



Optimal feature selection-based diabetic retinopathy detection using improved rider optimization algorithm enabled with deep learning

Ambaji S. Jadhav¹ · Pushpa B. Patil² · Sunil Biradar³

Received: 26 November 2019 / Revised: 9 March 2020 / Accepted: 21 March 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

This proposal tempts to develop automated DR detection by analyzing the retinal abnormalities like hard exudates, haemorrhages, Microaneurysm, and soft exudates. The main processing phases of the developed DR detection model is Pre-processing, Optic Disk removal, Blood vessel removal, Segmentation of abnormalities, Feature extraction, Optimal feature selection, and Classification. At first, the pre-processing of the input retinal image is done by Contrast Limited Adaptive Histogram Equalization. The next phase performs the optic disc removal, which is carried out by open-close watershed transformation. Further, the Grey Level thresholding is done for segmenting the blood vessels and its removal. Once the optic disk and blood vessels are removed, segmentation of abnormalities is done by Top hat transformation and Gabor filtering. Further, the feature extraction phase is started, which tends to extract four sets of features like Local Binary Pattern, Texture Energy Measurement, Shanon's and Kapur's entropy. Since the length of the feature vector seems to be long, the feature selection process is done, which selects the unique features with less correlation. Moreover, the Deep Belief Network (DBN)-based classification algorithm performs the categorization of images into four classes normal, earlier, moderate, or severe stages. The optimal feature selection is done by the improved meta-heuristic algorithm called Modified Gear and Steering-based Rider Optimization Algorithm (MGS-ROA), and the same algorithm updates the weight in DBN. Finally, the effectual performance and comparative analysis prove the stable and reliable performance of the proposed model over existing models. The performance of the proposed model is compared with the existing classifiers, such as, NN, KNN, SVM, DBN and the conventional Heuristic-Based DBNs, such as PSO-DBN, GWO-DBN, WOA-DBN, and ROA-DBN for the evaluation metrics, accuracy, sensitivity, specificity, precision, FPR, FNR, NPV, FDR, F1 score, and MC. From the results, it is exposed that the accuracy of the proposed MGS-ROA-DBN is 30.1% higher than NN, 32.2% higher than KNN, and 17.1% higher than SVM and DBN. Similarly, the accuracy of the developed MGS-ROA-DBN is 13.8% superior to PSO, 5.1% superior to GWO, 10.8% superior to WOA, and 2.5% superior to ROA.

Keywords Diabetic retinopathy diagnosis · retinal abnormalities · Optimal feature selection · Deep belief network · Modified gear and steering-based rider optimization algorithm

✉ Ambaji S. Jadhav
asjadhavec@gmail.com

Pushpa B. Patil
Pushpa_ms@rediffmail.com

¹ Department of Electrical and Electronics, B.L.D.E.A's V.P. Dr. P.G. Halakatti College of Engineering and Technology, (Affiliated to Visvesvaraya Technological University, Belagavi-590018), Vijayapur, Karnataka 586103, India

² Department of Computer Science and Engineering, B.L.D.E.A's V.P. Dr. P.G. Halakatti College of Engineering and Technology, (Affiliated to Visvesvaraya Technological University, Belagavi-590018), Vijayapur, Karnataka 586103, India

³ Ophthalmology, B.L.D.E.A's Medical College and Research Center, Vijayapur, India

Abbreviations

IRMA	Intra retinal microvascular abnormalities
ROA	Rider optimization algorithm
DBN	Deep belief network
DR	Diabetic retinopathy
MGS-ROA	Modified gear and steering-based ROA
ETDRS	Early treatment DR study
NPDR	Non-proliferative DR
FPR	False positive rate
PDR	Proliferative DR
HOG	Histogram of oriented gradients
SIFT	Scale invariant feature transform
FNR	False negative rate
LBP	Local binary pattern

Deep CNN	Deep convolutional neural network
SASG	Single annotations by single grader
NPV	Negative predictive value
SAMG	Single annotations from multiple graders
MAV	Multiple annotations by voting
FDR	False discovery rate
DAAD	Double annotations with adjudication of disagreement
ANN	Artificial neural network
PSO	Particle swarm optimization
MCC	Mathews correlation coefficient
MIL	Multiple instance learning
PCA	Principal component analysis
GWO	Grey wolf optimization
WOA	Whale optimization algorithm
CLAHE	Contrast limited adaptive histogram equalization
CNN	Convolutional neural network
NN	Neural network
TEM	Texture energy measurement
CPD	Cumulative probability distribution
SKIZ	Skeleton of influence zones
SVM	Support vector machine
LTE	Laws texture energy
KNN	K-nearest neighbour
RBM	Restricted Boltzmann machine
CD	Contrastive divergence

1 Introduction

DR is one of the major diseases of diabetes and the primary cause of vision impairment [1]. In order to avoid blindness, the early detection and treatment of DR are very essential. The first indication of DR is microaneurysms. As specified by ETDRS, the existence of some microaneurysms exhibits the sign of mild NPDR [2]. If there are more microaneurysms, then there is a huge risk of DR development. At present, ophthalmologists examine the color fundus images, where the monitoring and detection of microaneurysms are done manually which is a cyclic, tiring and error-prone process. Moreover, the manual monitoring of DR will not meet the detection requirements of severe DR. If DR is diagnosed in the early stages, then the development of vision impairment will be reduced [3].

DR consists of two major kinds such as NPDR, and PDR [4]. NPDR is usually seen in the starting phases of DR, which is later split into moderate, severe, and mild phases. The mild stage consists of one microaneurysm, a little circular red dot at the end of blood vessels. The microaneurysms bliss into deep layers and create a flame-shaped haemorrhage in the retina, which is seen in a moderate stage. Moreover, the severe stage consists of about 20 intraretinal

haemorrhages in each quadrant holding specific venous bleeding with well-known intraretinal microvascular abnormalities. PDR is usually seen in the advanced stages of DR that cause neovascularisation, and the usual development of new blood vessels in the form of functional microvascular networks, which develops inside the retina [5].

However, the macula detection of DR is a time-consuming process, which thus depends on trained clinicians for inspecting and validating the digital color fundus images of the retina. Consequently, the improvement of automatic DR detection approaches is based on eye screening is provoked and helped the researches in both medical and academic communities [6–8]. Therefore, the number of DR diagnostic models is dependent on image evaluation approaches. Moreover, the computer vision-based techniques are classified into end-to-end learning [9, 10], and hand-on engineering [11–13]. In end-to-end learning, the concealed rich features are learned automatically and hence produce the best classification. The hand-on engineering techniques take out the features by existing methods like HOG [14], SIFT [15], LBP [16], Gabor Filters [17], etc., that fails in encoding the dissimilarities of scale, rotation, and illumination. In addition, several end-on-end learning and hand-on engineering-based methodologies [8, 12] are employed for detecting DR in the Kaggle dataset, but any of the methods are not capable of detecting the mild stage. The mild stage detection is very crucial for early diagnosis. Thus, the above techniques are generally found to be defective because of a lack of interpretability.

Hereunder, the major contributions of the paper are given:

- The efficient DR diagnosis model is developed using optimal feature extraction and classification strategies.
- Next to the image pre-processing, the optic disc is removed with the help of open-close watershed transformation and blood vessel segmentation and removal are done by grey level thresholding.
- The methods called Top hat transformation and Gabor filtering is used to extract the abnormalities like hard exudates, haemorrhages, Microaneurysm, and soft exudates. Next, the features like LBP, TEM, Shannon, and Kapur's entropy are extracted and further subjected to select the optimal feature selection.
- The proposed MGS-ROA performs both feature selection and improved DBN classification in order to reduce the correlation between the features, and error difference between the predicted and actual detection outcome.

The entire paper is designed as follows:

The review on literal works is given in Sect. 2. Section 3 describes the proposed model for diagnosing diabetic retinopathy from retinal fundus images. Moreover, the abnormality segmentation and optimal feature selection for diabetic

retinopathy detection are discussed in Sect. 4. Section 5 specifies the objective model considered for proposed diabetic retinopathy detection. The results and discussions of the proposed paper are given in Sect. 6. Finally, Sect. 7 concludes the paper.

2 Literature review

2.1 Related works

In 2019, Qummar et al. [18] have presented different computer vision-based approaches for diagnosing DR automatically and the corresponding phases from retina images. Here, Kaggle dataset images were utilized for training a group of Deep CNN methods such as Inceptionv3, Dense121, Xception, Dense169, and Resnet50 for encoding the best features and enhance the classification of various phases of DR. From the test results, it has been revealed that the developed approach was performed well in detecting DR at all stages.

In 2018, Kar and Maity [19] had offered a novel and automatic lesion detection model that includes four key stages such as “pre-processing, candidate lesion detection, eradication of optic disc and vessel extraction, and post-processing”. At first, the blood vessels and optic disc were removed for further processing. In order to split the dark lesions from the poor illuminated retinal background, curvelet-based edge enhancement was performed whereas a wideband bandpass filter was used to improve the contrast among the bright lesions and the background. Moreover, the experimental analysis was done on various publicly accessible datasets and the results have shown that the developed model was outperformed conventional models in diagnosing different kinds of DR lesions.

In 2019, Wang et al. [20] have explored the viability of detecting the existence of DR-related features and severity levels in a two-step process. For the experiment, 1589 retinal fundus images categorized by four general ophthalmologists, and three retinal specialists were taken. If both grading costs and prediction performance were taken into consideration, then SAMG was the best option. Similarly, when considering only the prediction performance, the DAAD was having the highest priority. The results have revealed that the two-step process was the best model for detection purposes.

In 2019, Chakraborty et al. [21] have introduced ANN for attaining exact diagnoses results in the case of DR. By calculating multiple entities of conventional ANN, the accuracy of feed forward back propagation-based ANN was enhanced. Finally, the outcomes have shown that the suggested approach was well suited for detecting the DR in an efficient manner.

In 2018, Costa et al. [22] have developed a framework named MIL to frame the DR detection model. The major

theme of the developed approach was the image categorization phases and the joint optimization of the instance encoding. From that, the important mid-level representations of diseased images were acquired. The performance of the system was improved by the latest loss function implementing suitable mid-level representations, and instance. Hence, the suggested method has attained the best performance over conventional methods.

