Journal of Information Systems Engineering and Management
2025, 10(16s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

AI-Driven Diabetic Retinopathy Detection Using ILWOA-Enhanced Extreme Learning Machine on EyePACS and APTOS Datasets

Mrs Xma R Pote[1], Dr. Mahaveerakannan R[2], Dr. Priscilla Joy[3], Sheeba Santhosh[4], Dr. Narina Thakur[5], M. Vignesh[6]
[1] *Assistant Professor, Dept of Electrical Engg, Yeshwantrao Chavan College of Engineering, Nagpur Maharashtra*
[2] *Professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India.*
[3]*Assistant professor, Division of CSE, Karunya institute of technology and sciences, Coimbatore – 641035*
[4] *Assistant Professor Grade 1, Department of ECE, Panimalar Engineering College, Chennai, Tamil Nadu*
[5]*Assistant Professor, Computer Science & Software Engineering, University of Stirling, RAK Campus, United Arab Emirates*
[6]*Assistant Professor, Department of Artificial Intelligence and Data Science, Karpagam Institute of Technology, Coimbatore, Tamil Nadu, India*
[1] potexma@gmail.com, [2] mahaveerakannanr.sse@saveetha.com, [3] priscillajoy@karunya.edu, [4]drsheebas@panimalar.ac.in, [5] narinat@gmail.com, [6] m.viky07@gmail.com

ARTICLE INFO

Received: 03 Dec 2024

Revised: 26 Jan 2025

Accepted: 08 Feb 2025

ABSTRACT

Introduction: A serious consequence of diabetes, diabetic retinopathy (DR) damages the retinal tissues besides can cause severe vision loss or complete blindness in some people. The only way to slow or stop the worsening of this ailment is to catch it early. The minor signs of DR might be difficult to discern on one's own, making early detection a challenge. An innovative approach to DR detection is shown.

Objectives: In this study, which utilises an Extreme Learning Machine (ELM) that has been optimised utilising an Improved Logical Whale Optimisation Algorithm (ILWOA). In order to improve classification accuracy, convergence speed, and robustness, the ILWOA adjusts the parameters of the ELM model.

Methods: Two benchmark datasets, EyePACS and APTOS, are used to test the methodology, making sure the model can be used for various real-world data circumstances. Overcoming issues like overfitting and sensitivity to parameter initialisation, the suggested method uses sophisticated optimisation techniques to circumvent the shortcomings of conventional ELM models. The ILWOA-optimized ELM achieves better performance metrics and surpasses existing approaches in detecting different stages of DR, according to the experimental data. The approach is also flexible and scalable, so it can handle massive ophthalmic datasets.

Conclusions: This study makes a substantial influence to the area of automated healthcare diagnostics by developing an AI-driven system for the identification of diabetic retinopathy that is dependable, efficient, and scalable. To further enhance the proposed framework's applicability in clinical applications, future efforts should focus on studying real-time implementation and multi-modal data integration.

Keywords: Diabetic Retinopathy; Extreme Learning Machine; Improved Logical Whale Optimization Algorithm; Scalability; Large-scale ophthalmic datasets.

INTRODUCTION

Diabetic retinopathy is among the most serious eye illnesses associated with diabetes, which has led to a remarkable increase in the number of persons diagnosed with the disease in recent decades [1]. In addition, the majority of cases of blindness in middle-aged adults are caused by diabetic retinopathy [2]. DR can lead to diabetes. This is due to the fact that the retina of the eye can be damaged by certain eye injuries [3]. Digital retinal imaging can be used to identify

diabetic retinopathy [4], and fundus examination is thought to be a good way to find aberrant symptoms in the eyes of diabetes patients [5].

Despite ongoing efforts, even for highly qualified clinicians, early detection of diabetic retinopathy is a time-consuming process, which can lead to treatment delays, misunderstandings, etc. [6]. The need for a reliable automated system to diagnose diabetic retinopathy has been acknowledged. Diabetic retinopathy and normal retinal imaging are the primary foci of our research [7]. Efforts that used machine learning and picture feature extraction in the past were successful [8]. Classifiers utilised for the job support vector machine (SVM), k-nearest neighbours (KNN) procedure, etc., while features utilised for the task include things like blood vessel detection, micro-aneurysms, red lesions, and hard exudates [9]. A huge percentage of cases end up being normal, despite a lot of time spent diagnosing normal cases [10], and none of the handmade features can cover all the signs of diabetic retinopathy in the photos. As a result, the autonomous diagnostic system has few real-world clinical uses. Elevated blood sugar levels are a symptom of diabetes, a metabolic disorder. The World Health Organisation estimates that 642 million people in 2015 [11, 12].

There are five categories of DR based on the severity of the disease: normal, mild, moderate, severe, and proliferative. Minimal enlargement of the retinal blood vessels causes mild diabetic retinopathy, the first stage of the development of DR [13]. Swelling and distorting of the blood vessels linked to the retina causes moderate DR [14]. When the blood arteries that supply the retina become severely clogged, a condition known as severe DR develops. Retinal proliferative DR develops when tiny blood capillaries in the retina get clogged, leading to the formation of new vessels and potential retinal injury [15]. There has been prior work on using ML and DL to classify fundus images of DR. This dataset was taken from the DR1 and MESSIDOR databases. A resultant accuracy of 74.04% was achieved in the study [16].

Incorporating augmentation and doing continuous evaluations yield the greatest outcomes for the model when applied to unseen datasets that are distinct from the EyePACS and APTOS datasets. To determine the model's generalisability, it is vital to continuously test it on varied datasets. In order to detect possible overfitting and guide changes to training methods, regular performance evaluations are essential [17]. To further reduce the likelihood of overfitting, data augmentation during training can be used to expose the model to more variances, which could make it more resilient to data that has never been seen before. For applications such as diabetic retinopathy diagnosis, it is essential to evaluate a model's generalisability and robustness by testing it on unseen datasets that are different from the EyePACS and APTOS datasets [18].

