostic tool, facilitating early detection and management of DR, ultimately helping to protect the vision of millions of individuals at risk.

The remainder of this article is structured as follows: section 2 provides an in-depth examination of related work in DR classification. Section 3 details the architectural enhancements made to the VGG16 model and the methodology employed. Section 4 presents a detailed evaluation of the results. Section 5 discusses the findings within the existing landscape of DR classification techniques, highlighting the model's advancements over current methods. Finally, section 6 explores the implications of this research for future diagnostic practices and the broader field of medical imaging.

2. RELATED WORK

DR is a critical area of medical research due to its position as a leading cause of vision loss among diabetic patients. The advent of deep learning techniques in medical imaging has opened new avenues for early detection and accurate classification of DR, offering the potential for improved patient outcomes. Doshi *et al.* [13] were among the first to employ deep convolutional neural networks (DCNNs) for automating the classification of DR stages. Their groundbreaking work demonstrated the capability of CNNs to handle complex medical imaging tasks, providing accurate assessments of DR severity. However, they faced challenges related to the variability and complexity of retinal images, especially in distinguishing subtle features across different DR stages.

Building on this foundation, Qummar *et al*. [14] explored the effectiveness of ensemble DCNN models for DR stage classification using the Kaggle DR dataset. Their approach significantly enhanced the accuracy of diagnosing various DR stages by effectively encoding a broad range of retinal image features. Despite the improved performance, ensemble models increased computational complexity and required extensive training data, which may not be feasible in all clinical environments.

Kalyani *et al*. [15] investigated the use of capsule networks, an advanced neural architecture that preserves spatial hierarchies between features. Their work demonstrated precise classification of DR stages, highlighting the potential of such innovative architectures in medical diagnostics. However, capsule networks are computationally intensive and may not scale well with large datasets.

Asia *et al*. [16] conducted a comparative analysis of various CNN-based models, including ResNet-101, ResNet-50, and VGGNet-16, for detecting DR. Their study showed that ResNet-101 achieved the highest accuracy, outperforming other network models in DR classification tasks, thereby highlighting CNNs' potential in medical image interpretation and the importance of exploring deeper network architectures for improved sensitivity and classification performance. The performance of CNNs is significantly influenced by the selection of convolutional kernels, which are essential for feature learning [17]. Recent studies have focused on optimizing kernel sizes to minimize redundant or noisy features while effectively capturing crucial patterns. Strategies like adaptive kernel selection and feature map refinement have shown promise in enhancing CNN performance, leading to more efficient feature extraction tailored to specific medical imaging tasks.

Advancements in kernel technologies have introduced dynamic filter modules into CNNs [18]. By incorporating a decoupled dynamic filtering layer alongside a filter-generating network, convolutional filters can adjust dynamically to changing input data. This development represents a substantial step toward creating more adaptable and responsive network architectures, increasing the sensitivity of CNNs to specific characteristics of medical images.

Esquivel *et al*. [19] advanced the concept of adaptability in CNNs by proposing adaptive convolutional kernels whose weights change dynamically based on the input image characteristics. Their method improved memory efficiency and reduced training times and computational requirements, which is particularly beneficial in resource-limited settings. However, they did not fully explore integrating adaptive kernels within established architectures like VGG16 for DR classification.

Maharjan *et al*. [20] expanded the application of deep learning in medical imaging by developing a CNN model using a modified softmax loss function with regularization for brain tumor detection from magnetic resonance imaging (MRI) images. Their approach minimized overfitting risks while providing multi-class classification, improving accuracy and reducing processing time. This work emphasizes the importance of precise mathematical formulations in enhancing medical diagnostic models.

Despite these advancements, several gaps remain in the current literature. A common limitation is the use of fixed kernel sizes throughout convolutional layers, which may not effectively capture the multi-scale features inherent in retinal images. This shortcoming affects the models' ability to accurately classify different DR stages, especially when lesions vary significantly in size and appearance. Al-Antary and Arafa [21] addressed this issue by proposing a multi-scale attention network (MSA-Net) that employs a multi-scale feature pyramid and attention mechanisms to capture retinal structures at various scales. While their approach improved classification performance, it primarily focused on attention mechanisms without fully integrating adaptive kernel techniques into established architectures like VGG16.

Additionally, although advanced architectures like ensemble models and capsule networks have improved classification accuracy, they often introduce greater computational complexity and resource demands. This makes them less practical in resource-constrained clinical environments. Alzubaidi *et al*. [22] discussed these challenges, highlighting that the computational intensity and scalability issues associated with deep learning models can impede their deployment in real-world settings, particularly where hardware resources are limited.

Moreover, DR datasets often suffer from class imbalance, where certain DR stages are underrepresented. This imbalance leads to biased models that perform poorly in minority classes. Addressing dataset imbalance is crucial for developing models that perform reliably across all DR stages. Saini and Susan [23] investigated the impact of class imbalance in DR datasets, conducting extensive comparative analyses using various deep learning models. They found that traditional models often struggle with imbalanced data, resulting in poor classification performance for minority classes. Their study underscores the necessity of employing techniques such as data resampling, appropriate evaluation metrics, and tailored architectures to mitigate the effects of class imbalance.

To bridge these gaps, our research integrates adaptive kernel techniques into the VGG16 architecture to enhance DR stage classification. By incorporating multiple convolutional branches with varying kernel sizes and adaptively merging them using a SoftMax-weighted combination, our model more effectively captures multi-scale features without significantly increasing computational complexity. Additionally, we employ SMOTE before training to address dataset imbalance. This ensures equitable representation of all DR stages, enhancing the model's predictive performance across both majority and minority classes.