In 2017, Zhou et al. [23] have recommended a new unsupervised classification approach on the basis of sparse PCA for detecting microaneurysms. The class imbalance issue was averted, as the proposed model was not considering the non-microaneurysms training set. In the meantime, an individual T2 statistic was developed, and the control limit was defined for differentiating true microaneurysms from false candidates automatically. Finally, the test results were performed on retinopathy online challenge competition dataset, and the results have shown that the developed model was efficient.

In 2019, Hemanth et al. [24] have used an approach that includes the image processing using histogram equalization, and the CLAHE methods. Later, the diagnosis was carried out by the categorization of CNN. With the help of 400 retinal fundus images, the proposed technique was evaluated in the MESSIDOR database using various performance metrics. By utilizing the image processing approaches and deep learning techniques, the suggested mechanism was superior to the conventional mechanisms.

In 2019, Xu et al. [25] have suggested a framework and the whole method consisting of two distinct ways in the view of microaneurysms turnover and compulsive risk factors for detecting the improvement of DR. The experiments on Grampian diabetes database has revealed that the recommended framework was attained the best sensitivity, specificity. Finally, it was confirmed that the proposed model was provided a novel and the non-invasive detection approach for early diagnosis of DR.

2.2 Review

Even though there are many mechanisms for detecting DR in an effective manner, still there are some defects with those methods so that new models need to be introduced in the future for further improvement in detecting DR at an early stage. Table 1 specifies the benefits and limitations of existing DR methods. Among them, Deep CNNs [18, 24] has high performance, there is no need of feature extraction, these are reliable and work well on image data, and pre-training of features is not required. Yet, there are few disadvantages such as if GPU was not good, the training is very slow, and still it utilizes black box testing for improving the detection system at an early stage. Differential Evolution [19] improves the capability of local search, and less

Table 1 Features and challenges of conventional diabetic retinopathy detection methods

Authors [citations]	Methodology	Features	Challenges
Qummar et al. [18]	Deep CNN	Has high performance There is no need for feature extraction	If the GPU was not good, the training is very slow
Kar and Maity [19]	Differential Evolution	Improves the capability of local search Less parameter settings have the ability to handle composite optimization problems	Convergence is not steady There is a possibility of falling into local optima
Wang et al. [20]	DAAD	It has high prediction performance It is less expensive	The performance is not good when considering grading along with prediction
Chakraborty et al. [21]	ANN	It designs compound and non-linear relationships ANNs are easy, once learned from the previous inputs and the corresponding relationships; it is assumed that unseen relationship on hidden data as well	There is no specific restriction for defining the structure of ANN These will only work with numerical information
Costa et al. [22]	MIL	It has high accuracy It is directly utilized for instance classification tasks	It won't work when the instances have no specific classes
Zhou et al. [23]	PCA	The images grouping with PCA decreases complexity Decreases the noise	It is difficult to evaluate the covariance matrix accurately Even the simple invariance is not confined
Hemanth et al. [24]	Deep CNN	These are reliable and work well on image data Pre-training of features is not required	It still utilizes a black box testing mechanism
Xu et al. [25]	SVM	It is relatively memory efficient It is highly efficient in high dimensional spaces	It is not suitable for huge datasets

parameter settings have the ability to handle composite optimization problems. However, there are few conflicts such as the convergence is not steady, and there is a possibility of falling into local optima. DAAD [20] has high prediction performance, and it is less expensive. Still, it is having some disadvantages like when both grading costs and prediction performance are considered, it should not be preferred. ANN [21] designs compound and non-linear relationships, and ANNs are easy, once learned from the previous inputs and the corresponding relationships; it is assumed that unseen relationship on hidden data as well. However, there are some conflicts like there is no specific restriction for defining the structure of ANN, and these will only work with numerical information. MIL [22] has high accuracy, and it is directly utilized for instance classification tasks. But, it won't work when the instances have no specific classes. The images grouping with PCA [23] decreases the complexity and Decreases the noise. However, there are few conflicts like it is difficult to evaluate the covariance matrix accurately, and even the simple invariance is not confined. SVM [25] is relatively memory efficient, and it is highly efficient in high dimensional spaces. Still, it is not suitable for huge datasets. Hence, the above-specified cons are utilized for developing new methodologies for early diagnosis of DR.

3 Proposed model for diagnosing diabetic retinopathy from retinal fundus images

3.1 Proposed methodology

The main intent of the proposed model is to detect DR automatically from the abnormalities present in the retinal fundus images. Most of the eye-related diseases that lead to blindness are DR, glaucoma, and age-related macular degeneration. In order to detect the eye infections easily and effectively, image processing is employed, which helps ophthalmologists to monitor the patients and perform a clinical review. The abnormalities like hard exudates, haemorrhages, microaneurysms, and soft exudates present in the retina provide detailed information regarding the changes followed by these retina related diseases, especially DR. The architecture of the developed DR detection system is shown in Fig. 1.

The proposed model includes phases such as image pre-processing, optic disc removal, blood vessel removal, abnormality segmentation, feature extraction, optimal feature selection, and classification. In the pre-processing phase, the retinal fundus image is given as input, where the RGB image is initially converted into a green channel, and

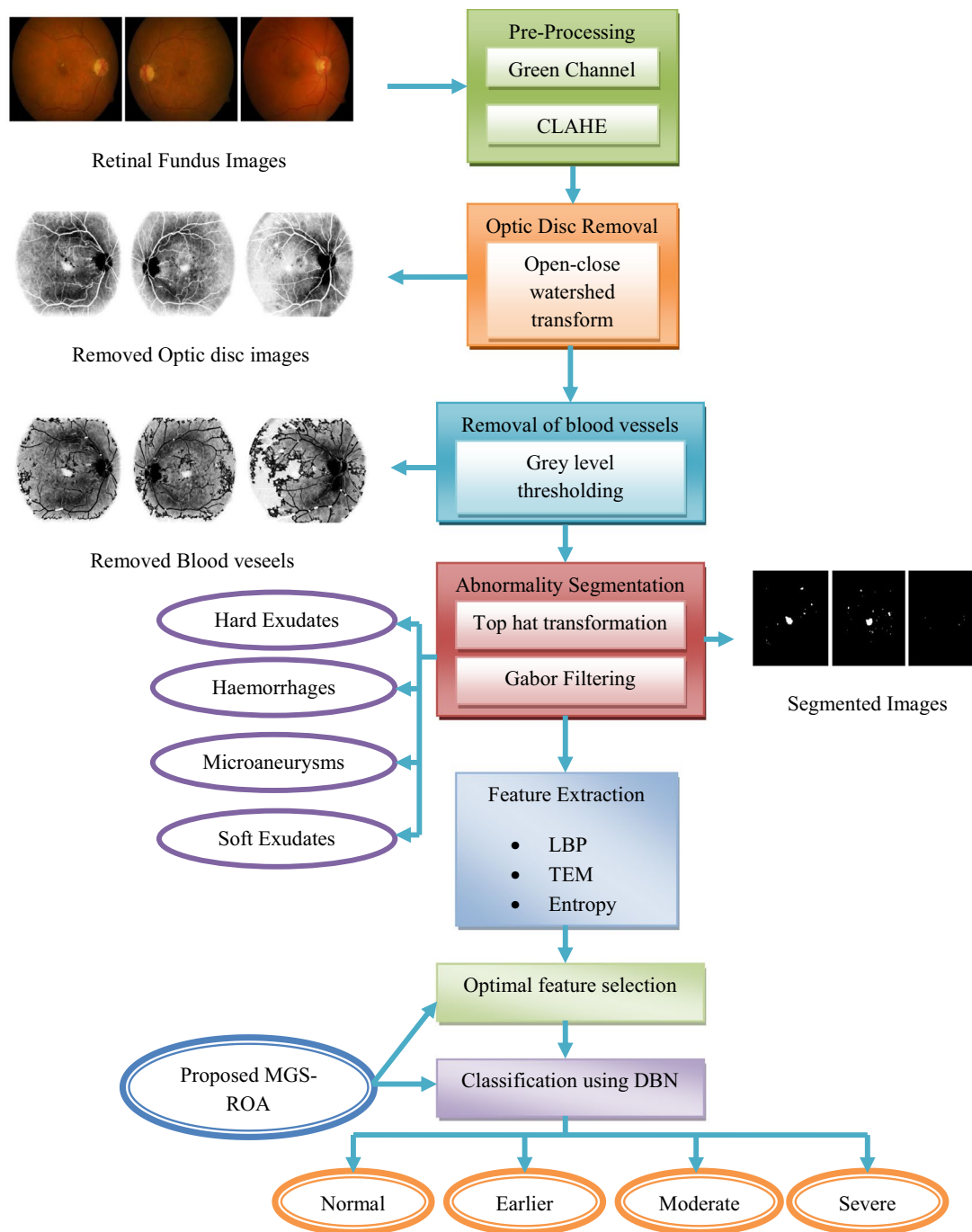


Fig. 1 Diabetic retinopathy detection model using the segmentation of retinal abnormalities

the image enhancement is done using CLAHE. From the contrast-enhanced image, the segmentation and removal of the optic disc are done by the open-close watershed transform. This watershed transform is used to split touching objects present in an image and recognizes the catchment basins and watershed ridgelines in an image by considering it as a surface, in which the light pixels are more, and the dark pixels are less. After removing the optic disc, the

blood vessels are segmented and removed using Grey level thresholding. Next to the removal of blood vessels and the optic disc, the abnormality segmentation is performed by top hat transformation followed by Gabor filtering. The obtained abnormality segmented image is subjected further for extracting features such as LBP, TEM, Shannon and Kapur’s entropy. LBP is used to extract the texture features of an image, whereas TEM is used to find the energy.

Shannon's and Kapur's entropy quantifies the expected value of data present in the image. The length of the feature vector is found to be very large so that the feature selection process is performed that selects the individual features with less correlation. Optimal feature selection is adopted in the proposed methodology for selecting the unique features. The improved meta-heuristic algorithm named MGS-ROA is used to perform the feature selection process. In addition, the optimally collected features by proposed MGS-ROA are given to the optimized DBN-classifier for detecting DR. Here in DBN classifier, the weight function is updated by the same MGS-ROA-based optimization algorithm. Moreover, the objective of the proposed feature selection and detection for DR diagnosis is to minimize the correlation of selected features and detection errors, respectively. Hence, the final classification model gives the outcome regarding the DR diagnosis whether it is normal, earlier, moderate, or severe stage.