DR is a major cause of avoidable worldwide; this project aims to develop a highly precise and efficient approach for detecting it. In order to intervene promptly and enhance patient outcomes, early detection of DR is essential. Problems with scalability, parameter sensitivity, and overfitting are some of the obstacles that current methods must overcome before they may be used effectively in clinical settings. Utilising an Extreme Learning Machine (ELM) that has been fine-tuned using the Improved Logical Whale Optimisation Algorithm (ILWOA), this research presents a new framework. By improving convergence speed, accuracy, and robustness, the ILWOA aims to overcome the drawbacks of conventional optimisation methods. The generalisability of the methodology is ensured by evaluating it on the EyePACS and APTOS datasets, which are well-known for their diversified and large-scale ophthalmic data. The goal of this work to improve the accessibility of trustworthy DR screening tools and obtain better performance in categorising different phases of DR by fine-tuning the parameters of the ELM model. This will contribute to the advancement of automated healthcare diagnostics.

RELATED WORKS

The goal of Minarno et al. [19] is to use CNNs for DR classification; specifically, they want to optimise their performance by adopting the InceptionV3 architecture. Using the APTOS 2019 Blindness Finding dataset as a case study, this research investigates how various data augmentation and preprocessing strategies affect classification accuracy. To improve model generalisability and reduce overfitting, data pretreatment and augmentation are essential deep learning procedures. The research trains the InceptionV3 model with the help of preprocessing and data augmentation. The model outperforms models trained without data augmentation, with results showing an accuracy of 86.5% on training data and 82.73% on test data. Performance graphs showing a significant drop in test accuracy compared to training accuracy further prove that overfitting occurs in the lack of data augmentation. The

significance of customised preprocessing and augmentation methods in strengthening the predictive capacity and resilience of CNN models for DR finding is emphasised in this study.

The ResViT FusionNet model, developed by Ikram and Imran [20], combines the thorough knowledge offered by Vision Transformers (ViTs) with the powerful feature extraction capabilities of Convolutional Neural Networks (CNNs), notably ResNet50. Several preprocessing methods, including data augmentation, were used to make sure the datasets were balanced and to improve the presentation of the models. In an effort to make the model more generalisable and resilient, these methods rescaling pixel values, and rotation. Furthermore, to utilised Explainable AI (XAI) practices to enhance the model's forecasts by making them more visible and interpretable, particularly in clinical contexts. In order to understand the model's forecasts, LIME was employed. Then, heatmaps were generated which showed which parts of the fundus images were most important for the classification decisions. Healthcare providers can better grasp the elements impacting the forecasts and have more faith in the model with the help of these graphic explanations. When compared to top CNNs and baseline ViT models, ResViT FusionNet performs far better in experiments. With scores of 0.8944 for MCC, 0.8935 for Kappa, 0.9300 for Recall, 0.9275 for F1 Score, 0.9301 for Accuracy, and a Jaccard Index of 0.8749, the model demonstrated remarkable performance in evaluation measures.

An innovative massive retinal feature aggregation network (MRFANet) for precise multi-lesion diabetic retinopathy segmentation was suggested by Zhou and Zhang [21]. It is capable of making end-to-end pixel-level forecasts for various sorts of lesions by aggregating huge retinopathy features. To used a strategy to improve semantic representations, sequential dilated linkages to achieve hierarchical context feature extractions, and last, to aggregated the huge features to forecast the final segmentations. When approaches, the suggested method performs admirably in experiments run on the IDRiD and DDR datasets, with AUC scores of 0.7566 and 0.3286 for soft exudate segmentation on the IDRiD dataset and 0.3286 for the DDR dataset, respectively. The method also achieved top performance in all test scenarios when evaluating generalisation performance.

In order to effectively detect DR in both sequential and non-sequential fundus pictures, the model presented by Henge et al. [22] has 172 weighted layers. A multi-layered transfer learning method is employed, with 86 layers dedicated to processing colour fundus images and another 86 layers devoted to processing greyscale images. In order to handle different data matrices and incorporate global besides specialised features, the model goes through rigorous pre-processing and testing processes, with eight layers of level. The performance is enhanced by the chi-square testing process, which refines cases. By utilising multi-decision hybrid approaches, the model surpasses other current models with a detection accuracy of 98.1%.

Using StyleGAN3, Das, & Walia [23] have created synthetic DR1 pictures that are diverse and characterised by microaneurysms. The goal is to improve supervised classifier performance and deal with data shortage. To trained the model using a dataset of 2,602 DR1 pictures and then evaluated it using a battery of quantitative metrics, such as FID, KID, and EQ-T and EQ-R, which measure equivariance with respect to translation and rotation, respectively. Human Turing tests were one kind of qualitative evaluation; in these tests, qualified ophthalmologists would judge how lifelike computer-generated pictures appeared. The image quality was further confirmed using spectral analysis. Compared to the mean FID of 21.18 (95 percent obtained by bootstrap resampling, the model's final FID score of 17.29 was better. The model was able to generate very accurate visuals in human Turing tests, with the exception of some small artefacts along the edges. This research provides strong evidence that synthetic DR1 images created by StyleGAN3 can greatly improve the accuracy of DR early detection when used to enrich training datasets.

2.1. Problems statement

A. Performance on Unseen Datasets

1) Generalization Ability

By successfully identifying key features associated with diabetic retinopathy on unseen datasets, a model trained on EyePACS and APTOS datasets has learnt to do more than just memorise the training data. The model seems to have captured the fundamental characteristics of the disease if it can generalise.

