Advanced CNN Deep Learning Model for Diabetic Retinopathy Classification

Noor Ali Sadek (PhD)[1]*[iD], Ziad Tarik Al-Dahan (PhD)[2], Suzan Amana Rattan (PhD)[3], Ahmed F Hussein (PhD)[4], Brendan Geraghty (PhD)[5], Ahmed Kazaili (PhD)[6]

ABSTRACT

Background: Diabetic Retinopathy (DR) is one of several retinal microvascular complications of Diabetes Mellitus (DM), a disease of increasing global prevalence. However, early detection and treatment can reduce or even prevent DR progression. In this work, Deep Learning (DL) techniques are used to grade DR from an early stage using either binary or multiclass classification as a clinical aid to help reduce the risk of patient vision loss.

Objective: The primary objective of this research is to develop a low-cost, fast, and accurate automated system using DL for the early detection and classification of DR from retina fundus images.

Material and Methods: This cross-sectional study employed three DL models, namely Convolutional Neural Networks (CNNs), decision tree, and logistic regression, to categorize three distinct clinically graded datasets, namely the Iraqi dataset, the Indian Diabetic Retinopathy Image Dataset (IDRiD) and the Eyepacs dataset, according to DR severity.

Results: Evaluation of the DL model results showed that logistic regression emerged as the most effective, where accuracies of 99%, 99.3%, and 99.4% were achieved for the Iraqi, IDRiD, and Eyepacs datasets, respectively. Conversely, the decision-tree model achieved the lowest accuracy across the three datasets with 95.2%, 95.9%, and 96.0%, respectively.

Conclusion: The logistic regression model demonstrated the highest overall accuracy of the three models for the classification of DR, with the Iraqi dataset with the highest accuracy of the three datasets.

Keywords

Introduction

Diabetes Mellitus (DM) is an increasingly prevalent disease that can lead to microvascular malfunctions of the retinal blood vessels, namely Diabetic Retinopathy (DR) [1-3]. The primary assessment of DR is carried out via examination of fundus images by an ophthalmologist. Early diagnosis of DR is crucial in mitigating the visual impairment and diminished visual acuity associated with this condition. Undiagnosed DR can progress to the subsequent severity levels; mild non-proliferative, moderate non-proliferative, severe non-proliferative,

[1]Department of Biomedical Engineering, Al-Nahrain University, Baghdad, Iraq
[2]Department of Biomedical Engineering, College of Engineering, Al-Bayan University, Baghdad, Iraq
[3]AlKindy College of Medicine, University of Baghdad, Baghdad, Iraq
[4]Artificial Intelligent and Robotics Engineering, College of Engineering, Al-Nahrain University, Baghdad, Iraq
[5]Department of Musculoskeletal and Ageing Science, Institute of Life Course and Medical Sciences, University of Liverpool, L7 8TX, UK
[6]School of Engineering, University of Liverpool, Liverpool, L69 3GH, UK

*Corresponding author:
Noor Ali Sadek
Department of Biomedical Engineering, Al-Nahrain University, Baghdad, Iraq
E-mail: nooralisadek@gmail.com

and the most severe level, proliferative DR [4-6]. Because of the increased number of diabetic patients coupled with the limited numbers of ophthalmologists globally, the development of an automated system to detect and grade DR would reduce the workload on clinicians and improve patient prognosis. Automated detection of DR has received increasing interest over the last two decades, especially after the rise of Deep Learning (DL), which eliminated the need to extract specific features from diseased images manually [7].

A Convolutional Neural Network (CNN) algorithm was used to detect very small lesions on preprocessed and augmented data obtained from the EyePacs dataset, after optimizing the CNN architectures, which ultimately achieved an accuracy of 95% [8]. Another study proposed the extraction of blood vessels from the fundus image requiring preprocessing steps, namely image size normalization by scaling and boundary removal, to gain an accuracy of 74% [9]. Another study proposed a novel approach for feature extraction, including vascular area, exudate areas, and texture features with the use of a Support Vector Machine (SVM) [10]. Other studies [11-13] proposed various comprehensive models that focused on utilizing CNN to classify target lesions within fundus images at different scales.

Priya Henry et al. [14] introduced thirteen filters, such as smoothing and sharpening, to enhance the Indian Diabetic Retinopathy Image Dataset (IDRiD). The Gaussian, Median, Wiener, Bilateral, and partial differential equation filters were utilized to enhance the detection performance of their algorithm, and outstanding results were achieved. In a recent study, Levi et al. [15] utilized the GoogLeNet CNN to identify and classify DR, achieving a detection accuracy of 88% for binary classification. Salma et al. [16] suggested a CNN-based technique for DR classification, utilizing both GoogLeNet and transfer learning. The experiment categorized DR into three levels: No DR, mild, and severe. The approach

achieved a sensitivity of 95% and specificity of 96%. Whereas Lam et al. [17] employed a CNN model for multi-classification of DR severity levels and demonstrated an accuracy of up to 93%. Further work was carried out by Butt et al. [18] using a hybrid system involving a CNN with a transfer learning algorithm. Their technique showed an enhanced technical performance; however, it achieved an overall accuracy of 89.29%, which was lower than previous studies. Kang et al. [19] used four ophthalmic imaging modalities to diagnose DR with a DL system. This work showed a practical approach to diagnosing DR with impressive accuracy results for the four models as 98.4%, 97.7%, 96.7%, and 96.3%, respectively. In another investigation, Adem [20] used a CNN built on DL methods to classify retinal fundus images from the public Kaggle dataset; there was a 75% accuracy, 95% specificity, and 93% sensitivity in DR detection. Using four CNN techniques, Xu et al. [21] used a similar method and classified DR images from Kaggle with an accuracy of 94%.

