IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 October 2023, accepted 2 November 2023, date of publication 6 November 2023, date of current version 9 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3330436

RESEARCH ARTICLE

Segmentation Using the IC2T Model and Classification of Diabetic Retinopathy Using the Rock Hyrax Swarm-Based Coordination Attention Mechanism

B. N. JAGADESH[1], (Senior Member, IEEE), M. GANESH KARTHIK[2], D. SIRI[3], S. K. KHAJA SHAREEF[4], SRIHARI VARMA MANTENA[5], AND RAMESH VATAMBETI[1]

[1]School of Computer Science and Engineering, VIT-AP University, Vijayawada 522237, India
[2]Department of Computer Science and Engineering, GITAM School of Technology, GITAM University, Bengaluru Campus, Bengaluru 561203, India
[3]Department of CSE, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad 500090, India
[4]Department of Information Technology, MLR Institute of Technology, Hyderabad 500043, India
[5]Department of Computer Science and Engineering, SRKR Engineering College, Bhimavaram 534204, India

Corresponding author: B. N. Jagadesh (nagajagadesh@gmail.com)

ABSTRACT Diabetic Retinopathy (DR) evaluations are increasingly being automated using artificial intelligence. Diabetes-related retinal vascular disease is a major cause of blindness and visual impairment worldwide. Therefore, automated DR detection devices would greatly aid in reducing visual impairment due to DR through early screening and treatment. Researchers have provided many techniques for picking out abnormalities in retinal images during the past several years. In the past, automated methods for diagnosing diabetic retinopathy required a human to extract information from retinal images before passing them on to a classifier. This study takes a novel two-pronged approach to automated DR classification to solve the issues. Due to the low positive instance percentage of existing asymmetric, we segment O.D.s and B.V.s with an enhanced version of an improved contoured convolutional transformer (IC2T). We develop a contoured optical disc (OD), a blood vessels (BV) detection module, and a dual convolutional transformer block that combines local and global contexts to make trustworthy associations. A second-stage Improved Coordination Attention Mechanism (ICAM) network is trained to recognize retinal biomarkers for DR such as microaneurysms (M.A.), haemorrhages (H.M.), and exudates (EX). With an average accuracy of 96%, 97%, and 98% on EyePACS-1, Messidor-2, and DIARETDB0, respectively, the suggested technique has proven itself to be at the field's cutting edge. Comprehensive testing and comparisons to established methods support the proposed strategy.

INDEX TERMS Contoured detection module, diabetic retinopathy, dual convolutional transformer block, improved coordination attention mechanism, improved contoured convolutional transformer, optic disk segmentation.

I. INTRODUCTION

A. BACKGROUND OF DIABETIC RETINOPATHY

In the Western working-age population, diabetes is the most prevalent metabolic illness, and DR is the most common consequence of diabetes. Consistent DR screening has been shown to cut the risk of severe DR-related visual loss by 90% [1]. Retinal specialists routinely use a disease severity scale [2] to provide a numerical value for the degree to which patients are affected by DR in clinical settings. When defining the level of damage caused by DR, the International Clinical Diabetic Retinopathy (ICDR) disease severity scale is the gold standard [3]. From no DR (level 0) to proliferative DR (level 4) on a five-point scale, the

The associate editor coordinating the review of this manuscript and approving it for publication was Joewono Widjaja.

standard proposes non-proliferative DR (level 3) [4], [5], [6]. Microaneurysms (M.A.) are indicative of a mild case of the illness. The presence of more than M.A. or haemorrhages (HEM) is used to define level 2; however, level 3 is indicated by the presence of evident venous beadings or substantial intraretinal microvascular abnormalities (IRMA) [7]. Active or pan-retinal photocoagulation-treated neovascularizations (N.V.) or vitreous haemorrhages suggest a level 4 diagnosis. Level 4 DR can cause permanent visual loss without treatment. Hard exudates (HE) and cotton wool spots (CWS) are also common lesions found in the retina of people with DR [8].

B. ICDR CLASSIFICATION FOR MACULAR EDEMA

The ICDR (International Clinical Diabetic Retinopathy Disease Severity) classification is used to determine the severity of diabetic retinopathy, an eye condition caused by diabetes. While diabetic macular edema (DME) is a particular disorder that can develop with diabetic retinopathy, it is not explicitly classified by the ICDR. The ICDR classification system, on the other hand, is primarily concerned with diabetic retinopathy as a whole.

The ICDR categorization system divides diabetic retinopathy into numerous stages, which are as follows:

No Diabetic Retinopathy (No DR): There are no indications of diabetic retinopathy at this stage.

Mild Non-proliferative Diabetic Retinopathy (Mild NPDR): Microaneurysms are present at this stage. Microaneurysms are microscopic, balloon-like swellings in the retina's tiny blood capillaries.

Moderate Non-proliferative Diabetic Retinopathy (Moderate NPDR): More severe abnormalities in the retinal blood vessels, such as haemorrhages and hard exudates (lipid deposits), become visible at this stage.

Severe Non-proliferative Diabetic Retinopathy (Severe NPDR): This stage is characterised by more substantial obstructions in the retinal blood vessels, which can result in decreased blood supply to regions of the retina.

Proliferative Diabetic Retinopathy (PDR): The most advanced stage of diabetic retinopathy involves the formation of aberrant blood vessels in the retina. These new blood vessels can be unstable, resulting in bleeding into the vitreous gel of the eye and visual loss.

C. ALLENGES IN FUNDUS IMAGES

Images of the retina captured with a fundus camera are used to assess DR. Image capture presents a number of difficulties [9], including light noise and low contrast, both of which negatively impact performance. Variations in size, shape, and colour make DR lesion segmentation a difficult process [10]. Another obstacle in this field is the identification of the optic disc (OD), whose circular form is similar to that of retinal lesions. Because of this, the area is frequently misidentified as a lesion. In order to address these issues, a method is described for identifying and categorizing retinal lesions [11].

D. CHALLENGES IN DL TECHNIQUES

Ophthalmologists have a difficult and error-prone job of manually detecting DR. Thus, an automated approach is necessary for fast and accurate identification. Several algorithmic approaches to DR lesion identification have been presented in the literature [12]. Hough Transform Algorithms (HTAs) and convolutional neural networks (CNNs) are employed for EX identification [13]. In addition, smart edge recognition and histogram equalization are used to boost image quality. The O.D. is an anatomical term. Thus, this also prevents any inference from it [14], [15].

Like classification, automatic medical picture segmentation has benefited substantially from the development of deep learning techniques. CNNs are algorithms that ''learn'' from training data that has been tagged. This requires analyzing more sophisticated visual features before ultimately classifying entire pictures as desired when dealing with imaging data. Recent research [17] shows that the accuracy of these algorithms is on par with or even higher than that of human specialists. CNNs are getting close to human levels of performance in automatic DR detection and grading, but they still can't be easily understood [18]. The need for more transparency in the findings obtained by deep learning algorithms is a potential roadblock to their use in actual clinical settings [19]. Networks that have been taught to identify the outliers employed by human experts in the grading process might be used to solve this issue. Many retinal abnormalities are very tiny or have characteristics that make them difficult to diagnose [20], [21] appropriately. Not only is it difficult to collect enough photos to train algorithms capable of this task, but manually annotating each important pixel in those images is both labour-intensive and time-consuming [22].

E. CONTRIBUTION TO THE SEGMENTATION MODEL

Combining the advantages of the transformer and the U-shaped architecture may fully realise both benefits. Combining these two cutting-edge methods strengthens segmentation by making greater use of both local traits and global contextual information. As a result, the following contributions are highlighted:

- The study created an advanced segmentation method for medical images called IC2T. A dual convolutional (DC) basis makes up the proposed model's U-shaped architecture.
- The suggested approach may combine local and global settings to build trustworthy relationships. Convolutional kernels of various sizes capture multi-scale information for use by the DC transformer blocks.
- The interpretability of the model is improved by combining short- and long-range attention techniques to extract local characteristics and capture long-range interdependence.
- Using conventional CV methods, the contour detection module may zero in on regions of interest and hone in on superfluous contoured segmentation data.

- To further improve segmentation accuracy, a hybrid grasshopper-based optimization approach (G-BAT) is also presented.

F. CONTRIBUTION TO THE SEGMENTATION MODEL

Integral framework of Segmentation and Classifier is shown in Figure 1.