3. MATERIALS AND METHODS

3.1. Dataset acquisition and image preprocessing

We focused on Kaggle 2015 dataset [24] due to its diversity and real-world applicability, recognizing its widespread acceptance as a benchmark in DR research. This choice allows for a direct comparison with multiple studies, including state-of-the-art models evaluated on the same dataset. Although exploring multiple datasets could broaden our study, the depth and challenge presented by this dataset offer significant insights and advancements in DR diagnosis, justifying its exclusive use. Our findings, which leverage this single dataset, contribute valuable advances in this field, demonstrating the improved diagnostic accuracy of our model within an established research context. In addressing the composition of the data set and our categorical selection, it is important to emphasize the diversity of the data set and its clinical relevance. The dataset includes 35,126 images, reflecting a broad spectrum of DR stages. This diversity is crucial for developing a model capable of recognizing and classifying the nuanced progression of DR. Each category represents a specific stage of DR, from no DR to mild, moderate, severe non-proliferative DR, and proliferative DR, as clinically recognized. This categorical division aligns with medical standards, facilitating potential clinical application of the model.

The primary goal of the dataset is to facilitate predictions across five categorically distinct stages of DR, reflecting its progression from no apparent condition to advanced stages. These categories are not chosen arbitrarily but are rooted in clinical practice, representing medically recognized stages of DR that are crucial for diagnosis and treatment planning. This systematic classification, detailed in Table 1, is essential for developing a nuanced understanding and detection capability of DR at various stages. The dataset initially presents an imbalanced representation, with class 1 (No DR) constituting approximately 73% of the entries. This imbalance is systematically addressed in the preprocessing phase using the SMOTE technique to ensure a balanced learning environment and improve the predictive performance of the model across all DR phases. By focusing on these five categories, the research aligns with clinical diagnostic criteria, enhancing the model's applicability and potential utility in medical settings.

A meticulous preprocessing routine was implemented to prepare the dataset for deep learning analysis. At first, all images were uniformly adjusted to a fixed resolution of 256×256 pixels. The standardization step was essential to ensure consistency in the input size throughout the dataset. Through the act of resizing the images, we achieved two significant outcomes: enhancing the efficiency of the training process and diminishing the computational requirements on memory. Following the resizing process, we implemented normalization techniques to adjust the pixel intensity values of all images, ensuring they fall within a standardized range of 0 to 1. Normalization is crucial for deep learning models, as it expedites convergence during training and enhances the model's ability to learn effectively from the dataset.

To address the challenge of the imbalanced nature of the original dataset, which had a significant bias towards the 'No DR' stage, we employed SMOTE [25]. SMOTE played a crucial role in artificially expanding the dataset by generating synthetic images for the underrepresented classes, using the similarities in the feature space of existing samples. The synthetic images were created to represent the underrepresented stages of DR accurately and inclusively, thus ensuring a balanced dataset for training purposes. After undergoing the preprocessing phase, the dataset attained a state of balance, with each DR stage being equally represented by 2324 images. The balance was crucial in maintaining an impartial training regimen for the deep learning model, enabling a thorough and fair evaluation of the model's predictive performance across all DR classes.

Table 1. Distribution of DR stages in the dataset

Stages of DR	Descriptions	Number of images
Normal (No DR)	– Without any abnormalities.	25810
Mild non-proliferative diabetic retinopathy (NPDR)	– Presence of microaneurysms only.	2443
Moderate NPDR	– Microaneurysms are present but in smaller amounts as compared to severe NPDR.	5292
Severe NPDR	– Venous beading in two or more regions. – Prominent intraretinal microvascular-abnormality (IRMA) in one or more regions.	873
Proliferative DR	– Vitreous/pre-retinal hemorrhage. – Neovascularization.	708

3.2. State of the art solution: VGG16

The VGG16 model, created by the Visual Geometry Group at the University of Oxford, is a significant milestone in the advancement of deep learning architectures for image classification. Its architecture is distinguished by the use of small 3×3 convolutional kernels throughout, which are combined with 2×2 max-pooling layers, allowing for increased network depth and improved feature extraction capabilities. This model comprises 16 convolutional and fully connected layers that systematically extract

and process a hierarchical variety of features from images, leading to its wide adoption in diverse domains and demonstrating exceptional performance on various image classification benchmarks [26], [27].

Figure 1 shows the VGGNet architecture, emphasizing its profound structure that enables the extraction of complex features, starting from simple edges and textures in the initial layers and progressing to more advanced patterns in the deeper layers. Particularly notable within this family are VGG16 and its extension, VGG19, which includes three additional convolutional layers for a total of 19. VGG19 seeks to improve the model's ability to discern even more complex image features, leveraging the depth of the architecture to achieve refined image classification results [28].

Although VGG16 and VGG19 are effective, they encounter difficulties when dealing with images of different scales and complexities, which are often encountered in medical imaging. In addition, the deep structures of these models require significant computational resources, posing implementation challenges in environments where such resources are limited [28]. In subsection 3.3, we will introduce the adaptations made to the VGG architecture in our work. Our proposed enhancements aim to maintain the depth and comprehensive feature learning capabilities inherent in VGG16, while addressing their limitations regarding scalability and computational efficiency. The purpose of these modifications is to guarantee that the models not only perform exceptionally well in academic environments but are also suitable for real-world applications, particularly in challenging image classification tasks like diagnosing medical conditions.