The logistic regression model has been widely used in various fields to classify data objects into groups. Our logistic regression technique aimed to accurately characterize the relationship between the target variable and predictor factors, an approach, which was previously used by Zhu et al. [22]. The decision tree was a model to display classifiers and regressions. As the name implies, this tree was made up of a number of nodes and branches [23]. This study focused on the three DL algorithms for an automated system for early-stage diagnosis of DR. Three DL algorithms were developed and compared, namely CNN, Decision Trees, and Logistic Regression. Each algorithm was responsible for classifying fundus images into different grades of DR. Additionally, three distinct datasets, namely IDRiD, EyePACS, and a newly collected dataset (the Iraqi dataset), which contained a large number of labeled images, were employed to evaluate the DR diagnosis. The comparison of these algorithms is

a crucial aspect of the study to determine the most effective method for DR classification.

Material and Methods

In this cross-sectional study, a novel DL algorithm was implemented to automatically detect and classify fundus images of DR. The algorithm was used to classify DR into healthy, mild, moderate, severe, and Proliferative Diabetic Retinopathy (PDR) according to the Early Treatment Diabetic Retinopathy Study (ETDRS) classification standards of severity level as indicated in Figure (1). The DR classes are described as follows: healthy indicates a healthy fundus image without DR and does not show any abnormality, such as microaneurysms or any other DR lesions. Mild is the initial stage of Non-Proliferative Diabetic Retinopathy (NPDR) of the disease. At this stage, the only abnormalities detected in the examinations are microaneurysms. In moderate NPDR, dot blot haemorrhages or microaneurysms appear in at least one quadrant, with or without cotton-wool spots, venous beading, or intraretinal microvascular abnormalities. Severe NPDR includes any of the following cases; 20 or more intraretinal haemorrhages (dot blot haemorrhages) in each of the four quadrants, definite venous beading in 2 or more quadrants, and prominent intraretinal microvascular abnormality in one or more quadrants. The term "4:2:1 rule" is used to refer to these three points since they necessitate the presence of abnormalities in at least four,

two, and one quadrants of the retina. PDR is the highest level of the disease's progression. During this stage, the retina or optic nerve experiences the growth of newly formed, delicate, and abnormal blood vessels. These blood vessels have the potential to leak, which can impact the quality of vision. Examinations detect either a clear presence of neovascularization, pre-retinal, or vitreous haemorrhages.

After the initial collection of the Iraqi dataset and downloading of the IDRiD and EyePACS datasets, all three datasets were first preprocessed to enhance the images as an input to the DL algorithm, including; cropping, resizing, and contrast enhancement. DL algorithms based on three different models (CNN, Decision Tree, and Logistic Regression) were applied to classify the images in all three datasets, as explained below. The classification results of the algorithms were assessed based on the accuracy, precision, logistic loss, and Area Under the Curve (AUC) [23]. These results were then compared with all other results obtained from other datasets [24, 25] to identify the best performance of the classifier, as shown in the block diagram in Figure (2).

Datasets

In this study, the three datasets used are described below. Samples of used image data are shown in Figure 3 as follows:

1. EyePACS dataset

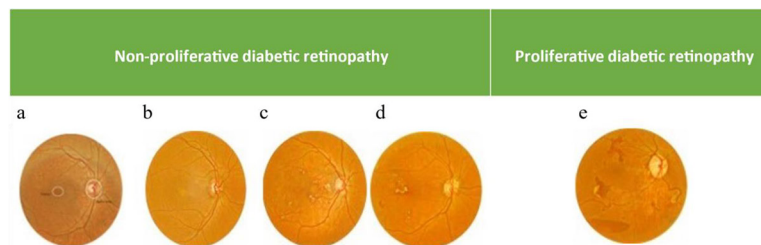The EyePACS dataset [24] contained 35000 images; a sample of each DR stage is shown

Figure 1: Diabetic Retinopathy (DR) progression. This Figure describes the classification stages **a**) is the normal eye, **b**) is mild (3-5 years), **c**) represents moderate (5-10 years), **d**) refers to severe stage (10-15 years). **b**, **c**, and **d**) are also known as Non-Proliferative Diabetic Retinopathy (NPDR). **e**) represents the Proliferative Diabetic Retinopathy (PDR) level.
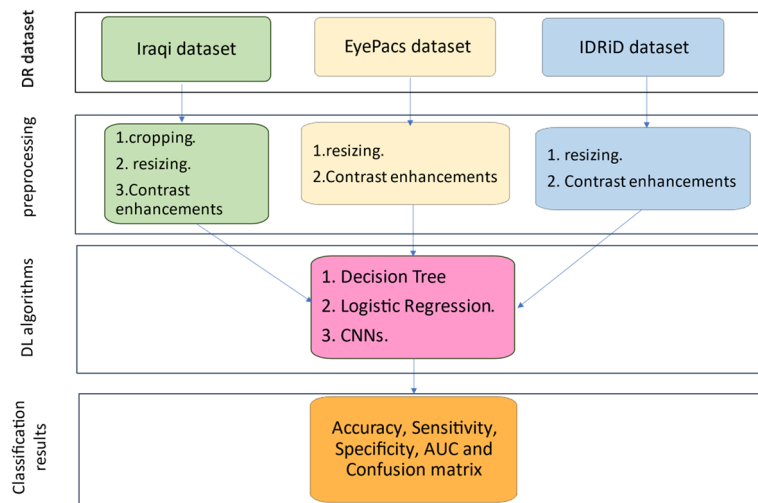
Figure 2: General workflow chart for Diabetic Retinopathy (DR) detection: Dataset selection, preprocessing, and Deep Learning classification. It illustrates the main steps of the method, beginning with dataset selection followed by preprocessing, including image cropping, resizing, and contrast enhancement. Subsequently, three Deep Learning (DL) systems (Convolutional Neural Network (CNN), Decision Tree, and Logistic Regression) were chosen for the training and classification of DR. The results were then statistically analyzed using evaluation metrics such as accuracy, sensitivity, specificity, Area Under the Curve (AUC), and the confusion matrix.
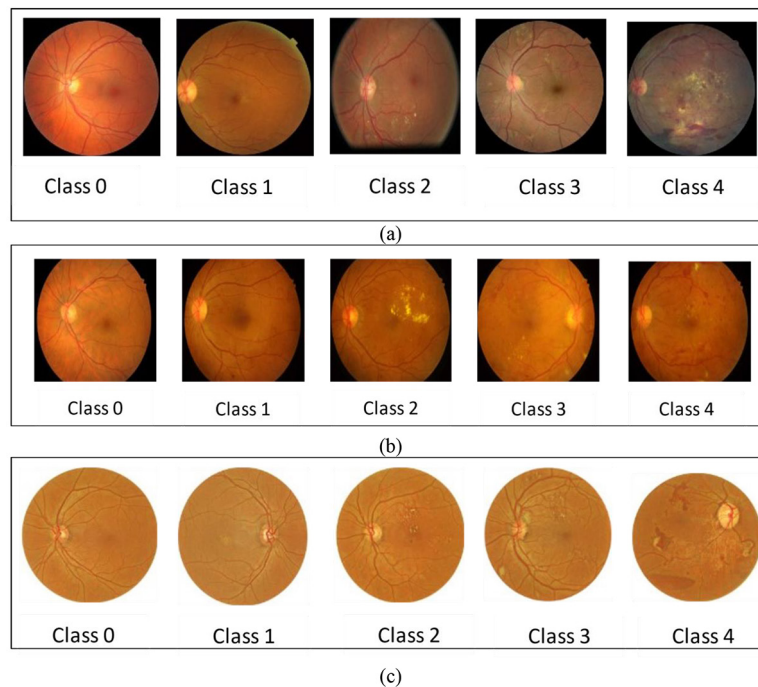