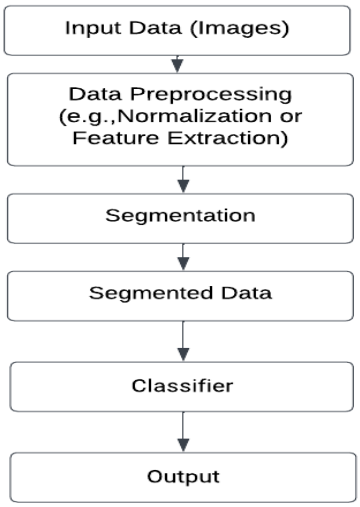

FIGURE 1. Integration of segmentation and classifier.

Here's a quick rundown of each component in the diagram:

Input Data: This is where your raw data enters the framework, such as photos or sensor data.

Data Preprocessing: Various data preprocessing operations including as normalisation, data augmentation, and feature extraction are performed in this step to prepare the data for segmentation and classification.

Segmentation: The segmentation component finds regions of interest within the input data using techniques such as Convolutional Neural Networks (CNNs) or Region Proposal Networks (RPNs).

Segmented Data: This section contains the segmentation step's output, which includes the regions or items of interest detected in the input data.

Classifier: The classifier employs machine learning or deep learning models to classify and predict the segmented data as input.

Output: The framework's final output, which may include flood or earthquake detection predictions, judgements, or alarms.

In the first stage, innovative Gaussian space-scale generic augmentation settings were used to pre-process and augment the data.

The research recommends an enhanced coordination attention mechanism network (ICAMNet) that takes into account the spatial location relationship.

A convolution-coordinated attention mechanism is one of the three primary building blocks of a CAMNet. In the first place, spatial and spectral characteristics are completely extracted using CNN in the convolution module. Second, the linear module intends to provide a feature map with more data. The developed CAM also takes into account both axes of spatial information.

The Rock Hyrax Swarm Optimization (RHSO) model handles the hyper-parameter tuning procedure.

The suggested approach has been trained using 11841 retinal fundus pictures from three publicly available datasets.

G. PAPER ORGANIZATION

In the first section, we discuss the history of DR, its difficulties, and the role that research has played in addressing these issues. In Section II, we provide a problem definition and a summary of related research on DR segmentation and classification. Sections III and IV detail the resources and the short approach used. In Section V, we see the experimental design with the validation analysis. Section VI wraps up the findings and discusses what comes next.

II. RELATED WORKS

A reformed capsule network is built for the diagnosis and categorization of diabetic retinopathy by Kalyani et al. [23]. Fundus pictures have their characteristics extracted using a convolution and main capsule layer, and then the class capsule layer and softmax layer are used to assess the likelihood that the image belongs to a certain class. Using the dataset, we verify the efficacy of the proposed reformed network across four performance metrics. On healthy retina, stage 1, stage 2, and stage 3 fundus pictures, the accuracy of the created capsule network is 97.98%, 97.65%, 97.65%, and 98.64%, respectively.

Zbay [24] provided a novel multi-layer architecture for automated DR stage detection. In the ADL system's preprocessing step, a threshold value is chosen based on the results of the image histogram to help find retinal lesions. The picture is then segmented using this method. To further automate the process of extracting segmented retinal characteristics, a tag-efficient architecture called ADL-CNN has been designed. This scheme operates in two phases. In the first, pictures are picked to learn either straightforward or advanced retinal features based on the accuracy labels included in the training set. Second, the important characteristics of the lesions and the segments of interest within the retinal picture are presented as masks. The suggested ADL-CNN model's performance is measured against that of the state-of-the-art approaches on the same dataset. Statistics, including the F-measure, sensitivity, specificity, and classification accuracy, are used to evaluate the system's performance. Using the EyePacs dataset, which consists of 35,122 retinal pictures, the ADL-CNN model was able to achieve 99.66% accuracy, 93.76% S.E., 96.71% S.P., and 94.58% F-measure. In this regard, it is fair to say that the suggested technique performs admirably when applied to a wide variety of fundus pictures for the purpose of identifying DR lesions and grading their severity.

Uppamma and Bhattacharya [25] have zeroed in on information vital to the forecasting of diabetic retinal illness.

Complete blindness is the end result of diabetic retinopathy, a chronic illness induced by diabetes. Early detection of the condition is essential for preventing eyesight loss. The dataset utilized is an open-source one obtained through the IEEE data port. Before performing lesion segmentation, we pre-processed the data with the median filtering approach. The most important characteristics were supplied to the SqueezeNet classifier, which predicted, and the data was then exposed to the Taylor African Vulture Optimization (AVO) method for hyper-parameter tuning. The EHR administrator then had access to the final outcome stored in the blockchain architecture, guaranteeing that only authorized parties could view the prediction results and any associated patient data. By comparing the classifier's performance to that of previous studies, we find that the suggested model outperforms the state-of-the-art models in terms of accuracy.

Ohri and Kumar [26] have applied transfer learning to the downstream job of diabetic retinopathy (DR) severity identification by fine-tuning the network when different subsets of medical data are available. The experimental results show that the efficiency of the medical image classifier, when trained on complete training data, is greatly improved by supervised pre-training on ImageNet, followed by fine-tuning pictures. Less clear, however, is how different data subsets impact fine-tuning performance and whether or not this learning is label efficient. Therefore, we examine the model's efficacy when trained on 20%, 40%, 60%, and 80% of the total labelled data for the DR classification task, with findings indicating that supervised fine-tuning performs poorly when the model is trained in the low-data regime. When trained on all available data, the suggested model has a high level of performance (test accuracy of 0.8010, AUC of 0.86, F1 score of 0.6477, and Cohen's kappa score of 0.7007) but suffers in the low data regime. This point to the limitations of supervised learning when only a small amount of labelled data is available for model training. Our approach therefore provides a springboard for future investigation into improving performance in low-data settings.

Parthiban and Kamarasan [27] have presented an intelligent coyote optimisation algorithm with a deep learning-based DR detection and grading (ICOA-DLDRD) model using retinal fundus pictures. The ICOA-DLDRD method is meant to detect DR in fundus photographs of the eye. At its core, the ICOA-DLDRD algorithm uses a noise-reduction method based on Gabor filters (G.F.) and a method for finding the best way to divide up a large image into smaller ones. In addition, the glowworm swarm optimization (GSO) algorithm is used to generate appropriate main seed locations and thresholds for the region-expanding segmentation method. For feature extraction, we also create a SqueezeNet with a class attention learning (CAL) layer. Lastly, we use COA with a deep extreme learning machine (DELM) classifier to find and grade DR. We use COA to optimise the DELM model's penalty parameter C and kernel parameter gamma. We tested how well the ICOA-DLDRD method worked using the MESSIDOR dataset as a reference. The results showed that our method was more accurate than other state-of-the-art methods, with a maximum accuracy of 99.65%.

Bansode et al. [28] hope to create a new method of DR detection that relies on deep learning. In this case, we utilize the incorporation of 'Optimized Iterative Thresholding (O-IT)' for precise segmentation of blood vessels. The first innovative aspect of this study is the development of a hybrid meta-heuristic Shark Smell-Jaya Optimisation (SS-JO) method for optimizing the thresholding strategy to improve blood vessel segmentation and classification. A deep learning framework known as optimized long short-term memory (LSTM) is used in place of CNN. The second originality of this study is the suggested model's use of SS-JO to optimize LSTM parameters, reducing the network's complexity and making it more suitable for use in practical applications.

Fundus images can help identify the phases of DR in patients, and Naik et al. [29] have developed a distinctive suggested model. It can analyze a fundus picture for signs of retinopathy and make an educated guess as to which stage of the illness it represents. Nonproliferative are the two main categories under which DR can be placed. The suggested model is capable of differentiating between no DR, mild DR, moderate DR, severe DR, and PDR thanks to its training. Researchers and clinicians alike will benefit from rapid DR patient identification. The current state of affairs makes it highly time-consuming for physicians to manually analyze each individual nerve cell from fundus images. Therefore, in this study, we present a convolutional neural network-based model with data augmentation for DR classification from fundus pictures. Models like DenseNet121, DenseNet169, ResNet50, and InceptionV3 may all be trained on powerful GPUs with the help of this enriched dataset. These models attained accuracy levels of 96.64, 95.95, 95.71, and 94.73%. DenseNet121 has the highest reported accuracy, at 96.64 percent, when compared to other SOTA models.

