

^{1,2}Anamika Raj¹Noor Maizura
Mohamad Noor¹Rosmayati
Mohamad¹Noor Azliza Che
Mat²Shahid Hussain

An Efficient Feature Selection and Extraction using Metaheuristic Technique for Diabetic Retinopathy



Abstract: - Diabetic retinopathy (DR) is associated with diabetes, which causes harm to the retina as a result of persistent elevated blood sugar levels, resulting in symptoms such as impaired vision. Regular eye examinations are crucial for the timely identification of any issues. The process of selecting characteristics in DR involves optimizing important features using a combination of Hybridized Weight-Optimized Particle Swarm Optimization and Whale Optimization Algorithm (HWOPSO-WOA). This leads to an enhancement in the accuracy of detection, which is crucial for appropriate diagnosis and therapy. Pre-processing encompasses the utilization of linear filters such as Sobel and Prewitt for image manipulation, mean filters for the reduction of noise, and Gaussian filters for the purpose of smoothing. Feature selection employs an objective function that considers significant performance measure, which utilizes binary encoding and conducts fitness evaluation. The methodology is utilized for the analysis of DR utilizing the Indian Diabetic Retinopathy Image Dataset. The results comprise convergence graphs, and a comparative analysis, which emphasize the greater accuracy and efficiency of the Proposed technique. Visual representations, such as fundus images with selected features, highlight the importance of the chosen features in detecting DR. The proposed HWOPSO-WOA achieves highest accuracy of 97.3% and minimal processing time 98.45seconds that outperforms the state-of-art techniques.

Keywords: Optimization, convergence, feature selection, accuracy, processing time, and diabetic retinopathy.

Introduction

Computer-Aided Diagnosis (CAD) is extensively utilized in clinical environments to identify and forecast different diseases through the analysis of medical images. Scientists utilize several imaging techniques to enhance the early identification, evaluation of treatment effectiveness, and implementation of therapeutic approaches for serious diseases. The interdisciplinary field of study has garnered significant interest from scientists and physicians, resulting in advancements in computerized systems for medical imaging and analysis.

Glaucoma is a retinal disease marked by elevated pressure within the eye, leading to harm to the optic nerve and eventual loss of vision. Glaucoma affects the optic nerve head during its early stages [1]. The cup-to-disc ratio is a diagnostic measure used to evaluate the probability of getting glaucoma. The average cup to disc ratio is 0.3. An expanded cup is a clear indication of the presence of glaucoma. The cup is synonymous with the middle pale region of the optic disc [2].

Diabetic Retinopathy (DR) is the primary cause of visual impairment in working-age persons in developed countries, stemming from complications related to diabetes mellitus [3]. While diabetes does not necessarily lead to vision loss, approximately 2% of individuals with the condition eventually experience blindness, and 10% observe a deterioration in their eyesight after 15 years. The global prevalence of diabetes is projected to increase from 2.8% in 2000 to 4.4% by 2030, leading to a significant increase in the number of individuals affected, from

¹ Faculty of Ocean Engineering Technology and Informatics, University Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia

²Department of Computer Science, Applied College Al Mahala, 61421, King Khalid University, Saudi Arabia

P5301@pps.umt.edu.my,

maizura@umt.edu.my ,

rosmayati@umt.edu.my ,

azliza@umt.edu.my ,

shosain@kku.edu.sa

171 million to 366 million. Timely identification and immediate intervention are crucial in averting visual impairment in this progressive ocular condition [4].

Timely recognition and assessment of *DR* are essential for safeguarding the patient's visual acuity. Automated techniques for identifying *DR* have the capacity to replace manual approaches by significantly diminishing the level of manual labor needed throughout the screening procedure [5]. To enhance the efficiency of screening a larger population, the system can be programmed to differentiate between normal and abnormal cases, hence removing the necessity of manually examining each image. Multiple automated techniques have been recorded for quantifying the changes in the morphology of retinal blood vessels (such as width and tortuosity), which could potentially suggest the existence of retinal or cardiovascular problems. *DR* is the most common eye condition associated with diabetes and a leading cause of blindness [6].

Feature extraction is a method of reducing the number of dimensions in an image by efficiently representing its noteworthy components as a concise feature vector [7]. This method is advantageous for handling big image sizes, with the goal of efficiently performing tasks such as image matching and retrieval by leveraging compact feature representations [8]. It integrates feature detection, extraction, and matching to successfully tackle computer vision tasks such as object identification, recognition, content-based image retrieval, face detection, recognition, and texture classification. Feature extraction is performed prior to applying classifier algorithms on the transformed data in feature space [9].

Feature selection and extraction are significant steps in medical image analysis, especially in tasks namely identifying *DR* from fundus images [10]. The necessity for employing these strategies emerges due to the inherent difficulties presented by high-dimensional data, wherein the multitude of characteristics can result in computational inefficiencies, overfitting, and diminished interpretability [11]. By meticulously choosing or extracting relevant characteristics, these techniques enable the reduction of dimensions, simplifying computational processes and enhancing the effectiveness of machine learning models. Moreover, feature selection improves the accuracy and interpretability of the models by prioritizing the most useful features of the data that are important to *DR* [12]. Reducing dimensionality also helps to decrease noise and handle collinearity, ensuring that the chosen features have a significant impact on the study. Furthermore, in settings with limited resources, such as those frequently seen in medical imaging, careful selection of features helps to preserve vital computational resources. Feature selection and extraction are crucial steps in the preprocessing pipeline, serving a fundamental function in refining the input data for precise and quick diagnosis of *DR*.

Metaheuristic algorithms provide an alternate approach to conventional machine learning methods for the selection and extraction of features in the context of *DR* fundus images. Metaheuristic techniques, including *Genetic Algorithms (GA)*, *Particle Swarm Optimization (PSO)*, and *Simulated Annealing (SA)*, are used to systematically explore the feature space and select a subset of unique traits [13]. These strategies enhance the selection process by assessing the performance of several feature subsets using a predetermined fitness function. In addition, metaheuristic methods can be expanded to include feature extraction by refining the parameters of image processing algorithms or by adjusting the configuration of texture and form descriptors. Researchers intend to improve the effectiveness of *DR* diagnosis by utilizing the optimization capabilities of metaheuristic algorithms. This approach enhances the discriminatory strength of selected characteristics and does not rely on traditional machine learning methodologies [14].