3.2 Image description

Let the input retinal fundus image is denoted as In_{img} . The pre-processed retinal fundus image is indicated by In_{Pr} . Moreover, the image obtained after removing the optic disc is In_{ODR} , and the term In_{BVR} is the image acquired next to the removal of blood vessels. The final abnormality segmented image is indicated by In_{ABN} . The input features from the segmented abnormalities are denoted by Fe_{inp} , where $inp = 1, 2, \dots, NF$, and NF indicates the number of features. The optimal selection of features is represented by Fe_{inp}^* where $inp = 1, 2, \dots, NF^*NF^*$ refers to the number of optimal features.

3.3 Pre-processing of retinal fundus images

Here, the pre-processing of the input retinal fundus image is performed by Green channel conversion and CLAHE.

(a) *Green channel* The input retinal fundus image is an RGB image that includes three channels such as red, green, and blue, which are transformed into green channel images. The retinal images are often low contrast images. In this green channel, the abnormalities are clearly seen because of high contrast.

(b) *CLAHE* [26] It is a well-known procedure utilized for improving the brightness of the image. The algorithm for improving the contrast of the image using the CLAHE approach is as follows:

(1) Every image is split into a count of non-overlapping contextual regions of a similar size of 8×8 blocks; each block refers to the neighborhood of 64 pixels.

(2) Measure the intensity histogram for every contextual region.

(3) In order to clip the histograms, set the clip limits, which is a threshold parameter to change the brightness of the image in an efficient manner. The brightness of the local image is enhanced by higher clip limits so that it should be set to a minimum optimal value.

(4) By choosing transformation functions, each histogram is modified.

(5) Each histogram is changed by not exceeding the chosen clip limit. The numerical equation for modified gray levels for the benchmark CLAHE approach with uniform distribution is given in Eq. (1), in which the maximum pixel value is given by gr_{max} , gr_{min} is the minimum pixel value, gr is the measured pixel value, and $PD(f)$ is the CPD.

$$gr = [gr_{max} - gr_{min}] * PD(f) + gr_{min} \quad (1)$$

To improve the exponential distribution, the gray level is determined by Eq. (2), where α is the clip parameter.

$$gr = gr_{min} - \left(\frac{1}{\alpha}\right) * \ln [1 - PD(f)] \quad (2)$$

(6) By using bilinear interpolation, the neighboring tiles were merged and the grayscale values of the image are changed with respect to modified histograms.

Thus, after contrast enhancement, the pre-processed image is represented as In_{Pr} and it is further applied to watershed transform for segmenting and removing the optic disc.

3.4 Optic disc removal

In order to segment the optic disc and to remove it from the pre-processed image, open-close watershed transformation [27] is carried out. To successfully remove the optical disc, watershed transformation undergoes few steps i.e., initially, the image is to be read for segmenting. After reading the image, the structure element of that image is determined. For the structured element, morphological dilation followed by opening-closing is performed. In addition, the background markers are measured and those are subjected to watershed transformation. Later, the function is given to perform the quantitative analysis.

Morphological opening or closing makes simple by eradicating the dark components, which are not appropriate in the structural element. A close-open or open-close watershed mechanism is used for dealing with the bright and dark elements. The numerical equation of the morphological opening of the image In_{Pr} by a structuring element F is expressed in Eq. (3), and the corresponding closing function is given in Eq. (4).

$$In_{Pr} \circ F = (In_{Pr} \ominus F) \oplus F \quad (3)$$

$$In_{Pr} \cdot F = (In_{Pr} \oplus F) \ominus F \tag{4}$$

The watershed line is nothing but a function, which is the group of points of a function that doesn't belong to any other catchment basin, which is given in Eq. (5). Here, $su(f)$ is the support to the function f and $CB(rm_t)$ is the catchment basin with the regional minima (rm_t).

$$Wa_{shd}(f) = su(f) \cap \left[\bigcup_t (CB(rm_t)) \right]^o \tag{5}$$

By evaluating the distance transformation over the foreground image, SKIZ is measured. The geodesic distance zone equation is denoted in Eq. (6), which $gd_G(i, H_t)$ is the geodesic distance from a point i to the group H . Moreover, few points of pixels G doesn't contribute to any influence zone. Therefore, those points make SKIZ of H in G and the corresponding equation is given in Eq. (7), where $SKIZ_G(H)$ are the points that doesn't belong to any of the influence zones. The mathematical equation $DZ_G(H)$ is indicated in Eq. (8).

$$tdz_G(H_t) = \{i \in G, \forall s \in [1, r] \setminus \{t\}, gd_G(i, H_t) < gd_G(i, H_s)\} \tag{6}$$

$$SKIZ_G(H) = G \setminus DZ_G(H) \tag{7}$$

$$DZ_G(H) = \bigcup_{t \in [1, r]} dz_G(H_t) \tag{8}$$

Moreover, after calculating the watershed ridgelines, the intensity of the image is changed by morphological reconstruction so that it will be having the regional minima in the necessary locations. The resultant image In_{ODR} is further segmented for removing the blood vessels effectively.

3.5 blood vessel removal

The segmentation and removal of blood vessels from the image In_{ODR} are done using grey-level thresholding [28], which operates on the basis of a presumption that the image has a bimodal histogram. Thus, from the background, the object might be taken using an easy procedure, which contrasts the image with the threshold TL value, which separates into two modes and the resultant threshold image is shown in Eq. (9), which is a binary image. Here, the pixel with intensity value 1 refers to the objects, and 0 refers to the background.

$$In_{BVR} = \begin{cases} 1 & \text{if } Im(x, y) > TL \\ 0 & \text{if } Im(x, y) \leq TL \end{cases} \tag{9}$$

Moreover, the image is represented as In_{BVR} after removing the blood vessels, and it is further subjected to abnormality segmentation.

4 Abnormality segmentation and optimal feature selection for diabetic retinopathy detection

4.1 Abnormality segmentation

The abnormality segmentation is performed next to the removal of the optic disc as well as blood vessels, which is done by top hat transform and Gabor filtering.

Top hat transform [29] It is determined as the difference between the image opening and the original image. The key objective of the top hat transform is to light objects on a dark background. This results in the enhancement of the contrast of the image. Moreover, top hat transform of the segmented image In_{BVR} is given in Eq. (10), where C is the structuring element.

$$T(In_{BVR}) = In_{BVR} - (In_{BVR} \circ C) \tag{10}$$

Gabor filtering [30] This filter is created from two components named sinusoidal and Gaussian. This mechanism is capable of linking the optimal representation of the spatial domain and the orientation direction. The mathematical formulation expressing the Gabor filtering of the image is depicted in Eq. (11), in which the cosine wave frequency is denoted by fr , along with u and v axes, σ_u and σ_v is the fixed the distance from the Gaussian properties, respectively, and θ be the orientation direction. Moreover, the representation of u_θ and v_θ is given in Eq. (12), and Eq. (13), respectively.

$$GF(u, v; \theta, fr) = \exp \left\{ -\frac{1}{2} \left[\frac{u_\theta^2}{\sigma_u^2} + \frac{v_\theta^2}{\sigma_v^2} \right] \right\} \cos(2\pi fr u_\theta) \tag{11}$$

$$u_\theta = u \cos \theta - v \sin \theta \tag{12}$$

$$v_\theta = u \sin \theta + v \cos \theta \tag{13}$$

Finally, the abnormality segmented image In_{ABN} is applied to the feature extraction phase for extracting reliable features.

4.2 Feature extraction

Here feature extraction is done using three techniques (1) LBP, (2) LTE, and (3) Entropy.

LBP [31] It is recommended as an efficient and robust texture descriptor, which could be implemented in a wide range of applications from texture segmentation to face recognition. By thresholding, the neighborhood of every pixel with the center value, the LBP operator categorizes the image pixel and this thresholding output is taken as the binary value. When each and every pixel is categorized with respect to LBP codes, the histogram of the labels is calculated and employed as a texture descriptor. The histogram of the labeled image $In_{ABN}(x, y)$ is

utilized as a descriptor and it is given in Eq. (14), in which $I(C) = 1$, when C is true, and $I(C) = 0$ when C is false, and m is the number of distinct labels provided by the LBP operator.

$$Hm_{img} = \sum_{x,y} I(In_{ABN}(x,y) = img), \quad img = 0, \dots, m - 1 \quad (14)$$

Moreover, a circular neighborhood is taken around a pixel. On the circumference of the circle, Po points are selected with radius Ra . The uniformity metric (U) is determined using Eq. (15) by taking $LBP_{Po,Ra}$ as Po -bit binary number $(k_{Po-1}, k_{Po-2}, \dots, k_1, k_0)$ for acquiring a rotation-invariant uniform pattern with finer angular quantization.

$$UM(LBP_{Po,Ra}) = |k_{Po-1} - k_0| + \sum_{Po=1}^{Po-1} |k_{Po} - k_{Po-1}| \quad (15)$$

The rotation invariant uniform pattern by U value less than or equal to 2 is given in Eq. (16), where the grey value of the centre pixel is denoted by gv_{cp} , and the grey values of Po points is indicated by $gv_{pPo}, pPo = 0, \dots, Po - 1$. By selecting circles with various radii around the center pixels, a multi-scale evaluation is done using LBP and creating a separate LBP image for every scale. Moreover, the energy and the entropy of the LBP image is built over various scales $Ra = 1, 2, \text{ and } 3$ with corresponding pixel count $Po = 8, 16, \text{ and } 24$ are employed as the feature descriptor.

$$LBP_{Po,Ra} = \begin{cases} \sum_{p=0}^{Po-1} s(gv_{pPo} - gv_{cp}) & \text{if } U(LBP_{Po,Ra}) \leq 2 \\ Po + 1 & \text{Otherwise} \end{cases} \quad (16)$$