Qaid et al. [30] presented the automated diagnosis system's accuracy in detecting DR and its severity, and they analyzed the results. Segmentation of fundus images using fuzzy entropy multi-level thresholding is the primary topic of this article. The research set out to find the best possible settings for detecting DRs and their severity. To begin, we built a model of the retina as an image, one that can be used across a wide range of retinal and image properties. The retinal model was used to create 45,000 photos in total. Second, we quantified the effectiveness of DR detection and the severity levels by conducting a feasibility and consistency study using a tailored Monte Carlo statistical approach. Guaranteed DR detection is achieved under the following conditions: Finally, the reliability of these circumstances was assessed by contrasting synthetic retinal pictures with those available to the public. Test findings showed that DR detection success might be predicted based on the analysis's conditions.

In this study, Nahiduzzaman et al. [31] propose an innovative automated method for DR detection. Contrast Limited Adaptive was used as a preprocessing step on the fundus images (F.I.s) to bring out the lesions. Features were extracted using a parallel convolutional neural network (PCNN), and DR was classified using an extreme learning machine (ELM) approach. The PCNN architecture employs fewer parameters and layers than the comparable CNN architecture, which reduces the time needed to extract unique features. Both the Kaggle DR 2015 competition (Dataset 1; 34,984 FIs) and the APTOS 2019 datasets (3,662 FIs) were used to assess the technique's efficacy, and the findings are encouraging. The suggested method achieved 91.78 and 97.27 percent accuracy on the two datasets, respectively. One of the study's corollary findings, however, was that the suggested framework was robust across a variety of dataset sizes and shapes, including both symmetrical and asymmetrical ones. In addition, the proposed method outperformed state-of-the-art models in terms of classifier performance metrics, model parameters and layers, and prediction time, which would be of great assistance to medical professionals in correctly detecting the DR.

In [32], Alwakid et al. offer a deep learning (DL) model that accurately detects all five phases of DR. A contrast-limited adaptive histogram is used to improve images in Case 1, whereas in Case 2, the images are not improved. The dataset was then augmented using the identical parameters used in both situations to ensure equality. The created model outperformed prior techniques for detecting the five phases of DR using Inception-V3-applied datasets, with an accuracy of 98.7% for case 1 and 80.87% for case 2. It was shown that including CLAHE and ESRGAN in a model enhances both its performance and its capacity for learning.

Using convolutional neural networks, Mercaldo et al. [33] present a method for automatically detecting the presence of diabetic retinopathy in ocular angiography. Specifically, two models are proposed: the first is meant to differentiate between normal and diseased eyes, while the second is meant to differentiate between nonproliferative and mild to moderate proliferative retinopathy. Our results show that the suggested models can be useful tools for clinicians, with an accuracy of 0.98 for the first model and 0.91 for the second. Also, the proposed method tries to find the disease in the angiography by using two different class activation mapping algorithms that show on the images the areas where the disease is showing itself. This is done to give doctors and patients more confidence in the model's diagnosis by giving them some way to explain it. We also offer a similarity score to measure the dissimilarity between heatmaps produced by the same model's class activation mapping algorithms in order to assess the stability of the model.

To better identify DR events and evaluate their developing phases, a unique Gannet approach is presented by Krishnamoorthy et al. [34]. There are six main stages of DR that may be detected and categorized using the GO-DBN-WKELM method: normal, mild DR, moderate DR, severe DR, and proliferative DR. First, the Deep Belief Network (DBN) model reduces the feature dimensions of the original datasets to extract the most pertinent information. The recovered pictures are then put through the suggested GO-DBN-WKELM classification model, which is able to reliably detect and categorize fundus images according to severity. Incorporating a wavelet into the G.O. algorithm is what contributes most to the improved detection performance. The convergence speed of the classifier is improved using the G.O. method, and the kernel parameters of the WKELM are optimized as well. Three datasets, MESSIDOR, DIARETDB1, and IDRiD, are used to evaluate the proposed classifier. Different performance measures, including accuracy, precision, recall, and F-measure, are used to evaluate the suggested classification model's efficacy in detecting DR The suggested GO-DBN-WKELM classifier performed very well in simulations, with an accuracy of around 98% on the MESSIDOR dataset and 97.8% on the DIARETDB1 dataset. Results like this highlight providing an effective, automated alternative to manual methods that can help eye specialists diagnose and treat patients in a timely fashion.

Lin and Jiang [35] have proposed a preprocessing method to improve the picture's characteristics. Based on the study's findings, preprocessing is a viable option for making more information accessible to the training model. As a result, this research enhanced the EfficientNet model to better classify data at the DR level. The outcomes also showed an improvement in model accuracy from 0.7727 to 0.7920 with regard to the categorization of DR phases. The improved EfficientNet also outperformed MobileNet (0.54) and the original EfficientNet (0.922) in terms of the average area under the ROC curve across all five classes, with a value of 0.926. Finally, an application programming interface (API) was used in this work to construct the suggested system, allowing users to input a fundus picture and receive the DR findings.

To categorize the DR picture, Venkaiahppalaswamy et al. [36] utilize a powerful hybrid binocular Siamese with a deep learning technique. To begin, the stage is implemented to filter out background noise. The use of a cross-guided bilateral filter (CGBF) is suggested for this purpose. Once an image has been preprocessed, the feature extraction step may be used to pull out certain characteristics from it. There is a new feature extraction approach called the wavelet-based Chimp optimization algorithm (WBCOA). The optical disc (O.D.) and blood vessels (B.V.) are then segmented using open-closed watershed management (OCWSM) after feature extraction. In this study, we suggest using AlexNet and GoogleNet with the SVM model, both of which are based on the binocular Siamese network architecture. The proposed hybrid DL network takes as input the O.D. and B.V. that have been segmented. At last, the extracted pictures are fused, and the SVM model is utilized to categorize them. The DIARETDB0 (DB0) and DIARETDB1 (DB1)

datasets are used to evaluate the suggested approach, which is implemented in Python. The accuracy of the projected hybrid DL network was 94% on DB0 and 94.83% on DB1. The results of the suggested model are compared to those of other methods as well. The suggested approach used mean and standard deviation (S.D.)-based statistical analysis on the DR picture to get useful results.

To identify and categorize retinopathy illness from retinal pictures, Maithili et al. [37] have presented a novel hybrid strawberry-based convolution neural framework (SbCNF). The retinal veins are cut using a variety of data sources. In this case, DRIVE datasets serve as the basis for the entire operation. This study is implemented in the Python programming language. In addition, this research shows how retinopathy detection software might be enhanced in the future. Traditional classification model techniques were used to verify the implementation results for things, etc. The results show that the devised method outperformed the competition in retinopathy recognition accuracy by virtue of its useful benefits, such as reduced computational complexity.

A. PROBLEM STATEMENT

There have been numerous great outcomes from research using deep CNN architectures for DR diagnosis. However, there are also several limits and restrictions. The following is a description of them.

- Instead of recognizing the location of DR lesions in fundoscopic pictures, current CNN models exclusively focus on grading DR from beginning to conclusion. Here, the pictures are sent straight to CNN, and the DR severity is determined by the images' outputs. Nevertheless, ophthalmologists place a premium on knowing specifics about the clinical presentation of DR lesions.
- In particular, a deep convolutional neural network (CNN) model requires large and well-annotated data collection. Getting it and making it usable are both very time-consuming and expensive processes.
- For some current CNN models, the complex lesion architecture that DR causes can be challenging to train. Many minor lesions, such as M.A. and HEM, were difficult to find in CNN's limited patch of fundus pictures because of their vague forms. Therefore, it is crucial for DR detection algorithms to learn lesions with fine details.

III. MATERIALS

Two of the most commonly utilized public datasets for evaluating the performance of the proposed system were Messidor-2 and DIARETDB0.

A. EYEPACS-1

About 8980 retinal pictures were included in the EyePACS-1 dataset [38]. There are 7552 healthy subjects, 842 mild cases, 545 moderate cases, 54 severe cases, and 95 PDR cases in the EyePACS-1 dataset.

B. MESSIDOR-2

There are 1748 photos of the back of the eye (the retina) in the Messidor-2 collection [39]. There are only 1017 "normal" photos and 270 "mild" and "PDR" images in the collection. The Topcon digital F.I. camera used to take the digital F.I.s for the Messidor-2 dataset has a 45-degree field of view.