2) Variability in Data

If the unseen datasets differ significantly from the training datasets in terms of image quality, acquisition methods, or demographic characteristics (e.g., age, ethnicity), the model's performance can suffer. The model's properties may not be applicable across all populations or imaging settings, which could explain this drop.

B. Evidence of Overfitting

1) Training Versus Validation Performance

If the unseen datasets differ significantly from the training datasets in terms of image quality, acquisition methods, or demographic characteristics (e.g., age, ethnicity), the model's performance can suffer. The model's properties may not be applicable across all populations or imaging settings, which could explain this drop.

2) Learning Curves Analysis

Examining learning curves can provide insights into overfitting. A significant gap between accuracy, particularly when the training accuracy continues to improve while suggests overfitting.

3) Performance Metrics

Metrics ROC curve on unseen datasets can help evaluate the model's effectiveness. A notable drop in these metrics compared to training metrics may signal overfitting.

PROPOSED METHODOLOGY

In this section, detection of DR is carried out by advanced deep learning model and it is graphically shown in Figure 1, where each of its block is mentioned in the upcoming sub-section.

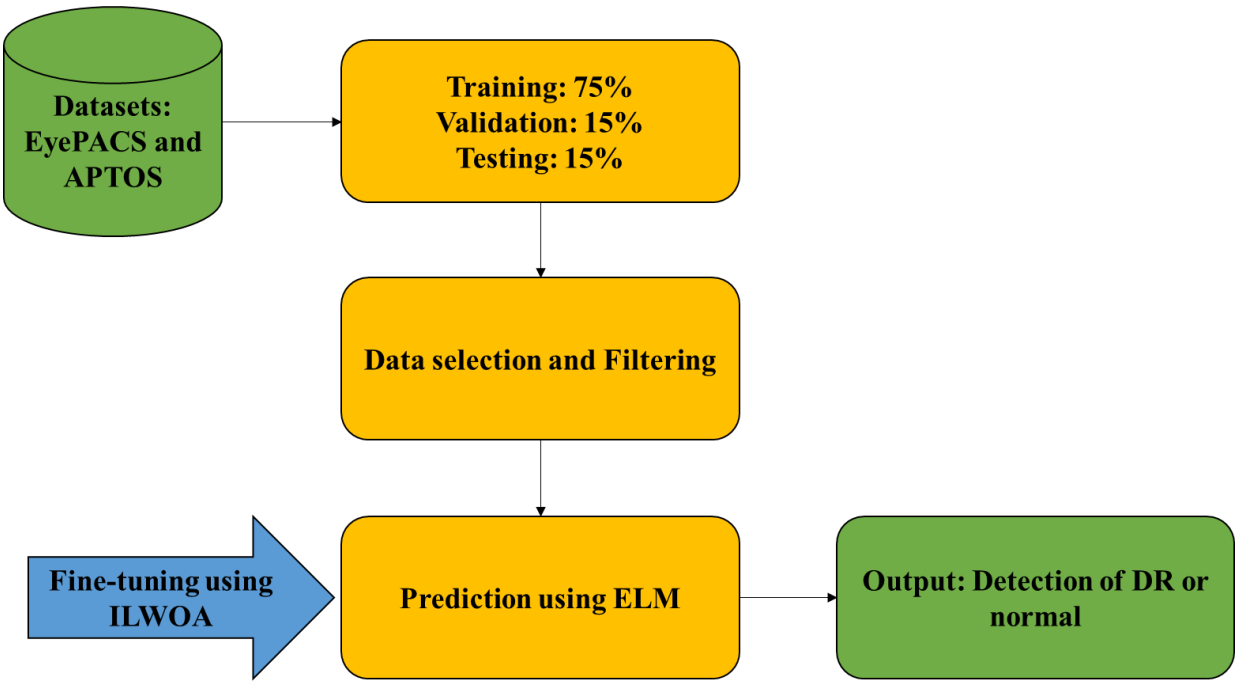

Figure 1: Workflow of the Research Model

3.1. Dataset Integration

Providing comprehensive details about the division of the EyePACS and APTOS datasets [24], the specific training cum testing splits, filtering criteria, preprocessing steps, and final dataset statistics is essential for replicability and understanding the model's presentation which will help to enhance the reliability of the findings and supports future research efforts in diabetic DR.

The EyePACS dataset fundus images used for DR detection, typically labeled according to harshness levels ranging from 0 which considered as no-DR to 4 which considered as proliferative-DR. It contains a diverse set of images representing various demographics and clinical conditions. The APTOS dataset consists of labeled fundus images for diabetic retinopathy detection. It includes images with a similar grading system but may vary in quality and acquisition methods.

3.1.1. Training, Testing and Validation Sets and Subsets

This section articulated divisions of datasets based on training, and testing sets and supporting subsets along with its validations. Often, a common tactic is to use 70-80% of the data for training, 10-15% for validation, besides the residual 10-15% for testing. In sampling, it has ensured that each subset reflects the overall distribution of diabetic retinopathy grades, stratified sampling employed, which maintains the proportion of each class in all subsets.

3.1.2. Image Selection and Filtering

In training and testing, the specific criteria initiated for selecting images from the EyePACS and APTOS through main and sub-clusters-based quality control and exclusion of anomalies. Images are controlled the filtering process based on quality metrics such as resolution, clarity, reflection-rate and lighting. In the basic stage of input, exclude anomalies such as artifacts, poor lighting, and low-resolution issues and applied augmentation and normalization techniques to the images before training the data. In normalization adjusted pixel values to a standard range, typically (0, 1) to ensure consistency across images. In augmentation, the dual-image multilayer ResNet algorithmic sequences are used to implement the image pre-processing techniques. The fundus-colored and black-and-white images are resized by 312×312×3 data matrices for each image, and various data augmentation techniques, including zooming (z), trimming (t), rotary motion (rm), horizontal spin (hs), vertical spin (vs), breadth move (bm), and height move (hm), are integrated.