Research Motivation

The motivation behind feature selection and extraction arises from the difficulties presented by high-dimensional data, namely in the field of medical image analysis, such as the identification of *DR* from fundus images. The extensive array of characteristics included in these datasets can result in computational complexities, delineates processing durations, and the potential for overfitting. Hence, the main objective is to decrease the number of dimensions, thereby enhancing the computing efficiency of following studies. The primary objective of this technique is to optimize the performance of machine learning models by prioritizing the most pertinent features, hence enhancing their accuracy in tasks such as diagnosing *DR*. Furthermore, the process of selecting and extracting features enhances the interpretability of models, hence facilitating a more comprehensive comprehension of the elements that influence diagnostic findings. The importance of these techniques is further

emphasized by their ability to reduce noise, handle collinearity, and conserve resources. This is particularly relevant in healthcare settings, where computational resources are frequently limited. The primary rationale behind feature selection and extraction is their combined capacity to enhance and optimize the data, facilitating more streamlined and impactful studies in medical image-based applications.

Research Objective

The research objectives are

- Create a systematic approach for selecting features with the goal of identifying the best subset of characteristics from the dataset on *DR*.
- Enhance the feature selection process by merging *Weight-Optimized Particle Swarm Optimization (WOPSO)* and *Whale Optimization Algorithm (WOA)* using a hybrid technique.
- Develop and apply the *Weight-Optimized Particle Swarm Optimization (WOPSO) algorithm*, which involves adjusting particle locations and velocities using a weight optimization mechanism. Concurrently, utilize the *Whale Optimization Algorithm (WOA)* to revise the placements of whales, taking advantage of their ability to explore and exploit.
- Revise the feature subsets using the adjusted positions derived from the *hybridized WOPSO-WOA* algorithm.
- Incorporate a convergence check by monitoring criteria such as a predefined maximum number of iterations or reaching sufficient fitness levels. Cease the process if the specified conditions are satisfied.

The remainder of the article is organized as follows: overview, research objective, and motivation is given in Section 1, the related works with research gap is detailed in Section 2, the proposed *Hybridizing Weight Optimized Particle Swarm Optimization and Whale Optimization Algorithm (HWOPSO-WOA)* for feature selection and extraction is illustrated in Section 3, the outcome of the *HWOPSO-WOA* and existing *state-of-art* technique is illustrated in Section 4, and the research is concluded in Section 5.

Related Works

Gu et al. (2023) presented a smart *DR* classification model for fundus pictures. All the stages of severe, and proliferative *DR* can be detected with the proposed approach. There deployed two primary modules into the concept. The fundus images were primarily processed by the *Feature Extraction Block (FEB)*, while the *Grading Prediction Block (GPB)* was developed to be responsible for categorising the five stages of *DR* based on their appearance. The transformer in the *FEB* has finer-grained attention and can focus more on areas of retinal haemorrhage and exudate. The *GPB's* residual attention was well-suited in identifying the areas of space inhabited by various types of objects. Extensive studies using *DDR* datasets proved that the technique is superior, and when compared to the gold standard, it performed comparably [15].

Three step preprocessing to improve image, including noise filtering, artefact removal, and contrast enhancement, (Nage et al. 2023) have detected *DME* at an earlier stage. Then blood vessel segmentation was done using improved *Convolutional Neural Network (CNN)* called *Masked CNN* to improve the accuracy and precision rate of Diabetic Retinopathy and Diabetic Macular Edema (*ME*). At last, *VGG-16* was used to extract structural features, colour features, and orientation information, before being put to use in a classification process. The projected *VGG-16* uses the retrieved features to sort the image into the normal, *DR*, and *DME* categories. Conditional entropy was then used to categorise the disease severity into mild, moderate, and severe after *DR* and *DME* that have been detected. The proposed work was tested on the *IDRiD* and *MESSIDOR* datasets. The proposed model has proved 80.7% of accuracy when compared to existing methods [16].

Bajwa et al. (2023) have created and field-tested a deep learning model that has been put to real-world use at *SIOVS*. The quality of the test images were evaluated using the sophisticated model. The deployed model classified the assessed images as *DR-Positive* or *DR-Negative*. Also, clinical experts analyzed the data to determine the model's efficacy. A total of 398 patients were screened over the course of five weeks. This model uses test data annotated by clinical specialists in *DR* and achieved 93.72% of accuracy, 97.30% of sensitivity and 92.90% of specificity [17].

Kundu et al. (2022) use a nested *U-Net* to segment red lesions and subsequently filter out false positives using a sub-image classification method. Researchers investigated the potential for lowering the false positive rate by experimenting with various sub-image sizes for categorization. The network was able to collect semantic characteristics and fine details because of the dense by skip connections between down sampling and up sampling channels. As a result of the efforts, the number of false-positive detections was greatly reduced. The *DIARETDB1* dataset showed that the proposed framework was 88.79% sensitive, 71.50% precise, and 79.21% F1-Score [18].

Skouta et al. (2022) detailed a modified *CNN UNet* architecture for detecting retinal haemorrhages in the *IDRiD* dataset to segment and detect potential regions that may harbour retinal haemorrhages. Experimental outcomes were also verified using the freely available *IDRiD* and *DIARETDB1* datasets. By employing preprocessing, it was able to improve the image quality and increase the amount of data, both of which are essential when defining the complicated features employed in the segmentation process. In the end, the trained neural network was noticeably more effective, correctly segmenting the bleeding with an 80.49 percent sensitivity, a 99.68 percent specificity, and an accuracy of 98.68 percent. Experiments also showed an *IoU* of 76.61 percent and a Dice value of 86.51 percent, indicating the reliability of the network's predictions and its potential to workloads. These results demonstrated a significant advancement in the ability to diagnose *DR*, a disease of the retina that is particularly important to monitor [19].

Diabetic Retinopathy (DR) is a serious eye condition that occurs as a result of long-term diabetes, leading to structural and functional alterations in the retina. The timely identification of retinal abnormalities and classification of *DR* grades is of utmost importance, and this study suggests the use of a sophisticated *Deep Neural Network (DNN)* combined with segmentation techniques for this purpose. The process of accurately segmenting lesions involves the use of saliency detection, the use of structural tensor, and the approximation of active contours. The severity levels are determined by calculating ratios, and a *VGG-19* obtains a sensitivity of 82% and an accuracy of 96% in classifying *DR* grades using the *KAGGLE* fundus picture dataset [20].