LTE [32] It is one of the most utilized texture descriptors, which is employed in different applications consisting of medical image analysis. The whole masks are acquired from a one-dimensional vector with five-pixel lengths $L5$ indicates the level detection, $S5$ refers to spot detection, $E5$ denotes the edge detection, $R5$ represents the ripple detection, and denotes the wave detection. The image is convoluted with every two-dimensional mask for extracting the texture data from $In_{ABN}(x,y)$. When the filter $L5E5$ is utilized, the acquired texture image is shown in Eq. (17). All the two dimensional masks excluding $L5L5$ had zero mean based on the laws. Therefore, the texture image TEX_{L5L5} is employed for normalizing the contrast of the remaining texture images $TEX(x,y)$, which is given in Eq. (18).

$$TEX_{L5E5} = In_{ABN}(x,y) \otimes L5E5 \quad (17)$$

$$Normalize(TEX_{(x,y)}) = \frac{TEX_{(x,y)}}{TEX_{(x,y)}^{L5L5}} \quad (18)$$

The suitable convolution of these masks provides 25 various combinations. The results are sent to TEM, which includes moving non-linear window average of absolute

values is shown in Eq. (19). By uniting 25 TEM descriptors, a fourteen rotationally invariant TEM (TR) is obtained, and it is given in Eq. (20).

$$TEM_{(x,y)} = \sum_{a=-7}^7 \sum_{b=-7}^7 |TEX_{(x+ay+b)}| \quad (19)$$

$$TR_{E5L5} = \frac{TEM_{E5L5} + TEM_{L5E5}}{2} \quad (20)$$

Entropy [33] It is defined as the uncertainty linked with randomness. Here, two various kinds of entropy are considered such as Shannon and Kapur’s entropy models. Assume, the input segmented image $In_{ABN}(x,y)$ consisting of N_{dg} distinct gray values. Thus, the normalized histogram is determined for the specific region of interest of size $(ne \times me)$ is given in Eq. (21). In addition, the mathematical equation of Shannon entropy is given in Eq. (22). The Kapur’s entropy equation is represented in Eq. (23), which is having a more dynamic range than Shannon entropy.

$$Ent_{dg} = \frac{N_{dg}}{ne \times me} \quad (21)$$

$$SE = - \sum_{dg=0}^{M-1} Ent_{dg} \log_2(Ent_{dg}) \quad (22)$$

$$K_{\alpha,\beta} = \frac{1}{\beta - \alpha} \log_2 \frac{\sum_{dg=0}^{M-1} Ent_{dg}^\alpha}{\sum_{dg=0}^{L-1} Ent_{dg}^\beta} \quad (23)$$

The overall features extracted are finally termed as Fe_{inp} , which is denoted in Eq. (24).

$$Fe_{inp} = LBP_{Po,Ra} + TEM_{(x,y)} + SE + K_{\alpha,\beta} \quad (24)$$

Thus, Eq. (24) is the final features extracted and further used for selecting the optimal feature using the proposed MGS-ROA algorithm.

4.3 Optimal feature selection

The optimal feature selection of the input features Fe_{inp} is done by the developed MGS-ROA algorithm, which intends to select the optimal features Fe_{inp}^* .

Conventional ROA The algorithm ROA [34] is inspired based on the rider’s cluster, which is moving towards the target. Assume, a few groups of riders are moving in the direction of a similar destination in order to win the race. There are four clusters of riders are taken into consideration, where the number of riders is equal in all the clusters. The four rider’s clusters are bypass rider, followers, overtakers, and attackers. Each group is having its particular mechanism

to accomplish the destination. Moreover, the overtaker follows his individual location to reach the destination, which is the closest position of the leading rider. The attacker comes with high speed and occupies the location of the rider for reaching the target location. This algorithm has a few steps that each rider should follow are specified below.

(1) *Parameter initialization for group and rider* Four groups of riders are initialized, which is indicated by *RG* and their positions are initialized at random. The cluster initialization is represented in Eq. (25). Here, the number of riders is indicated by *RN*, which is the same as the group of riders (*RG*). The count of co-ordinates or dimension is indicated by *CN*. Moreover, the position of c^{th} rider is represented as $L_{it}(c, d)$. The overall count of riders is known by computing the number of riders present in each group, and the corresponding equation is denoted in Eq. (26).

$$L_{it} = \{L_{it}(c, d)\}; 1 \leq c \leq RN; 1 \leq d \leq CN \tag{25}$$

$$RN = ByR + Fol + Ovr + Att \tag{26}$$

In Eq. (26), *ByR*, *Fol*, *Ovr*, and *Att* are the “bypass rider, follower, overtaker, and attacker”, respectively. Moreover, steering, gear, accelerator, and brake are the rider parameters, which are initialized after the cluster initialization. The steering angle of the rider’s vehicle at a time *it* is given in Eq. (27), which $ST_{c,d}^{it}$ is the steering angle of c^{th} rider’s vehicle, and the mathematical equation is denoted in Eq. (28).

$$ST_{it} = \{ST_{c,d}^{it}\}; 1 \leq c \leq RN; 1 \leq d \leq CN \tag{27}$$

$$ST_{c,d} = \begin{cases} \theta_c; & \text{if } d = 1 \\ ST_{c,d-1} + \varphi; & \text{if } d \neq 1 \ \& \ ST_{c,d-1} + \varphi \leq 360 \\ ST_{c,d-1} + \varphi - 360; & \text{otherwise} \end{cases} \tag{28}$$

In addition, from Eq. (28), θ_c and φ is the location angle and the co-ordinate angle of the c^{th} rider’s vehicle. Based on the number of riders and the maximum angle 360° , the position angle of c^{th} rider’s vehicle is determined, which is given in Eq. (29). In order to find the steering angle, a co-ordinate angle is used and it is denoted in Eq. (30).

$$\theta_c = c * \frac{360^\circ}{RN} \tag{29}$$

$$\varphi = \frac{360^\circ}{CN} \tag{30}$$

(2) *Determining the success rate* Based on the distance, the success rate is defined i.e., computed between the target and the position of the rider as shown in Eq. (31), in which the position of the target is denoted by Lo_{ta} . Moreover, the

distance must be declined in order to improve the success rate so that the success rate of the rider is measured by the reciprocal of the distance.

$$sur_c = \frac{1}{\|L_c - Lo_{ta}\|} \tag{31}$$

(3) *Determining the leading rider* In order to find the leading rider, the success rate is considered majorly. The rider, who is having more success rate, will be assumed as the leading rider because the rider is nearer to the target. The location of the target changes every time so that the leading rider also changes at every time period. Thus, it is stated that the rider who is having more success rate will be the leading rider.

(4) *Update the location of the riders* For finding the leading rider and the winner in the race, the position of the riders in each cluster is updated every time. Based on the features, the location update of the rider in each cluster is done by following the below steps.

(a) *Bypass update* Since these riders bypass the usual path by not going in the path of leading riders, the bypass rider’s update process is denoted in Eq. (32). Here, δ and β are the random values ranging in between 0 and 1. Moreover, η and ξ are the values that choose the random value ranging from 0 to *RN*. The size β is $1 \times CN$. Therefore, each rider in the group updates its location and turn out as the winner.

$$L_{it+1}^{ByR} = \delta [L_{it}(\eta, d) * \beta(d) + L_{it}(\xi, d) * [1 - \beta(d)]] \tag{32}$$

(b) *Follower update* This follows the path of the leading rider, and updates its location thus the follower reaches the destination quickly and effectively. In addition, the follower is based on the co-ordinate selector, such that the location is updated for the selected values existing in *CN*, and the corresponding update equation is given in Eq. (33).

$$L_{it+1}^{Fol}(c, crs) = L^{Ler}(Ler, crs) + [cso(ST_{c,crs}^{it}) * L^{Ler}(Ler, crs) * dst_c^{it}] \tag{33}$$

In Eq. (33), the location of the rider is given by L^{Ler} , the co-ordinate selector is indicated by *crs*. The steering angle of the c^{th} rider in cs^{th} co-ordinate is denoted by $ST_{c,cs}^{it}$, and the distance traveled by the c^{th} rider is given by dst_c^{it} , which is computed by the product of the velocity of the rider with the rate of off time, and the corresponding equation is given in Eq. (34). Here, the maximum time is represented as Ti_{off}^{it} and ver_c^{it} is the velocity of c^{th} rider.

$$dst_c^{it} = ver_c^{it} * \left(\frac{1}{Ti_{off}^{it}}\right) \tag{34}$$

Moreover, the velocity is directly proportional to the speed of the vehicle and the rider parameters except for the steering angle, and the corresponding velocity equation is denoted in Eq. (35).

$$ver_c^{it} = \frac{1}{3} [ge_c^{it} * sp_c^{ge} + sp_{max}^c * acc_c^{it} + (1 - br_c^{it}) * sp_{max}^c] \tag{35}$$

In the above equation, ge_c^{it} , br_c^{it} , and acc_c^{it} are the gear, brake, and accelerator of the c^{th} rider's vehicle, respectively. The speed limit of the gear of c^{th} rider is given by sp_c^{ge} .

(c) *Overtaker update* Based on three factors like relative success rate, direction indicator, and co-ordinate selector, the overtaker is updated. The mathematical formula is given in Eq. (36), in which the location of c^{th} rider in $crsth$ crs^{th} co-ordinate is given by $L_{it}(c, crs)$, the direction indicator of c^{th} rider at time it is denoted by $di_{it}(c)$. Moreover, it measures the direction by the relative success rate, which is given in Eq. (37) using Eq. (38). Here, the relative success rate of c^{th} rider at time it is given by $(Res_{it}^{Rs}(c))$.

$$L_{it+1}^{Ovr}(c, crs) = L_{it}(c, crs) + [di_{it}(c) * L^{Ler}(Ler, crs)] \tag{36}$$

$$di_{it}(c) = \left[\frac{2}{1 - \log(Res_{it}^{Rs}(c))} \right] - 1 \tag{37}$$

$$Res_{it}^{Rs}(c) = \frac{sur_{it}(c)}{\max_{c=1}^{Rs} sur_{it}(c)} \tag{38}$$

(d) *Attacker update* The attacker who is trying to occupy the location of the leader will follow the similar procedure of the follower. The update process of the attacker is denoted in Eq. (39), in which the location of the leading rider is given by $L^{Ler}(Ler, d)$ and the steering angle of c^{th} rider in d^{th} co-ordinate is denoted as $ST_{c,d}^{it}$.

$$L_{it+1}^{Att}(c, d) = L^{Ler}(Ler, d) + \left[\cos(ST_{c,d}^{it}) * L^{Ler}(Ler, d) \right] + dst_c^{it} \tag{39}$$

(5) *Determining the success rate* Each rider's success rate is computed after the completion of the update process based on the location of the rider who is first in the race is replaced by the new rider's position such that the success of each rider is maximum.