C. DIARETDB0

For research on DR detection and classification, you can use the DIARETDB0 dataset [40]. There are 130 fundus photos in the DIARETDB0 dataset; 110 are categorized as DR and 20 are categorized as normal F.I.s. A digital F.I. camera with an unspecified field of view (about 50 degrees) was used to capture these images. The information is applicable to real-world situations and may be used to evaluate the efficacy of diagnostic methods.

Table 1 shows the distributions of DR severity grades among the datasets Messidor-2, EyePACS-1, and DIARETDB0. Figure 2 displays some representative photos taken from the available data sets.

TABLE 1. Dataset delivery of DIARETDB0.

Severity Grade	EyePACS-1	Messidor-2	DIARETDB0
Normal	7552	1017	20
Mild	842	270	50
Moderate	545	347	35
Severe	54	75	15
PDR	95	35	10

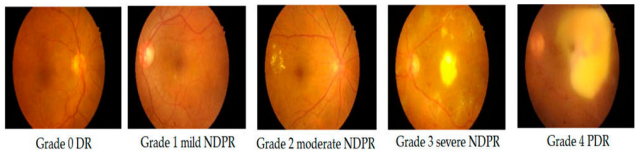

FIGURE 2. Sample images from the datasets.

IV. METHODOLOGY

In this study, we introduce a unique approach to DR detection based on retinal fundus pictures. DIARETDB0 are the three open-source databases employed here. Preprocessing methods include picture scaling, GCE (Green Channel Extraction), and top-bottom hat transformation to improve the quality of the F.I.s. Hybrid optimization is then used to fine-tune the model's parameters, and a convolutional transform network is used to segment the OD. Finally, a better picture dataset is utilized in conjunction with the refined model of the attention process to diagnose DR. A number of performance criteria, including sensitivity, accuracy, precision, F1-score, specificity, and area under the curve, are employed to assess the viability of the proposed method.

A. PREPROCESSING AND DATA AUGMENTATION

1) PREPROCESSING

The suggested method was evaluated using Messidor-2, DIARETDB0, and EyePACS-1. In all, 10,966 retinal fundus

images (from the DIARETDB0-130 databases) are taken into account here. The datasets' locations are shown in Table 1 for EyePACS-1, DIARETDB0, and Messidor-2. The size of the F.I. may have an effect on how well a deep learning model performs. To address this issue, we uniformly resized all of the photos to 256 pixels on each side. The loss of the optic disc and other major blood vessels makes direct resizing of F.I.s challenging. Bicubic interpolation was used to resize the retinal F.I.s while preserving their aspect ratio. Figure 3 shows that, compared to the red and blue channels, the green channel in F.I.s conveys more information, making it a good fit for our study. Retinal images benefit from a top-to-bottom hat modification. Figure 4 shows how various preparatory measures are taken.

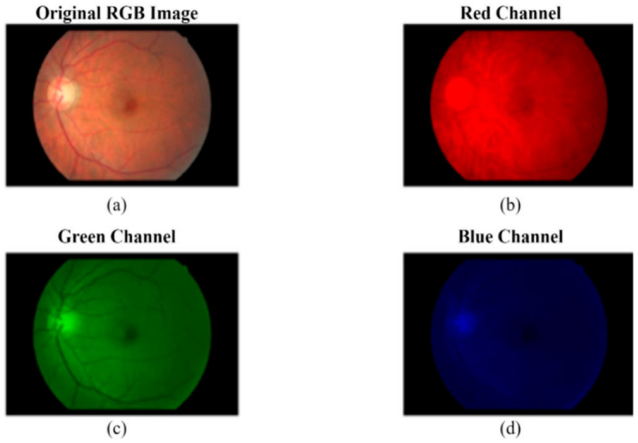

FIGURE 3. FI RGB channel. (a) OI (Unique Image) (b) RCH; (c) GCH, (d) BCH.

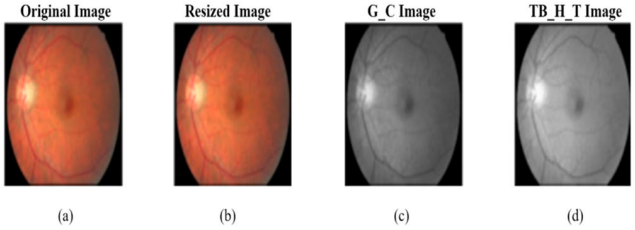

FIGURE 4. Preprocessing stages. (a) Unique Image, (b) Resized Image, (c) G_C Image, (d) TB_H_T Image.

2) DATA AUGMENTATION

One of the most important considerations for efficient DL model processing is the quantity of the training dataset. Thus, a large dataset is necessary for training deep learning networks to prevent overfitting and generalization issues. The dataset has a highly skewed distribution across classes, with the vast majority of pictures being in category 0 (normal). There is a risk of misclassification because of the high imbalance in this dataset. To clean up the fundus pictures and increase the size of the retinal dataset, we employed data augmentation methods. Here is a rundown of the main data augmentation procedures we ran.

- Images were randomly rotated between 0 and 360 degrees.
- Shearing: Sheared at an arbitrary angle between twenty and two hundred degrees.
- Vertically and horizontally flipped images were used.
- Images were randomly zoomed by a factor of between (1/1.3 and 1.3).
- Images were cropped at random to be 85–95% of their unique size.
- Randomly shifting images between the range of −25 and +25 pixels served as translation.

Several instances of post-augmentation images are exposed in Figure 5.

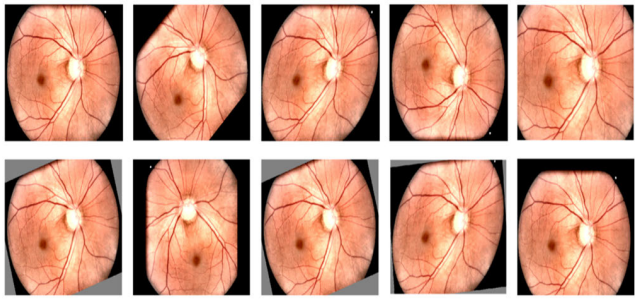

FIGURE 5. Using pre-processed images, multiple augmentation procedures were functional to augment the retinal dataset.

B. OPTIC DISC (O.D.) AND BLOOD VESSEL (B.V.) SEGMENTATION

If IC2T is a specific model or framework for image segmentation, I recommend referring to the official documentation or research papers associated with it for detailed information on how to use it for segmentation tasks. Additionally, you may want to check for any code repositories or GitHub projects related to IC2T for implementation examples and guidance.

For image segmentation in general, there are several well-known models and frameworks, such as U-Net, Mask R-CNN, FCN (Fully Convolutional Network), and SegNet, which are commonly used for tasks like semantic segmentation and instance segmentation.

Three parts of the suggested approach are elaborated upon in this research. The basic framework is shown first. Second, the component's deeper features are handled with better precision because of the dual convolutional transformer's architecture. Thirdly, our proposed model includes a contour detection module, which helps the network better extract edge features that aren't immediately apparent.

1) OVERALL ARCHITECTURE

a: ATTENTION MECHANISM

The "attention mechanism" may be employed to focus on important regions or features within the segmented retinal images when making a classification decision. This can help improve the accuracy of diabetic retinopathy classification.

In practice, the segmentation component (IC2T Model) would be used to preprocess retinal images, identifying and

isolating the regions of interest. The segmented regions or features would then be fed into the classification component (Rock Hyrax Swarm-Based Coordination Attention Mechanism) for diagnosing diabetic retinopathy based on the coordinated analysis of these features.

To gain a detailed understanding of this specific approach, including the architecture, training process, and performance metrics, you would need to refer to the original research paper or documentation associated with the model. Additionally, if code or pre-trained models are available, you can use them to implement and experiment with the system for your specific application.

b: ICDR

The abbreviation ''ICDR'' is not a widely recognized standard or classification system in the context of diabetic retinopathy diagnosis. Therefore, it's essential to clarify what ''ICDR'' stands for and how it relates to your research. It's possible that ''ICDR'' could be a novel or specific classification system or dataset created for your research, and you should provide a detailed explanation of it in your manuscript.

c: ETDRS

The term ''ETDRS'' stands for ''Early Treatment Diabetic Retinopathy Study.'' ETDRS is a widely recognized and established grading system and classification for diabetic retinopathy used in clinical evaluation, research, and clinical trials. It is often considered a gold standard for assessing diabetic retinopathy severity. If ETDRS or any other established grading system is relevant to your work, you should discuss its role and relevance in your research, especially if you are comparing or benchmarking your results against it. Regarding the phrase ''proposes non-proliferative DR (level 3),'' it's necessary to provide context and clarification. In the context of diabetic retinopathy, ''DR'' typically stands for ''Diabetic Retinopathy,'' and ''non-proliferative'' refers to the stage of the disease that precedes the more severe proliferative diabetic retinopathy (PDR). The term ''level 3'' likely refers to a specific severity level within the non-proliferative diabetic retinopathy stage, but without additional context or explanation, it may be unclear to readers.