Figure 3: Sample of the dataset with Diabetic Retinopathy (DR) shows the main classification grades: class 0 represents the normal or healthy image; class 1 represents the mild stage of Diabetic retinopathy (DR); class 2 represents the moderate stage; class 3 shows the severe non-proliferative DR; and class 4 shows the last Proliferative Diabetic Retinopathy (PDR) stage for: (**a**) EyePACS dataset [24]. (**b**) IDRiD dataset [25]. (**c**) Iraqi dataset [26].

in Figure (3a). The dataset contained left and right eye images from each patient with an image size of 1024×1024 pixels in jpeg format. The dataset exhibited a significant imbalance, with 65,343 healthy images, 6,195 mild class, 13,153 moderate class, 2,087 severe class, and 1,914 images belonging to the PDR class.

2. IDRiD dataset

The IDRiD dataset was recently published [25] and consisted of 413 images. These images were divided into two categories: patients with and without signs of DR. The images had a size of 4288×2848 pixels in jpeg format. The dataset included 135 images classified as healthy, 20 as mild, 142 as moderate, 63 as severe, and 53 as PDR. Figure (3b) displays a sample of the IDRiD images.

3. Iraqi dataset

This dataset was collected at the main ophthalmology teaching center in Baghdad, where patients from all cities in Iraq were admitted. It consisted of 700 retinal fundus images that were collected and labeled under the supervision of specialized ophthalmologists. The data was then classified according to ETDRS standards into five stages depending on the severity level of DR. This dataset included 153 images classified as healthy, 59 as mild, 304 as moderate, 99 as severe, and 85 as PDR. The collected images were stored in the hospital server and anonymized for teaching and research after obtaining ethical permission [26]. The images were captured with a resolution of 3507×2480 pixels and stored in jpeg format. A sample of this dataset is shown in Figure (3c).

Preprocessing

Preprocessing of the three sets of data was used to standardize image properties delivered to the DL algorithms. The preprocessing focused on resizing and contrast enhancements to make foreground details in fundus images, such as the optic disc, blood vessels, and lesions more recognizable and easier to identify by distinguishing them from the background. This aimed to reduce the time required for DR image training and classification testing. Preprocessing steps were performed using R2021b MATLAB software with image processing code and enhancements techniques. The EyePACS dataset was resized from 1024×1024 pixels to 332×314 pixels, as shown in Figure 4 (a and b). The IDRiD dataset was resized from 4288×2848 to 753×500 pixels, as shown in Figure 4 (c and d). The Iraqi dataset was cropped and then labeled before resizing from 3507×2480 to 320×314 pixels. The contrast is enhanced using adaptive binarization, as shown in Figure 4 (e, f, g, h), by converting the fundus images into gray binary images. This method is effective in enhancing and distinguishing the important features of DR, which can have varied illumination in fundus images. Adaptive binarization has many advantages, such as handling nonuniform illumination, reducing the shadows and highlights on the images, and also enhancing the local contrast, which is essential in identifying small and significant changes in the image features. However, it also has some disadvantages, such as potential data loss and computational intensity that can be reduced or handled through parameter optimization to preserve important features. They can also maintain the balance between enhanced features and preserved necessary information.

