

RESEARCH

Open Access



Automatic multi-disease classification on retinal images using multilevel glowworm swarm convolutional neural network

Rupali Chavan^{1*} and Dnyandeo Pete¹

*Correspondence:
crupali0502@gmail.com

¹ Department of Electronic Engineering, Datta Meghe College of Engineering, Navi Mumbai, India

Abstract

In ophthalmology, early fundus screening is an economical and effective way to prevent blindness from eye diseases. Because clinical evidence does not exist, manual detection is time-consuming and may cause the situation to be delayed clinically. With the development of deep learning, a wide variety of eye diseases have shown promising results; however, most of these studies focus on only one disease. Therefore, focusing on multi-disease classification based on fundus images is an effective approach. Consequently, this paper presents a method based on the multilevel glowworm swarm optimization convolutional neural network (MGSCNN) for the classification of multiple diseases. It is proposed that the proposed system has two stages, namely preprocessing and classification. In the beginning, the images are normalized, smoothed, and resized to prepare them for preprocessing. After pre-processing, the images are fed to the MGSCNN classifier to classify an image as normal or abnormal (covering 39 different types of diseases). In the CNN classifier, with the help of Glowworm Swarm Optimizer (GSO), we optimally detect the structure and hyperparameters of CNN simultaneously. This approach achieves an excellent accuracy of 95.09% based on various metrics.

Keywords: Fundus screening, Deep learning, Multi-disease, Classification, MGSCNN, Glowworm Swarm Optimizer

Introduction

Globally, fundus diseases account for the majority of blindness [1]. In addition to glaucoma, diabetic retinopathy (DR), age-related macular degeneration (AMD), and cataract, there are several other types of ophthalmic diseases [2]. Furthermore, chronic diseases such as diabetes can also lead to ocular damage, infectious diseases of the eye, and diabetic retinopathy [3]. Although this is the case, it is important to remember that people of all ages may suffer from vision loss. The risk of becoming blind or suffering from vision impairment increases with age. An image of the retinal fundus can reveal lesions or other abnormalities that may indicate a disease [4]. However, early detection of a disease can save a person's sight. The recommendation is therefore to conduct a comprehensive pathological examination on an annual basis. The retinal fundus image typically shows the retinal background,

blood vessels, macula, and fovea [5]. A fundus image can be used to identify specific diseases within the retina. In addition to diabetes mellitus, systemic hypertension, leukemia, AIDS, and many others, cottonwool spots can result from a variety of medical conditions [6]. In the standard structure, the optic disc and cup may be affected by certain diseases.

The number of people with DR is expected to increase by more than 400 million by 2030, and the number of people with glaucoma will increase by 80 million by 2020 [7]. According to the World Health Organization, China has one of the highest rates of visual impairment in the world [8]. There is less than 10% screening for diabetic retinopathy in Chinese people between the ages of 20 and 65, which is the leading cause of blindness among people between the ages of 20 and 65. It is estimated that by 2050, 50% of the global population will be affected by myopia, which is higher than the global average. Retinal tears and retinal detachments are the most common causes of blindness worldwide.

Eye diseases can cause irreversible blindness due to their irreversible nature. By early detection of vision disorders, about 80% of vision disorders can be prevented [9]. But the number of ophthalmologists is not balanced with the number of patients. Additionally, fundus screening is a manual procedure that requires a lot of experience on the part of the ophthalmologist [10]. As a result of these factors, conducting large-scale fundus screening is difficult. It has already been demonstrated that some deep learning models are capable of achieving significant performance in the diagnosis of eye diseases [11]. However, most identification models focus on only one eye disease. To overcome the limitations listed above, this paper proposes a multi-disease classification method based on a multilevel glowworm swarm-optimized convolutional neural network (MGSCNN).

The following contributions are made to this work.

- Initially, images are collected from the RFMiD dataset and then the images are pre-processed using the following stages, normalization, smoothing, and resizing.
- After pre-processing, images are fed to the MGSCNN classifier to classify the image as normal or abnormal (covering 39 types of diseases/pathologies).
- In the classification phase, the Glowworm Swarm Optimizer simultaneously finds the optimal structure and high parameter of the CNN.
- For experimental analysis, the Retinal Fundus Multi-Disease Image Dataset (RFMiD) is used and performance is analyzed with different metrics.

This article is organized as follows: The “[Literature survey](#)” section reviews some multi-disease classification models. CNN is used to automatically classify retinal images for multiple diseases as outlined in the “[Proposed MGSCNN-based multi-disease detection model](#)” section. The “[Results and discussion](#)” section discusses the effectiveness of MGSCNN. The results of the experiment are presented in the “[Conclusions](#)” section of the paper.

Literature survey

For this work, we reviewed several existing works. Some of them are given below; Smitha and Jidesh [12] have developed an end-to-end fundus image analysis system. To classify retinal fundus images into multiple categories, a semi-supervised generative adversarial network (GAN) was employed. In addition, non-spatial Retinex architecture was utilized to improve the fundus images without over-smoothing. The study

utilized a huge collection of raw fundus data from several eye hospitals. The average accuracy of this method was 87% in comparison with the transfer learning method.

Bhati et al. [13] proposed a discriminative kernel convolution network (DKCNet) that analyzes discriminating features in regions without incurring additional computing costs. Two modules comprise DKCNet: a module for attention and a module for pressing and stimulating (SE). As a result of the attention block, discriminative feature attention maps were generated from the backbone network. Through the SE module, channel interdependence is improved by taking discriminative feature maps. Based on the study of ODIR-5K fundus images, DKCNet showed the best performance with an Inception-Resnet backbone network, achieving an AUC of 96.08, an F1-score of 94.28, and a Kappa score of 0.81.

Chen et al. [14] proposed a network consisting of several branches that are based on attention to classify diseases across four different subject groups. An integrated module for multi-scale feature fusion as well as an integrated module for dual focus is used in this method. A multiscale feature fusion module was used to identify small-scale lesions. Combining the dual attention module and the global attention map allowed for a deeper exploration of the acquired features. To validate the performance of this model, extensive validations were conducted on private and public datasets. The method was found to be accurate.

Casado-García et al. [15] proposed analyzing retinal fundus images for the diagnosis of diabetic retinopathy and glaucoma. A model for detecting ERM automatically was developed by investigating several deep learning frameworks and various training methods. Thus, 86.82% F1 was obtained as a result of developing appropriate models.

Li et al. [16] developed a dataset of fundus image annotations that included multiple diseases to reduce the lack of benchmark datasets preventing the automated classification of clinical fundus images. The dataset consists of 10,000 images that were collected from the left and right eyes of patients in 5000 clinical trials. The dataset was also used to evaluate some existing deep learning models, which may prove useful for future research in this field. Multiple deep networks were combined in the experiments to enhance classification performance more than increasing the depth of a neural network alone.

Zhang et al. [17] proposed DeepUWF-Plus as a set of supplementary screening methods combining deep learning and UWF imaging technology. A fundus screening subsystem, a fundus abnormality detection subsystem in four key fundus locations, and a fundus disease detection subsystem are included in the service. Experimentally, two-level and one-level classification strategies were examined to overcome severe class disparities and homogeneity between classes. The results of DeepUWF-Plus tests demonstrate that it is effective at detecting minor diseases, especially when used in a two-stage approach.