The innovative model's architecture is based on Swin-Unet [41], and it takes the form of a U made up of connections. The encoder and decoder parts of the U-shaped architecture extract and reconstruct information, respectively, so that both local and global characteristics may be captured. The skip connections, which allow for the integration of low-level and high-level information, improve the model's accuracy and robustness. Multiple blocks and modules handle the processing of sequences of varying resolutions. These manipulations provide the model with a comprehensive means of extracting characteristics and capturing prospective data. In sum, this architecture improves the model's ability to extract and apply characteristics from the input data.

In the encoder stage, the input picture is written down as InputR (H,W,C), where H and W stand for the input image's height and width, respectively, and C stands for the input image's channel count. Based on this research, C is set at 3. The standard approach to patch embedding [42] includes segmenting an input picture into 4 × 4 pieces that do not overlap. Single-size convolution kernel models, on the other hand, are easily overfit and fail to capture all of an image's information because of their limited generalization capacity. Multi-scale sampling is a method for fixing these issues by first sampling the image at several scales and then piecing the results back together. With this method, you can save nuanced details while still getting the job done. Convolutional operations with four distinct kernel sizes (4 4, 8 8, 16 16, and 32 32) are used in the multiple-scale sampling procedure. The patch-embedding method results in a patch that is H/4 W/4 C in size.

In Section IV-B2, we describe the gated module's batch convolution, ReLU, and sigmoid layers in detail and present formulaic expressions for the proposed dual convolutional (D.C.) transformer structure. It does this by merging feature maps from the present and higher layers, filtering out noise in the process to get richer feature data. In the gated module, edge information from contour detection is combined with the encoder's one-dimensional features to generate two-dimensional features. The gated module receives the fusion information and combines it with additional modules, both of which provide intermediate data. The gated module's output will be integrated into the decoder's middle stages. The computation is written as a suggested block with two layers.

$$\hat{X}_i = Reshape(X_i) \tag{1}$$

$$C_i = Conv(\hat{X}_i) \tag{2}$$

$$token = Gated\left(C_i, contour_{token_i} + \hat{X}_i\right) \tag{3}$$

$$output = DC(token) \tag{4}$$

where X_i represents derived from the contour detection module and gated stands in for the *gate module*($i = 1, 2, 3$).

Up-sampling is accomplished with the help of a linear extension at the decoder step. All the collected characteristics are then used to produce the final results, which require mapping sequences linearly to a high-dimensional space.

2) DUAL CONVOLUTIONAL (D.C.) TRANSFORMER BLOCK

Figure 6 depicts the layout of the D.C. transformer chunk.

The Swin Transformer block [43] served as inspiration for the dual convolutional (D.C.) transformer block, which aims to capture features at various scales while also expanding the receptive field of the suggested method. To improve the convolutional operation of the attention mechanism for medical picture segmentation, a revised parallel convolutional attention mechanism was made to work with self-attention. With this method, data may be gathered in both the channel and spatial dimensions at the same time. Using the CBAM [44] method, we first produce features throughout the spatial dimension as input. $Z^{l-1} \in R^{\frac{H}{4} \times \frac{W}{4} \times C}$. The data is then sent across an interconnected system. Following careful

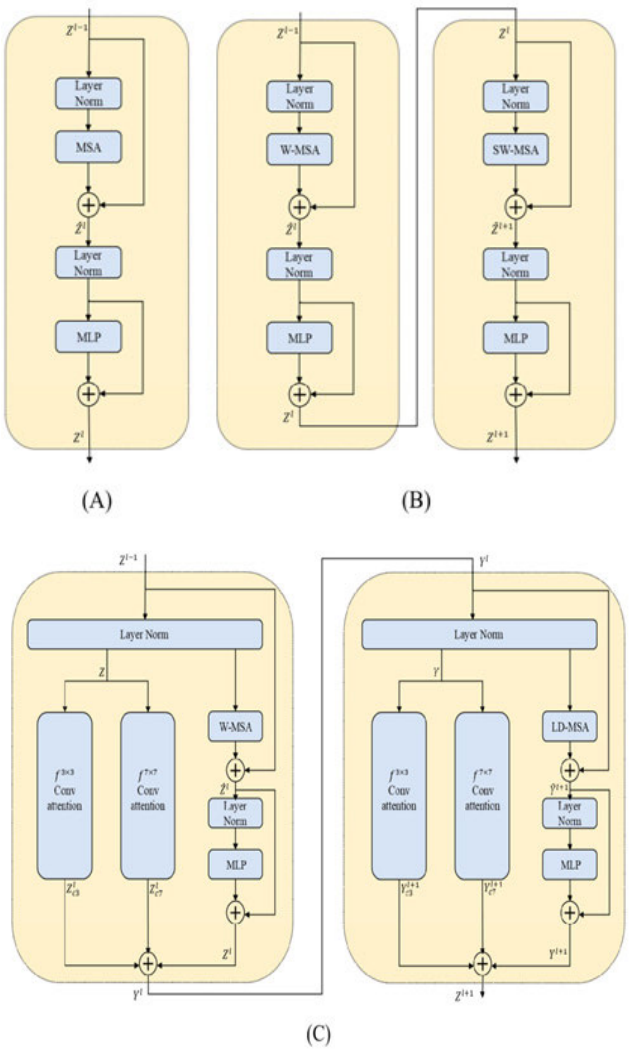

FIGURE 6. D.C. Transformer Block Architecture A) Basic transformer; B) Swin Transformer; C) Projected transformer.

consideration of the spatial and channel dimensions, the following is the result.

$$Z = Z^{l-1} \otimes \sigma\left(MLP\left(Z_{max}^{l-1}\right) + MLP\left(Z_{avg}^{l-1}\right)\right) \quad (5)$$

$$\Upsilon = \Upsilon^{l} \otimes \sigma\left(MLP\left(\Upsilon_{max}^{l}\right) MLP\left(\Upsilon_{avg}^{l}\right)\right) \quad (6)$$

The notation max signifies the extreme pooling operation, avg the average pooling process, and s the sigmoid function. When it comes to processing local information like texture and details, smaller convolutional kernels (f(33)) are preferable, whereas bigger convolutional kernels (f(77)) are preferable when dealing with global features like shape and contour. Convolutional kernels of varying sizes allow for improved capture of multi-scale information. The following are the formulas needed to calculate the convolution:

$$Z_{c3}^{l} = Z \otimes \sigma\left(f^{3\times3}\left(\left[Z_{max}; Z_{avg}\right]\right)\right) \quad (7)$$

$$Z_{c7}^{l} = Z \otimes \sigma\left(f^{7\times7}\left(\left[Z_{max}; Z_{avg}\right]\right)\right) \quad (8)$$

$$\Upsilon_{c3}^{l+1} = \Upsilon \otimes \sigma\left(f^{3\times3}\left(\left[\Upsilon_{max}; \Upsilon_{avg}\right]\right)\right) \quad (9)$$

$$\Upsilon_{c7}^{l+1} = \Upsilon \otimes \sigma\left(f^{7\times7}\left(\left[\Upsilon_{max}; \Upsilon_{avg}\right]\right)\right) \quad (10)$$

We also make extensive use of two mechanisms for multi-head attention: Relevant linkages between local regions can be developed using the W-MSA technique. For feature modelling, it partitions the input data into M M windows. Tokens are used to efficiently collect hidden details across several areas and include fine-grained internal characteristics. I stand in for the sample interval in the LD-MSA process. Unsampled sections of the picture were subjected to masking resulting in $\frac{H}{I} \times \frac{H}{I}$ groups. Feature within each group to obtain $\frac{H}{I} \times \frac{H}{I}$ feature maps. On the other hand, the W-MSA technique establishes interdependence by sampling nearby picture blocks. The final feature representation is derived from a combination of the two systems' outputs. Here's how you can figure out your W-MSA and LD-MSA:

$$head = Attention\,(Q, K, V) = softmax\left(\frac{QK^{T}}{\sqrt{d}} + B\right) V \quad (11)$$

$$W - MSAorLD - MSA\,(Q, K, V) = Concat\,(head_1, \ldots, head_n)\, W^{O} \quad (12)$$

where W^{O} represents the matrix, Q characterizes the query, K characterizes the key, V represents the value, and B involves W-MSA and LD-MSA.