Deep Learning System

A DL algorithm based on three models was used to analyze the datasets and then trained to identify and classify features based on specific clinical criteria. Feature selection and training were done using pooling and convolution operations over multiple layers in the network architecture. In the DL-based CNN model, the initial layers were used for feature extraction, specifically through the use of convolution layers. This process produces a feature map that identifies corners and edges in the images. To prevent overfitting during the classification stage, a dropout layer was used to reduce the size and complexity of the model. A dropout
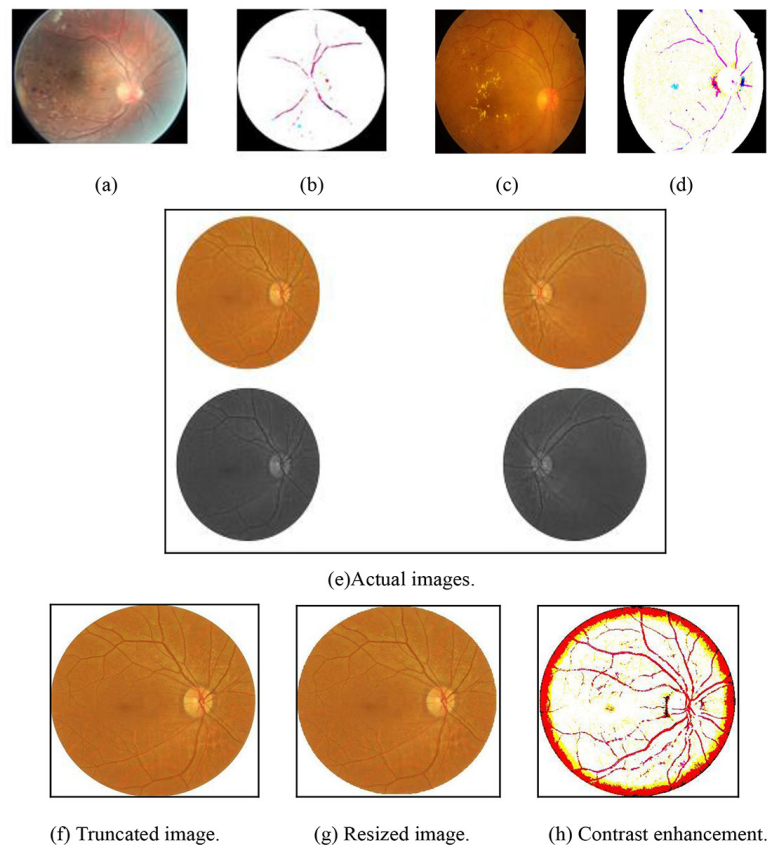
(a) (b) (c) (d)

(e)Actual images.

(f) Truncated image. (g) Resized image. (h) Contrast enhancement.

Figure 4: Diabetic Retinopathy (DR) image enhancement steps. (**a**) The EyePACS resized image (**b**) The same image with contrast enhancement. (**c**) The IDRiD resized image. (**d**) The same image was enhanced by adaptive binarization. (**e**) Iraqi dataset, including the original unenhanced DR image, (**b**) the cropped image (**c**) the resized image (**d**) the contrast-enhanced image.

rate of 0.01 is used in the proposed model to achieve the best accuracy. The activation function was an important parameter in the CNN model as it determined, which information should be passed to the next layer and which was less useful. The Rectified Linear Unit (ReLU) activation function was utilized due to its common usage in machine learning models. It was responsible for processing weighted inputs and helping deliver an output, and was typically used in the hidden layers of a neural network to add non-linearity. The softmax activation function was typically used in the last layer of a neural network to predict the class of an input image. The Hyperbolic Tangent function (tanH) gave better performance for multi-layer neural networks while sigmoid functions introduced non-linearity to the model and

helped the algorithm learn complex functions, as shown in Figure (5).

This research used the Iraqi dataset of DR patients and classified them to distinguish and grade DR. The results were then compared to the EyePACS and IDRiD datasets. Three DL algorithms were used to classify DR into five grades. These algorithms, the CNN, decision tree, and logistic regression model, were applied to the datasets. The decision tree algorithm splits the data into nodes based on class purity, while the CNN model is a Multi-layer Perceptron (MLP) algorithm capable of learning both linear and non-linear models. The CNN model, with a hundred neurons in the hidden layer, used the ReLU as an activation function and the Stochastic Gradient Descent (SGD) as an optimizer with maximum

iterations of 250 and a learning rate of 0.01. On the other hand, logistic regression is a statistical method to describe the relationship between a dependent variable and one or more independent variables.

Evaluation metrics [27] were utilized to assess the efficacy of the DL models for classification tasks. These metrics include accuracy, sensitivity, specificity, and the confusion matrix. The evaluation was based on finding True Negative (TN) instances, which are the cases correctly diagnosed as healthy. The True Positive (TP) instances indicate patients correctly diagnosed with DR. False Negative (FN) denotes cases that were incorrectly identified as healthy, while False Positive (FP) represents cases of patients mistakenly diagnosed as healthy individuals. The confusion matrix serves as a tabular representation of correct and incorrect predictions.

Accuracy reflects the proportion of correctly predicted observations, encompassing both positive and negative outcomes, relative to the total observations. The confusion matrix serves as a concise summary of predictive outcomes, detailing TP, TN, FP, and. Each row in the matrix corresponds to the actual class, while each column represents the predicted class. The matrix's diagonal indicates correct predictions, while off-diagonal elements signify errors. Furthermore, the AUC value was calculated to assess the classifier's capability to distinguish between different classes [27].

Results

The accuracy, AUC, specificity, and sensitivity of the classification on the Iraqi dataset using different models are shown in Table (1). The highest classification accuracy was found in the logistic regression (99%), while the lowest was seen in the decision tree model (95.2%).

The confusion matrix of the decision tree model for the Iraqi dataset showed that 118 images were correctly classified as healthy, 55 as mild, 298 as moderate, 87 as severe, and 77
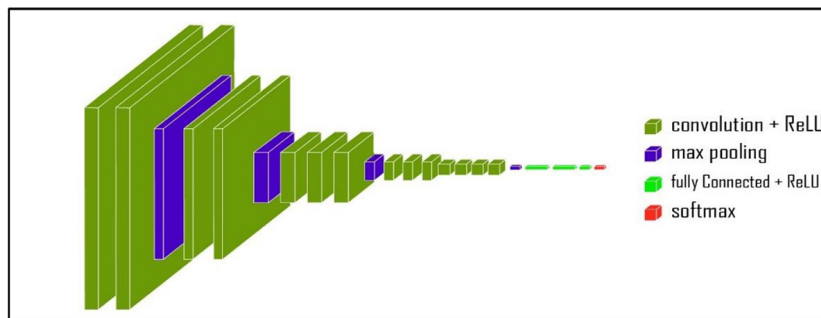

Figure 5: Schematic representation of Convolutional Neural Network (CNN) architecture shows the sequence of four filtration layers (convolutional layer, rectified linear unit layer, maxpooling layer, fully connected layer, and SoftMax layer).