4.4 Proposed MGS-ROA

The conventional ROA is based on the group of riders reaching the target, which is a new computing methodology. The idea of the ROA approach is accelerating the rider's location towards the leading riders at each time step. In that case, local minima are prevented by little local neighbourhood i.e., handled by the attacker, yet quick convergence is acquired by a more global neighbourhood, which is performed by the overtaker. Rather than having such benefits, this algorithm will often be stuck in solving

discrete optimization problems. In order to further improve its performance, the proposed MGS-ROA modifies the updated pattern of gear and steering angle. It is determined based on the fitness function of the current solution F_C and the maximum fitness values F_{MAX} . If $F_C < F_{MAX}$, then the gear ge_c^{it} is updated based on maximum gear (5), and steering angle is updated based on Eq. (40). On the other hand, gear ge_c^{it} is updated based on minimum gear (1), and the steering angle is updated based on Eq. (41).

$$ST_{it+1} = ST_{it} + Distn \tag{40}$$

$$ST_{it+1} = ST_{it} - Distn \tag{41}$$

The term *Distn* is a distance function, which is calculated by Eq. (42).

$$Distn = \frac{abs(L_{it+1}^* - L_{it})}{\max(L_{it+1}^*)} \tag{42}$$

In Eq. (42), L_{it+1}^* is the best position (leader position) and L_{it} is the current position. The pseudo-code of the proposed MGS-ROA is shown in Algorithm 1.

<p>Algorithm 1: Pseudocode of proposed MGS-ROA</p> <p>Random locations of the riders are initialized as RG_{it}</p> <p>Population is initialized</p> <p>Declare the rider parameters such as steering angle ST, gear ge, accelerator acc, and brake br</p> <p>Success rate need to be determined by sur_{it}</p> <p>while ($it < T_{i_{off}}$)</p> <p style="padding-left: 20px;">for $c = 1$ to RN</p> <p style="padding-left: 40px;">Location of the bypass reader is updated by Eq. (32)</p> <p style="padding-left: 40px;">Location of the follower is updated by Eq. (33)</p> <p style="padding-left: 40px;">Location of the overtaker is updated by Eq. (36)</p> <p style="padding-left: 40px;">Location of the attacker is updated by Eq. (39)</p> <p style="padding-left: 40px;">Rank the riders based on sur_{it}</p> <p style="padding-left: 40px;">Choose the rider who is having maximum sur_{it} as the leading rider</p> <p style="padding-left: 40px;">Update ST, ge, acc, and br</p> <p style="padding-left: 40px;">If $F_C < F_{MAX}$</p> <p style="padding-left: 60px;">If $ge_c^{it} = 5$</p> <p style="padding-left: 80px;">$ge_c^{it+1} = ge_c^{it}$</p> <p style="padding-left: 60px;">Else if</p> <p style="padding-left: 80px;">$ge_c^{it+1} = ge_c^{it} + 1$</p> <p style="padding-left: 60px;">End if</p> <p style="padding-left: 40px;">Update the steering angle using Eq. (40)</p> <p style="padding-left: 40px;">Else if</p> <p style="padding-left: 60px;">If $ge_c^{it} = 1$</p> <p style="padding-left: 80px;">$ge_c^{it+1} = ge_c^{it}$</p> <p style="padding-left: 60px;">Else if</p> <p style="padding-left: 80px;">$ge_c^{it+1} = ge_c^{it} - 1$</p> <p style="padding-left: 60px;">End if</p> <p style="padding-left: 40px;">Update the steering angle using Eq. (41)</p> <p style="padding-left: 40px;">End if</p> <p style="padding-left: 40px;">return RG^{Ler}</p> <p style="padding-left: 40px;">$it = it + 1$</p> <p style="padding-left: 20px;">end for</p> <p>end while</p>
--

4.5 Optimized DBN-based detection

The optimal trained DBN is used for diagnosing the DR. DBN [35] is a kind of deep NNs consisting of several layers. Each layer includes visible neurons in the input layer and the hidden neurons in the output layer. The major building blocks of DBNs are Boltzmann Machines for attaining the best results. The output of DBN is given by OP , which is a binary value and is denoted in Eq. (43), which $Pb(\lambda)$ is the probability function and the corresponding equation is given in Eq. (44). Here, $ptmp$ is the pseudo temperature, where $ptmp > 0$ is used to control the noise level in the probability. The stochastic model is depicted in Eq. (45)

$$OP = \begin{cases} 1 & \text{with } Pb(\lambda) \\ 0 & \text{with } 1 - Pb(\lambda) \end{cases} \quad (43)$$

$$Pb(\lambda) = \frac{1}{1 + e^{\frac{-\lambda}{ptmp}}} \quad (44)$$

$$\lim_{ptmp \rightarrow 0^+} Pb(\lambda) = \lim_{ptmp \rightarrow 0^+} \frac{1}{1 + e^{\frac{-\lambda}{ptmp}}} = \begin{cases} 0 & \text{for } \lambda < 0 \\ \frac{1}{2} & \text{for } \lambda = 0 \\ 1 & \text{for } \lambda > 0 \end{cases} \quad (45)$$

The main objective of the Boltzmann network model is to determine input patterns precisely according to the Boltzmann distribution. Moreover, the energy of the Boltzmann machine for configuring the neuron states st is represented in Eq. (46), in which $W_{g,h}$ indicates weights among the neurons, and θ_g is the biases of the neurons. In addition, the effect of a single unit's state st_g on global energy is measured by Eq. (47).

$$En(st) = - \sum_{g < h} W_{g,h} st_g st_h - \sum_g \theta_g st_g \quad (46)$$

$$\Delta En(st_g) = \sum_h W_{g,h} st_h + \theta_g \quad (47)$$

This energy is utilized in the training procedure with a gradient descend approach for determining the least possible energy of the model for the provided input. The main reason for this is the interconnections among the visible and the hidden neurons that make the neuron states depend on each other. By removing the connection, the joint configuration of visible and the hidden neurons will provide new energy definitions, which are given in Eq. (48), Eq. (49), and Eq. (50), in which the biases are given by p_g and q_h . The binary states of visible and hidden neurons are denoted by z_{vi_g} and z_{hm_h} .

$$En(\vec{z}_{vi}, \vec{z}_{hm}) = - \sum_{(g,h)} W_{g,h} z_{vi_g} z_{hm_h} - \sum_g z_{vi_g} p_g - \sum_h z_{hm_h} q_h \quad (48)$$

$$\Delta En(z_{vi_g}, \vec{z}_{hm}) = \sum_h W_{g,h} z_{hm_h} + p_g \quad (49)$$

$$\Delta En(\vec{z}_{vi}, z_{hm_h}) = \sum_g W_{g,h} z_{vi_g} + q_h \quad (50)$$

Moreover, the RBM is not dependent on the visible and the hidden neurons while calculating the energy difference. These are trained using unsupervised learning. In addition, RBMs learn by encoding the probability distribution of input data into weight parameters. The RBM allows a probability for each possible pair of a visible and the hidden vector using the energy function defined in Eq. (51), and the partition function is denoted in Eq. (52).

$$RP(\vec{z}_{vi}, \vec{z}_{hm}) = \frac{1}{Pa} e^{-En(\vec{z}_{vi}, \vec{z}_{hm})} \quad (51)$$

$$Pa = \sum_{\vec{z}_{vi}, \vec{z}_{hm}} e^{-En(\vec{z}_{vi}, \vec{z}_{hm})} \quad (52)$$

The time utilized by RBM for attaining the link by the developed approach is composite, therefore CD model is considered, which provides the fast convergence in system distribution. The corresponding algorithm is summarized below.

1. Assume a training set z_{vi} and merge it into visible neurons.
2. The probabilities are calculated for the hidden neurons $RP_{z_{hm}}$ by multiplying the visible vector z_{vi} with the weights matrix w . $RP_{z_{hm}} = \sigma(z_{vi} \cdot w)$.
3. Sample the hidden states z_{hm} from the probabilities $RP_{z_{hm}}$.
4. Calculate the outer product of vectors $RP_{z_{hm}}$ and z_{vi} , and name it as a positive gradient.
5. Reconstruction of visible states $z_{vi'}$ is sampled from the hidden states z_{hm} . After that, resample the hidden states from the renovation of visible states.
6. Now, calculate the outer product of reconstructed visible and hidden states and called a negative gradient. $\varphi^- = z_{vi'} \cdot RP_{z_{hm}}^T$.
7. By subtracting the positive gradient from a negative gradient, the weight updates are computed $\Delta w = \eta(\varphi^+ - \varphi^-)$.
8. The weights should be updated using the new values $W_{g,h}^* = \Delta W_{g,h} + W_{g,h}$.

Here, the training of DBN is done by both RBM and MGS-ROA in order to minimize the error function.

5 Objective model considered for proposed diabetic retinopathy detection

5.1 Objective function

There are two objective models employed for DR detection by segmenting the retinal abnormalities. The first one focuses on minimizing the correlation between the features, and the second one focuses on minimizing the error difference between the predicted and target outcome of DBN classification. Initially, the features have to be selected in such a manner that the correlation between the features should be minimum. If the correlation between the features is minimum, then it is having a possibility of determining the more accuracy. The correlation between the two features a and A^e is shown in Eq. (53), which NP is the number of feature pairs. Hence, the first objective function is the minimization of the correlation between features as shown in Eq. (53).

$$Corr = \frac{NP \sum ab - \sum a \sum b}{\left(\sqrt{NP} \sum a^2 - (\sum a)^2\right) - \left(NP \sum b^2 - (\sum b)^2\right)} \quad (53)$$

$$OBJ1 = Min(Corr) \quad (54)$$

The actual output vector of DBN is indicated by A^e , and the predicted outcome is denoted by B^e . The error function among the actual and the predicted outcome is shown in Eq. (55), which has to be minimized as the second objective function, as represented in Eq. (56).

$$MER = A^e - B^e \quad (55)$$

$$OBJ1 = Min(MER) \quad (56)$$

This objective model is attained by the proposed MGS-ROA so that the best detection accuracy is gained.