The standard $Attention(Q, K, V)$ uses a matrix of queries, keys, and values to perform operations. However, there is no denying the self-attention mechanism's flaws when it comes to modelling immediate dependence. We use the W-MSA to consolidate the affiliations as a solution to this problem. The W-MSA may be determined using the following formula when using the window-partitioning method:

$$\hat{Z}^{l} = W - MSA\left(LN\left(Z^{l-1}\right)\right) + Z^{l-1} \quad (13)$$

$$Z^{l} = MLP\left(LN\left(\hat{Z}^{l}\right)\right) + \hat{Z}^{l} \quad (14)$$

where $\hat{Z}^{l}$ is for the lth layer's multilayer perceptron (MLP) output, while Zl stands for the lth layer's weighted mean squared error (W-MSA) output. The LD-MSA technique was developed with the transformer's cross-scale attention mechanism [45] in mind, allowing it to effectively capture the associated properties among various tokens. In order to facilitate self-attention within the acquired groups and enhance the interaction among the info elements in each window, the LD-MSA maps along a certain length and breadth. Here's how we may characterize the LD-MSA:

$$\hat{\Upsilon}^{l+1} = LD - MSA\left(LN\left(\Upsilon^{l}\right)\right) + \Upsilon^{l} \quad (15)$$

$$\Upsilon^{l+1} = MLP\left(LN\left(\hat{\Upsilon}^{l+1}\right)\right) + \hat{\Upsilon}^{l+1} \quad (16)$$

where $\hat{\Upsilon}^{l+1}$ denotes the output consequences the $(l+1)$-th layer, while Υ^{l+1} denotes the output MLP of the $(l+1)$-th layer. Lastly, Υ^{l} characterizes the D.C. transformer block,

while Z^{l+1} indicates the output of the block.

$$\Upsilon^{l} = Z^{l}_{c3} + Z^{l}_{c7} + Z^{l} \tag{17}$$

$$Z^{l+1} = \Upsilon^{l+1}_{c3} + \Upsilon^{l+1}_{c7} + \Upsilon^{l+1} \tag{18}$$

3) CONTOUR DETECTION MODULE

An integral part of our suggested design is the contour discovery component. The module's goal is to maximize the model's interpretability and boost the efficiency of medical picture segmentation through the use of contour information included in the images. The model is given more reference information to improve segmentation accuracy, and the contour detection module is built to keep track of all contours (both internal and exterior). Our method is able to successfully acquire and utilize more detailed contour information than conventional segmentation methods that simply focus on outward contours. This improves the model's ability to comprehend medical picture limits and structure, which in turn yields more precise segmentation outcomes. The ''find Contours'' function, a variation of the Suzuki-Beck technique, is used in the OpenCV module of Python to recognize the items in a medical picture.

First, the image is processed to get rid of the testing couch or whatever else could be in the way. If there is no such disruption, you can skip this procedure. After that, we identify and extract the contoured features. In order to successfully extract contour information from pictures while keeping local features intact, CV methods are needed to complete the detection process, otherwise known as contour detection. Tasks requiring processing benefit greatly from both the contoured and local features. Downsampling is then used in the contour detection module to provide consistent sizing across all channels.

4) HYPER-PARAMETER TUNING PROCESS

The vast majority of researchers recommended using optimisation techniques to determine the best scaling factor. In particular, in recent years, the NIO algorithms have been proposed, such as the evolutionary algorithms ABC, grasshopper, bat, and firefly. However, it is also important for NIO to strike a balance between exploratory and exploitative search behaviours. Our suggested method uses a modified version of the grasshopper-BAT (G-BAT) optimization algorithm to strike a balance between exploratory and exploitative search styles.

The grasshopper optimisation (G.O.) algorithm is a quick and simple NIO technique that draws inspiration from the swarm intelligence of actual grasshoppers [46]. By not becoming stranded in local optima, G.O. maintains a healthy balance between exploration and exploitation in its search behaviour. Grasshoppers may get to their safe zones more quickly, but the swarm doesn't converge on a central node. Thus, the search is inaccurate, and it happens too soon. The suggested method combines the G.O. algorithm with the BAT algorithm to fix the flaws of the G.O. algorithm and its accuracy. Similarly, to NIO [47], BAT is a swarm intelligence optimization technique. The BAT algorithm is one of a kind because it strikes a balance between exploratory and exploitative behaviours and has the benefit of enabling automated switching between exploration and exploitation to obtain the ideal solution, as opposed to relying on the fixed and predefined algorithmic dependent limits utilized by many NIO procedures. The primary parameters for optimizing the hybrid G-BAT are shown in Table 2 [48].

TABLE 2. List of basic limits for hybrid G-BAT.

Parameter	Initial Value
Freq-min	0
Maximum Number of iterations (MNI)	30
Swarm size (S.Z.)	25
Swarm_minval (min)	0.001
Swarm cohort	SP = min + (rand(SZ,1)
Termination disorder	MNI
Loudness (l)	1
Pulse rate (r0)	1

C. CLASSIFICATION METHODOLOGY

In this section, the study introduces the three mechanisms of CAMNet in detail: the coordination attention mechanism (CAM) and the linear unit. This is likely an acronym or shorthand for the name of the specific model or system used for classification. It could represent a combination of initials or words related to the model's design or purpose.

Classification of Diabetic Retinopathy: This part of the description indicates the main task that the model is designed for, which is the classification of diabetic retinopathy. Diabetic retinopathy is a medical condition affecting the eyes of individuals with diabetes.

n the relevant section of your manuscript, clearly define the different severity levels within your diabetic retinopathy classification system. For example:

Level 0: No diabetic retinopathy (DR) features observed.

Level 1: Mild non-proliferative DR (NPDR) with [specific features].

Level 2: Moderate NPDR with the presence of any hemorrhages.

[Additional severity levels as applicable].

1) OVERALL FRAMEWORK OF CAMNET

In the case of a segmented picture, Z = X + Y, where X is the collection of image pixels and Y is the collection of labels associated with those pixels. Processing and filling the input picture pixel by pixel yields N cubes of size S.R. (HWL), which are used for learning edge features. Cube space size (H, W) and number of spectral bands (L) are shown below. The three primary components of the intended CAMNet are as follows: At first, a convolution module is used to get the picture from the input. Second, a coordinated attention method is developed to completely take into account the space and spectrum of the input picture. Following feature extraction, a linear module is developed in the spirit of the

ghost module to extract complex characteristics with more precision. The layer is used to produce the final classification results. After that, we'll go over CAMNet's overall design and the principles behind each individual module.

2) CAM

Convolutional neural networks (CNNs) can improve their feature discrimination and picture classification abilities by employing an attention mechanism. The key to improved classification, however, lies in feature extraction, namely in the efficient extraction of spatial and spectral dimension data. As a result, we suggest a technique for coordinated attention to be employed in investigating the feature-distance connection. The spatial and spectral attention masks are obtained by the process in accordance with the long-distance connection.

There are two components to the CAM: spectral attention and spatial coordination. In order to better discern between spectral bands and achieve more precise spatial correlations, backdrops. The formula for F_{out}, if FR(HWL) is fed into CAM, goes as follows:

$$F_{out} = F.M_H\,(F)\,.M_W\,(F)\,.M_L(F) \tag{19}$$

where F and F_{out} characterize the input and output of CAM correspondingly. $M_W.H.(\cdot)$ characterizes the attention map in direction H, and the output scope is $H \times 1 \times 1$. $M_W.W.(\cdot)$ characterizes the attention map in direction W, and the output scope is $1 \times W \times 1$. Similarly, $M_L(\cdot)$ signifies the attention map in direction L, and the output scope is $1 \times 1 \times$ L. $M_H(\cdot)$ and $M_W(\cdot)$ are obtained by seeing the vertical and info, so as dependent information. Exactly, F obtains $F^H \in R^{H\times 1\times 1}$ in the vertical direction and $F_W \in R^{1\times W\times 1}$ horizontally through a layer of pooling data from all across the world and then cascading the findings. The vertical and horizontal long-distance dependencies are extracted by sending the cascaded results to the layer, the batch normalization layer (B.N.), and the nonlinear activation layer. The h_swish [49] activation function is used in the nonlinear activation layer; it requires few parameters but improves neural networks' representational power. The h_swish function may be uttered as,

$$f(x) = x \cdot sigmoid(ax) \tag{20}$$

where a is a trainable limit. Lastly, the obtained results are unglued and attention map $M_H(\cdot)$ and the horizontal map $M_W(\cdot)$.