Retinal fundus image analysis is employed as an initial screening technique to determine the stage of *DR*, a significant eye problem associated with diabetes. This study utilizes deep learning techniques, specifically fine-tuning the *InceptionV3* convolutional neural network architecture with pre-trained weights, to identify and categorize the existence and intensity of *DR*. The fine-tuning method improves the ability of the model to learn, resulting in shorter training time and higher accuracy in categorization. This assists ophthalmologists in making diagnoses that are both efficient and accurate [21].

The current body of research on the classification and segmentation of *DR* exposes various significant deficiencies. Although a study has developed a complex model for classifying *DR*, there is a lack of research in investigating attention mechanisms, specifically in the *Grading Prediction Block (GPB)*. Additional exploration of attention mechanisms has the potential to improve the model's capacity to interpret and classify various phases of diabetic retinopathy. A different study suggests a three-step preprocessing methodology for the early diagnosis of *Diabetic Macular Edema (DME)*. However, there is a research gap in the requirement for a more thorough investigation of preprocessing methods. The current state of research on deep learning models lacks comprehensive analysis in real-world settings, resulting in a gap in knowledge on the models' effectiveness in practical situations, user input, and their ability to adapt to different patient populations. The study suggests using a nested *U-Net* model for segmenting lesions. However, there is a lack of research on how to optimize the sizes of sub-images for categorization. Additional exploration of sub-image dimensions could reduce the occurrence of incorrect identifications and improve the precision of picture segmentation. A different study introduces a Convolutional Neural Network (*CNN*) called *UNet* for detecting ocular hemorrhages. However, there is a lack of research in investigating the influence of preprocessing on image quality and the use of data augmentation. Furthermore, the study aims to investigate the efficacy of a *Deep Neural Network (DNN)* and segmentation approaches on *KAGGLE* fundus images. However, there is a lack of research on the extent to which this methodology can be used to various datasets. Rectifying these deficiencies would enhance and progress the approaches used for detecting diabetic retinopathy.

Feature Selection and Feature Extraction Hybridized Weight Optimized Particle Swarm Optimization and Whale Optimization

The process of *Hybridizing Weight Optimized Particle Swarm Optimization and Whale Optimization Algorithm (HWOPSO-WOA)* for feature selection and extraction entails merging the advantageous aspects of both metaheuristic algorithms. This section explains the process of *HWOPSO-WOA*.

Pre-Processing

A linear filter, commonly referred to as a convolution filter, is a basic filter used in image processing. The process entails performing a convolution operation on the input image using a kernel or filter mask. The kernel comprises coefficients that are multiplied by the pixel values in the vicinity of each pixel in the input image. Linear filters possess versatility and can be specifically engineered to execute many functions, including blurring, sharpening, edge detection, and noise reduction. Notable instances comprise the Sobel and Prewitt filters utilized for the purpose of detecting edges. The mean filter, alternatively referred to as a box filter or averaging filter, is a linear filter that substitutes the value of each pixel in the image with the average value of the surrounding pixels. This filter is highly efficient in eliminating noise and generating a smoothing effect on the image. Nevertheless, it can also lead to a diminishment of intricate particulars. The mean filter is computationally efficient and commonly employed in preprocessing stages.

The Gaussian filter is a type of filter used for smoothing data, which is based on the mathematical concept of the Gaussian distribution. The purpose of this technique is to apply a Gaussian kernel to the image, resulting in image blurring and noise reduction. The Gaussian filter has a response that resembles a bell curve, and its ability to blur is determined by the standard deviation (σ) of the Gaussian distribution. An increased value of σ leads to a wider filter and a more pronounced blurring effect. Gaussian filters are frequently used as a preliminary step before to more intricate image analysis jobs in order to diminish noise and augment the visibility of characteristics.

Feature Selection

The objective function, $f(X)$, quantifies the quality of feature subsets and is designed to strike a balance between maximizing classification performance and minimizing the number of selected features. A common formulation involves a weighted sum of classification metrics and a penalty term for the feature count in Equation 1.

$$f(X) = w_1 \cdot Accuracy + w_2 \cdot Sensitivity + w_3 \cdot Specificity - w_4 \cdot Penalty(X) \text{-----(1)}$$

where X represents the binary feature subset vector, $Penalty(X)$ penalizes the number of selected features, and it can be defined as $Penalty(X)=k \cdot Count(X)$, where k is a penalty coefficient, and $Count(X)$ is the number of selected features. Accuracy, Sensitivity, and Specificity are the performance metrics obtained from a classifier trained on the selected features. The objective is to maximize $f(X)$ by optimizing the feature subset X .

For both the *Weight Optimized Particle Swarm Optimization (WOPSO)* and *Whale Optimization Algorithm (WOA)*, initialize the population with random solutions. Each solution represents a potential feature subset, encoded as a binary vector in Equation 2.

$$X_i = [x_{i1}, x_{i2}, \dots \dots \dots, x_{in}] \text{-----(2)}$$

where X_i is the binary feature subset for the i th solution in the population, and x_{ij} is the binary value (0 or 1) indicating the presence or absence of the j -th feature in the subset.

Initialize populations for *WOPSO* and *WOA* in Equation 3 and Equation 4.

$$P_{WOPSO} = [X_1, X_2, \dots \dots \dots, X_{pop_size}] \text{-----(3)}$$

$$P_{WOA} = [X_1, X_2, \dots \dots \dots, X_{pop_size}] \text{-----(4)}$$

The size of the populations (pop_size) is determined based on the specific requirements of the optimization task. The goal is to evolve these populations using the *hybridized WOPSO-WOA* algorithm to converge towards an optimal feature subset that maximizes the defined objective function. Binary Encoding represent feature subsets as binary strings, where each bit corresponds to the presence (1) or absence (0) of a feature in the subset. For a feature subset of size n , the binary encoding in Equation 2.

$$X_i = [x_{i1}, x_{i2}, \dots, x_{in}]$$

where X_i is the binary feature subset for the i^{th} solution in the population and x_{ij} is the binary value indicating the presence (1) or absence (0) of the j^{th} feature in the subset.