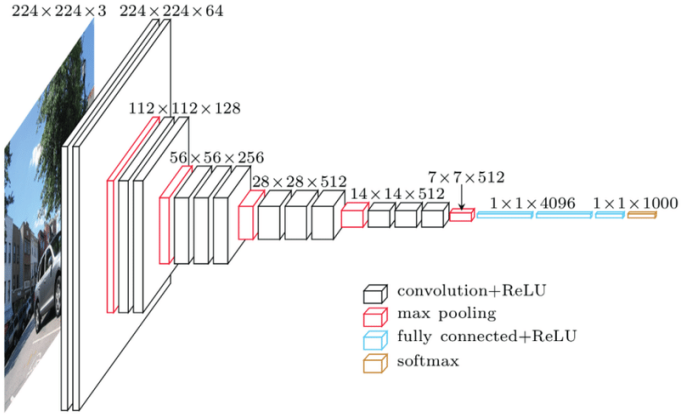

Figure 1. Architecture of VGG [29]

3.3. Model architecture and deep integration of adaptive kernels

The structure of our model is an advancement of the traditional VGG16 framework [26], which has been widely employed in image classification because of its deep layers and strong ability to extract features. The architecture of our model preserves the fundamental structure of VGG16, consisting of a sequence of convolutional layers that commonly employ 3×3 kernels. These layers are then followed by max-pooling layers, which serve to decrease dimensionality and capture the most significant features. The basic design has been enhanced to include a new multi-branch configuration that starts after the "block5_conv1" layer, where high-level features have started to develop in the network's hierarchy.

Figure 2 illustrates the customized configuration with multiple branches. Every branch is furnished with convolutional layers that have different kernel sizes—3×3, 5×5, and 7×7. The reason for choosing these particular sizes is to examine the retinal images at multiple scales concurrently. The smallest 3×3 kernel is highly effective in capturing complex details such as subtle edges and textures, which are crucial in detecting early-stage DR. The 5×5 kernels offer an intermediate perspective, striking a balance between detail and overall understanding, which is crucial for detecting characteristics such as exudates and hemorrhages. The largest 7×7 kernels are intended to encompass larger regions of the image, enabling the detection of broader patterns and structural alterations in the retina that are indicative of more advanced stages of DR.

The convolutional layers are arranged in these branches, enabling the network to construct a sophisticated hierarchy of features at various scales. The output of each layer is used as input for the next layer, gradually enhancing the level of abstraction of the visual information. After performing convolution, batch normalization is applied to normalize the activations, which helps speed up the training process and improves the overall stability of the model. Non-linear activation functions, such as rectified linear units (ReLUs), are used to introduce non-linearity into the network. This allows the network to learn and represent more intricate patterns.

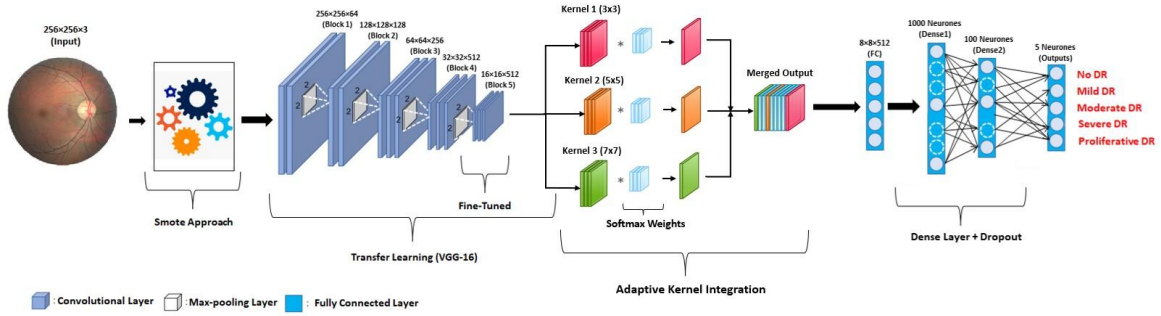

Figure 2. Augmented VGG16 model with multi-branch convolutional layers and adaptive kernel integration

3.4. Adaptive kernel integration technique

After processing feature maps through their respective branches with varying kernel sizes, we employ an adaptive kernel integration technique [19] to merge the multi-scale features into a unified representation. This technique allows the network to dynamically prioritize features from different scales based on their relevance to the input image. Which is crucial for accurately classifying DR stages with varying lesion sizes.

Firstly, the output feature maps from each branch—corresponding to kernel sizes 3×3, 5×5, and 7×7—are flattened into one-dimensional vectors, denoted as $F_{3\times3}$, $F_{5\times5}$, and $F_{7\times7}$, respectively. We introduce trainable parameters w_i for each branch, where $i \in \{1, 2, 3\}$ corresponds to the three kernel sizes. A softmax function is applied to these parameters to compute adaptive weights α_i, ensuring that the weights are positive and sum to one. The adaptive weights are calculated as follows:

$$\alpha_i = \frac{e^{w_i}}{\sum_{j=1}^{3} e^{w_j}}, \ for\ i\ =\ 1, 2, 3. \tag{1}$$

These adaptive weights enable the network to emphasize or de-emphasize features from each branch during training, based on their relevance to the specific input image. The flattened feature vectors are then combined using the adaptive weights through a weighted sum.

$$F_{integrated}\ =\ \alpha_1\, F_{3\times3} + \alpha_2\, F_{5\times5} + \alpha_3\, F_{7\times7} \tag{2}$$

This integration effectively merges the multi-scale features into a single comprehensive feature vector $F_{integrated}$, capturing both fine-grained details and broader structural information pertinent to the stages of DR.

Subsequently, the integrated feature vector is passed through fully connected layers for classification. It first enters a dense layer with 256 neurons utilizing the ReLU activation function. To prevent overfitting, a dropout layer with a rate of 0.45 is applied. This is followed by a second dense layer comprising 128 neurons, also with ReLU activation. The final output layer consists of a dense layer with 5 neurons corresponding to the five DR classes, employing a softmax activation function to produce class probabilities.

By employing this adaptive kernel integration technique, our model dynamically adjusts the importance of features from different scales for each input image. This adaptability is essential in medical imaging, where lesions can vary significantly in size and appearance. The integration method enables the network to prioritize the most relevant features, enhancing its ability to accurately classify the stages of DR. This approach not only improves classification performance but also contributes to a more nuanced understanding of the retinal images, which is critical for effective diagnosis and treatment planning.