Likewise, F passes through the layer to get $F_L \in R^{1\times 1\times L}$, and then the gotten result permits finished the unit the layer to find the map $M_L(F)$. The application procedure of CAM is exposed in Procedure 1.

3) CONVOLUTION MODULE

In terms of feature extraction, CNNs are quite capable. In particular, CNN's convolution and pooling procedures may be used to extract more nuanced insights from raw data. In order to prevent any loss of information, a CNN may be used to segment pictures to maintain the correlation between data

Algorithm 1 Facts of CAM

Input:
Features: $F \in R^{H\times W\times L}$.
Output:
Feature of CAM: $F_{out} \in R^{H\times W\times L}$.
Initialization:
Initialize all weight parameters of convolutional kernels.
F passes through L Avgpool, H AvgPool, and W AvgPool layers to generate $F_L \in R^{1\times 1\times L}$, $F_H \in R^{H\times 1\times 1}$, *and* $F_W \in R^{1\times W\times 1}$, *respectively;*
Reshape the size of feature F.H. to $1 \times H \times 1$ *and cascade with F.W. to generate* F_{HW};
Convolute F_{HW} *with the 3D unit convolution kernel and the results through regularization and nonlinear a:*
  Activation function layer to generate F_{HW};
Split F'_{HW} *and convolute the results with convolution kernel to generate F.H.' and F.W.';*
Normalize $F.H.'$ *and* $F.W.'$ *with the sigmoid function to generate the attention features M.H. (F)*$\in R^{H\times 1\times 1}$ *and* $M_W(F) \in R^{1\times W\times 1}$;
Convolute F_L *with the convolution kernel to generate F.L.';*
Normalize F'_L *with the sigmoid function to generate the attention feature* $M_L(F) \in R^{1\times 1\times L}$;
  Finally, the attention features $M_H(F) \in R^{H\times 1\times 1}$, $M_W(F) \in R^{1\times W\times 1}$, *and* $M_L(F) \in R^{1\times 1\times L}$ *are added to the input feature F to*
obtain $F_{out} \in R^{H\times W\times L}$.

pixels. Additionally, DR classification is still centred on the efficient extraction of spatial and spectral information from input pictures.

In this research, we present a space-spectrum convolution block for efficient extraction of spatial-spectral characteristics. Based on how Inception V3 [49] used a smaller convolution kernel to learn spatial-spectral characteristics, the convolution layer also uses this kernel to lower the parameters. Figure 7 depicts the internal construction of the convolution module.

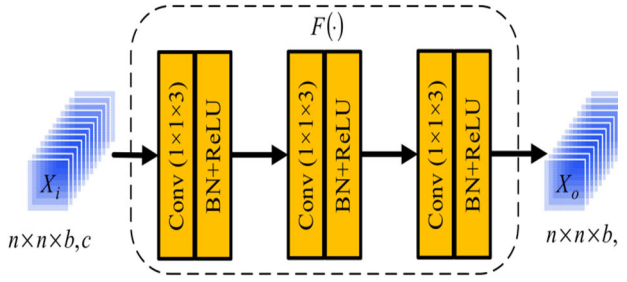

FIGURE 7. Convolution unit construction diagram.

As can be seen from Figure 7, input X_i entails the size of $n \times n \times b$. X_o is convolution, which can be spoken as,

$$X_0 = F(X_i) \tag{21}$$

In which F() is a composite nonlinear function. A three-layer neural network with convolution, batch normalization (B.N.), and a recurrent linear unit (ReLU) as its activation function. In the convolutional layer, the convolution kernel size is 1 1 3. With the ReLU function, nonlinear relationships between layers of a neural network may be strengthened, allowing the network to do its complicated duties.

$$g_{activate}(x) = \begin{cases} x & others \\ 0 & x \le 0 \end{cases} \tag{22}$$

where x characterizes $g_{activate}(x)$ symbolically stands for the nonlinear activation functions. In addition, a ReLU to normalize the data, which helps with the dispersion issue and speeds up the convergence time. Here is the formula for normalization:

$$\hat{x}^{(i)} = \frac{x^{(i)} - E\left[x^{(i)}\right]}{\sqrt{Var\left[x^{(i)}\right]}} \tag{23}$$

where $E\left[x^{(i)}\right]$ characterizes the regular of each neuron, and $\sqrt{Var\left[x^{(i)}\right]}$ represents the typical nonconformity of the input rate of each neuron.

4) LINEAR MODULE

The best way to boost classification performance in the DR classification assignment is to extract as much feature information as feasible. The ghost module [50] served as inspiration for the linear module used in this paper. Based on the features provided by the combined CAM and convolution modules, the linear module creates a more detailed feature map. Figure 8 depicts the linear module's internal construction. To generate the output y_o, we first linearly convolve the input y_i to produce y_m, and then we cascade the feature map y_m with the input y_i. The following expression defines the linear convolution output y_m.:

$$y_m = \varphi(y_i) = v_{i,j}^{x,y,z} \tag{24}$$

$$v_{i,j}^{x,y,z} = \sum_C \sum_{a=0}^{h_i-1} \sum_{\beta=0}^{w_i-1} \sum_{\gamma=0}^{l_i-1} K_{i,j,C}^{\alpha,\beta,\gamma,} \cdot v_{(i-1),C}^{(x+\alpha),(y+\beta),(z+\gamma)} + b_{i,j} \tag{25}$$

where $\varphi(\cdot)$ is a function, $v_{i,j}^{x,y,z}$ characterizes the neuron at the site (x, y, z) of the j-th map on the i-th layer, h_i, w_i, and l_i characterize the dimension, correspondingly, and C is the index of $(i-1)$ map. In totalling, $K_{i,j,C}^{\alpha,\beta,\gamma,}$ characterizes the j-th kernel on (a, b, g) at the C-th map position of layer i. $v_{(i-1),C}^{(x+\alpha),(y+\beta),(z+\gamma)}$ characterizes the value of the neuron at $(x+a, y+b, z+g)$ of layer $(i-1)$, and $b_{i,j}$ is the partiality term.

5) HYPER-PARAMETER TUNING PROCESS

The proposed classifier is tuned using RHSO (Rock Hyrax Swarm Optimisation), a meta-heuristic motivated by the social behaviour of rock hyraxes. The RHSO algorithm represents the collective foraging strategy and point of view of

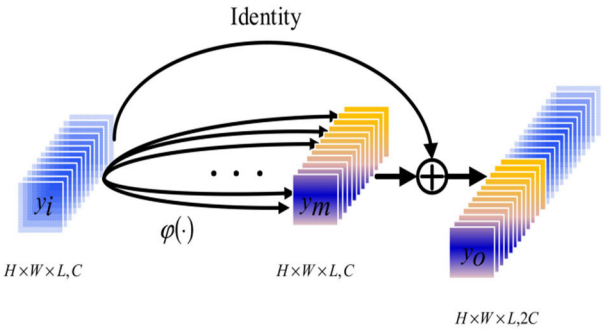

FIGURE 8. Construction diagram of linear unit.

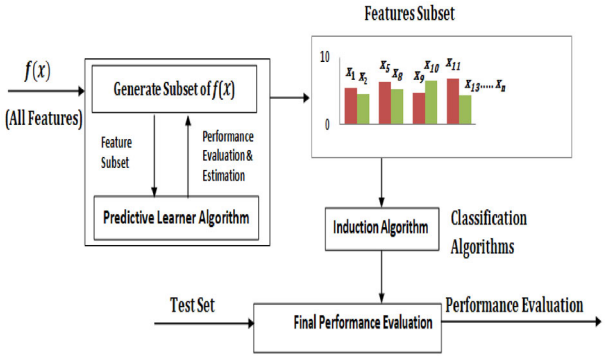

FIGURE 9. Feature selection replicas block diagram.

rock hyraxes. Rock hyraxes live in colonies or groups, each of which is overseen by a dominant male. The algorithm looks for the best options by integrating local heuristics with historical data to determine which attributes will be the most useful in the categorization process [51].