This binary encoding facilitates the exploration of different combinations of features within the population. To evaluate the fitness of each solution in the population, train a classifier on the selected features using a machine learning algorithm. Common deep learning classifier is suitable for the specific task. Let $C(X_i)$ denote the classifier trained on the i^{th} feature subset X_i . Utilize classification performance metrics (e.g., accuracy, sensitivity, specificity) that is obtained from the trained classifier to quantify the fitness of each solution. The objective function defined in the problem formulation step (Equation 1) is used for this purpose. The fitness value represents the quality of the feature subset, considering both the classification performance and the penalty for feature count. Repeat this process for all solutions in the population to obtain their fitness values. This fitness evaluation step is crucial for guiding the evolutionary process, where solutions with higher fitness values are more likely to be selected for reproduction and crossover operations in the subsequent steps of the *hybridized WOPSO-WOA* algorithm.

Update each particle's position X_i and velocity V_i using the WOPSO algorithm. The position and velocity updates are determined by the following Equations 5 and 6.

$$V_{ij}^{t+1} = w \cdot V_{ij}^t + c_1 \cdot r_1 \cdot (P_{ij}^t - X_{ij}^t) + c_2 \cdot r_2 \cdot (G_j^t - X_{ij}^t) \text{ -----(5)}$$

$$X_{ij}^{t+1} = X_{ij}^t + V_{ij}^{t+1} \text{ -----(6)}$$

where X_{ij}^t is the j^{th} bit of the i^{th} particle's position at iteration t , V_{ij}^t is the j^{th} bit of the i^{th} particle's velocity at iteration t , P_{ij}^t is the j^{th} bit of the personal best position of the i^{th} particle at iteration t , G_j^t is the j^{th} bit of the global best position among all particles at iteration t , w is the inertia weight, c_1 and c_2 are acceleration coefficients, r_1 and r_2 are random numbers between 0 and 1.

Introduce a weight optimization mechanism to adjust the inertia weight w dynamically. The weight optimization equation can be defined in Equation 7.

$$w = w_{min} + (w_{max} - w_{min}) \cdot \frac{t}{\max_iterations} \text{ -----(7)}$$

where w_{min} and w_{max} are the minimum and maximum inertia weights, respectively. t is the current iteration, and $\max_iterations$ is the maximum number of iterations.

Whale Optimization Technique is emulated from the bubble net hunting policy and it is derived from the humpback whales. The hunting policy is composed of prey encircling, exploitation, and exploration. The target prey is identified from the candidates in the current best solution and the updating is initiated towards the best search agent where the best search agent is acquired. To formulate the process mathematically, the equation is defined in Equation 8 and 9.

$$\vec{S} = |\vec{X} \cdot \vec{C}^*(T) - \vec{C}(t)| \text{ ----- (8)}$$

$$\vec{S}(t + 1) = \vec{C}^*(t) - \vec{Z} \cdot \vec{S} \text{ ----- (9)}$$

where \vec{Z} and \vec{X} indicates the coefficient vectors, the iteration is indicated by t , the optimized value acquired from the solution space is C^* along the position vector of \vec{C} . For every iteration, the solution is updated once it identifies the best solution. The vector values are calculated by the following Equation 10 and 11.

$$\vec{Z} = 2\vec{z} \cdot \vec{r} - \vec{z} \text{ ----- (10)}$$

$$\vec{X} = 2 \cdot \vec{r} \text{ ----- (11)}$$

The value of \vec{z} vector is decreased linearly from the value two to zero and the random value is indicated by \vec{r} within the duration of [0,1] in the exploitation and exploration phase. The updating process by spirally and shrinking of encircling is occurred due to the behavior of the bubble net. The mechanism is equated as in 12.

$$\vec{C}(t + 1) = \vec{S}.e^{bl}.\cos(2\pi l) + \vec{C}^*(t) \text{ ----- (12)}$$

where the distance among the best solution and its space is indicated as $\vec{S} = |\vec{C}^*(T) - \vec{C}(t)|$, the random number and constant are indicated as l and b respectively that lies in the period of [-1, 1].

In the phase of exploration, search agents are elected randomly and the locations are updated that is in contrast to the exploitation process. The global search process is permitted by the following Equations 13 and 14.

$$\vec{S} = |\vec{C}_{rand} - \vec{C}| \text{ ----- (13)}$$

$$\vec{S}(t + 1) = \vec{C}_{rand} - \vec{Z}.\vec{S} \text{ ----- (14)}$$

where the random whale from the current population is indicated by \vec{C}_{rand} .

Let S_{WOPSO} represent the subset of particles selected from WOPSO for information exchange. The selection process can be based on various criteria, such as the fitness values of particles. The subset selection from WOPSO is done by Equation 15.

$$S_{WOPSO} = SelectSubset(P_{WOPSO}) \text{ ----- (15)}$$

Update the positions or velocities of WOA whales in PWOA based on the selected characteristics from SWOPSO . The update function can depend on the specific characteristics being exchanged and given in Equation 16.

$$P'_{WOA} = UpdatePosition(P_{WOA}, S_{WOPSO}) \text{ ----- (16)}$$

where P'_{WOA} represents the new population of WOA whales after the information exchange.

Let SWOA represent the subset of whales selected from WOA for information exchange is given in Equation 17.

$$S_{WOA} = SelectSubset(P_{WOA}) \text{ ----- (17)}$$

Update the positions or velocities of WOPSO particles in PWOPSO based on the selected characteristics from SWOA is given in Equation 18.

$$P'_{WOPSO} = UpdatePosition(P_{WOPSO}, S_{WOA}) \text{ ----- (18)}$$

where represents the new population of WOPSO particles after the information exchange.

The functions SelectSubset and UpdatePositions would be defined according to the requirements of the hybridization strategy, considering factors such as diversity, convergence, and exploration capabilities of the algorithms.

Determine convergence by evaluating predetermined criteria, such as reaching a certain maximum number of iterations or attaining a satisfactory level of fitness. Upon meeting the termination conditions, move to the subsequent step. Alternatively, iterate the hybridization procedure. This hybrid technique leverages the synergistic advantages of WOPSO and WOA, hence improving the efficiency and efficacy of feature selection and extraction for the identification of diabetic retinopathy or comparable tasks. The process of robust optimization is facilitated by the dynamic modification of weights, maintaining a balance between exploration and exploitation, and exchanging information.