3.5. Training procedure

The architecture of the model was built using TensorFlow and Keras. We utilized the VGG16 network as the base of our model, which was already equipped with pre-trained weights derived from training on the ImageNet dataset. The decision to freeze the network's layers up to 'block5_conv1' was a strategic decision aimed at preserving pre-existing powerful feature extraction capabilities. This approach ensures that the model retains the ability to recognize a wide range of image features, which is crucial given the diverse nature of retinal images in the diagnosis of DR. By freezing these layers, we harness the generalizability of VGG16's lower-level feature detectors, which have been optimized for a wide variety of image types, allowing our model to effectively adapt to the specific nuances of retinal imaging. This method not only enhances the model's accuracy by using proven feature extraction methods, but also simplifies the training process, as only the layers beyond 'block5_conv1' need to be trained from scratch, reducing both the computational overhead and the complexity of the optimization process.

Our adaptation involved adding convolutional branches with varying kernel sizes (3×3, 5×5, 7×7) after 'block5_conv1' of VGG16, aimed at capturing a diverse range of features critical for DR stage classification. Smaller kernels excel in detecting fine details crucial for early DR stages, while larger kernels capture broader pathological features of advanced DR. This design ensures our model adeptly identifies features across all DR stages. Combining branch outputs through a softmax-weighted combination allows dynamic weighting of each branch's significance, optimizing feature integration for accurate DR classification. This method enhances both the versatility and precision of our model, significantly improving its diagnostic efficacy.

After merging these branches, the resulting output was compressed and processed through a sequence of dense layers and dropout layers (with a dropout rate of 0.45) to avoid overfitting. The last layer consisted of a dense layer with softmax activation, specifically designed to generate the probability distribution across the five DR classes. The model was compiled utilizing the stochastic gradient descent (SGD) optimizer, employing an initial learning rate of 5e-5 and a momentum of 0.9. The loss function employed was categorical cross-entropy, which is suitable for multi-class classification tasks.

In order to enhance the efficiency of the training process, we utilized multiple callback functions:

- ModelCheckpoint: Used to store the model weights after each epoch in which there was an enhancement in validation accuracy.
- ReduceLROnPlateau: This function decreases the learning rate when the validation accuracy reaches a plateau, which aids in refining the model.
- The CSVLogger is used to record the training and validation metrics for each epoch, making it easier to analyze the model's performance over time.

The model completed training on the preprocessed dataset for a total of 60 epochs, employing a custom training generator for data input. The model was validated using a separate validation generator.

3.6. Validation and performance metrics

During the development phase of our diagnostic tool for DR, we divided the augmented dataset of 11,620 synthesized fundus images into two main categories in a strategic manner. The training set consisted of a substantial majority of 10,458 images, which effectively facilitated the learning algorithm of our model. The remaining subset, consisting of 1,162 images, was designated as the validation set, serving as a crucial tool for assessing the effectiveness of our diagnostic methodology.

The validation set played a crucial role when evaluating the model's performance using a range of statistical metrics. We have implemented the next set of evaluation criteria:

- Accuracy: This metric provides a measure of the model's overall classification accuracy by determining the proportion of correct results, including both positive and negative outcomes, out of all the cases examined.
- Recall, also referred to as the true positive rate, is a metric that quantifies the model's ability to correctly identify actual positive instances and serves as an indicator of its sensitivity and effectiveness. Ensuring comprehensive case identification is of utmost importance in clinical diagnostics.
- Precision: This metric measures the proportion of correctly predicted positive instances out of all the positive predictions made by the model. A high precision score indicates a lower number of false positives, which is essential for ensuring clinical reliability.
- The F1 score is a composite metric that combines recall and precision. It provides a single measure that represents a balance between the two. The F1 score is particularly useful when dealing with imbalanced class distributions. This highlights the significance of both detecting and accurately identifying positive occurrences.

The metrics that underpin our evaluation strategy are established by (3) through (6). These formulas are fundamental to evaluating classification models and are widely cited in literature, including but not limited to [30], [31] for their application in medical imaging and machine learning contexts. Collectively, they provide a thorough evaluation of our model's performance, encompassing not only its accuracy but also its ability to consistently detect true positive cases while maintaining a balanced consideration of both sensitivity and precision.

$$Accuracy = \frac{(True\ Positive + True\ Negative)}{(True\ Positive + False\ Positive + True\ Negative + False\ Negative)} \quad (3)$$

$$Recall = \frac{True\ Positive}{(True\ Positive + False\ Negative)} \quad (4)$$

$$Precision = \frac{True\ Positive}{(True\ Positive + False\ Positive)} \quad (5)$$

$$F1\ score = 2 \times \frac{Recall \times Precission}{(Recall + Precission)} \quad (6)$$

By conducting this comprehensive validation procedure, our goal is to confirm the accuracy of the model and provide evidence of its strength in correctly categorizing the different stages of DR. This will further support its potential for reliable use in clinical settings.

4. RESULTS AND DISCUSSION

Despite advances in deep learning applications for DR classification, accurately identifying all stages of the disease—particularly intermediate stages like Moderate NPDR—remains challenging. Previous models, such as traditional CNN architectures, often struggled with several key limitations: they were unable to effectively capture multi-scale features within retinal images, leading to poor differentiation of subtle pathological signs in intermediate DR stages. Additionally, class imbalances in datasets were inadequately addressed, resulting in models that disproportionately favored the classification of more dominant stages while underperforming in the detection of rarer, yet clinically critical, stages like Moderate NPDR.