Rodriguez et al. [18] described the use of fundus images from different sources to diagnose multiple retinal diseases using a multi-label classification method. MuReD was constructed by combining several publicly available datasets for fundus disease classification. Following the acquisition of the image data, several post-processing procedures were implemented to ensure the quality of the data as well as the range of diseases present. The architecture of the system has been improved through several experiments.

Based on AUC scores, it outperformed the state-of-the-art by 7.9% and 8.1%, respectively, in the diagnosis and classification of diseases.

Müller et al. [19] investigated the performance impact of ensemble learning techniques: augmenting, stacking, and packing. It includes nine deep convolutional neural network architectures, as well as sophisticated methods for preprocessing and enhancing images. The algorithm was applied to four popular medical imaging datasets of different complexity levels. Several different pooling functions were examined, ranging from unweighted averaging to support vector machines, which are more complex learning functions. Stacking improved F1scores by up to 13% according to our results.

He et al. [20] developed a multidimensional feature extraction module for fundus images. Further, the OCT image contains large areas of background that make diagnosis impossible. To encode features of the retinal layer, we used a region-guided focus block, which ignored the background of the OCT images. To create a multi-model feature, a multi-model retinal image classification network is trained based on specific features. By combining the advantages of fundus imaging and optical coherence tomography, the model was able to provide an accurate diagnosis. A clinically acquired multimodal retinal image dataset (fundus and OCT) was used to demonstrate the effectiveness of this MSAN (mode-specific attention network) in comparison to other well-known single-modal and multi-modal retinal image classification algorithms.

Proposed MGSCNN-based multi-disease detection model

This work proposes a multi-disease classification of retinal images based on MGSCNN. Figure 1 presents the overall architecture for the proposed MGSCNN-based multi-disease classification model. Initially, the images are pre-processed using normalization, smoothing, and rescaling. Following pre-processing, the images are fed into the MGSCNN classifier for classification as normal or abnormal. In this classification model, with the help of GSO, MGSCNN can select the classifier structure and hyperparameter.

Preprocessing

Preprocessing is required to prepare image data for model input. In general, all images must be presented in equal-sized sequences for convolutional neural networks to fully connect. Additionally, model preprocessing can reduce the time required to train a model and speed up the rate at which it can be inferred. As a pre-processing step, we use techniques like normalization, smoothing, and resizing.

Normalization: During normalization, the range of intensity values of pixels is linearly changed.

Smoothing: Smoothing is used to smooth out input images that are noisy and/or broken. To achieve a smooth shape, some pixels must be added to the image.

Resizing: To determine the effects of different sizes on the recognition process, images are rescaled to different sizes. The best image size should be determined by carefully exploring the differences between different image sizes. To speed up processing time, image resizing reduces the data size of the image. Image sizes vary from 0.1 to 0.9 depending on the amount of resizing. If a small amount of resizing is per-

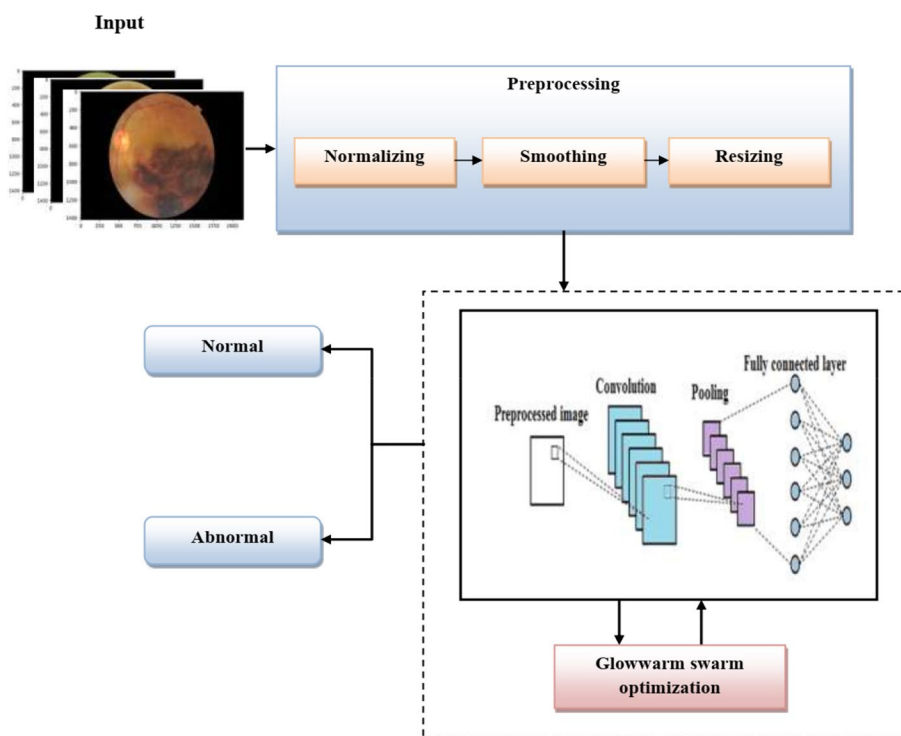


Fig. 1 Architectural diagram of proposed MGSCNN

formed on an image, many important features may be lost, particularly if image texture is used during classification.

MGSCNN

MGSCNN is a combination of CNN and GSO optimization algorithms. Here, the CNN acts as a classifier, and the GSO technique is used to improve the accuracy of CNN by optimizing the structure and hyperparameter of the CNN.

Glowworm swarm optimization (GSO)

The GSO is a two-dimensional workspace in which each artificial glow, or agent, is carrying a light and has its view, known as the local resolution range. To determine Luciferin's position, it is necessary to consider its objective value. Agents with higher levels of intelligence are more likely to fly to better positions (have a higher objective value). A neighbor with luciferin intensity greater than its magnitude within the local decision range will be fanned toward by the agent when it detects a neighbor with luciferin intensity greater than its magnitude. According to the number of neighbors, there is a different local decision limit. There is an increase in threshold when there are fewer neighbors, and a decrease in threshold when there are more neighbors. Regardless of which neighbor is selected, the agent always changes its direction of movement. As luciferin levels increase, a neighbor becomes more attractive. In addition, most agents are located in several locations at the same time. There are three main phases to GSO: the luciferin update phase, the motion phase, and the decision threshold phase.

Luciferin update phase: Luciferin is updated according to the value of the gloss state function. All glowworms begin with the same level of luciferin, but luciferin levels vary with the level of activity in the glowworm’s current state even though all glowworms have the same level of luciferin in the initial iteration. An individual’s perception of the temperature and radiation levels at a particular location determines the value of luciferin. Each glowworm increases its level of luciferin in addition to its previous level. Luminescence values are subtracted from previous luminescence values to simulate decay in the glow. The updated rule of Luciferin value is as follows:

$$L_g(d + 1) = (1 - \rho)L_g(d) + \gamma J_g(d + 1) \tag{1}$$

Luciferin level of a glowworm is represented by $L_g(d)$, in which ρ represents luciferin decay constant $0 < \rho < 1$, γ represents luciferin enhancement constant, and J_g represents an objective function at the location of agent i at time t .