3.1.3. Dataset Size, Diversity and Final Dataset Composition

In test cases, the total sum of images used from each dataset and their distribution across different. Data about the demographic diversity such as age, ethnicity of the images provided. The final composition of dataset initiated with balancing classes and ensured uniformity in preprocessing, and considered the number of training 70-80%, validation 10-15%, besides testing 10-15% of images, as well as their corresponding labels and classes.

3.2. Prediction using ELM

In this work, the prediction of DR is carried out by optimized ELM model [25], where the classic ELM model's basic design, which consists of three crucial layers. These layers consist of an output layer (y) with m neurons to represent the m output variables, a hidden layer with l neurons, and an input layer (x) with n neurons to characterise the n input variables. W_{ij} stands for the connection weight that neuron in the input neuron in the hidden layer in this context, whereas β_{jk} stands for the connection weight that connects layer [26]. The training dataset containing "Q" samples, the "X" and "Y" input and output matrices, and the "b" bias for the hidden layer are further essential elements. These components' mathematical formulations are defined by equations (1) through (3), which together make up the activation function abbreviated as $g(\cdot)$.

$$\omega = \begin{bmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1n} \\ \omega_{21} & \omega_{22} & \cdots & \omega_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{l1} & \omega_{l1} & \cdots & \omega_{ln} \end{bmatrix}_{l\times n} \quad (1)$$

$$\beta = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{l1} & \beta_{l1} & \cdots & \beta_{ln} \end{bmatrix}_{l\times m} \quad (2)$$

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_l \end{bmatrix}_{l\times 1} \quad (3)$$

Equation (4) may be used to determine the ELM neural network's output Y:

$$\Upsilon = \begin{bmatrix} \sum_{i=1}^{l} \beta_{i1} g(\omega_i x_j + b_i) \\ \sum_{i=1}^{l} \beta_{i2} g(\omega_i x_j + b_i) \\ \vdots \\ \sum_{i=1}^{l} \beta_{im} g(\omega_i x_j + b_i) \end{bmatrix}_{m \times l} \quad (4)$$

Where H is the ELM neural network's hidden layer output matrix, as illustrated in Equation (5):

$$H = \begin{bmatrix} g(\omega_1 . x_1 + b_1) & g(\omega_2 . x_1 + b_2) & g(\omega_l . x_1 + b_l) \\ g(\omega_1 . x_2 + b_1) & g(\omega_2 . x_1 + b_2) & g(\omega_l . x_2 + b_l) \\ \vdots & \cdots & \vdots \\ g(\omega_1 . x_Q + b_1) & g(\omega_2 . x_Q + b_2) & g(\omega_l . x_Q + b_l) \end{bmatrix}_{Q \times l} \quad (5)$$

Equation (6)'s least squares solution leads to the connection weights that join the hidden layer and output layer.

$\min_{\beta} \|H\beta - \Upsilon^T\|$ (6)

The response may be stated as follows:

$\beta = H^{+}\Upsilon^T$ (7)

Algorithm 1 provides an illustration of the ELM algorithm.

Algorithm 1: ELM algorithmic procedure

Step 1: There are different numbers of neurons in the three layers symbolised by the letters "n," "l," and "m," which are the layers, respectively. Additionally, random numbers are utilised to initialise the weight connection (***ω***) between the l and the n as well as the bias (b) of the l.

Step 2: The hidden layer output matrix, abbreviated as "H," is generated when the activation function for the hidden layer neurons, indicated as "g(x)," has been specified.

Step 3: The connection weight that connects an output layers to the layer that is hidden is calculated using the equation ***β***= H+YT.

Neural network technology, a ground-breaking development in artificial intelligence, is used to estimate the severity of the DR. Rapid learning, effective processing, and strong generalisation skills are just a few of the benefits provided by the ELM, a single hidden-layer feedforward neural network. However, ELM's variable number of l neurons and the setting of connection weights and thresholds at random may make predictions less precise. Our main goal has been to improve ELM's DR prediction performance by using the improved whale optimisation approach in order to overcome these restrictions.

3.3. Hyper-parameter tuning using ILWOA

An innovative algorithm for optimising population intelligence, the whale optimisation algorithm (WOA) [27] takes its cues from the three primary categories of humpback whale foraging behaviour: The first is swimming and foraging; the second is surrounding prey; and the third is attacking. The regulating coefficient A determines foraging behaviour and surrounding predators, with A taking values in the variety [−2, 2]. A s When 0 ≤ |A| ≤ 1, the algorithm achieves a predation apparatus; when 1 ≤ |A| ≤ 2, the procedure achieves a mechanism.

(1) Swim away encirclement. It is assumed in the mathematical model that all individual whales swim towards the optimal location, and that the target prey is the optimal solution in the present generation of whales.

$X(t+1) = X^*(t) - A \times |C \times X^*(t) - X(t)|$ (8)

where X(t) X^* (t) represents the ideal placement of whales in populace, A and C stand for the adjustment coefficient, and represents the position of separate whales in the first-generation populace. Here are the particular expressions:

$A = 2a \times r_1 - a$ (9)

$C = 2 \times r_2$ (10)

where a represents the junction factor that reductions linearly from 2 to 0, and r_1, r_2 symbolises a random number among [0, 1]. The specific appearance is as shadows:

$a = 2 - 2 \times \frac{t}{t_{max}}$ (11)

where t signifies the number of current recapitulations. t_{max} represents the extreme sum of iterations.