5.2 Solution encoding

The proposed MGS-ROA is used for selecting the optimal features as well as optimizing the weight function in DBN classification. The solution encoding for optimal feature selection is shown in Fig. 2, and the optimal selection of weight in DBN is shown in Fig. 3.

After optimizing the features and weights in DBN by the proposed MGS-ROA, the optimal features Fe_{inp}^* are obtained with the updated weight, which could be highly suitable for improving the classification accuracy.

Figure 4 shows the flowchart of the proposed MGS-ROA-DBN. Initially, the input retinal image is subjected to pre-processing. Then, the next phase performs the optic disc removal, which is carried out by open-close watershed

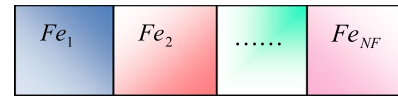


Fig. 2 Encoding of solution under optimal feature selection

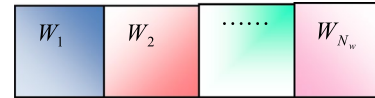


Fig. 3 Encoding of solution under weight updating in DBN

transformation. Afterthat, the segmentation and removal of blood vessels is done fgrey-level thresholding, in which, the pixel with intensity value 1 refers to the objects, and 0 refers to the background. Once the optic disk and blood vessels are removed, segmentation of abnormalities takes place. Then, the features are extracted from the image and the optimal feature selection is done by the improved meta-heuristic algorithm. Finally the classification is done using the DBN.

6 Results and discussions

6.1 Experimental procedure

The developed DR detection approach was implemented in MATLAB 2018a, and the performance evaluation of the proposed model was carried out. Here, the experiment was done on the dataset DIARETDB1, which includes 89 color fundus images. From that images, the experts have marked the regions belong to microaneurysms, haemorrhages, hard exudates, and soft exudates. In order to evaluate the feature selection, the population size was fixed as 10 and the maximum number of iterations was considered as 100. The performance of the developed MGS-ROA-based DBN was compared over the conventional PSO-DBN [36], GWO-DBN [37], WOA-DBN [38], and ROA-DBN [34] by evaluating the “performance measures such as accuracy, sensitivity, specificity, precision, FPR, FNR, NPV, FDR, F1 score, and MCC”.

6.2 Performance measures

In this paper, ten performance measures are taken into consideration to evaluate the performance and the description of each performance measure is given below.

(a) *Accuracy* It is a “ratio of the observation of exactly predicted to the whole observations”. The formula for accuracy is shown in Eq. (57), where $TR^P TR^N$ denotes the “true positives,

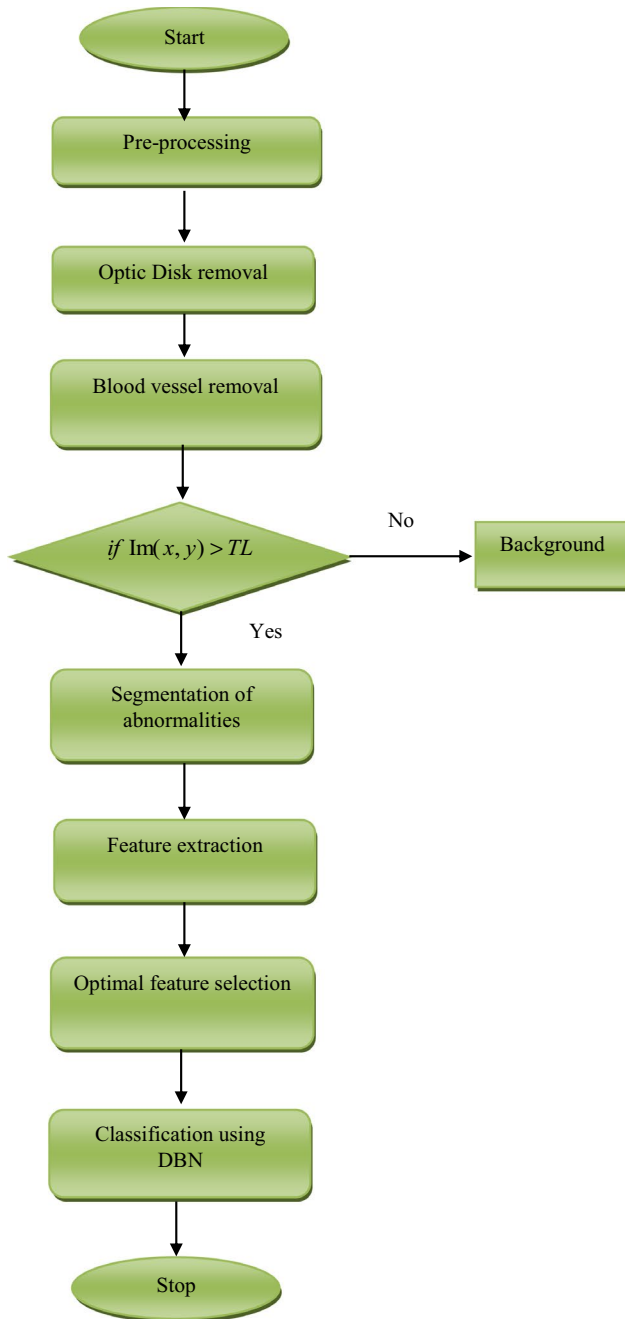


Fig. 4 The flowchart of the proposed MGS-ROA-DBN

and true negatives”, respectively. Moreover, FA^P and FA^N are the “false positives and false negatives”, respectively.

$$Acc = \frac{TR^P + TR^N}{TR^P + TR^N + FA^P + FA^N} \quad (57)$$

(b) *Sensitivity* It measures “the number of true positives, which are recognized exactly”. It is represented in Eq. (58).

$$Sen = \frac{TR^P}{TR^P + FA^N} \quad (58)$$

(c) *Specificity* It measures “the number of true negatives, which are determined precisely”. It is given in Eq. (59).

$$Spe = \frac{TR^N}{FA^P} \quad (59)$$

(d) *Precision* It is “the ratio of positive observations that are predicted exactly to the total number of observations that are positively predicted”. It is mathematically represented in Eq. (60).

$$Pre = \frac{TR^P}{TR^P + FA^P} \quad (60)$$

(e) *FPR* It is computed as “the ratio of the count of false-positive predictions to the entire count of negative predictions”, which is shown in Eq. (61)

$$FPR = \frac{FA^P}{FA^P + TR^N} \quad (61)$$

(f) *FNR* It is “the proportion of positives which yield negative test outcomes with the test”. The numerical equation is denoted in Eq. (62).

$$FNR = \frac{FA^N}{TR^N + TR^P} \quad (62)$$

(g) *NPV* It is the “probability that subjects with a negative screening test truly don’t have the disease”. It is shown in Eq. (63).

$$NPV = \frac{FA^N}{FA^N + TR^N} \quad (63)$$

(h) *FDR* It is “the number of false positives in all of the rejected hypotheses”. The numerical formula for FDR is given in Eq. (64).

$$FDR = \frac{FA^P}{FA^P + TR^P} \quad (64)$$

(i) *F1 score* It is defined as the “harmonic mean between precision and recall. It is used as a statistical measure to rate performance”. It is given in Eq. (65)

$$F1score = \frac{Sen \cdot Pre}{Pre + Sen} \quad (65)$$

(j) *MCC* It is a “correlation coefficient computed by four values”. It is denoted in Eq. (66).

$$MCC = \frac{TR^P \times TR^N - FA^P \times FA^N}{\sqrt{(TR^P + FA^P)(TR^P + FA^N)(TR^N + FA^P)(TR^N + FA^N)}} \quad (66)$$

6.3 Segmentation results

Here, the results of the pre-processing, optic disc removal, blood vessel removal, and segmented images are shown in Fig. 5.

6.4 Performance analysis

The performance analysis of the proposed MGS-ROA-DBN compared over the conventional heuristics-based DBN, which is graphically represented in Fig. 6. From Fig. 6a, the accuracy of the developed MGS-ROA-DBN is computed effectively when compared with other conventional algorithms. At a learning percentage of 45%, the suggested MGS-ROA is 2.1% better than WOA and 6.6% better than PSO-DBN. Moreover, the specificity of the developed MGS-ROA-DBN is exactly predicting the true negative observations from the false-positive observations from Fig. 6c. Therefore, the specificity of the developed MGS-ROA-DBN is 3.3% superior to ROA, 6.8% superior to GWO, 9.4% superior to WOA, 12.7% superior to DBN, and 16.2% superior to PSO-DBN at learning percentage 100%. The precision of the presented approach correctly determined the positive values, which is shown in Fig. 6d. It is 4.4% improved than WOA, and 16.2% improved than PSO-DBN at learning percentage 45%. In addition, at a learning percentage 65%, the NPV of the recommended MGS-ROA-DBN is 3.1%, 4.2%, 8.8%, and 10.1% enhanced than ROA, WOA, DBN, and PSO, respectively and it is given in Fig. 6g. From Fig. 6j, the MCC of the improved MGS-ROA-DBN is 8.2% better than WOA and 27.7% better than PSO-DBN at learning percentage 50%. Finally, it is proved that the implemented MGS-ROA-DBN is superior in detecting DR effectively.

The overall performance of the developed model is tabulated in Table 2. From Table 3, the accuracy of the implemented MGS-ROA-DBN has precisely determined the positive observations from the whole observations. Thus, the accuracy of the developed MGS-ROA-DBN is 13.8% superior to PSO, 5.1% superior to GWO, 10.8% superior to WOA, and 2.5% superior to ROA. Moreover, the precision of the improved MGS-ROA-DBN correctly defined the positive observations from the negative values. Therefore, the precision of the proposed MGS-ROA-DBN is 35.7% enhanced than PSO, 11.7% enhanced than GWO, 26.6% enhanced than WOA, and 5.5% enhanced than ROA. Similarly, the NPV of the modified MGS-ROA-DBN is 8.6% better than PSO, 3.2% better than GWO, 6.7% better than WOA, and 1.6% better than ROA. Thence, it is confirmed that the developed MGS-ROA-DBN is outperforming the conventional heuristics-based DBN in detecting DR effectively.