Movement phase: The movement phase involves each glowworm determining the movement of a neighbor with a higher level of luciferin than itself through the use of a probabilistic mechanism. To attract glowworms, neighbors who emit a brighter glow are attractive. The probability of each glowworm g moving toward a neighbor h can be calculated as follows:

$$P_{gh} = \frac{L_h(d) - L_g(d)}{\sum_{n \in K_g(d)} L_n(d) - L_g(d)} \tag{2}$$

where $h \in K_g(d)$, $K_g(d) = \{h; E_{g,h}(d) < r_E^g(d); L_g(d) < L_h(d)\}$ is a set of neighborhood of glowworm g at time d . $E_{g,h}(d)$ denotes the Euclidean distance between glowworms g and h at time d , and $r_E^g(d)$ denotes the variable neighborhood range associated with glowworms g at time d . Assume glowworm g selects a glowworm $h \in K_g(d)$ with $P_{gh}(d)$ given by (2). Therefore, glowworm movements can be described as follows:

$$y_g(d + 1) = y_g(d) + S(y_h \left(\frac{y_h(d) - y_g(d)}{\|y_h(d) - y_g(d)\|} \right)) \tag{3}$$

where S represents the step size, and $\| \cdot \|$ represents the Euclidean norm operator.

Decision range update: Every agent is associated with a neighborhood whose radial range r_e^g is dynamic in nature $0 < r_e^g < r_s$. The luciferin sensor has a radial range referred to as r_s . It must be justified why there is no fixed neighborhood range. When glowworms are reliant only on local information for movement, the number of peaks captured will vary according to the range of the radial sensor. Agents whose sensors are capable of covering the entire search space will move to the global optimum. There is no consideration of the local optimum in this case. As a result, determining a neighborhood range that is suitable for different function landscapes can be difficult without a priori knowledge of the objective function (e.g., peak number,

inter-peak distance, etc.). As a general rule, objective functions with minimum inter-peak distances greater than r_e are preferred over those with minimum interpeak distances lower than r_e . Therefore, GSO detects multiple peaks by utilizing an adaptive neighborhood range in a multimodal function landscape. When the following rule is applied, performance appears to be significantly reduced:

$$r_e^g(d + 1) = \min \left\{ rS, \max \left\{ 0, r_e^g(d) + \beta(kd - |Kg(d)|) \right\} \right\} \tag{4}$$

This equation is expressed as a function of the constant parameter β and the number of neighbors parameter kd .

Convolution neural network (CNN)

There are several types of artificial neural networks (ANNs), including convolutional neural networks (CNNs). CNNs are capable of automatically learning hierarchies of features from input image matrices as opposed to handcrafted features extracted by elaborate algorithms. CNN models have achieved several revolutionary advances in computer vision in recent years, including classification, segmentation, and object tracking. As a result of fewer connections and fewer parameters, CNN architecture is capable of sharing and pooling weight parameters.

Figure 2 illustrates the architecture of CNN. A typical CNN architecture consists of several convolution layers nested within one another, followed by a fully connected layer. This type of network can be presented as follows in a simplified manner:

Input: The input of a CNN typically consists of matrices of 3-channel color or 1-channel gray images containing intensity values at each position.

Conv (convolution layer): The output of the last layer is filtered in a small region by each of the convolution layers. Moreover, the filters are usually small learnable matrices of 3×3 or 5×5 dimensions. Using parameter sharing, one filter is convolved across all spatial dimensions to extract one feature from an image.

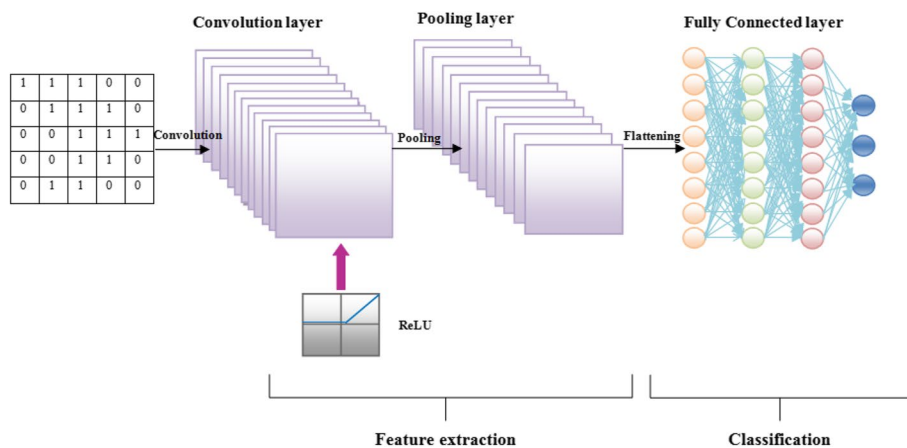


Fig. 2 Architecture of CNN

ReLU (Rectified Linear Units): In convolutional and fully connected layers, ReLUs are commonly used as activation functions for introducing non-linear transformations. The formula for this function is $f(x) = \max(0, x)$. It has been demonstrated experimentally that ReLU is superior to conventional sigmoid-like activation functions in the development of deep networks in recent years as it prevents training convergence and gradient saturation while maintaining as much original value as possible.

Pool: It is possible to reduce the spatial size of the output by downsampling both spatial dimensions in a nonlinear manner by using the pooling layer. To reduce the computation costs and parameters of the network, it is necessary to reduce its parameters. When the input feature map is placed between two successive convolution layers, max pooling can produce the maximum value.

FCL (fully connected layer): The FCLs are the endpoints of an artificial neural network. There is an interconnection between each neuron in the FCL at the last layer. There are N neurons in the last FCL of this network that generate the output from all the input labels. The probability of appearing for each label in the N -dimensional output is calculated using the softmax function.

$$P(z_i) = \frac{\exp(z_i)}{\sum_{i=1}^N \exp(z_i)} \quad (5)$$

In the second last layer, $P(z_i)$ denotes the probability of predicting the i^{th} value. To make decisions, all layers must be stacked together to form a CNN.

The proposed MGSCNN

This section provides an overview of the proposed MGSCNN, along with flow diagrams, algorithms, and architecture. Using multiple swarms, MGS constructs the CNN structure and its higher parameters. Glowworms represent possible configurations of the CNN. To determine the probability that samples from each class will appear in a CNN, a softmax classification layer is applied as the final layer. The merit value of each glowworm is determined by the accuracy of the results obtained. To determine the best configuration for a CNN, GSO optimizes the hyperparameters. An optimal set of high criteria can be used in the design of the CNN framework to allow it to be trained in one step using a large number of training samples. To classify unknown samples, an optimized CNN based on trained parameter values is used. The MGSCNN basic architecture is given in Fig. 3.