Parameter tuning is an essential process for enhancing the control parameters of *Hybridized Weight-Optimized Particle Swarm Optimization and Whale Optimization Algorithm (HWOPSO-WOA)*. Sensitivity analysis is utilized to comprehend the impact of different parameters on the overall efficacy of the hybrid strategy. This procedure entails methodically altering the values of some parameters while maintaining the constancy of others in order to observe the resultant influence on the optimization process.

WOPSO relies on several critical criteria, such as the inertia weight, cognitive and social components, and the maximum velocity. The behavior and convergence characteristics of the particle swarm are determined by these parameters. Sensitivity analysis entails the individual adjustment of these parameters and the subsequent observation of the ensuing changes in both convergence speed and solution quality.

Similarly, in the context of WOA, the algorithm's performance is significantly influenced by factors such as the exploration and exploitation coefficients, the spiral updating formula, and the maximum number of iterations. Sensitivity analysis is used to assess how the hybrid technique is affected by changes in these factors. The goal of parameter tuning is to determine the most effective combination of parameter values that increases both the speed at which the hybridized algorithms converge and the quality of their solutions. This procedure may entail employing methodologies such as grid search, random search, or metaheuristic optimization algorithms to methodically investigate the parameter space and identify the configuration that produces the most optimal outcomes.

The efficacy of the hybrid technique depends on achieving a harmonious equilibrium between exploration and exploitation, so assuring a suitable compromise between global and local search capabilities. By doing sensitivity analysis and parameter modification, the hybridized method can be optimized to reach the best possible performance in tackling intricate optimization problems, such as selecting features for detecting diabetic retinopathy.

Result and Discussion

The *Indian Diabetic Retinopathy Image Dataset (IDRiD)* is obtained from genuine clinical examinations performed at an eye clinic in India. The dataset consists of 516 images in jpg format, each with a field of view of 50° and a resolution of 4288×2848 pixels. The images are categorized into five classes for *Diabetic Retinopathy (DR)* and three classes for *Diabetic Macular Edema (DME)*, following international clinical standards. The dataset has professional annotations of *DR* lesions, normal retinal components, and levels of disease severity for each image. The severity levels for *DR* and *DME* are specified, along with the distribution in the training and testing datasets for different types of lesions. The curated dataset comprises three common diabetic retinopathy anomalies, as depicted in Figure 1.

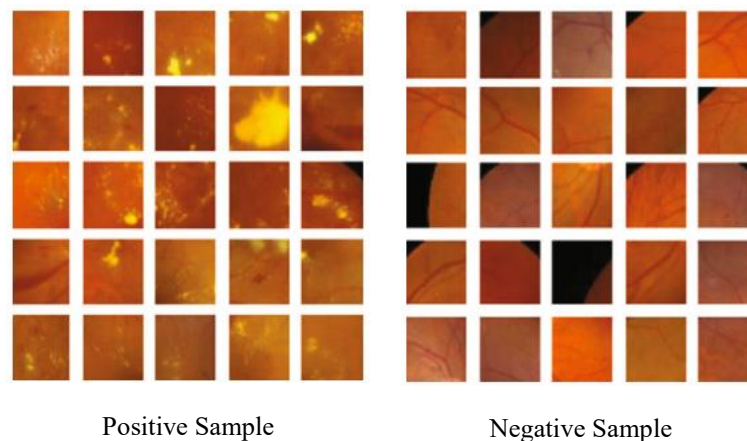


Figure 1. Colour Fundus Image with Positive and Negative Samples

Linear filters, such as Sobel and Prewitt, are essential in image processing as they utilize convolution to carry out operations including blurring, sharpening, and edge detection. The mean filter, which calculates the average of surrounding pixel values, is an excellent method for reducing noise. However, it may result in the loss of intricate details. Gaussian filters, which employ a bell-shaped response, are essential for reducing noise in images prior to complex image analysis tasks. The pre-processing applied images are illustrated in Figure 2.

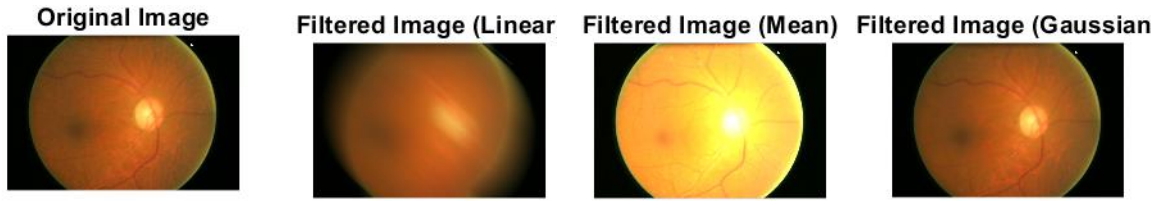


Figure 2. Pre-processing

The convergence graph of Weight Optimized Particle Swarm Optimization (WOPSO) and Whale Optimization Algorithm (WOA) provides a visual depiction of their optimization efficiency during the iterations. The vertical axis represents the value of the objective function or fitness, which indicates the algorithm's advancement in tasks such as selecting features for detecting diabetic retinopathy. The x-axis displays the number of iterations, illustrating the progression of the objective function value over time. A gradual and sharp decrease in the curve indicates a quick coming together, while consistency in subsequent iterations implies adjustment for the best possible answers. An analysis of the convergence graphs of WOPSO and WOA facilitates the assessment of their respective performance, enabling the determination of which algorithm achieves faster convergence or attains a superior final solution. The graph also aids in identifying potential problems, such as premature convergence or divergence, and informs judgments on when to stop early. In the context of image analysis tasks, the convergence graph serves as a significant tool for evaluating the efficiency and efficacy of optimization techniques. The convergence attained by the optimization techniques is given in Figure 3.