6.5 Analysis on different classifiers

The performance of the proposed MGS-ROA-DBN is analyzed and compared with the existing classifiers and it is shown in Table 3. In Table 3, the accuracy of the suggested MGS-ROA-DBN is defined precisely. It is 30.1% improved than NN, 32.2% improved than KNN, and 17.1% improved than SVM and DBN. Similarly, the precision of the recommended MGS-ROA-DBN calculated the positive values exactly. Therefore, the precision of the presented MGS-ROA-DBN is 55.4% better than NN, 52.6% better than KNN, and 46.1% better than SVM as well as DBN. Hence, it is concluded that the improved MGS-ROA-DBN is producing the best performance in detecting DR abnormalities.

7 Conclusion

The paper has presented a novel approach for developing automated DR detection by evaluating the retinal abnormalities such as hard exudates, haemorrhages, microaneurysms, and soft exudates. Here, the processing stages of DR included pre-processing, optic disc removal, blood vessel removal, segmentation of abnormalities, and classification. At first, the input retinal fundus image was enhanced using CLAHE. In the next stage, the removal optic disc was done using open-close watershed transformation. Later, the gray level thresholding was performed for removing and segmenting the blood vessels. After the removal of blood vessels and the optic disc, the segmentation of abnormalities was done by top-hat transformation and Gabor filtering. Next, the features such as LBP, TEM, Shannon's, and Kapur's entropy were extracted. As the feature vector seemed to very large the optimal feature selection was performed to select the individual features with less correlation. In addition, the categorization of images into four classes such as normal, earlier, moderate, or severe stages was done using the DBN-based classification algorithm. Here, the optimal feature selection and the weight update in DBN was done by the modified algorithm named MGS-ROA. From the results, it states that the accuracy of the suggested MGS-ROA-DBN is 30.1% improved than NN, 32.2% improved than KNN, and 17.1% improved than SVM and DBN, which proves the superior performance of the proposed model. The future extension of this work will consider the segmentation of abnormal features, such as cotton wool, spot drusen, and so on. In addition, the classification could be extended by improved machine learning or deep learning algorithms with new variant of optimization algorithm for attaining the high detection rate.

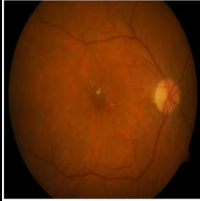


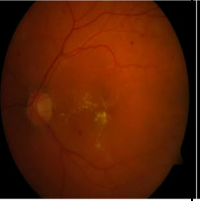
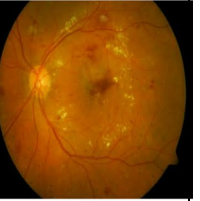
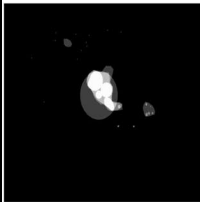
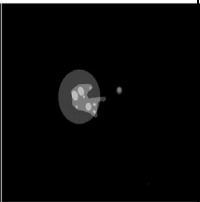
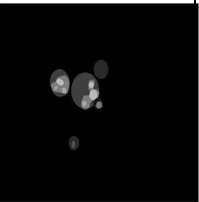
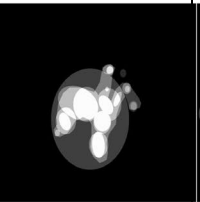
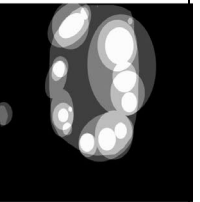
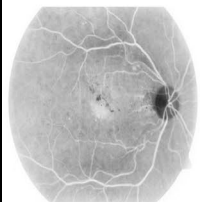
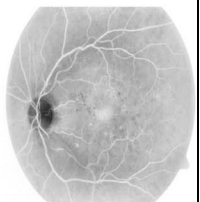
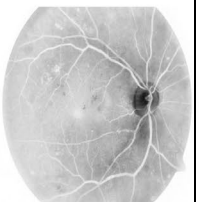
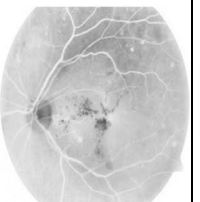
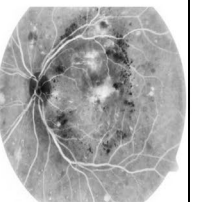
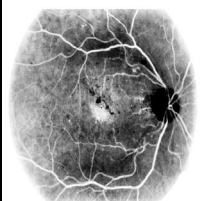
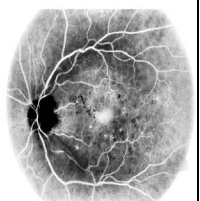
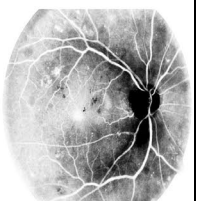

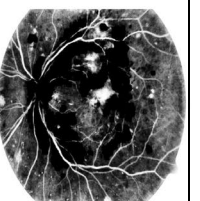
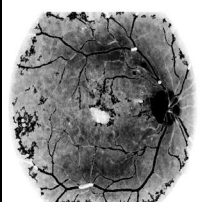
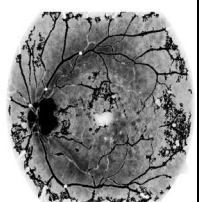
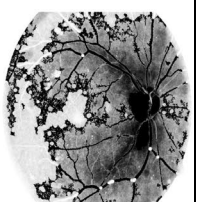
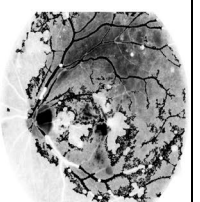
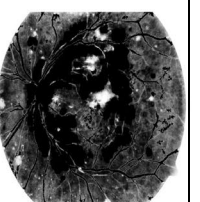
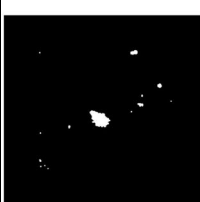
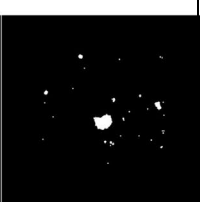
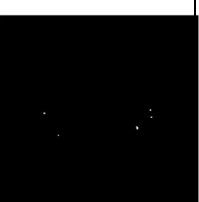

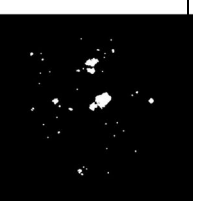
	Image 1	Image 2	Image 3	Image 4	Image 5
Original Image					
Ground Truth Images					
Pre-processing Images					
Optic disc removal Images					
Blood Vessels removal Images					
Segmented Images					

Fig. 5 Experimental results of pre-processed, optic disk removal, blood vessel removal and segmentation of abnormalities

Fig. 6 Performance analysis of proposed and conventional heuristic-based DBN for Diagnosing DR showing **a** accuracy, **b** sensitivity, **c** specificity, **d** precision, **e** FPR, **f** FNR, **g** NPV, **h** FDR, **i** F1 score, and **j** MCC

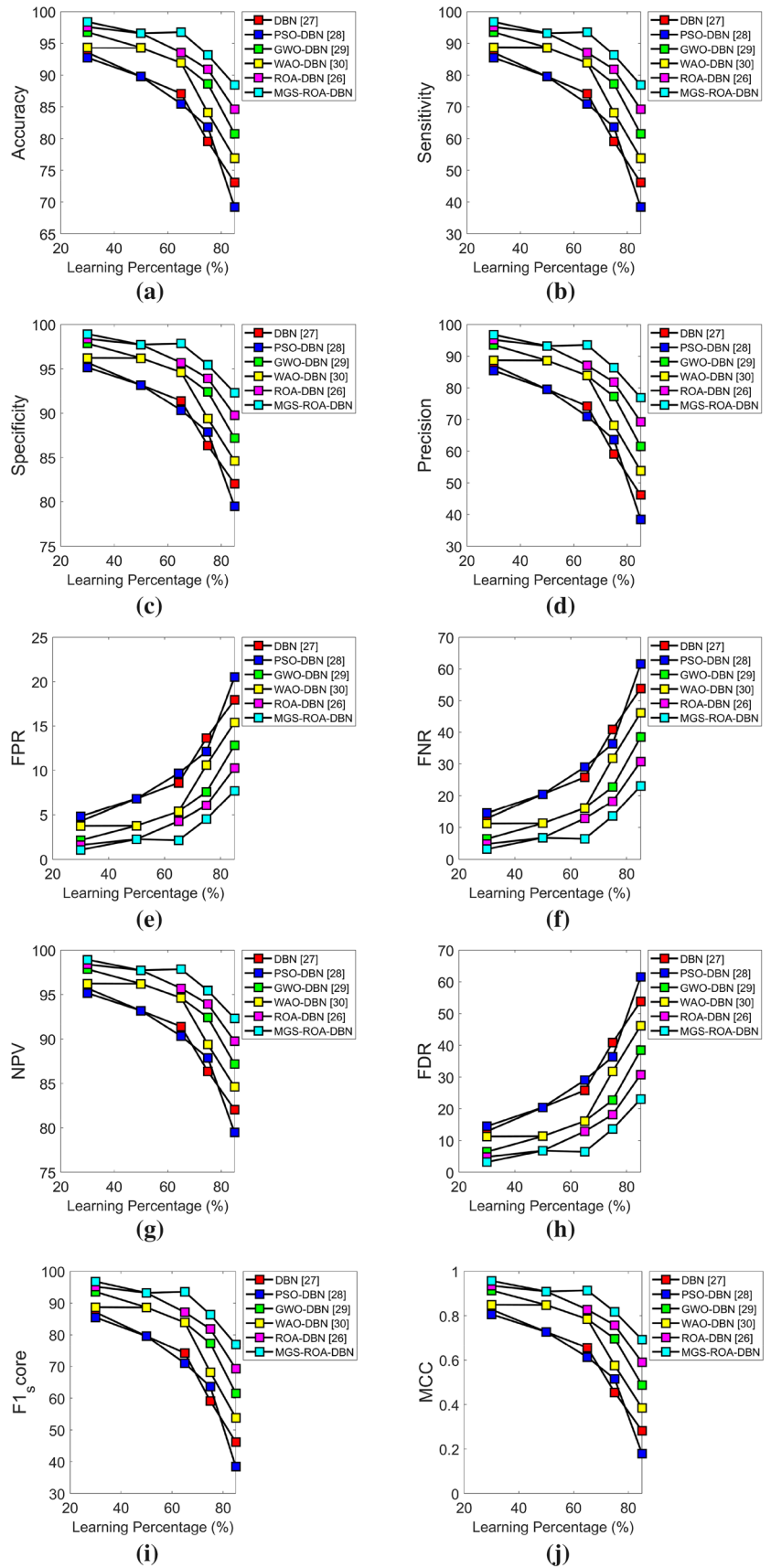


Table 2 Overall Performance analysis of proposed and conventional-based DBN for DR detection