Figure 4 illustrates the workflow diagram of MGSCNN. Stage 1 initializes a set of layers $[G_1, G_2, \dots, G_i]$ with random values for pooling, convolution, and fully connected layers. At swarm level 2, multiple swarms $([G_{11}, G_{12}, \dots, G_{1j}], [G_{21}, G_{22}, \dots, G_{2j}], \dots, [G_{i1}, G_{i2}, \dots, G_{ij}])$ consisting of j glowworms each are set up. In swarm level 2, glowworms are randomly initialized by the number of filters, stride size, filter size, padding for the convolution layer requirements for the pooling layer, and the number of output neurons for the fully connected layers.

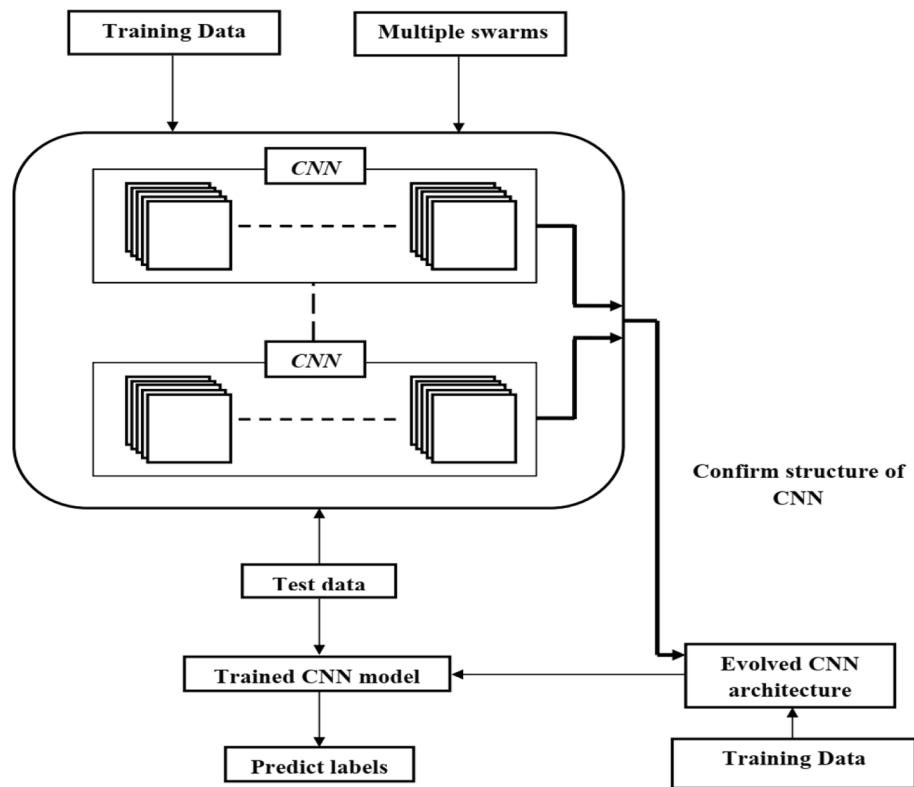


Fig. 3 MGSCNN basic architectural diagram

Using CNN, features are extracted from each glowworm of level 2, and the softmax layer calculates the accuracy of each glowworm (fitness value).

It is necessary to repeat this procedure for the first level of swarming and then for the second level of swarming to achieve maximum accuracy for the CNN based on the given search space. It is represented by $[G_m, G_{mk}]$ that the glowworms traversed so far have been traversed with the lowest error value, where G_m is the number of layers of the swarm level-1 glowworm and G_{mk} represents the amount of layers of every type in an evolved CNN, and G_{mk} provides all hyperparameters required at each layer. The workflow diagram of MGSCNN is presented in Fig. 4 and step-by-step process is explained below section.

Step 1: Swarms initialization in hyperparameter search space A convolutional neural network (CNN) is optimized by a method known as GSO, which optimizes 11 hyperparameters. In a glowworm, the first level consists of three hyperparameters: number of convolution layers (Cn), number of fully connected layers (Fn), number of pooling layers (Pn), and the second level of a swarm contains eight hyperparameters: size of filter/kernel in the convolutional layer (s_{fc}), number of filters in the convolutional layer (n_{fc}), size of stride in the convolutional layer (s_{sc}), padding (valid or same) requirement in the convolutional layer (p_{pc}), size of stride in the max-pooling layer (s_{sp}), size of a filter in the max-pooling layer (s_{fp}), padding pixels in pooling layer (p_{pp}), and the number of output neurons in the fully connected layer (n_{of}). An optimal CNN

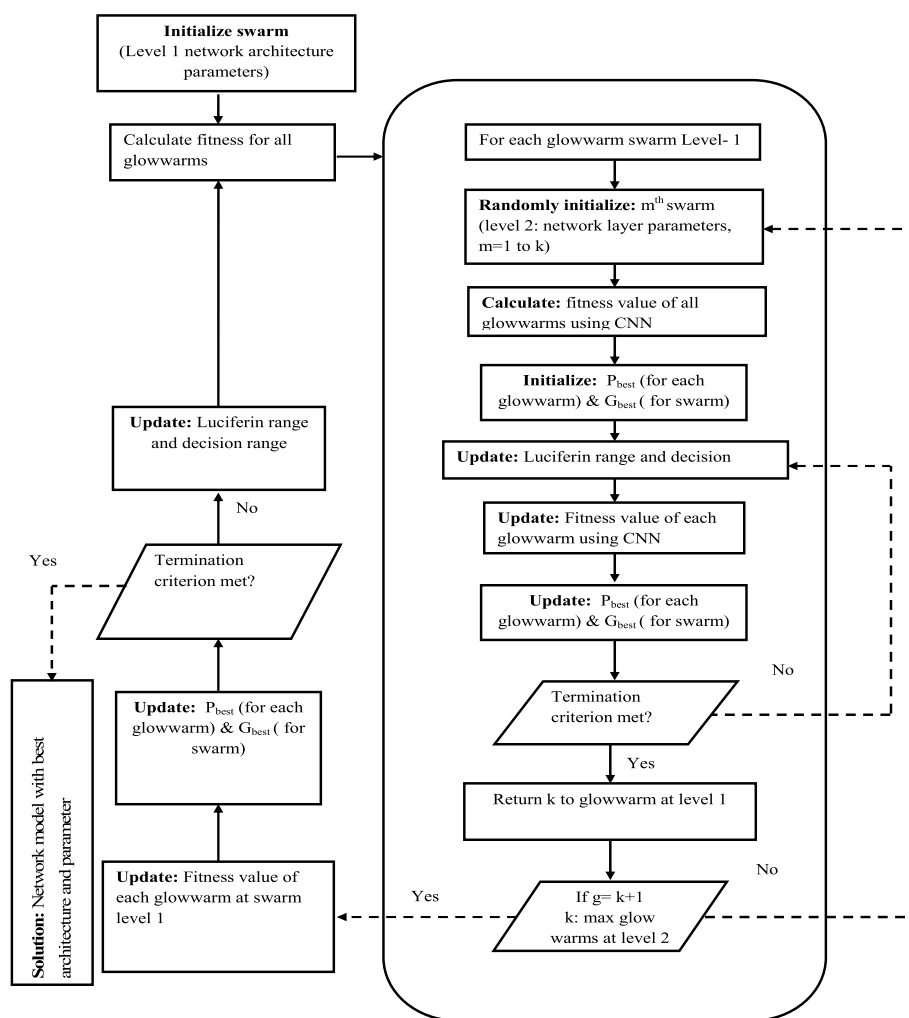


Fig. 4 Workflow diagram for MGSCNN

hyperparameter set is determined by randomly initializing glowworms within the specified range. According to Table 1, the dimensions of the glowworms are controlled by hyperparameters whose minimum and maximum values are described.