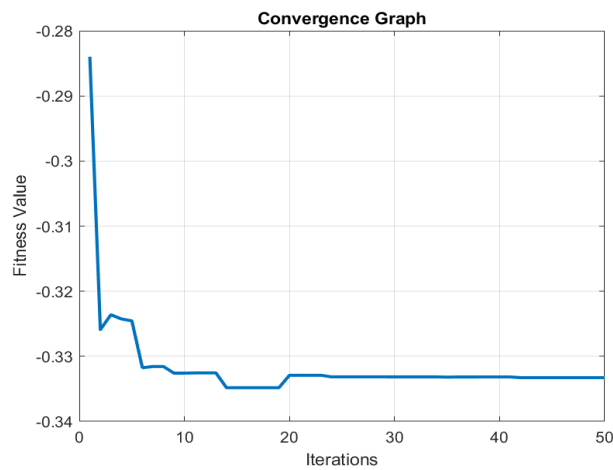


Figure 3. Convergence of HWOPSO-WOA

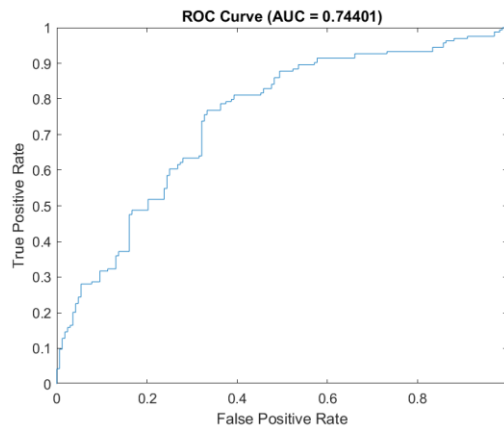


Figure 4.ROC Curve

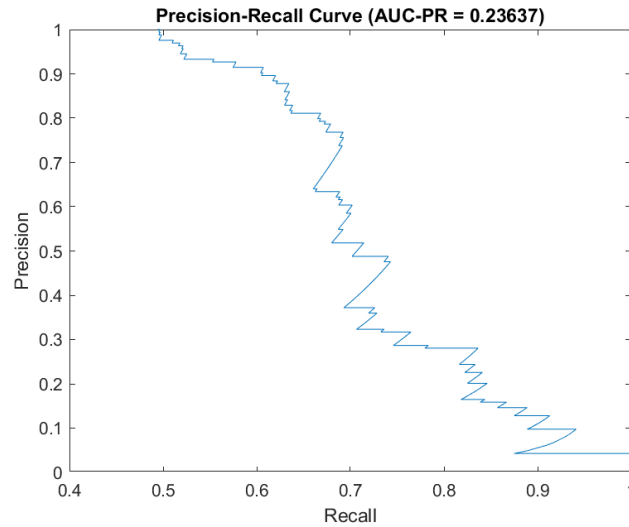


Figure 5. Precision-Recall Curve

The accuracy of feature selection refers to the effectiveness of a method in discovering and keeping relevant features that have a substantial impact on the overall performance of a predictive model. Within the domain of machine learning and data analysis, feature selection refers to the meticulous process of selecting a subset of pertinent features from the original set, guided by particular criteria. The main goal is to improve the accuracy of the model, reduce the computational complexity, and minimize the risk of overfitting. The evaluation of accuracy in feature selection entails the establishment of crucial metrics, the execution of the feature selection process, the training and testing of the model, and the comparison of its performance with a baseline model that employs all the characteristics at hand. The efficacy of the feature selection procedure is measured by enhancements in accuracy or other pertinent metrics. To enhance the reliability and effectiveness of the selected feature selection method, cross-validation techniques and iterative refinement can be utilized. This will ultimately lead to the development of more efficient and accurate machine learning models.

The amount of time required for feature selection is a crucial factor in the creation of machine learning models. The procedure, crucial for improving model efficiency, is impacted by factors such as the quantity of the dataset, the complexity of the algorithm, and the availability of computer resources. Filter methods, which assess features individually, are more efficient than wrapper approaches, which include repeated model training and evaluation. Embedding methods, which include incorporating feature selection into the process of model training, have a significant effect on the overall duration of training. In order to address temporal limitations, techniques such as parallelization or distributed computing can be utilized. Optimizing the machine learning pipeline requires a key balance between computing cost and the desired accuracy of feature selection. The performance for both best and worst case for Accuracy in % and Time Consumption for feature selection in seconds is given in Table 1.

Table 1. Performance Comparison

Technique	Accuracy (%)		Time (s)	
	Best	Worst	Best	Worst
VGG19	91.9	89.8	91.11	229.86
Inception V3	90.4	88.3	89.55	297.46
Fusion	93.8	74.2	240.34	444.76
Selection	96.4	94.8	88.78	140.24
Proposed	97.3	95.6	98.45	120.25