Our research addresses these gaps by integrating adaptive kernel techniques into the VGG16 architecture, enhancing feature extraction across multiple scales and improving classification performance across all DR stages. The adaptive kernels allow the model to effectively capture both fine and coarse details in retinal images, improving detection across all DR stages, particularly the more challenging intermediate stages. Furthermore, to address the issue of class imbalance, we utilized the SMOTE technique during the data preprocessing phase, which ensures a more balanced classification across all stages.

Our augmented VGG16 model, enhanced with adaptive kernels, demonstrated exceptional performance on the Kaggle DR dataset. The classification report in Table 2 highlights the model's effectiveness across various stages of DR, and our approach consistently achieved high metrics across classes. In terms of overall results, the model reached an accuracy of 90%, with both the macro-averaged and weighted-averaged metrics also at 0.89. This strong performance indicates balanced classification results, ensuring that no particular DR stage is favored at the expense of others.

Table 2. Classification report for DR stages

Class	Precision	Recall	F1-score
Normal (No DR)	0.81	0.96	0.88
Mild NPDR	0.86	0.90	0.88
Moderate NPDR	0.90	0.67	0.77
Severe NPDR	0.98	0.97	0.97
Proliferative DR	0.96	0.97	0.97

Notably, the model exhibited outstanding precision and recall in detecting advanced stages of DR. For Severe NPDR (Class 3), it achieved a precision of 0.98 and a recall of 0.97. Similarly, for Proliferative DR (Class 4), the precision and recall were 0.96 and 0.97, respectively. Accurate identification of these critical stages is vital for timely clinical interventions that can significantly improve patient outcomes.

However, the model faced challenges in classifying Moderate NPDR (Class 2), with a recall of 0.67. This lower recall suggests difficulty in identifying all true positive cases in this category, possibly due to the subtle symptoms characteristic of this intermediate stage. The confusion matrix in Figure 3 illustrates that misclassifications in Class 2 often occurred, with instances being incorrectly labeled as 'No DR' or 'Mild NPDR'. This pattern indicates the need for further refinement of the model or additional training data to better distinguish the nuances between early DR stages. To contextualize our model's performance, we compared it with other significant studies in the field, as detailed in Table 3. Our model's accuracy surpasses that of existing methods, demonstrating its superior efficacy in DR classification.

Our model's superior performance can be attributed to the integration of adaptive kernels, which allows effective analysis of complex medical images by capturing features at multiple scales. Compared to traditional methods like the CNN with weight matrix-based approach by Lam *et al.* [32], which achieved 74.1% accuracy, our model shows a significant improvement. Even when compared to more sophisticated architectures, such as the Inception-ResNet-v2 model used by Gangwar and Ravi [33] with an accuracy of 82.18%, our approach outperforms them while utilizing a relatively simpler architecture. This suggests that strategically enhancing the VGG16 model with adaptive kernels is more effective than merely increasing model complexity.

Furthermore, our model surpasses ensemble approaches like that of Qummar *et al.* [14], who combined multiple DCNN models and achieved 80.8% accuracy. Achieving higher accuracy with a single-model architecture highlights the efficiency of our method. Despite these successes, the difficulty in accurately classifying Moderate NPDR indicates an area for future investigation. Possible improvements include integrating a broader range of training data to expose the model to more variations of moderate DR

symptoms or exploring hybrid models that combine the strengths of different architectures to enhance feature extraction capabilities.

Additionally, incorporating clinical data alongside image features could provide a more holistic approach to DR classification. Multimodal models that include patient history, genetic factors, and other relevant clinical indicators may capture a broader spectrum of diagnostic information, potentially improving classification accuracy, especially for challenging stages like Moderate NPDR. In summary, our study contributes significantly to medical image analysis by accurately classifying DR stages using an enhanced VGG16 model with adaptive kernel integration. The model not only achieves high accuracy but also maintains balanced performance across different DR stages, setting a new benchmark in this critical area of ophthalmological research.

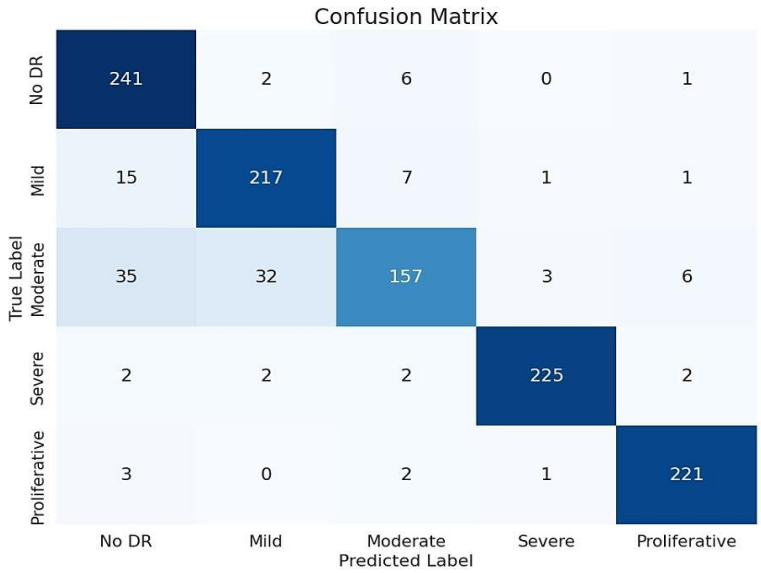

Figure 3. Confusion matrix of DR stage classification by the augmented VGG16 model