Step 2: Structure of swarm Figures 5 and 6 illustrate the architecture of multilevel, multidimensional swarms at swarm levels 1 and 2 to provide a deeper understanding of their behavior. At level 1, each type of layer is represented by a swarm with five glowworms. A CNN can reach a level-2 extension of hyperparameters at the end of its evolution. A glowworm is represented in stage 1 by a swarm of layers. In stage 2, we explore five swarms based on the number of layers within each swarm. Stage 2 will result in the creation of five swarms. There are five glowworms in each swarm of level 2, and each glowworm has a dimension of $C_n \times 8$. As a result, the swarm at stage 2 has a dimension of $5 \times C_n \times 8$. There are eight parameters to optimize in a convolutional layer and fully connected layer: filters, padding bits, strides, padding bits, and output neurons.

Table 1 Range of hyperparameters

Layers	Hyper parameters	Range	
		Minimum value	Maximum value
Convolution	1. Number of convolution layers (Cn)	1	5
Pooling	2. Number of pooling layers (Pn)	1	5
Fully connected	3. Number of fully connected layers (Fn)	1	5
Convolution	1. Number of Filters (n_fc),	1	64
	2. Size of filter/kernel (s_fc) (odd),	1	13
	3. padding pixels (p_pc),	0 (valid)	1 (same)
	4. size of stride (s_sc) (<s_fc),	1	$g = \frac{s_fc - 1}{2}$
Pooling	5. Filter size (s_fp), (odd)	1	13
	6. Stride size (s_sp),	1	5
	7. padding pixels (p_pp)(<s_fp)	0 (valid)	1 (same)
			$g = \frac{s_fp - 1}{2}$
Fully connected	8. number of output neurons (n_of)	1	1024

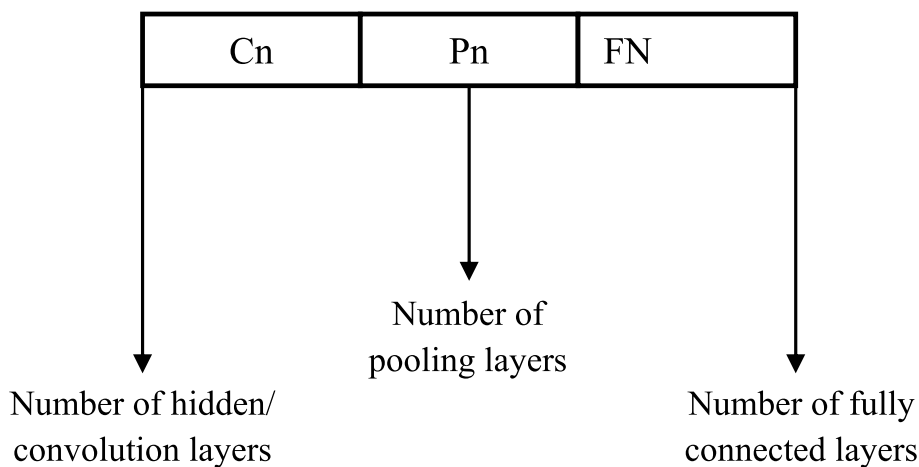


Fig. 5 Structure of glowworm at level 1

Step 3: Fitness evaluation Glowworm fitness is evaluated using a CNN followed by a softmax layer in MGSO. Comparing CNN hyperparameters with those from another glowworm that provides a lower level of accuracy, the CNN configuration that provides the best accuracy is the most optimal. It can be defined using Eq. (7)

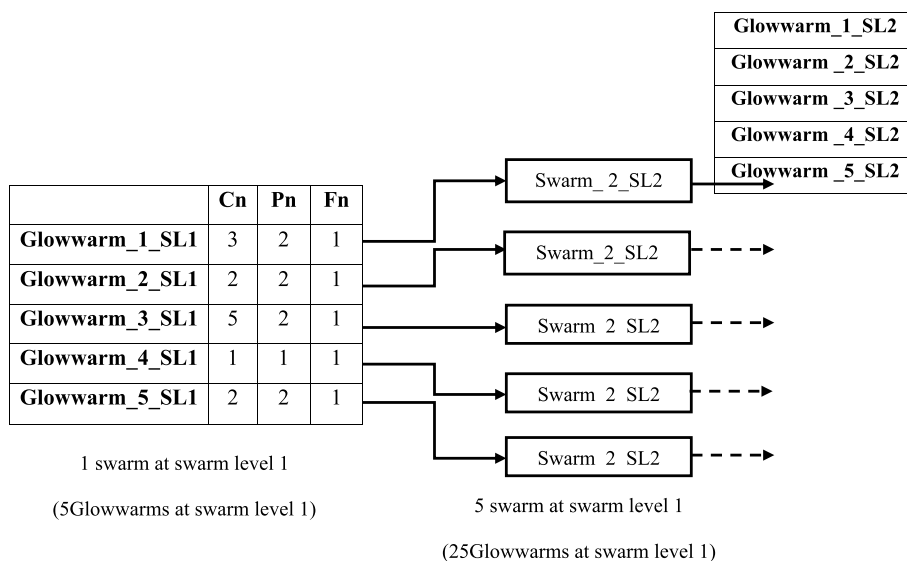


Fig. 6 Level 1 and level 2 architectures of swarms

$$Fitness = \max[accuracy] \tag{7}$$

Step 4: Update the solution After fitness calculation, the solutions are updated using the GSO algorithm. In this stage, luciferin range, movement phase, and the decision range are updated using Eqs. (1), (3), and (4).

Step 5: Termination To terminate the GSO, the best luciferin range, moment, and decision must be obtained to fulfill the termination criteria. The algorithm will be terminated once the solution has been obtained.

Results and discussion

This work proposes MGSCNN-based multilevel classification using fundus images. The evaluation parameters for the proposed work are shown in Table 2. An Intel Core i7 processor with 8 GB of memory was used with Windows 10 as the operating system for simulating the proposed multi-stage disease.

Dataset description

It contains approximately 3200 retinal fundus images interpreting 46 conditions. The data is included in the Retinal Fundus Multi-Disease Image Dataset (RFMiD). In our proposed work, 39 conditions are detected from these 46 conditions. There are a variety of diseases that can be found in routine clinical settings in RFMiD, which is a publicly available dataset. Due to this challenge, general models for retinal screening are being developed, in contrast to previous efforts focused on detecting specific diseases. Experimental used sample images are listed in Fig. 7. There are five classes of diseases represented by Fig. 7: age-related macular degeneration (ARMD), central retinal vein occlusion (CRVO), optic disc center (ODC), diabetic retinopathy (DR), and branch retinal vein occlusion (BRVO).