Performance measures	PSO-DBN [36]	GWO-DBN [37]	WOA-DBN [38]	ROA-DBN [34]	MGS-ROA-DBN
Accuracy	0.81818	0.88636	0.84091	0.90909	0.93182
Sensitivity	0.63636	0.77273	0.68182	0.81818	0.86364
Specificity	0.87879	0.92424	0.89394	0.93939	0.95455
Precision	0.63636	0.77273	0.68182	0.81818	0.86364
FPR	0.12121	0.075758	0.10606	0.060606	0.045455
FNR	0.36364	0.22727	0.31818	0.18182	0.13636
NPV	0.87879	0.92424	0.89394	0.93939	0.95455
FDR	0.36364	0.22727	0.31818	0.18182	0.13636
F1-score	0.63636	0.77273	0.68182	0.81818	0.86364
MCC	0.51515	0.69697	0.57576	0.75758	0.81818

Table 3 Algorithmic analysis of the proposed and conventional models for detecting DR

Performance measures	NN [39]	KNN [40]	SVM [41]	DBN [35]	MGS-ROA-DBN
Accuracy	0.71591	0.70455	0.79545	0.79545	0.93182
Sensitivity	0.22727	0.40909	0.59091	0.59091	0.86364
Specificity	0.87879	0.80303	0.86364	0.86364	0.95455
Precision	0.38462	0.40909	0.59091	0.59091	0.86364
FPR	0.12121	0.19697	0.13636	0.13636	0.045455
FNR	0.77273	0.59091	0.40909	0.40909	0.13636
NPV	0.87879	0.80303	0.86364	0.86364	0.95455
FDR	0.61538	0.59091	0.40909	0.40909	0.13636
F1-score	0.28571	0.40909	0.59091	0.59091	0.86364
MCC	0.12943	0.21212	0.45455	0.45455	0.81818

References

- Klein BE (2007) Overview of epidemiologic studies of diabetic retinopathy. *Ophthalmic Epidemiol* 14(4):179–183
- Cusick M, Chew EY, Hoogwerf B, Agrón E, Wu L, and Lindley A (2004) Risk factors for renal replacement therapy in the early treatment diabetic retinopathy study (ETDRS), early treatment diabetic retinopathy study report no. 26, *Kidney Int*, vol 66, no 3, pp 1173–1179
- Rohan TE, Frost CD, Wald NJ (1989) Prevention of blindness by screening for diabetic retinopathy: a quantitative assessment. *BMJ* 299(6709):1198–1201
- Memon WR, Lal B, Sahto AA (2017) Diabetic retinopathy. *Prof Med J* 24(2):234–238
- Wu L, Fernandez-Loaiza P, Sauma J, Hernandez-Bogantes E, Masis M (2013) Classification of diabetic retinopathy and diabetic macular edema. *World J Diabetes* 4(6):290–294
- Singh N, Tripathi RC (2010) Automated early detection of diabetic retinopathy using image analysis techniques. *Int J Comput Appl* 8(2):18–23
- Akram M, Khan SA (2011) Automated detection of dark and bright lesions in retinal images for early detection of diabetic retinopathy. *J Med Syst* 36(5):3151–3162
- Abramoff M, Reinhardt J, Russell S, Folk J, Mahajan V, Niemeijer M, Quéllec G (2010) Automated early detection of diabetic retinopathy. *Ophthalmology* 117(6):1147–1154
- Gargeya R, Leng T (2017) Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* 124(7):962–969
- Quéllec G, Charrière K, Boudi Y, Cochener B, Lamard M (2017) Deep image mining for diabetic retinopathy screening. *Med Image Anal* 39:178–193
- Abramoff MD, Lavin PT, Birch M, Shah N, Folk JC (2018) Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med* 1(1):39
- Seoud L, Hurtut T, Chelbi J, Cheriet F, Langlois JMP (2016) Red lesion detection using dynamic shape features for diabetic retinopathy screening. *IEEE Trans Med Imaging* 35(4):1116–1126
- Welikala RA, Fraz MM, Dehmeshki J, Hoppe A, Tah V, Mann S, Williamson TH (2015) Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy. *Comput Med Imaging Graph* 43:64–77
- Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR), vol 1, pp 886–893
- Lowe DG (1999) Object recognition from local scale-invariant features. *Proceedings of the seventh IEEE international conference on computer vision (ICCV)* 2:1150–1157
- Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: application to face recognition. *IEEE Trans Pattern Anal Mach Intell* 28(12):2037–2041
- Jones JP, Palmer LA (1987) An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J Neurophysiol* 58(6):1233–1258

18. Qummar S, Khan FG, Shah S, Khan A, Shamshirband S, Rehman ZU, Khan IA, Jadoon W (2019) A deep learning ensemble approach for diabetic retinopathy detection. *IEEE Access* 7:150530–150539
19. Kar SS, Maity SP (2018) Automatic detection of retinal lesions for screening of diabetic retinopathy. *IEEE Trans Biomed Eng* 65(3):608–618
20. Wang J, Bai Y, Xia B (2019) Feasibility of diagnosing both severity and features of diabetic retinopathy in fundus photography. *IEEE Access* 7:102589–102597
21. Chakraborty S, Jana GC, Kumari D, Swetapadma A (2019) An improved method using supervised learning technique for diabetic retinopathy detection. *Int J Inf Technol* 1–5
22. Costa P, Galdran A, Smailagic A, Campilho A (2018) A weakly-supervised framework for interpretable diabetic retinopathy detection on retinal images. *IEEE Access* 6:18747–18758
23. Zhou W, Wu C, Chen D, Yi Y, Du W (2017) Automatic microaneurysm detection using the sparse principal component analysis-based unsupervised classification method. *IEEE Access* 5:2563–2572
24. Jude Hemanth D, Deperlioglu O, Kose U (2019) An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network. *Neural Comput Appl* 1–15
25. Xu J, Zhang X, Chen H, Li J, Zhang J, Shao L, Wang G (2018) Automatic analysis of microaneurysms turnover to diagnose the progression of diabetic retinopathy. *IEEE Access* 6:9632–9642
26. Sahu S, Singh AK, Ghrera SP, Elhoseny M (2019) An approach for de-noising and contrast enhancement of retinal fundus image using CLAHE. *Opt Laser Technol* 110:87–98
27. Madhumitha S, Manikandan M (2018) Quantitative analysis of marker-based watershed image segmentation. Department of Electronics Engineering, Madras Institute of Technology, Anna University, vol 114, no 5, 10 Mar 2018.
28. Dey S, Bhattacharyya S, Maulik U (2014) Quantum inspired genetic algorithm and particle swarm optimization using chaotic map model based interference for gray level image thresholding. *Swarm and Evol Comput* 15:38–57
29. Yahya AA, Tan J, Hu MA (2013) A novel model of image segmentation based on watershed algorithm. *Adv Multimed*. <https://doi.org/10.1155/2013/120798>
30. Khaleefah SH, Mostafa SA, Mustapha A, Nasrudin MF. (2019) The ideal effect of Gabor filters and uniform local binary pattern combinations on deformed scanned paper images. *J King Saud University—Comput Inf Sci*
31. Liao S, Law MWK, Chung ACS (2009) Dominant local binary patterns for texture classification. *IEEE Trans Image Process* 18(5):1107–1118
32. Gupta R, Undrill PE (1995) The use of texture analysis to delineate suspicious masses in mammography. *Phys Med Biol* 40:835–855
33. Mookiah MR, Acharya UR, Martis RJ, Chua CK, Lim CM, Ng EY, Laude A (2013) Evolutionary algorithm based classifier parameter tuning for automatic diabetic retinopathy grading: a hybrid feature extraction approach. *Knowl-Based Syst* 39:9–22
34. Binu D, Kariyappa BS (2018) RideNN: a new rider optimization algorithm-based neural network for fault diagnosis in analog circuits. *IEEE Trans Instrum Meas* 68(1):2–26
35. Liu Y, Zhou H, Tsung F, Zhang S (2019) Real-time quality monitoring and diagnosis for manufacturing process profiles based on deep belief networks. *Comput Ind Eng* 136:494–503
36. Pedersen MEH, Chipperfield AJ (2010) Simplifying particle swarm optimization. *Appl Soft Comput* 10(2):618–628
37. Mirjalili S, Mirjalili SM, Lewis A (2014) (2014) Grey wolf optimizer. *Adv Eng Softw* 69:46–61
38. Mirjalili S, Lewis A (2016) The whale optimization algorithm. *Adv Eng Softw* 95:51–67
39. Fernández-Navarro F, Carbonero-Ruz M, Becerra Alonso D, Torres-Jiménez M (2017) Global sensitivity estimates for neural network classifiers. *IEEE Trans Neural Netw Learn Syst* 28(11):2592–2604
40. Chen Y, Hu X, Fan W, Shen L, Zhang Z, Liu X, Du J, Li H, Chen Y, Li H. (2019) Fast density peak clustering for large scale data based on kNN. *Knowl-Based Syst*, Available online 3 July 2019.
41. Yu S, Tan KK, Sng BL, Li S, Sia ATH (2015) Lumbar ultrasound image feature extraction and classification with support vector machine. *Ultrasound Med Biol* 41(10):2677–2689

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.