(2) Surround predation. The following expression updates the position when a humpback whale spirals up to attack after seeing prey.:

$X(t+1) = X^*(t) + D_p \times e^{b \times l} \times cos(2 \times \pi \times l)$ (12)

the shape of the defined coefficient b, besides l is a random sum between -1 and 1. D_p denotes the distance among a whale current optimal individual, as uttered by the following appearance:

$D_p = |X^*(t) - X(t)|$ (13)

The whale must swim away from its target in order to encircle it as it spirals around it. In order to depict this procedure, the researchers from Mirjalili hypothesised that the mathematical model represents the 50% chance of a humpback whale selecting -surround and surround predation.:

$$X(t+1) = \begin{cases} X^*(t) - A \times |C \times X^*(t) - X(t)| & p < 0.5 \\ X^*(t) + D_p \times e^{b \times l} \times cos(2 \times \pi \times l) & p \geq 0.5 \end{cases} \quad (14)$$

(3) Searching for prey. To circle and encircle their prey, whales move probabilistically when the absolute value of A is between 0 and 1. A typical mathematical follows: when 1 < |A| ≤ 2, the humpback whale chooses a random populace as prey for roundup instead of ideal those:

$X(t+1) = X_{rand}(t) - A \times |X_{rand}(t) - X(t)|$ (15)

where $X_{rand}(t)$ shows where one whale from the t generation population happens to be at any given time. Using chaotic mapping to initialise tactics for populations. Using better logistic chaos mapping, the whale population strategy is initiated. Using a random distribution of whale populations in the original algorithm made the initial populations less diverse and more unevenly distributed, which slowed down the algorithm's convergence and reduced its convergence accuracy. This, in turn, reduced the algorithm's search performance and made robot path planning slower and less accurate. To start the whale population out on the right foot, increase its diversity, and boost the algorithm's complete particular exploration capacity within a certain range, to apply chaotic mapping with ergodicity [27]. The algorithm is subjected to logistic chaos mapping, a technique commonly employed to initialise populations in intelligent bionic algorithms. There is a noticeable lack of improvement in population diversity, and the evenly distributed around the two ends of the interval [0, 1]. In contrast, the distribution.

There is less variation between the percentages of the various regions on the interval [0, 1] and the enhanced logistic chaotic mapping values, and the initialised particles are more uniformly dispersed. To enhance the original distribution of whale populations, the first stage of the whale optimisation method introduces the chaotic arrangement produced by the enhanced logistic chaos mapping into persons. This sequence then generates the individual positions.

Using nonlinear convergence factors as a balancing method is next covered. The initial whale optimisation method controlled the algorithm's local and global search competences with the balance parameter A. Expanding the algorithm's portion of global exploration improves its global exploration capability, which was previously lacking due to the original algorithm's inadequate global search capabilities. The algorithm's global exploration capability has been strengthened after multiple experiments were conducted to increase the A-value to 1.3, resulting in a 15% improvement in the percentage. The linear factor determines the value of A. Weak global exploration ability, accuracy, and sluggish algorithm convergence are caused by the linear reduction of a from 2 to 0. In order to slow down the rate of decrease of the improved a value compared to the original a value at the beginning of the iteration,

a nonlinear convergence method is utilised. Iteratively decreasing the iteration enhances the local search competence and speeds up the convergence of the procedure, while using larger values at the repetition increases the global exploration capability and search accuracy.

While the nonlinear factorial balancing strategy does enhance the procedure's global search performance, it hinders the algorithm's local search capability and reduces the line accuracy of the robot path planning due to the random numbers affecting the value of A in Equations (9) and (10). Hence, to enhance the algorithm's local search capabilities, a weighting approach is incorporated into the site update formula for both global and local nonlinear factors. The next step is to modify the location update approach using the Corsi variant. After the aforementioned strategy improvement, the WOA is updated with the new position. However, instead of actively updating the target position after each iteration, it relies on the new position. As a result, the robot is prone to falling into local optimum in late iterations and eventually resorting to local planning. In light of this, the Corsi variation technique is implemented to update the target position through random perturbations; this enhances the algorithm's search capabilities and accuracy while preventing it from sliding into a local optimum.

RESULTS AND DISCUSSION

Research for the proposed classical was conducted using Python's deep learning toolbox and Google Colab. The NVIDIA Quadro P4000, a graphics card with 8 GB of RAM, was utilised as the GPU for both testing and training. To assess the suggested models, to divided the benchmark datasets into a training set besides a test set, and then used a validation method to train and test the models. Several hyper-parameters need to be set in order for the suggested architecture to be employed throughout the prediction process. In order to get the most out of the architecture, this is necessary. These hyperparameters include batch size, learning rate, and epochs.

4.1. Performance Metrics

Popular assessment criteria such as accuracy, ROC, and F1 score were used to assess our approach. These measurements are defined in connection to TPs, FPs, TNs, and FNs, or true positives and false negatives, correspondingly.

The percentage of instances that are accurately predicted sum of instances is called accuracy. It can be stated mathematically in the following way:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (16)$$

Accuracy is the percentage of true positives relative to the total sum of positives (including false positives and correctly classified samples), where n is the number of positives. It can be stated mathematically in the following way:

$$Precision = \frac{TP}{TP+FP} \quad (17)$$

The percentage of identified relative to the total sum of positive samples is called recall. Here is the mathematical expression:

$$Recall = \frac{TP}{TP+FN} \quad (18)$$

The F1 score, which is usually falls between 0.0 and 1.0. An improved model's performance is score, which shows a better balance among recall and precision. It can be stated mathematically in the following way:

$$F1 = \frac{2\times Precision\times Recall}{Precision+Recall} = \frac{2\times TP}{2\times TP+FP+FN} \quad (19)$$

An important measure for assessing model presentation, ROC AUC shows how well the model performs in normal and attack conditions when it comes to categorisation. A model is considered effective if its ROC value is high. Additionally, to take into account sensitivity and specificity, which are the same as recall.

4.2. Validation analysis of proposed classical on Two datasets

Table 1 besides Figure 2 delivers the experimental study of proposed model on EyePACS in terms of different metrics.