Table 3. Comparative analysis of DR classification models across various studies

Study	Proposed solution	Data set	Number of images used	Performance measure	Results (%)
Lam *et al*. [32]	CNN and weight matrix-based method	MESSIDOR-1	36,200	Accuracy	74.1
Gangwar and Ravi [33]	Inception-ResNet-v2 and CNN-based model	MESSIDOR and APTOS 2019	4862	Accuracy	82.18
Nagaraj *et al*. [34]	CNN and VGG16 network-based framework	EyePACS	35,126	Accuracy	73.72
Lin *et al*. [2]	CNN-based architecture for entropy images	EyePACS	33,000	Accuracy	86.10
Qummar *et al*. [14]	Ensemble approach which consists of five different DCNN models that include Inception-V3, Resnet50, Dense-121, Dense-169, and X-ception	Kaggle	35,126	Accuracy	80.8
Oh *et al*. [35]	Residual network with 34-layer (ResNet-34)-based model	Custom-developed at Catholic Kwandong University International St. Mary's Hospital, South Korea	11,734	Accuracy	83.38
Kanungo *et al*. [36]	Inception-V3-based architecture	EyePACS	40,000	Accuracy	88
Khan *et al*. [37]	VGG16, spatial pyramid pooling layer (SPP), and network-in-network (NiN)-based model	EyePACS	88,702	Accuracy	85

5. CONCLUSION AND FUTURE WORK

This study introduced an enhanced VGG16 model with adaptive kernel integration for the classification of DR stages. The model achieved an overall accuracy of 90%, demonstrating balanced performance across different DR stages, which is vital for effective patient diagnosis and treatment. Its strong ability to detect advanced stages of DR highlights its potential as a reliable tool in ophthalmology, contributing to timely clinical interventions and improved patient outcomes. Beyond the immediate results, this research contributes to the broader context by showcasing how integrating adaptive kernels into CNNs can enhance feature extraction in medical image analysis. This approach can be applied to other medical imaging tasks where capturing multi-scale features is crucial, potentially advancing diagnostic methods in various specialties. However, challenges remain in accurately classifying the intermediate stage of Moderate NPDR, as evidenced by the lower recall rate for this category. Addressing this limitation is important not only for DR classification but also for other conditions where early detection is essential. Future work will focus on enhancing the model's sensitivity to subtle features characteristic of this stage. This may involve exploring hybrid architectures, incorporating additional clinical data, or utilizing advanced image processing techniques. Validating the model on larger and more diverse datasets will also be a priority to improve its robustness and applicability across different clinical settings. By refining the model in these ways, we aim to contribute further to the field of medical image analysis and ultimately support better healthcare outcomes for patients with DR.

ACKNOWLEDGEMENTS

The authors sincerely thank their research supervisors and the members of the Tims Laboratory at Abdelmalek Essaadi University in Tetuan, Morocco, for their invaluable insights, mentorship, and encouragement throughout this doctoral investigation. The faculty and administrative staff also deserve recognition for creating an environment conducive to rigorous inquiry and exploration.

REFERENCES

[1] M. Z. Atwany, A. H. Sahyoun, and M. Yaqub, "Deep learning techniques for diabetic retinopathy classification: a survey," *IEEE Access*, vol. 10, pp. 28642–28655, 2022, doi: 10.1109/ACCESS.2022.3157632.

[2] G. M. Lin *et al.*, "Transforming retinal photographs to entropy images in deep learning to improve automated detection for diabetic retinopathy," *Journal of Ophthalmology*, vol. 2018, pp. 1–6, 2018, doi: 10.1155/2018/2159702.

[3] B. Tymchenko, P. Marchenko, and D. Spodarets, "Deep learning approach to diabetic retinopathy detection," in *International Conference on Pattern Recognition Applications and Methods*, SCITEPRESS - Science and Technology Publications, 2020, pp. 501–509, doi: 10.5220/0008970805010509.

[4] S. I. Pao, H. Z. Lin, K. H. Chien, M. C. Tai, J. T. Chen, and G. M. Lin, "Detection of diabetic retinopathy using bichannel convolutional neural network," *Journal of Ophthalmology*, vol. 2020, pp. 1–7, 2020, doi: 10.1155/2020/9139713.

[5] M. W. Nadeem, H. G. Goh, M. Hussain, S. Y. Liew, I. Andonovic, and M. A. Khan, "Deep learning for diabetic retinopathy analysis: a review, research challenges, and future directions," *Sensors*, vol. 22, no. 18, 2022, doi: 10.3390/s22186780.

[6] M. Voets, K. Møllersen, and L. A. Bongo, "Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *PLoS ONE*, vol. 14, no. 6, 2019, doi: 10.1371/journal.pone.0217541.

[7] P. D. Jani, L. Forbes, A. Choudhury, J. S. Preisser, A. J. Viera, and S. Garg, "Evaluation of diabetic retinal screening and factors for ophthalmology referral in a telemedicine network," *JAMA Ophthalmology*, vol. 135, no. 7, pp. 706–714, 2017, doi: 10.1001/jamaophthalmol.2017.1150.

[8] W. Lu, Y. Tong, Y. Yu, Y. Xing, C. Chen, and Y. Shen, "Applications of artificial intelligence in ophthalmology: general overview," *Journal of Ophthalmology*, vol. 2018, pp. 1–15, 2018, doi: 10.1155/2018/5278196.

[9] L. Balyen and T. Peto, "Promising artificial intelligence–machine learning–deep learning algorithms in ophthalmology," *Asia-Pacific Journal of Ophthalmology*, vol. 8, no. 3, pp. 264–272, 2019, doi: 10.22608/APO.2018479.

[10] Y. Fujinami-Yokokawa *et al.*, "Prediction of causative genes in inherited retinal disorders from spectral-domain optical coherence tomography utilizing deep learning techniques," *Journal of Ophthalmology*, vol. 2019, pp. 1–7, 2019, doi: 10.1155/2019/1691064.