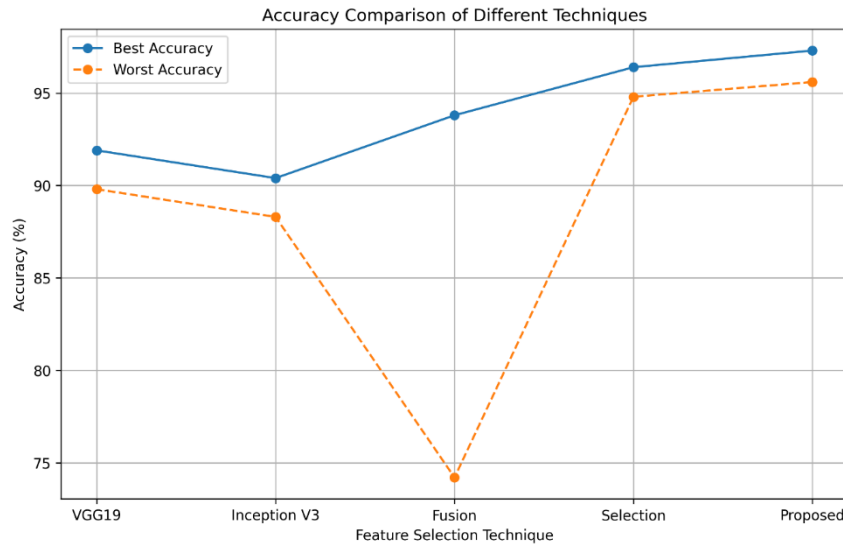


Figure 6. Comparison of Accuracy

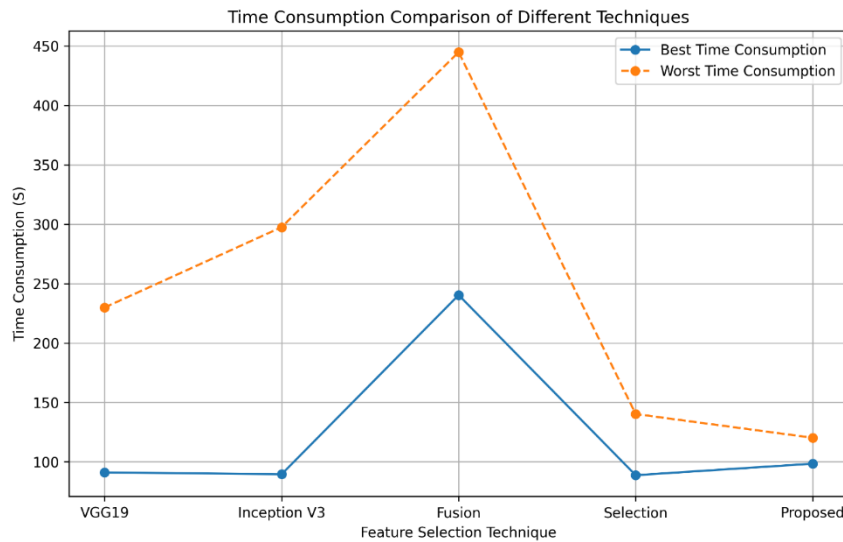


Figure 7. Comparison of Feature Selection Time

The comparative investigation of several strategies, such as VGG19, Inception V3, Fusion, Selection, and the Proposed method, reveals significant performance features. The Proposed technique stands out as the best performer, with an accuracy of 97.3% and showcasing efficiency with a processing time of 98.45 seconds. The fusion method, although providing acceptable accuracy, has the longest processing time of 444.76 seconds, which may restrict its suitability for real-time applications. VGG19 and Inception V3 demonstrate comparable performance in terms of accuracy and processing time, with Inception V3 exhibiting a slightly faster best-case time. The Selection technique demonstrates exceptional accuracy, especially in its most unfavorable situation, although it exhibits a lengthier processing time in the best-case scenario compared to the Proposed approach. The selection of the most appropriate technique would be contingent upon the precise application requirements, achieving a harmonious equilibrium between precision and computational efficiency.

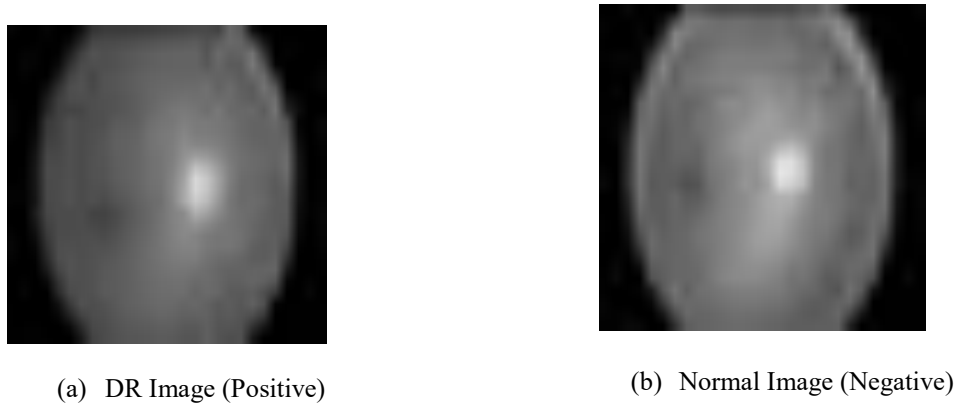


Figure 8. Feature Selected Fundus Image

Figure 1 displays a Color Fundus Image with Positive and Negative Samples, illustrating the visual depiction of the two categories of samples in the context of diabetic retinopathy detection. Linear filters, such as Sobel and Prewitt, are essential for a range of image processing applications, providing capabilities such as blurring, sharpening, and edge recognition. The mean filter, which reduces noise by averaging the values of surrounding pixels, is employed, recognizing its effectiveness despite the possibility of losing fine details. Gaussian filters are necessary for reducing noise before performing sophisticated image analysis tasks. They are also used in the preprocessing processes, as shown in Figure 2. The convergence graphs in Figure 3 illustrate the optimization efficiency of Weight Optimized Particle Swarm Optimization (WOPSO) and Whale Optimization Algorithm (WOA) over iterations. Both graphs offer useful insights into the convergence patterns of both algorithms. The convergence graphs serve as a tool to evaluate performance, allowing for the identification of the algorithm that delivers quicker convergence or a superior final solution. Furthermore, Figures 4, 5, and 6 display the ROC Curve, Precision-Recall Curve, and a Comparison of Accuracy, respectively, providing a comprehensive perspective on the classification performance. The Comparative Analysis in Figure 7 demonstrates that the Proposed method surpasses other strategies in terms of accuracy and feature selection time. It achieves a favorable balance between high accuracy and efficient processing time. Figure 8 showcases the Feature Selected Fundus Image, highlighting the importance of the chosen features in detecting diabetic retinopathy. In summary, this extensive analysis integrates visual representations and quantitative indicators to offer a comprehensive image of the techniques' performance in the assigned task.

Conclusion

Diabetic Retinopathy (DR) is linked to diabetes, presents a significant risk to eyesight, requiring prompt identification and care. This work utilizes a complete methodology that combines *Hybridized Weight-Optimized Particle Swarm Optimization and Whale Optimization Algorithm (HWOPSO-WOA)* to perform feature selection in DR analysis. By employing linear filters for pre-processing and a well stated objective function for feature selection, the suggested approach exhibits exceptional accuracy (97.3%) and efficiency (98.45 seconds processing time). The performance of the strategy is shown through convergence graphs and a comparative study, demonstrating its effectiveness in comparison to previous methods. Promising possibilities for DR analysis could be explored. The proposed methodology *HWOPSO-WOA*, can be expanded to handle bigger and more varied datasets, assuring both robustness and generalizability. Investigating the incorporation of sophisticated machine learning techniques or deep learning architectures could improve the ability to extract features and classify data.