Table 2 Parameter values of the proposed model

Parameters	Range
Input size	256 × 256 × 3
Output size	5
Epoch	50
Optimizer	Adam
Learning rate	0.0010340233 (using glowworm swarm optimization)
Loss function	Categorical_Crossentropy
Convolutional kernel size	3
Pooling kernel size	2
Activation function for activation layer	“relu”
Batch size	64
Activation function	Softmax

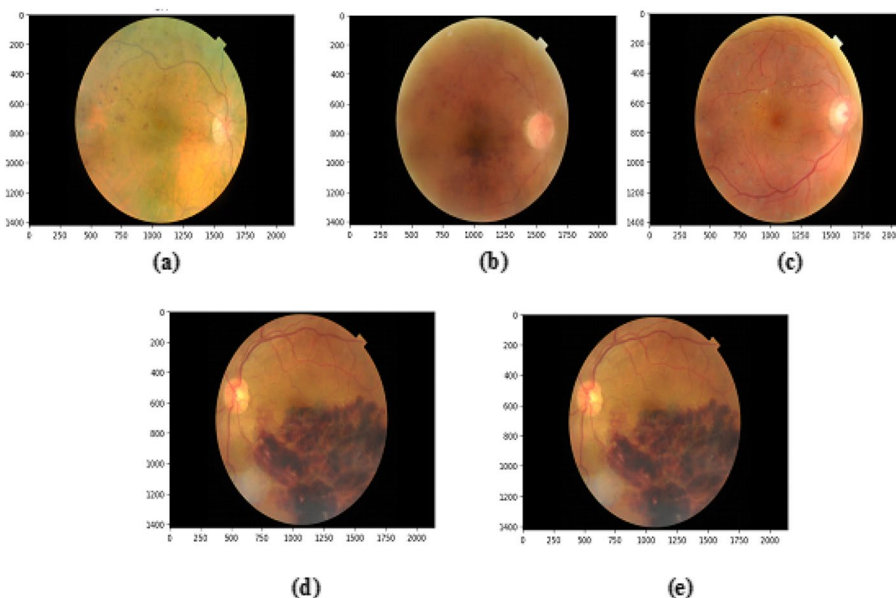


Fig. 7 Sample results for the proposed work with five classes: **a** DR, **b** ARMD, **c** ODC, **d** CRVO, and **e** BRVO

Experimental results

The results obtained from the proposed model are presented in this section. Table 3 shows the model comparison metrics evaluation for DR with existing techniques. Our proposed work achieved an accuracy of 94.02%. Our proposed task is improved by 3.67% from CNN, 4.6% from SVM, and 24.4% from LSTM. The sensitivity of our proposed method is 95.18%, which is improved by 3.46% from CNN, 4.55% from SVM, and 30.28% from LSTM. The specificity of the proposed MGSCNN is 92.98%, which is improved by 3.99%, 4.83%, and 17.45% from the existing CNN, SVM, and LSTM. The accuracy of our proposed method is 92.5%, which is improved by 3.24% from CNN, 4.35% from SVM, and 15.67% from LSTM. The recall rate of our proposed task is 95.16%, and our proposed task is improved by 3.46%, 4.55%, and 30.28 from the

Table 3 Comparative result evaluation for DR

	Accuracy	Sensitivity	Specificity	Precision	Recall	F-score
Proposed	94.02	95.18	92.98	92.5	95.18	93.82
CNN	90.35	91.72	88.99	89.26	91.72	90.47
SVM	89.42	90.63	88.26	88.15	90.63	89.37
LSTM	69.62	64.9	75.53	76.83	64.9	70.37

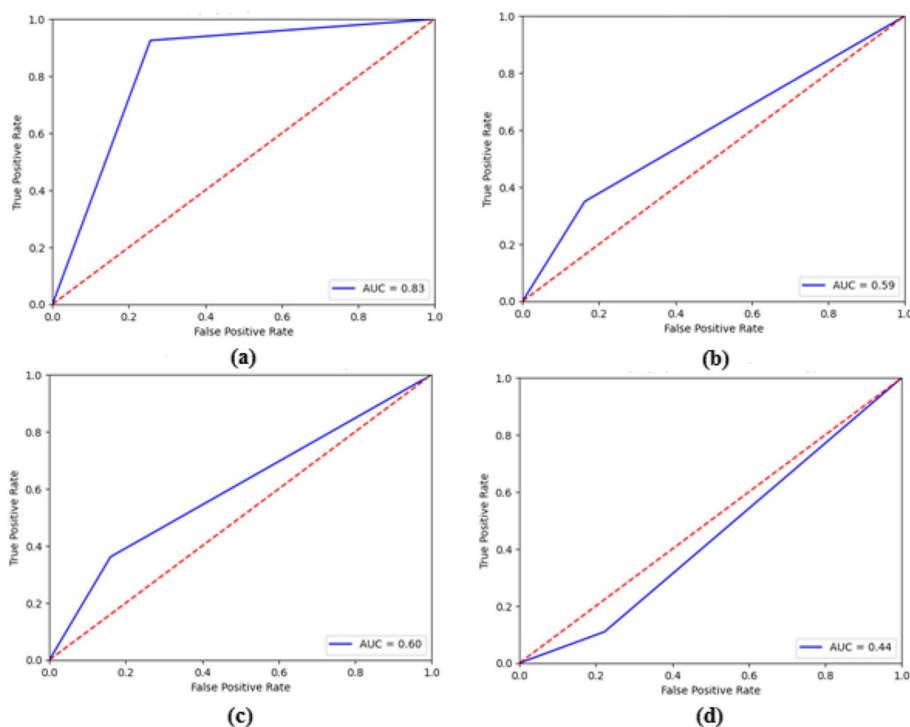
existing CNN, SVM, and LSTM techniques. We have an F -score of 93.82%, which is an improvement over CNN by 3.35%, an improvement over SVM by 4.45%, and an improvement over LSTM by 23.45%. From the results, it is clear that the presented model attained better results compared to the existing algorithms.

Figure 8 represents the receiver operating characteristic curves associated with the proposed and existing techniques. Figure 9 represents the accuracy, loss, and ROC curves associated with the proposed and existing techniques, and the confusion matrix is presented in Fig. 10.

Comparative analysis with published research works

The proposed work is compared with existing work to demonstrate its accuracy, sensitivity, specificity, precision, recall, and F1 score values.