[11] E. Vaghefi, S. Hill, H. M. Kersten, and D. Squirrell, "Multimodal retinal image analysis via deep learning for the diagnosis of intermediate dry age-related macular degeneration: a feasibility study," *Journal of Ophthalmology*, vol. 2020, pp. 1–7, 2020, doi: 10.1155/2020/7493419.

[12] W.-J. Huang, C.-H. Yeh, C.-C. Kuo, Y.-C. Cheng, and J.-Y. Lin, "Rating realism assessment for computer generated imagery," in *2017 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, IEEE, 2017, pp. 327–328, doi: 10.1109/ICCE-China.2017.7991128.

[13] D. Doshi, A. Shenoy, D. Sidhpura, and P. Gharpure, "Diabetic retinopathy detection using deep convolutional neural networks," in *International Conference on Computing, Analytics and Security Trends, CAST 2016*, IEEE, 2017, pp. 261–266, doi: 10.1109/CAST.2016.7914977.

[14] S. Qummar *et al.*, "A deep learning ensemble approach for diabetic retinopathy detection," *IEEE Access*, vol. 7, pp. 150530–150539, 2019, doi: 10.1109/ACCESS.2019.2947484.

[15] G. Kalyani, B. Janakiramaiah, A. Karuna, and L. V. N. Prasad, "Diabetic retinopathy detection and classification using capsule networks," *Complex and Intelligent Systems*, vol. 9, no. 3, pp. 2651–2664, 2023, doi: 10.1007/s40747-021-00318-9.

[16] A. O. Asia *et al.*, "Detection of diabetic retinopathy in retinal fundus images using CNN classification models," *Electronics*, vol. 11, no. 17, 2022, doi: 10.3390/electronics11172740.

[17] Z. Ye and J. Yu, "AKSNet: A novel convolutional neural network with adaptive kernel width and sparse regularization for machinery fault diagnosis," *Journal of Manufacturing Systems*, vol. 59, pp. 467–480, 2021, doi: 10.1016/j.jmsy.2021.03.022.

[18] J. Zhou, V. Jampani, Z. Pi, Q. Liu, and M. H. Yang, "Decoupled dynamic filter networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6643–6652, doi: 10.1109/CVPR46437.2021.00658.
[19] J. Z. Esquivel, A. C. Vargas, P. L. Meyer, and O. Tickoo, "Adaptive convolutional kernels," in *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019*, IEEE, 2019, pp. 1998–2005, doi: 10.1109/ICCVW.2019.00249.
[20] S. Maharjan, A. Alsadoon, P. W. C. Prasad, T. Al-Dalain, and O. H. Alsadoon, "A novel enhanced softmax loss function for brain tumour detection using deep learning," *Journal of Neuroscience Methods*, vol. 330, 2020, doi: 10.1016/j.jneumeth.2019.108520.
[21] M. T. Al-Antary and Y. Arafa, "Multi-scale attention network for diabetic retinopathy classification," *IEEE Access*, vol. 9, pp. 54190–54200, 2021, doi: 10.1109/ACCESS.2021.3070685.
[22] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00444-8.
[23] M. Saini and S. Susan, "Diabetic retinopathy screening using deep learning for multi-class imbalanced datasets," *Computers in Biology and Medicine*, vol. 149, 2022, doi: 10.1016/j.compbiomed.2022.105989.
[24] W. Cukierski, "Diabetic retinopathy detection," *Kaggle*, 2015. Accessed: Apr. 14, 2024. [Online]. Available: https://kaggle.com/competitions/diabetic-retinopathy-detection
[25] H. Mustafa, C. Mohamed, O. Nabil, and A. Noura, "Rapid detection of diabetic retinopathy in retinal images: a new approach using transfer learning and synthetic minority over-sampling technique," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 1, pp. 1091–1101, 2024, doi: 10.11591/ijece.v14i1.pp1091-1101.
[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv-Computer Science*, pp. 1–14, 2014.
[27] Y. Shen and F. Yu, "The influence of artificial intelligence on art design in the digital age," *Scientific Programming*, vol. 2021, pp. 1–10, 2021, doi: 10.1155/2021/4838957.
[28] Z. P. Jiang, Y. Y. Liu, Z. E. Shao, and K. W. Huang, "An improved VGG16 model for pneumonia image classification," *Applied Sciences*, vol. 11, no. 23, 2021, doi: 10.3390/app112311185.
[29] N. Rachburee and W. Punlumjeak, "Lotus species classification using transfer learning based on VGG16, ResNet152V2, and MobileNetV2," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 4, pp. 1344–1352, 2022, doi: 10.11591/ijai.v11.i4.pp1344-1352.
[30] L. Chaves and G. Marques, "Data mining techniques for early diagnosis of diabetes: A comparative study," *Applied Sciences*, vol. 11, no. 5, pp. 1–12, 2021, doi: 10.3390/app11052218.
[31] M. Panda, D. P. Mishra, S. M. Patro, and S. R. Salkuti, "Prediction of diabetes disease using machine learning algorithms," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 1, pp. 284–290, 2022, doi: 10.11591/ijai.v11.i1.pp284-290.
[32] C. Lam, D. Yi, M. Guo, and T. Lindsey, "Automated detection of diabetic retinopathy using deep learning," in *AMIA Summits on Translational Science Proceedings*, 2018, pp. 147–155.
[33] A. K. Gangwar and V. Ravi, "Diabetic retinopathy detection using transfer learning and deep learning," in *Advances in Intelligent Systems and Computing*, vol. 1176, 2021, pp. 679–689, doi: 10.1007/978-981-15-5788-0_64.
[34] G. Nagaraj, S. C. Simha, H. G. R. Chandra, and M. Indiramma, "Deep learning framework for diabetic retinopathy diagnosis," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE, 2019, pp. 648–653, doi: 10.1109/ICCMC.2019.8819663.
[35] K. Oh, H. M. Kang, D. Leem, H. Lee, K. Y. Seo, and S. Yoon, "Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images," *Scientific Reports*, vol. 11, no. 1, pp. 1–9, 2021, doi: 10.1038/s41598-021-81539-3.
[36] Y. S. Kanungo, B. Srinivasan, and S. Choudhary, "Detecting diabetic retinopathy using deep learning," in *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, IEEE, 2017, pp. 801–804, doi: 10.1109/RTEICT.2017.8256708.
[37] Z. Khan *et al.*, "Diabetic retinopathy detection using VGG-NIN a deep learning architecture," *IEEE Access*, vol. 9, pp. 61408–61416, 2021, doi: 10.1109/ACCESS.2021.3074422.