Furthermore, it would be beneficial to examine the practicality of implementing the suggested method in real-time and its ability to handle large-scale screening programs. Engaging in partnerships with clinical experts could provide additional validation for the efficacy of the methodology and investigate its possible incorporation into current healthcare systems. Finally, considering the interpretability and explainability features of the model outputs can improve its acceptance and reliability in clinical settings. Further investigation can focus on the ability to handle larger volumes, immediate implementation, and more extensive verification in clinical settings to promote wider acceptance.

Reference

1. Sungheetha, A. (2021). Design an early detection and classification for diabetic retinopathy by deep feature extraction based convolution neural network. *Journal of Trends in Computer Science and Smart Technology*, 3(2), 81-94.
2. Das, D., Biswas, S. K., & Bandyopadhyay, S. (2023). Detection of diabetic retinopathy using convolutional neural networks for feature extraction and classification (DRFEC). *Multimedia Tools and Applications*, 82(19), 29943-30001.
3. Usman, T. M., Saheed, Y. K., Ignace, D., & Nsang, A. (2023). Diabetic retinopathy detection using principal component analysis multi-label feature extraction and classification. *International Journal of Cognitive Computing in Engineering*, 4, 78-88.
4. Chen, B., Wang, L., Sun, J., Chen, H., Fu, Y., Lan, S., ... & Xu, Z. (2018). Diverse lesion detection from retinal images by subspace learning over normal samples. *Neurocomputing*, 297, 59-70.
5. Subramanian, B., Saravanan, V., Nayak, R. K., Gunasekaran, T., & Hariprasath, S. (2019). Diabetic Retinopathy–Feature Extraction and Classification using Adaptive Super Pixel Algorithm. *Int J Eng Adv Technol*, 9, 618-627.
6. Kandhasamy, J. P., Balamurali, S., Kadry, S., & Ramasamy, L. K. (2020). Diagnosis of diabetic retinopathy using multi level set segmentation algorithm with feature extraction using SVM with selective features. *Multimedia Tools and Applications*, 79, 10581-10596.
7. David, D. S. (2020). Retinal image classification system for diagnosis of diabetic retinopathy using morphological edge detection and feature extraction techniques. *Artech J. Eff. Res. Eng. Technol*, 1, 28-33.
8. Umapathy, A., Sreenivasan, A., Nairy, D. S., Natarajan, S., & Rao, B. N. (2019, January). Image processing, textural feature extraction and transfer learning based detection of diabetic retinopathy. In *Proceedings of the 2019 9th International Conference on Bioscience, Biochemistry and Bioinformatics* (pp. 17-21).
9. Nahiduzzaman, M., Islam, M. R., Islam, S. R., Goni, M. O. F., Anower, M. S., & Kwak, K. S. (2021). Hybrid CNN-SVD based prominent feature extraction and selection for grading diabetic retinopathy using extreme learning machine algorithm. *IEEE Access*, 9, 152261-152274.
10. Salamat, N., Missen, M. M. S., & Rashid, A. (2019). Diabetic retinopathy techniques in retinal images: A review. *Artificial intelligence in medicine*, 97, 168-188.
11. Gayathri, S., Gopi, V. P., & Palanisamy, P. (2020). Automated classification of diabetic retinopathy through reliable feature selection. *Physical and Engineering Sciences in Medicine*, 43, 927-945.
12. Ali, A., Qadri, S., Khan Mashwani, W., Kumam, W., Kumam, P., Naeem, S., ... & Sulaiman, M. (2020). Machine learning based automated segmentation and hybrid feature analysis for diabetic retinopathy classification using fundus image. *Entropy*, 22(5), 567.
13. Nahiduzzaman, M., Islam, M. R., Goni, M. O. F., Anower, M. S., Ahsan, M., Haider, J., & Kowalski, M. (2023). Diabetic retinopathy identification using parallel convolutional neural network based feature extractor and ELM classifier. *Expert Systems with Applications*, 217, 119557.
14. Frazao, L. B., Theera-Umpon, N., & Auephanwiriyakul, S. (2019). Diagnosis of diabetic retinopathy based on holistic texture and local retinal features. *Information Sciences*, 475, 44-66.
15. Gu, Zongyun, Yan Li, Zijian Wang, Junling Kan, Jianhua Shu & Qing Wang. 2023, 'Classification of Diabetic Retinopathy Severity in Fundus Images Using the Vision Transformer and Residual Attention', *Computational Intelligence and Neuroscience*, vol. 2023, Article ID 1305583, 12 pages, doi: 10.1155/2023/1305583.
16. Nage, P, Shitole, S & Kokare, M 2023, 'An Intelligent Approach for Detection and Grading of Diabetic Retinopathy and Diabetic Macular Edema Using Retinal Images', *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1-16, doi: 10.1080/21681163.2022.2164358.
17. Bajwa, Awais, Neelam Nosheen, Khalid Iqbal Talpur & Sheeraz Akram 2023, 'A Prospective Study on Diabetic Retinopathy Detection Based on Modify Convolutional Neural Network Using Fundus Images at Sindh Institute of Ophthalmology & Visual Sciences', *Diagnostics (Basel, Switzerland)*, Vol. 13, no. 3, p. 393.
18. Kundu, Swagata, Vikrant Karale, Goutam Ghorai, Gautam Sarkar, Sambuddha Ghosh & Ashis Kumar Dhara 2022, 'Nested U-Net for Segmentation of Red Lesions in Retinal Fundus Images and SubImage Classification for Removal of False Positives', *Journal of Digital Imaging*, vol. 35, no. 5, pp. 1111-1119.
19. Skouta, Ayoub, Abdelali Elmoufidi, Said Jai-Andaloussi & Ouail Ouchetto 2022, 'Hemorrhage Semantic Segmentation in Fundus Images for the Diagnosis of Diabetic Retinopathy by Using a Convolutional Neural Network', *Journal of Big Data*, vol. 9, no. 1, doi: 10.1186/s40537-022-00632-0.
20. Sudha, V., & Ganeshbabu, T. R. (2021). A Convolutional Neural Network Classifier VGG-19 Architecture for Lesion Detection and Grading in Diabetic Retinopathy Based on Deep Learning. *Computers, Materials & Continua*, 66(1).
21. Vijayan, T., Sangeetha, M., & Karthik, B. (2020). Efficient analysis of diabetic retinopathy on retinal fundus images using deep learning techniques with inception v3 architecture. *Journal of Green Engineering*, 10(10), 9615-9625.