Table 1: Validation analysis of projected classical on EyePACS

Models	Accuracy	Precision	Recall	F1 score
XGBoost	0.90	0.91	0.91	0.90
MLP	0.91	0.94	0.93	0.93
ELM	0.95	0.96	0.97	0.94
ELM-ILWOA	0.98	0.98	0.99	0.98

Validation analysis of EyePACS dataset, comparing classifiers based on Accuracy, Precision, Recall, besides F1-score. The XGBoost Precision of 0.91, Recall of 0.91, F1-score of 0.90, demonstrating a reliable but moderate presentation. The MLP model improves with an Accuracy of 0.91, Precision of 0.94, Recall of 0.93, and an F1-score of 0.93, indicating better classification effectiveness. The ELM classifier further Precision of 0.96, Recall of 0.97, and an F1-score of 0.94, reflecting a more robust predictive capability. The proposed ELM-ILWOA model outperforms all others, achieving the highest Recall of 0.99, besides an F1-score of 0.98, signifying superior classification presentation with minimal misclassification errors.

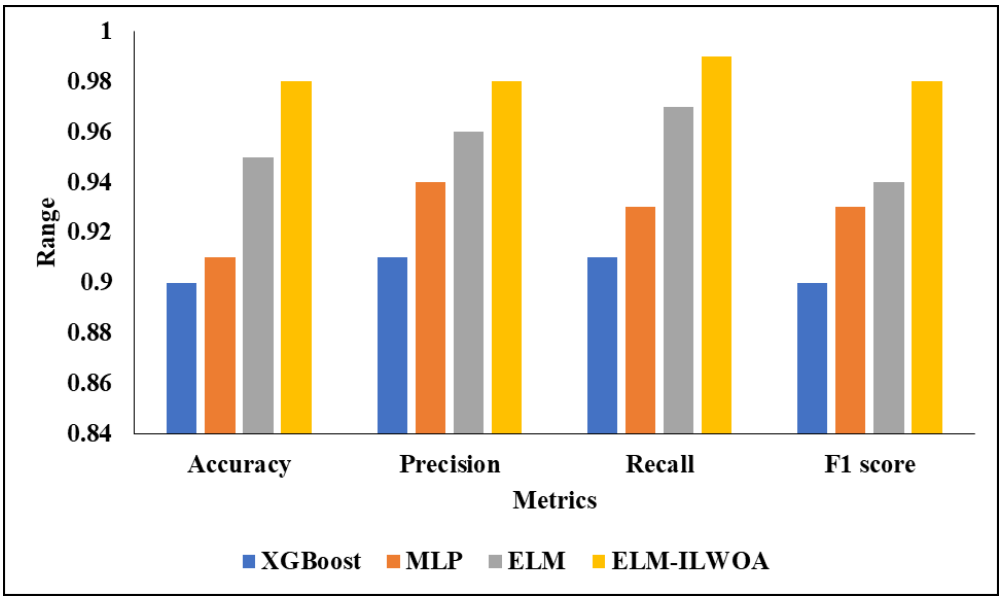

Figure 2: Graphical Comparison of wished-for classical on First dataset

Table 2 and Figure 3 provides the comparative investigation of projected model with existing techniques on APTOS.

Table 2: Validation Analysis of projected classical on proposed model on second dataset

Classifier	Accuracy	Precision	Recall	F1 score
XGBoost	0.9559	0.9258	0.9559	0.8956
MLP	0.9692	0.9418	0.9692	0.9059
ELM	0.9714	0.9634	0.9714	0.9192
ELM-ILWOA	0.9803	0.9704	0.9873	0.9336

Validation analysis of model on the second dataset, evaluating classifiers based on Accuracy, Precision, Recall, besides F1-score. The XGBoost classifier achieves an Accuracy of 0.9559, Precision of 0.9258, Recall of 0.9559, besides an F1-score of 0.8956, showing competitive but slightly lower performance. The MLP model improves upon XGBoost Accuracy of 0.9692, Precision of 0.9418, Recall of 0.9692, besides an F1-score of 0.9059, demonstrating enhanced classification capability. The ELM classifier further optimizes performance, achieving 0.9714 Accuracy, 0.9634 Precision, 0.9714 Recall, besides 0.9192 F1-score, indicating a more balanced and robust classification. The

proposed ELM-ILWOA model outperforms all others, Accuracy of 0.9803, Precision of 0.9704, Recall of 0.9873, besides F1-score of 0.9336, signifying superior predictive performance with minimal misclassification errors.

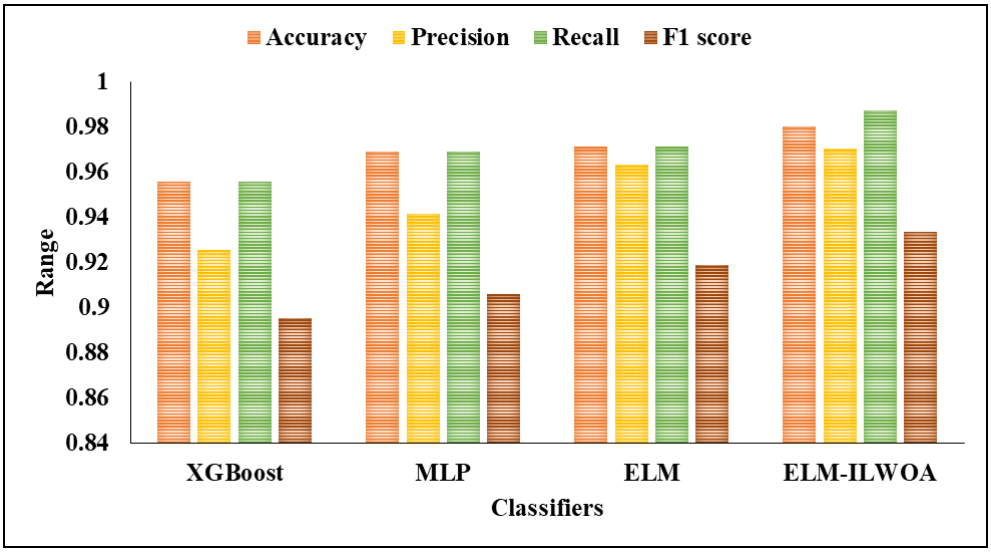

Figure 3: Visual Analysis of various models on APTOS

4.3. Analysis of proposed classical on combined dataset

Table 3 and Figure 4 gives the experimental enquiry of projected classical with existing techniques in terms of diverse metrics by combining two datasets such as EyePACS and APTOS.