An analysis of the proposed results in comparison with those of the present is shown in Table 4. A method for analyzing fundus images was developed by Smitha and Jidesh [12]. In this study, retinal fundus images were classified into multiple categories using

**Fig. 8** ROC curve for **a** proposed MGSCNN, **b** CNN, **c** SVM, and **d** LSTM

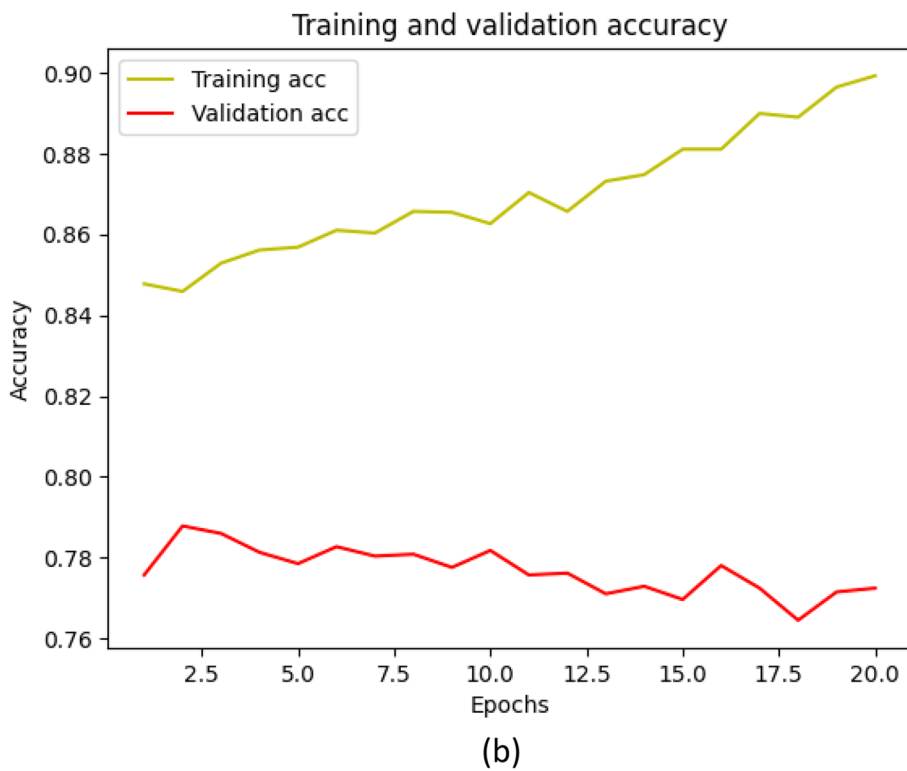
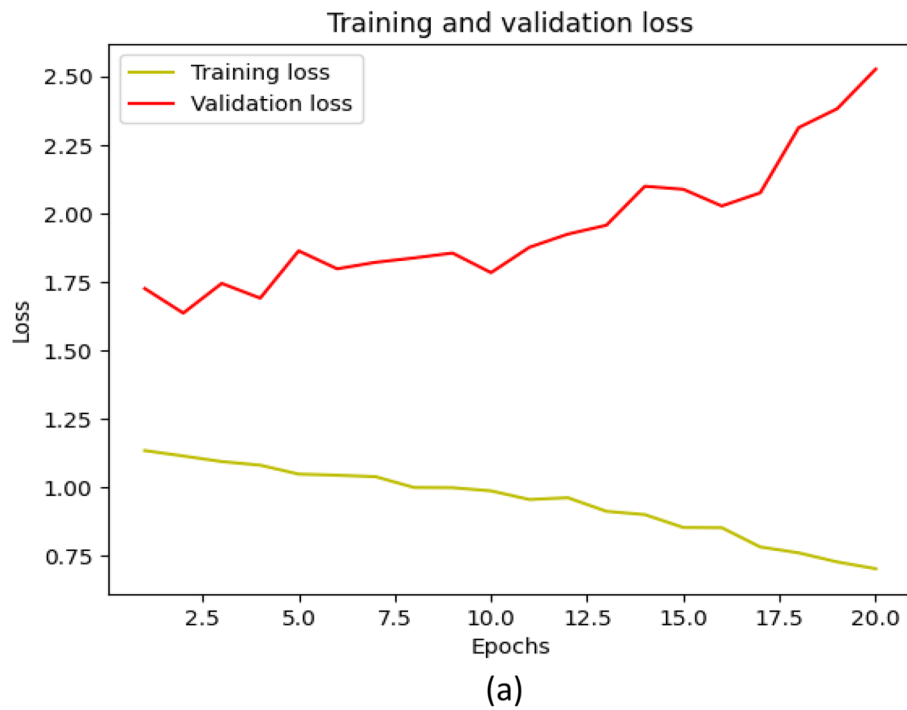


Fig. 9 Experimental results. **a** Training and validation accuracy. **b** Training and validation loss for the proposed work

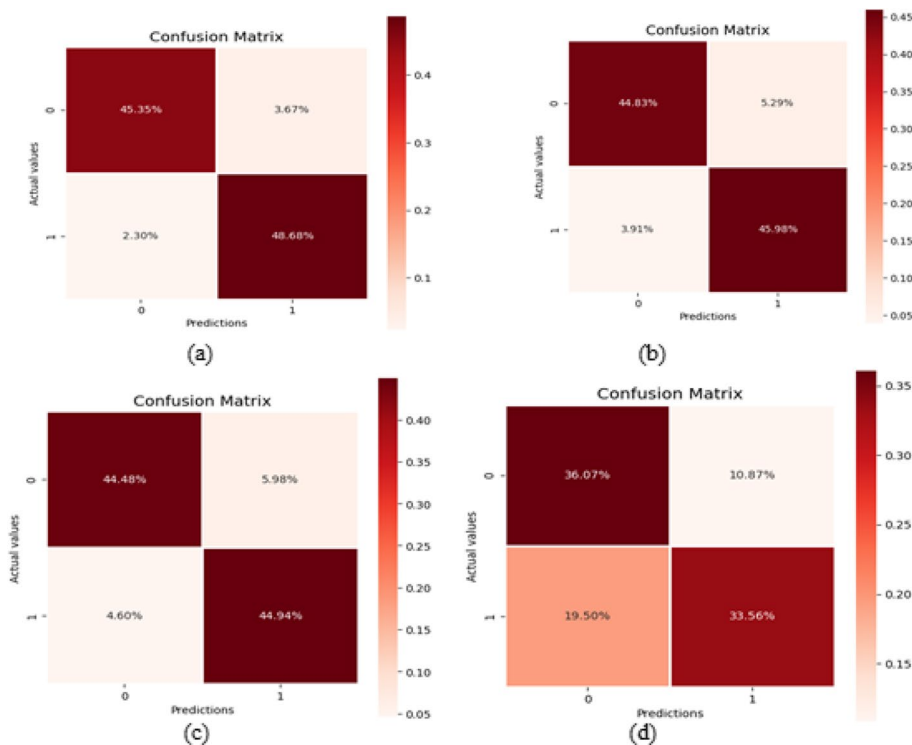


Fig. 10 Confusion matrix. **a** Proposed MGSCNN, **b** CNN, **c** SVM, and **d** LSTM

Table 4 A comparison of the proposed work with existing work is necessary

Ref. No	Accuracy	Sensitivity	Precision	Recall	Specificity	F1-score
[12]	87%	-	-	-	-	85%
[14]	92.3%	-	92.55%	93.36%	-	92.91%
[15]	-	-	-	-	-	86.82%
[20]	-	-	76.85%	71.59%	-	70.42%
Proposed	95.09%	96.59%	93.53%	96.67%	93.59%	95.07%

semi-supervised generative adversarial networks (GANs). Based on these results, the accuracy rate of this study was 87% and the F1 score was 85%. As a result of the proposed work, the accuracy and F1-score were improved by 8.09% and 10.07%, respectively. According to Chen et al. [14], a network with multiple branches corresponds to the attention given to classifying diseases in four different domains. Our proposed work improves precision, accuracy, recall, and F1 scores from this work by 2.79%, 0.98%, 3.31%, and 2.16%. For the diagnosis of diabetic retinopathy and glaucoma, Casado-García et al. [15] proposed analyzing retinal fundus images. Our proposed work improved the F1 score by 8.25% from this work. He et al. [20] developed a multidimensional feature extraction module for fundus images. The precision, recall, and F1 scores of the proposed work are improved by 16.68%, 25.08%, and 24.65% from this work. From the results, it is clear that the presented model attained better results compared to the existing algorithms.