Table 1: Evaluation results of classification performance using the Iraqi dataset and the three models.

Model	Accuracy (%)	AUC	Specificity	Sensitivity
Decision tree	95.2	0.998	0.977	0.953
CNN	98.5	0.991	0.993	0.985
Logistic regression	99	0.993	0.994	0.990

CNN: Convolutional Neural Network, AUC: Area Under Curve

as PDR. Furthermore, for the CNN model, it showed that 119 images were correctly classified as healthy, 56 as mild, 301 as moderate, 97 as severe, and 81 as PDR. For the logistic regression model, 118 images were correctly classified as healthy, 57 as mild, 303 as moderate, 98 as severe, and 84 as PDR, as shown in Figure 6 (a, b, c).

The accuracy, AUC, specificity, and sensitivity for the EyePACS dataset using decision tree, CNN, and logistic regression are shown in Table (2). The confusion matrix of the decision tree model for the EyePACS dataset

showed that 167 images were correctly classified as healthy 24 as mild, 153 as moderate, 80 as severe, and 50 as PDR. For the logistic regression model, it showed that 168 images were correctly classified as healthy, 24 as mild, 157 as moderate, 85 as severe, and 57 as PDR. For the CNN model, it showed that 166 images were correctly classified as healthy, 20 as mild, 154 as moderate, 84 as severe, and 57 as PDR, as shown in Figure 7 (a, b, c).

The accuracy, AUC, specificity, and sensitivity for the IDRiD dataset using decision tree, CNN, and logistic regression are shown
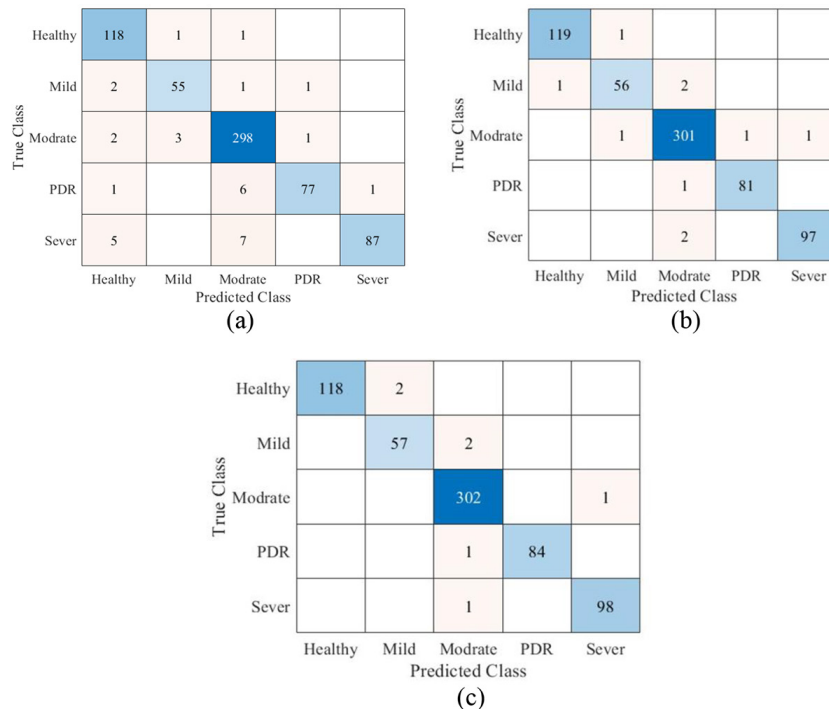

Figure 6: Confusion matrixes of the Iraqi dataset. (**a**) using the decision tree model. (**b**) by Convolutional Neural Networks (CNN) model. (**c**) with Logistic Regression. True class data were collected based on the physician's diagnosis. (PDR: Proliferative Diabetic Retinopathy)

Table 2: Performance evaluation of EyePACS dataset with three models.

Model	Accuracy (%)	AUC	Specificity	Sensitivity
Decision tree	0.96	0.997	0.988	0.961
CNN	0.974	0.998	0.974	0.974
Logistic regression	0.994	0.993	0.998	0.994

CNN: Convolutional Neural Network, AUC: Area Under Curve

in Table (3). The confusion matrix of the decision tree model of the IDRiD dataset showed that 167 images were correctly classified as healthy, 24 as mild, 153 as moderate, 80 as severe, and 50 as PDR. For the logistic regression model, it showed that 168 images were correctly classified as healthy, 24 as mild, 157 as moderate, 85 as severe, and 57 as PDR. For the CNN model, it showed that 166 images were correctly classified as healthy, 20 as mild, 154 as moderate, 84 as severe, and 57 as PDR, as shown in Figure 8 (a, b, c). Table (4) summarizes the performance of the three models with the three datasets.