Table 3: Comparative analysis of proposed model on combined datasets

Model	Accuracy	Precision	Recall	F1
XGBoost	0.94	0.93	0.93	0.94
MLP	0.96	0.95	0.94	0.95
ELM	0.97	0.97	0.96	0.97
ELM-ILWOA	0.99	0.99	0.98	0.98

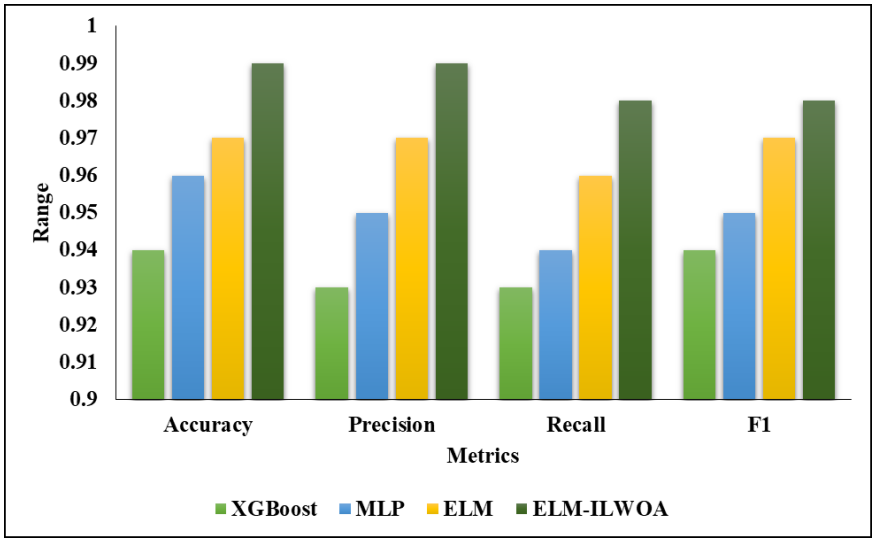

Figure 4: Visual Representation of various models on Combined dataset.

A comparative analysis of the projected model on combined datasets using key performance metrics: score. The XGBoost model achieves an Accuracy of 0.93, besides an F1-score of 0.94, representative strong presentation compared to other models. MLP improves upon XGBoost, attaining Recall of 0.94, besides an F1-score of 0.95, reflecting enhanced predictive capability. ELM further optimizes Accuracy of 0.97, Precision of 0.97, Recall of 0.96, and an F1-score of 0.97, demonstrating a classification. The proposed ELM-ILWOA model outperforms all others, achieving an Recall of 0.98, besides F1-score of 0.98, signifying superior classification presentation with minimal misclassifications.

CONCLUSION

The improved logical whale optimisation algorithm (ILWOA) fine-tuned Extreme Learning Machine (ELM) is used in this study to diagnose diabetic retinopathy (DR). Using the EyePACS and APTOS datasets for testing, the suggested strategy tackles the urgent requirement for effective and precise DR detection. The method improves the ELM classifier's capacity to distinguish between diverse phases of DR by using ILWOA, which allows for optimal parameter adjustment. The experimental consequences show that the ILWOA-optimized ELM model outperforms the conventional techniques in terms of resilience, convergence speed, and classification accuracy. Using the EyePACS and APTOS datasets demonstrates how well the model works with large-scale ophthalmic data in the actual world. The algorithm can be fine-tuned to optimise its hyperparameters, which makes it reliable and scalable for use in clinical settings. In order to improve patient outcomes and prevent vision loss, this work adds to the expanding area of AI-based healthcare solutions by providing a dependable tool for early DR identification. To increase the reach and influence of this groundbreaking approach, future studies can investigate ways to improve feature extraction, integrate data from several modalities, and use it in real-time.

REFRENCES

[1] Saranya, P., & Umamaheswari, K. M. (2024). Detection of exudates from retinal images for non-proliferative diabetic retinopathy detection using deep learning model. Multimedia Tools and Applications, 83(17), 52253-52273.

[2] Abushawish, I. Y., Modak, S., Abdel-Raheem, E., Mahmoud, S. A., & Hussain, A. J. (2024). Deep learning in automatic diabetic retinopathy detection and grading systems: a comprehensive survey and comparison of methods. IEEE Access, 12, 84785-84802.

[3] Costaner, L., Lisnawita, L., Guntoro, G., & Abdullah, A. (2024). Feature Extraction Analysis for Diabetic Retinopathy Detection Using Machine Learning Techniques. SISTEMASI, 13(5), 2268-2276.

[4] Muthusamy, D., & Palani, P. (2024). Deep learning model using classification for diabetic retinopathy detection: an overview. Artificial Intelligence Review, 57(7), 185.

[5] Ikram, A., Imran, A., Li, J., Alzubaidi, A., Fahim, S., Yasin, A., & Fathi, H. (2024). A systematic review on fundus image-based diabetic retinopathy detection and grading: Current status and future directions. IEEE Access.