Conclusions

The purpose of this work was to demonstrate how MGSCNN can be used to classify multi-diseases based on retinal images. Initially, the images were pre-processed using normalization, smoothing, and resizing. A preprocessed image is fed into the MGSCNN classifier to determine whether it is normal or abnormal. By tuning the parameters of the CNN, the GSO algorithm optimizes its structure and hyperparameters. To implement this work, Python is used. A performance evaluation of the proposed solution is based on RFMiD data. Our proposed work achieved overall performance in terms of 95.09% accuracy, 96.59% sensitivity, 93.59% specificity, 93.53% precision, 96.67% recall, and 95.07% *F* measure.

Abbreviations

DR	Diabetic retinopathy
GSO	Glowworm Swarm Optimizer
AMD	Age-related macular degeneration
GAN	Generative Adversarial Network
ACO	Approach to continuous optimization
FCL	Fully connected layer
CNN	Convolution neural network
SE	Stimulating
RFMiD	Retinal Fundus Multi-Disease Image Dataset
CRVO	Central retinal vein occlusion
ODC	Optic disc center
BRVO	Branch retinal vein occlusion

Acknowledgements

Not applicable.

Code availability

Code sharing is not applicable to this article because of its proprietary nature.

Authors' contributions

All authors read and approved the final manuscript.

Funding

The authors declare that they have competing interests and funding.

Availability of data and materials

The data used in the three case studies are online data that can be accessed.

Declarations

Ethics approval and consent to participate

As the public dataset is used in this approach for evaluation, "human participants" and/or "human data" are not involved directly in this research work.

Competing interests

The authors declare that they have no competing interests.

Received: 5 August 2023 Accepted: 10 December 2023

Published online: 29 January 2024

References

1. Fu H, Cheng J, Xu Y, Zhang C, Wong DWK, Liu J, Cao X (2018) Disc-aware ensemble network for glaucoma screening from fundus image. *IEEE Trans Med Imaging* 37(11):2493–2501
2. Kim JM, Kim SY, Chin HS, Kim HJ, Kim NR (2019) Relationships between hearing loss and the prevalences of cataract, glaucoma, diabetic retinopathy, and age-related macular degeneration in Korea. *J Clin Med* 8(7):1078
3. Markoulli M, Flanagan J, Tummanapalli SS, Wu J, Willcox M (2018) The impact of diabetes on corneal nerve morphology and ocular surface integrity. *Ocul Surf* 16(1):45–57
4. Yung M, Klufas MA, Sarraf D (2016) Clinical applications of fundus autofluorescence in retinal disease. *Int J Retina Vitreous* 2:1–25

5. Tan JH, Acharya UR, Bhandary SV, Chua KC, Sivaprasad S (2017) Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network. *J Comput Sci* 20:70–79
6. Panchal S, Naik A, Kokare M, Pachade S, Naigaonkar R, Phadnis P, Bhangre A (2023) Retinal Fundus Multi-Disease Image Dataset (RFMiD) 2.0: a dataset of frequently and rarely identified diseases. *Data* 8(2):29
7. Sengupta S, Singh A, Leopold HA, Gulati T, Lakshminarayanan V (2020) Ophthalmic diagnosis using deep learning with fundus images—a critical review. *Artif Intell Med* 102:101758
8. Xu Y, Wang A, Lin X, Xu J, Shan Y, Pan X, Ye J, Shan PF (2021) Global burden and gender disparity of vision loss associated with diabetes retinopathy. *Acta Ophthalmol* 99(4):431–440
9. Guo L, Normando EM, Shah PA, De Groef L, Cordeiro MF (2018) Oculo-visual abnormalities in Parkinson's disease: possible value as biomarkers. *Mov Disord* 33(9):1390–1406
10. Sommer AC, Blumenthal EZ (2020) Telemedicine in ophthalmology in view of the emerging COVID-19 outbreak. *Graefes Arch Clin Exp Ophthalmol* 258:2341–2352
11. Papandrianos NI, Feleki A, Papageorgiou EI, Martini C (2022) Deep learning-based automated diagnosis for coronary artery disease using SPECT-MPI images. *J Clin Med* 11(13):3918
12. Smitha A, Jidesh P (2022) Classification of multiple retinal disorders from enhanced fundus images using semi-supervised GAN. *SN Comput Sci* 3:1–11
13. Bhati A, Gour N, Khanna P, Ojha A (2023) Discriminative kernel convolution network for multi-label ophthalmic disease detection on imbalanced fundus image dataset. *Comp Biol Med* 153:106519
14. Chen F, Ma S, Hao J, Liu M, Gu Y, Yi Q, Zhang J, Zhao Y (2023) Dual-path and multi-scale enhanced attention network for retinal diseases classification using ultra-wide-field images. *IEEE Access* 11
15. Casado-García A, García-Domínguez M, Heras J, Inés A, Royo D, Zapata MA (2021) Prediction of epiretinal membrane from retinal fundus images using deep learning. *Advances in Artificial Intelligence: 19th Conference of the Spanish Association for Artificial Intelligence, CAEPIA 2020/2021, Málaga, Spain, September 22–24, 2021, Proceedings* 19. Springer International Publishing, Spain, pp 3–13
16. Li N, Li T, Hu C, Wang K, Kang H (2021) A benchmark of ocular disease intelligent recognition: one shot for multi-disease detection. *Benchmarking, Measuring, and Optimizing: Third BenchCouncil International Symposium, Bench 2020, Virtual Event, November 15–16, 2020, Revised Selected Papers* 3. Springer International Publishing, pp 177–193
17. Zhang W, Dai Y, Liu M, Chen Y, Zhong J, Yi Z (2021) DeepUWF-plus: automatic fundus identification and diagnosis system based on ultrawide-field fundus imaging. *Appl Intell* 51:7533–7551
18. Rodríguez MA, AlMarzouqi H, Liatsis P (2022) Multi-label retinal disease classification using transformers. *IEEE J Biomed Health Inform* 3
19. Müller D, Soto-Rey I, Kramer F (2022) An analysis on ensemble learning optimized medical image classification with deep convolutional neural networks. *IEEE Access* 10:66467–66480
20. He X, Deng Y, Fang L, Peng Q (2021) Multi-modal retinal image classification with modality-specific attention network. *IEEE Trans Med Imaging* 40(6):1591–1602

Publisher's Note

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