Discussion

The quality of life for diabetic patients can be significantly impacted by DR due to the resulting visual impairments and progressive development of DR symptoms. The shortage of highly skilled ophthalmologists combined with the high cost-effectiveness of eye screening methods will encourage the creation and use of artificially intelligent systems like DL models for early DR diagnosis and decrease the progression or manage the symptoms of
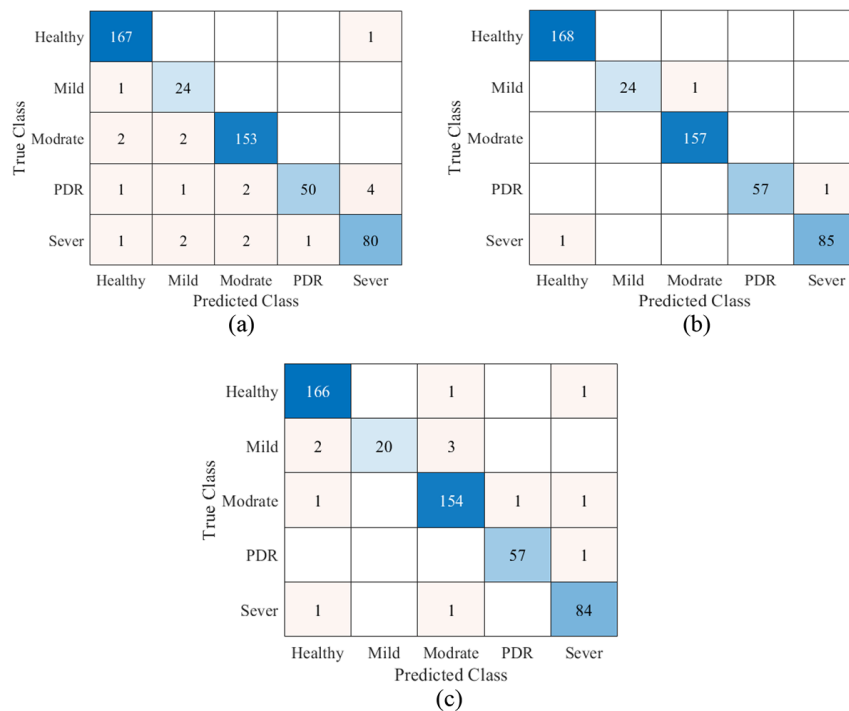

Figure 7: Confusion matrix of EyePACS (**a**) using decision tree model. (**b**) by logistic regression model. (**c**) with Convolutional Neural Network (CNN) model. (PDR: Proliferative Diabetic Retinopathy)

Table 3: Performance evaluation of the algorithm based on three models using the IDRiD.

Model	Accuracy (%)	AUC	Specificity	Sensitivity
Decision tree	95.9	0.991	0.960	0.961
CNN	98.8	0.996	0.995	0.988
Logistic regression	99.3	0.997	0.998	0.999

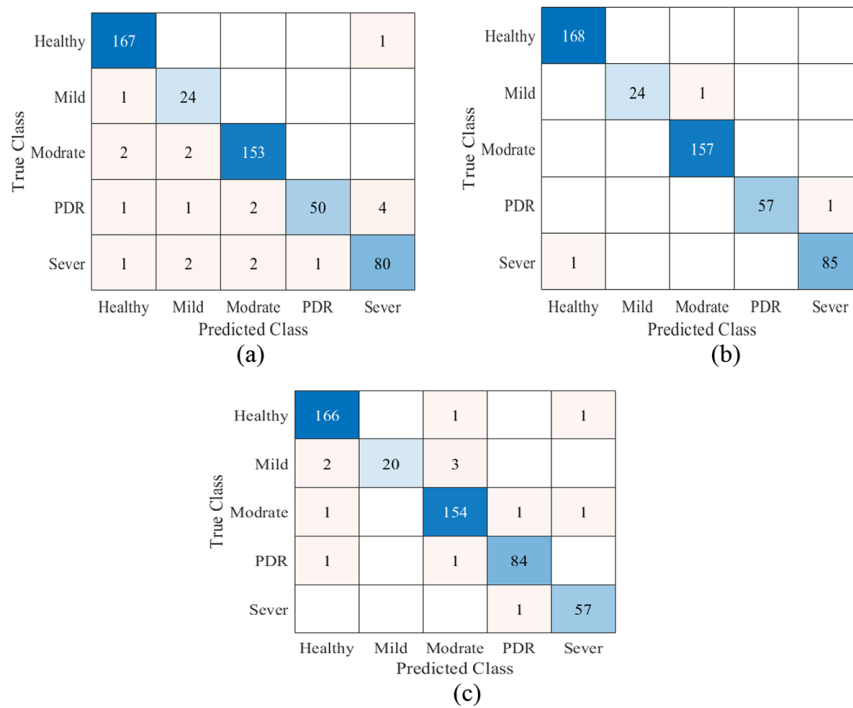CNN: Convolutional Neural Network, AUC: Area Under Curve

Figure 8: Confusion matrices of IDRiD, using **(a)** decision tree model, **(b)** logistic regression model. **(c)** Convolutional Neural Network (CNN) model. (PDR: Proliferative Diabetic Retinopathy)

Table 4: Summarization of the results of the three datasets with the three models.

Deep learning models	Datasets											
	Iraqi dataset				Eyepacs dataset				IDRiD dataset			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
Logistic regression	0.99	90.90	0.994	0.993	0.994	0.964	0.998	0.993	0.993	0.999	0.998	0.997
CNN	0.985	0.985	0.993	0.991	0.974	0.974	0.974	0.998	0.988	0.988	0.995	0.996
Decision tree	0.952	0.953	0.977	0.998	0.96	0.961	0.988	0.997	0.959	0.961	0.960	0.991

CNN: Convolutional Neural Network, AUC: Area Under Curve

DR. Therefore, the progress in the model architecture has resulted in the utilization of larger databases to test and evaluate these technologies in the diagnosis and grading.

The logistic regression, decision tree, and CNN models were used in this work and represent baseline models for the classification of DR because of their architectural diversity and simplicity. When interpretability, simplicity, and efficiency are required, logistic regression performs well and is ideal for datasets and circumstances requiring immediate insights. For clear decision-making processes and the capture of non-linear relations, decision trees are ideal and useful for analyzing the relevance of features. Simple CNNs work well for direct image analysis since they can quickly identify intricate patterns in large datasets and perform

well in image classification applications [27]. The aforementioned models based on DL were selected for this research because of their reliable performance.