[6] Gunapriya, M. B. B., Rajesh, T., Thirumalraj, A., & Manjunatha, B. (2023). LW-CNN-based extraction with optimized encoder-decoder model for detection of diabetic retinopathy. FRONTIER SCIENTIFIC PUBLISHING PTE. LTD, 1095.

[7] Shamrat, F. J. M., Shakil, R., Akter, B., Ahmed, M. Z., Ahmed, K., Bui, F. M., & Moni, M. A. (2024). An advanced deep neural network for fundus image analysis and enhancing diabetic retinopathy detection. Healthcare Analytics, 5, 100303.

[8] Kommaraju, R., & Anbarasi, M. S. (2024). Diabetic retinopathy detection using convolutional neural network with residual blocks. Biomedical Signal Processing and Control, 87, 105494.

[9] Latha, G., Priya, P. A., & Smitha, V. K. (2024). Enhanced diabetic retinopathy detection and exudates segmentation using deep learning: A promising approach for early disease diagnosis. Multimedia Tools and Applications, 1-24.

[10] Jabbar, A., Liaqat, H. B., Akram, A., Sana, M. U., Azpíroz, I. D., Diez, I. D. L. T., & Ashraf, I. (2024). A Lesion-Based Diabetic Retinopathy Detection Through Hybrid Deep Learning Model. IEEE Access.

[11] Gade, N. R., & Mahaveerakannan, R. (2024, October). Improving Access to Skin Care with a Multi-Factorial Strategy Using Vision Transformers. In 2024 4th International Conference on Sustainable Expert Systems (ICSES) (pp. 1189-1193). IEEE.

[12] Naik, S., Kamidi, D., Govathoti, S., Cheruku, R., & Mallikarjuna Reddy, A. (2024). RETRACTED ARTICLE: Efficient diabetic retinopathy detection using convolutional neural network and data augmentation: Efficient diabetic retinopathy detection using convolutional neural network and data... Soft Computing, 28(Suppl 2), 617-617.

[13] Guefrachi, S., Echtioui, A., & Hamam, H. (2024). Diabetic Retinopathy Detection Using Deep Learning Multistage Training Method. Arabian Journal for Science and Engineering, 1-18.

[14] Gupta, S., Thakur, S., & Gupta, A. (2024). Comparative study of different machine learning models for automatic diabetic retinopathy detection using fundus image. Multimedia Tools and Applications, 83(12), 34291-34322.

[15] Nazir, K., Kim, J., & Byun, Y. C. (2024). Enhancing Early-Stage Diabetic Retinopathy Detection Using a Weighted Ensemble of Deep Neural Networks. IEEE Access.

[16] Selvamuthukumaran, N., Aravinda, K., Manjunatha, B., & Thirumalraj, A. (2025). Breast Cancer Detection Using Mother Optimisation Algorithm Based Chaotic Map with Private AI Model. In Sustainable Development Using Private AI (pp. 278-294). CRC Press.

[17] Lalithadevi, B., & Krishnaveni, S. (2024). Diabetic retinopathy detection and severity classification using optimized deep learning with explainable AI technique. Multimedia Tools and Applications, 1-65.

[18] Mahaveerakannan, R., Rajakumar, B., Sharma, R. R., & Pushpa, S. (2024, September). Optimized Feature Selection and Transfer Learning for Accurate Fundus Image Classification in Glaucoma Detection. In 2024 3rd International Conference for Advancement in Technology (ICONAT) (pp. 1-6). IEEE.

[19] Minarno, A. E., Bagaskara, A. D., Bimantoro, F., & Suharso, W. (2025). Classification of Diabetic Retinopathy Based on Fundus Image Using InceptionV3. JOIV: International Journal on Informatics Visualization, 9(1), 23-28.

[20] Ikram, A., & Imran, A. (2025). ResViT FusionNet Model: An explainable AI-driven approach for automated grading of diabetic retinopathy in retinal images. Computers in Biology and Medicine, 186, 109656.

[21] Zhou, W., & Zhang, Q. (2025). MRFANet: Massive retinopathy feature aggregation network for pixel-level diabetes-induced lesion detection from fundus images. Biomedical Signal Processing and Control, 103, 107415.

[22] Henge, S. K., Viraati, N. R., Alhussein, M., Kushwaha, A. S., Aurangzeb, K., & Singh, R. (2025). Detection of Diabetic Retinopathy Using a Multi-Decision Inception-ResNet-Blended Hybrid Model. IEEE Access.

[23] Das, S., & Walia, P. (2025). Enhancing Early Diabetic Retinopathy Detection through Synthetic DR1 Image Generation: A StyleGAN3 Approach. arXiv preprint arXiv:2501.00954.

[24] Dao, Q. T., Trinh, H. Q., & Nguyen, V. A. (2023). An effective and comprehensible method to detect and evaluate retinal damage due to diabetes complications. PeerJ Computer Science, 9, e1585.

[25] Bilal, A., Imran, A., Liu, X., Liu, X., Ahmad, Z., Shafiq, M., ... & Long, H. (2024). BC-QNet: A quantum-infused ELM model for breast cancer diagnosis. Computers in Biology and Medicine, 175, 108483.

[26] Aluvalu, R., Sharma, T., Viswanadhula, U. M., Thirumalraj, A. D., Prasad Kantipudi, M. V. V., & Mudrakola, S. (2024). Komodo Dragon Mlipir Algorithm-based CNN Model for Detection of Illegal Tree Cutting in Smart IoT Forest Area. Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science), 17(6), 1-12.

[27] Mahaveerakannan, R., Choudhary, S. L., Dixit, R. S., Mylapalli, S., & Kumar, M. S. (2024, October). Enhancing Diagnostic Accuracy and Early Detection Through the Application of Deep Learning Techniques to the Segmentation of Colon Cancer in Histopathological Images. In 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 1809-1815). IEEE.