BIOGRAPHIES OF AUTHORS

Mustafa Hiri received a bachelor's degree in electronics from the University of Ibn Zohr (UIZ) in Agadir, Morocco. He then obtained a research master's degree in signal processing and machine learning from the National School of Applied Sciences at the University of Abdelmalek Essaadi (UAE) in Tetuan, Morocco, in 2017 and 2019, respectively. He is currently pursuing a Ph.D. at the Laboratory of Computer Science for Operational Research and Applied Statistics, which is part of the Faculty of Science at the University of Abdelmalek Essaadi, also in Tetuan, Morocco. His research focuses on developing models and architectures for predictive medicine. He can be contacted at email: mustafa.hiri@etu.uae.ac.ma or mustafa.hiri@gmail.com.

Nabil Ourdani obtained a state engineering degree in computer science from the National Institute of Statistics and Applied Economics (INSEA). He then received a master's degree in public management from the Higher Institute of Commerce and Business Administration (ISCAE) and earned multiple training certificates in cloud-computing, blockchain, big data, machine learning, and deep learning. With a vast experience spanning over 24 years in software engineering, IT management, system architectures, and data processing, he is now a doctoral researcher at the University of Abdelmalek Essaadi, delving into the integration of big data and machine learning in public finance. He has contributed to scientific literature with published papers and has been a proactive participant in conferences. He can be contacted at email: nabil.ourdani@etu.uae.ac.ma or ourdani@gmail.com.

Mohamed Chrayah has been a Professor of computer science at Abdelmalek Essaadi University in Tetuan, Morocco since 2014. His academic journey commenced with an engineering degree from the National Institute of Statistics and Economy Applied in Rabat, Morocco, in 2006. He then achieved a Ph.D. in computer science in 2013 from the same institution. Over the years, he has developed a keen interest in machine learning, big data, and e-learning, with a specialized focus on social media analysis. Apart from his teaching stint, he has also been an active contributor to international journals and has played a significant role in organizing workshops at premier international conferences. He had earlier served as a lecturer and research fellow at University Abdelmalek Essaadi between 2009-2014 and held positions as a senior developer and system administrator during 2006-2014. He can be contacted at email: m.chrayah@uae.ac.ma.

Abeer Alsadoon received her Ph.D. from the University of Technology, Baghdad. She is currently an Associate Professor at Education Centre of Australia (ECA)-Asia Pacific International College (APIC). In addition, she is an academic at Western Sydney University (WSU). She was a Dean of Scholarship and Research at Kent Institute Australia. She received twenty (20) awards for teaching and research excellence from different Australian Universities. She has been recognized nationally as one out of five finalists for the prestigious 2019 Australian Women's Agenda Leadership Awards in the category of Emerging Female Leader in the Government and Public Sector. She published more than 130 papers in A and B ranking journals, and has more than 100 conference papers published in different peer review IEEE conferences, and she has received four (4) best paper awards. She is the editor of three books in an Emerging Trends in Cybersecurity Applications, Cutting Edge Applications of Computational Intelligence Tools and Techniques, and Current and Future Trends in Health and Medical Informatics books published by Springer. She is also the editor of the Proceedings of the ICR'22 and ICR'23 International Conference on Innovations in Computing Research book published by Springer Book Series: Advances in Intelligent Systems and Computing, and Lecture Notes in Networks and Systems, and ACR'23 International Conference on Advances in Computing Research book published by Springer Book Series: Lecture Notes in Networks and Systems. She is the chairing of the annual International Conference on Health Informatics and Medical System (HIMS), Las Vegas, USA, since 2022. She is also the Program Chair of the annual International Conference on Innovations in Computing Research (ICR), Europe, since 2022, the annual International Conference on Advances in Computing Research (ACR), Europe, since 2023, and the annual International Conference on Innovative Technologies in Intelligent Systems & Industrial Application (CITISIA), Sydney, Australia, since 2020. She also chaired special sessions at different conferences. She is a Senior Member of the Institute of Electrical and Electronic Engineers (IEEE) and Association for Computing Machinery (ACM). She has many years of academic experience. She can be contacted at email: aalsadoon@studygroup.com.

Noura Aknin has been a Professor at Abdelmalek Essaadi University in Tetuan, Morocco since 2000. She began her academic path with a Ph.D. in electrical engineering which she obtained in 1998. Over the years, she has cultivated an interest in mobile and wireless communications, social web, e-learning, machine learning, and big data. Beyond her role as an educator, she is the co-founder of the IEEE Morocco Section and has been actively involved in several R&D projects, both national and international. Some of her key involvements include her role as the head of the Research Unit Information Technology and Modeling Systems, as well as the coordinator of the Bachelor of Network Administration and Security of Information Systems. She has played significant roles in organizing events in the field of mobile networks, social web, and information technologies and has a prolific output as an author, with several papers on telecommunications, mobile and wireless communications, and web 2 applications. Before her tenure as a professor, she held various research and academic positions and has been an active contributor to the IEEE Communications Society and Computer Society. She can be contacted at email: noura.aknin@uae.ac.ma.