Logistic regression provides clear insights into which features influence the outcome, making it more interpretable than decision trees and CNN models. Furthermore, the logistic regression model performs well even with smaller datasets, unlike CNN models that need larger datasets for effective training [27]. Logistic regression requires well-engineered features to perform well through the image preprocessing step. While all three models have these benefits, it is crucial to remember that the model selected will rely on the particular needs and limitations of the DR classification task, such as the size and makeup of the dataset, and the requirement for interpretability.

Based on the findings, the three datasets could yield impressive predictions when they underwent identical pre-processing procedures. In terms of classification performance for DR grading, the Iraqi dataset has shown promising results when compared to large datasets with different image sizes and resolutions, such as the EyePACS dataset because of its diversity and good quality images in addition to its correct labeling. The aforementioned reasons made the classification of the DR data using CNN, decision tree and logistic regression more valuable with results approaching real-world clinical diagnosis.

The logistic regression model exhibited superior performance across various datasets compared to prior studies [27-29] utilizing the preprocessed Pima Indians Diabetes (PID) dataset. In the current study, employing the Iraqi, EyePACS, and IDRiD datasets resulted in accuracies of 99%, 99.4%, and 99.3%, respectively, overtaking the 97% accuracy achieved with the PID dataset and its logistic regression model incorporating Principal Components Analysis (PCA) techniques [27]. Our proposed enhancement technique significantly

contributed to this improvement in predictive capability. Specifically, for the Iraqi dataset, logistic regression demonstrated an accuracy of 99%, an AUC of 99.3%, a specificity of 99.4%, and a sensitivity of 99%, as indicated in Table 2. Likewise, for the EyePACS dataset, logistic regression yielded outstanding classification results with a 99.4% accuracy, a 99.3% AUC, a 99.8% specificity, and a 99.4% sensitivity, as shown in Table 3. Similarly, Table 4 highlights the strong performance of the logistic regression model on the IDRiD dataset, with metrics of 99.3% accuracy, 99.7% AUC, 99.8% specificity, and 99.9% sensitivity.

Conclusion

Due to the increasing global frequency of DM, there is a growing need for developed algorithms to aid clinicians in the early detection, diagnosis, and classification of DR. In this work, we first dealt with the preprocessing of all datasets used (Iraqi, EyePACS, and IDRiD), which in turn impacts on maximizing the functionality of the DL system and underscores the importance of data preparation in maximizing the model functionality. Second, we compared three different DL models for their ability to classify DR with three different datasets to predict the best model according to the model performance, namely accuracy, precision, AUC, and sensitivity. The used datasets were specific to particular populations, which may limit the generalizability of the findings to other demographic groups. Also, the model needs clinical validation to be reliable and more robust in real-time clinical conditions. All of the DL models show a promising result for the ability to learn features that are required for DR fundus image classification. The logistic regression shows an optimal result of DR classification with high accuracy, precision, sensitivity, and confusion matrix for all of the datasets. Dataset quality and balance are very important and that is why the Iraqi dataset demonstrated greater accuracy than

other datasets in the disease classification of this work. The future direction for this work is to generate an optimization procedure with a multi-model setup and other enhancement techniques to have a comprehensive system for DR classification and also to collect new datasets for early diagnosis of pediatric DR.

Authors' Contribution

NA. Sadek conceived and planned the experiments, conducted all the experimental work, analyzed the data, and edited and wrote the manuscript. NA. Sadek with AF. Hussein helped in building a MATLAB code for the diagnosis and classification of DR, verified the analytical method, and edited the manuscript. ZT. Al Dahan with SA. Rattan supervised all the work. A. Kazaili and B. Geraghty reviewed, revised, refined the work, and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Ethical Approval

This study was conducted by the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration. Ethical approval was obtained from the Ministry of Health to include anonymized clinical human data for the Iraqi dataset.

Informed Consent

Informed consent was obtained from all individual participants included in the study. The participants were fully informed about the purpose, procedures, potential risks, and benefits of the study. They were assured that their participation was voluntary and that they could withdraw from the study at any time without any consequences. For participants who were unable to provide informed consent due to age or other reasons, consent was obtained from their legal guardians. All personal data collected during the study were anonymized to protect the privacy of the participants.

Conflict of Interest

None

References

1. Reguant R, Brunak S, Saha S. Understanding inherent image features in CNN-based assessment of diabetic retinopathy. *Sci Rep.* 2021;**11**(1):9704. doi: 10.1038/s41598-021-89225-0. PubMed PMID: 33958686. PubMed PMCID: PMC8102512.
2. Aguirre F, Brown A, Cho NH, Dahlquist G, Dodd S, Dunning T, et al. IDF Diabetes Atlas. 6th ed. Basel, Switzerland: International Diabetes Federation; 2013.
3. Ting DS, Tan KA, Phua V, Tan GS, Wong CW, Wong TY. Biomarkers of diabetic retinopathy. *Curr Diab Rep.* 2016;**16**:125. doi: 10.1007/s11892-016-0812-9.
4. Bourne RR, Stevens GA, White RA, Smith JL, Flaxman SR, Price H, et al. Causes of vision loss worldwide, 1990-2010: a systematic analysis. *Lancet Glob Health.* 2013;**1**(6):e339-49. doi: 10.1016/S2214-109X(13)70113-X. PubMed PMID: 25104599.
5. Cheung N, Mitchell P, Wong TY. Diabetic retinopathy. *Lancet.* 2010;**376**(9735):124-36. doi: 10.1016/S0140-6736(09)62124-3. PubMed PMID: 20580421.
6. Kroenke K. Telemedicine screening for eye disease. *JAMA.* 2015;**313**(16):1666-7. doi: 10.1001/jama.2015.107. PubMed PMID: 25919530.
7. Stolte S, Fang R. A survey on medical image analysis in diabetic retinopathy. *Med Image Anal.* 2020;**64**:101742. doi: 10.1016/j.me-

dia.2020.101742. PubMed PMID: 32540699.

8. Maswood MM, Hussain T, Khan MB, Islam MT, Alharbi AG. Cnn based detection of the severity of diabetic retinopathy from the fundus photography using efficientnet-b5. In 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON); Vancouver, BC, Canada: IEEE ;2020. p. 0147-50.

9. Yung-Hui L, Nai-Ning Y, Purwandari K, Harfiya LN. Clinically applicable deep learning for diagnosis of diabetic retinopathy. In: 2019 Twelfth International Conference on Ubi-Media Computing (Ubi-Media); Bali, Indonesia: IEEE: 2019. p. 124-9.

10. Paranjpe MJ, Kakatkar MN. Automated diabetic retinopathy severity classification using support vector machine. *Int J Res Sci Adv Technol.* 2013;**3**(3):86-91.

11. Liaw A. Classification and regression by randomForest. *R News.* 2002;**2**:18-22.

12. Lim G, Lee ML, Hsu W, Wong TY. Transformed representations for convolutional neural networks in diabetic retinopathy screening. In: Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence; Canada: 2014.

13. Bindu CH, Sravanthi GS. Retinopathy detection of EYE images in diabetic patients. *Indian J Sci Technol.* 2019;**12**(3):1-7. doi: 10.17485/ijst/2018/v12i3/141597.

14. Priya Henry AG, Jude A. Convolutional neural-network-based classification of retinal images with different combinations of filtering techniques. *Open Computer Science.* 2021;**11**(1):480-90. doi: 10.1515/comp-2020-0177.

15. Levi G, Hassner T. Age and gender classification using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); Boston, MA, USA: IEEE; 2015. p. 34-42.

16. Salma A, Bustamam A, Sarwinda D. Diabetic Retinopathy Detection Using GoogleNet Architecture of Convolutional Neural Network Through Fundus Images. *NST Proceeding.* 2021:1-6. doi: 10.11594/nstp.2021.0701.

17. Lam C, Yi D, Guo M, Lindsey T. Automated Detection of Diabetic Retinopathy using Deep Learning. *AMIA Jt Summits Transl Sci Proc.* 2018;**2017**:147-55. PubMed PMID: 29888061. PubMed PMCID: PMC5961805.

18. Butt MM, Iskandar DNFA, Abdelhamid SE, Latif G, Alghazo R. Diabetic Retinopathy Detection from Fundus Images of the Eye Using Hybrid Deep Learning Features. *Diagnostics (Basel).* 2022;**12**(7):1607. doi: 10.3390/diagnostics12071607. PubMed PMID: 35885512. PubMed PMCID: PMC9324358.

19. Kang EY, Yeung L, Lee YL, Wu CH, Peng SY, Chen YP, et al. A Multimodal Imaging-Based Deep Learning Model for Detecting Treatment-Requiring Retinal Vascular Diseases: Model Development and Validation Study. *JMIR Med Inform.* 2021;**9**(5):e28868. doi: 10.2196/28868. PubMed PMID: 34057419. PubMed PMCID: PMC8204240.

20. Adem K. Exudate detection for diabetic retinopathy with circular Hough transformation and convolutional neural networks. *Expert Systems with Applications.* 2018;**114**:289-95. doi: 10.1016/j.eswa.2018.07.053.

21. Xu K, Feng D, Mi H. Deep Convolutional Neural Network-Based Early Automated Detection of Diabetic Retinopathy Using Fundus Image. *Molecules.* 2017;**22**(12):2054. doi: 10.3390/molecules22122054. PubMed PMID: 29168750. PubMed PMCID: PMC6149821.

22. Zhu C, Idemudia CU, Feng W. Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques. *Inform Med Unlocked.* 2019;**17**:100179. doi: 10.1016/j.imu.2019.100179.

23. Karami A, Askarishahi M, Namiranian N. Comparison of Different Decision Tree Algorithms for Classification of Retinopathy Patients in Yazd City, Central Part of Iran. *Journal of Community Health Research.* 2022;**11**(3):158-64. doi: 10.18502/jchr.v11i3.10867.

24. Cuadros J, Bresnick G. EyePACS: an adaptable telemedicine system for diabetic retinopathy screening. *J Diabetes Sci Technol.* 2009;**3**(3):509-16. doi: 10.1177/193229680900300315. PubMed PMID: 20144289. PubMed PMCID: PMC2769884.

25. Prasanna P, Samiksha P, Ravi K, Manesh K, Girish D, Vivek S, Fabrice M. Indian diabetic retinopathy image dataset (idrid). IEEE Dataport; 2018. Available from: https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid.

26. Geraghty B, Kazaili A, Sadek NA, Al-Dahan ZT, Rattan SA, Faeq Hussein AF. Iraqi Retinal Fundus Diabetic Retinopathy Dataset (IRFDRD). 6th

International Scientific Conference titled "Modern Science and Technology: Foundations for Ensuring Community Development and Energy Utilization"; Duhok, Iraq: 2024. Available from: https://github.com/noor-aliz810/IRFDRD.git.

27. Maalouf M. Logistic regression in data analysis: an overview. *Int J Data Analysis Techniques and Strategies.* 2011;**3**(3):281-99. doi: 10.1504/IJ-DATS.2011.041335.

28. Xin Z, Hua L, Wang XH, Zhao D, Yu CG, Ma YH, et al. Reanalysis and External Validation of a Decision Tree Model for Detecting Unrecognized Diabetes in Rural Chinese Individuals. *Int J Endocrinol.* 2017;**2017**:3894870. doi: 10.1155/2017/3894870. PubMed PMID: 28638408. PubMed PMCID: PMC5468553.

29. Chang V, Bailey J, Xu QA, Sun Z. Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. *Neural Comput Appl.* 2022:1-17. doi: 10.1007/s00521-022-07049-z. PubMed PMID: 35345556. PubMed PMCID: PMC8943493.