All of the data is split into a training set and a testing set, as shown in Figure 9 of the RHSO's working model. The ideal features are found by feeding the training data (i.e., f(x)) into the optimization process. A classification model's efficacy may be evaluated by providing it with a training set, an evaluation set, and a feature subset (i.e., f(x)). Equation (26) might be used as a metaphor for the procedure of selecting the most desirable features. Equation (27) improves the accuracy of classification using the aforementioned features by reducing the error with each repetition.

The size of the population, the number of generations, starting social scaling factors, mutation rates, and crossover rates are all variables that may be tweaked to influence the outcome of a population-based algorithm. The optimization algorithm may get stuck in a local optimum stage with inappropriate parameter values, which would increase the computational cost of the optimization algorithm problem, and thus the performance of the optimization method. These problems can be solved by employing the RHOSFS technique. In order to streamline the classification process, this strategy allows for the selection of the most relevant input features. This section elaborates on the RHOSFS technique.

A random population must be generated, selected, and analyzed before the total number of F.S. input characteristics can be calculated. Make it such that each group of features

stands in for a different slice of the input population. The goal of this research is to identify the set of input qualities that, taken together, have the dual effect of reducing the model's fitness and increasing its accuracy.

$$err\,(x_i) = actual\_output(x_i) - model\_estimated\_output(x_i) \quad (26)$$

$$fitness\,(x) = \frac{\sum_{x=0}^{n} err(x)}{n} \quad (27)$$

$$leader = r_1 \times x(leader_{pos}, j) \quad (28)$$

where r1 is a random integer in the range [0, 1], x is the leader's prior position, leader_pos is the leader's old position, and j means "each reduction." Each member updates its position using Equation (29) once the leader's position has been modified.

$$x\,(i, j) = (x\,(i, j) - (circ \times x\,(i, j) + leader)) \quad (29)$$

To try to reproduce the circle system in Equation (30), where circ signifies circular motion, the following calculation is made:

$$circ = sqrt\left(n_1^2 + n_2^2\right) \quad (30)$$

$$n_1 = r_2 \times \cos(ang) \quad (31)$$

$$n_2 = r_2 \times \sin(ang) \quad (32)$$

In Eqs. (31) and (32), where radius is a chance number among [0,1], and ang is the angle of a motion and is a random number between [0, 360], respectively. The lower and upper bands of the random number generator (lb and up, respectively) are used to inform an update to the range that occurs with each generation.

$$dalta = random[lb, ub] \quad (33)$$

$$ang = ang + dalta \quad (34)$$

If the output goes beyond the specified range (greater than 360 or less than 0), the angle (ang) can be changed to 360 or 0. Only those people whose new fitness value are higher than or equal to their old value are selected, and their novel fitness rate is inverted. Only the least fit individuals are passed on to the following generation. At last, the algorithm decides which candidates is the best fit.

V. EXPERIMENTATION, RESULTS AND DISCUSSION

A. SETUP FOR IMPLEMENTATION

PyTorch (https://pytorch.org/), a universal Python library, was used to create the suggested architecture and conduct the experiments. This research's tests, including the training of the models, were conducted on an outfitted with an Intel Xeon E5-2620 CPU running at 2.4 GHz and three NVIDIA GPUs, each with 12 GB. of RAM.

B. PARAMETERS EVALUATION

Automatic identification of retinal images for early-stage diabetic retinopathy using the fundus camera needs basic preprocessing steps before image distribution steps can be made. The retinal fundus picture collection is preprocessed using a combination of different techniques, including contrast adjustment, standard strain, and adaptive sifting. The accuracy of the algorithmic method for describing the retina was evaluated by computing the PSNR. The PSNR has a logarithmic value in decibels. When compared to the original image, the modified version has a higher advanced PSNR value.

Disease-related statistics and non-disease-related statistics are the two main types of statistics in healthcare. Evaluations of comprehension and detail help determine the level of correctness of actions. Digital fundus pictures for DR are created by a computation of each image's empathy, and the area of medicine studies the importance of sensitivity in this context. Based on fundus images, the true identifies the pixels that contain lesions, whereas the true reveals the non-lesion pixels. In contrast, a false negative (F.N.) represents lesion pixels that were overlooked by the implies number of non-lesion pixels that were incorrectly followed by guidelines [52]. Area was used to assess the efficacy of the suggested approach.

TABLE 3. Consequences of segmentation technique using e-ophtha-EX datasets.

Lesions	mIoU	mDice	F1-Score	Precision (P)	Recall	Accuracy (Acc)
EX	0.94	0.97	0.98	0.94	0.99	0.96

TABLE 4. Consequences of segmentation technique using DIARETDB1 datasets.

Lesions	mDice	F1-Score	Precision (P)	mIoU	Recall	Accuracy (Acc)
SE	0.88	0.87	0.86	0.86	1.00	0.86
HM	0.83	0.72	0.87	0.87	0.99	0.87
HE	0.83	0.92	0.71	0.71	0.99	0.71
MA	0.83	0.72	0.87	0.87	0.99	0.87

C. VALIDATION ANALYSIS OF PROPOSED SEGMENTATION RESULTS

The proposed segmentation perfect is validated on three datasets, and its values are given in Tables 3, 4 and 5.

Table 3 above represents the consequences of the segmentation technique using e-aphtha-EX datasets. In the analysis of the consequences of the segmentation technique, the mIoU is 0.94 and the mDice is 0.97, and the F1-Score value is 0.98, the precision degree is 0.94, the recall rate is 0.99, and finally, the accuracy rate is 0.96, respectively.

The consequences of the segmentation technique using DIARETDB1 datasets are shown in Table 4 above. The experimental evaluation of the H.M. model yielded the following

results: mIoU rate of 0.87, mice degree of 0.83, F1-Score range of 0.72, precision range of 0.87, recall rate of 0.99, and accuracy rate of 0.87, respectively. The mIoU rate was 0.71, the mDice rate was 0.83, the F1-Score range was 0.92, the recall rate was 0.99, and the accuracy rate was 0.71 after the HE models reached those values, respectively. The MA model then achieved the following values: mIoU rate of 0.87, mDice rate of 0.83, F1-Score range of 0.72, precision range of 0.87, recall rate of 0.99, and accuracy rate of 0.87, respectively. Following the S.E. model's achievement of the mIoU rate of 0.86, the mDice rate of 0.88, the F1-Score range of 0.87, the precision range of 0.86, the recall rate of 1.00, and finally the accuracy rate of 0.86, respectively.

The consequences of the segmentation technique using IDRiD datasets are indicated in Table 5 above. The mIoU rate, mDice rate, F1-Score range, precision range, recall rate, and accuracy rate for the H.M. model were all reached at 0.86, 0.88, 0.86, and 1.00, respectively, in the assessment metrics. The mIoU rate, mDice rate, F1-Score range, precision range, recall rate, and accuracy rate were all reached by the HE models at respective values of 0.88, 0.84, 0.88, 1.00, and 0.88, respectively. Following the M.A. model, the following values were obtained: mIoU rate of 0.71, mDice rate of 0.83, precision range of 0.71, recall proportion of 1.00, and accuracy rate of 0.71, respectively. The O.D. model then reached the following values: mIoU rate of 0.86, mDice rate of 0.87, F1-Score range of 0.87, precision range of 0.86, recall rate of 1.00, and accuracy rate of 0.86, respectively. The S.E. model then achieved a mIoU rate of 0.84, a precision range of 0.87, a recall rate of 0.98, and an accuracy rate of 0.97, respectively.

TABLE 5. Consequences of segmentation technique using IDRiD datasets.

Lesions	Accuracy (Acc)	mDice	F1-Score	mIoU	Precision (P)	Recall
OD	0.86	0.87	0.87	0.86	0.86	1.00
HM	0.86	0.88	0.88	0.86	0.86	1.00
HE	0.88	0.84	0.81	0.88	0.88	1.00
MA	0.71	0.83	0.92	0.71	0.71	1.00
SE	0.97	0.83	0.82	0.84	0.87	0.98

D. VALIDATION ANALYSIS OF PROPOSED CLASSIFIER

Tables 6 to 8 present the analysis of the proposed classifier with existing models on three datasets. The existing models, such as Capsule Network [23], CNN [24], and SqueezeNet [25], DenseNet [29], EfficientNet [35], SbCNF [37], and AlexNet+GoogleNet with the SVM [36] use various datasets for DR. Hence, comparative techniques are implemented in our research data, and the results are averaged.

The results of the classifier method using the e-ophtha-EX datasets are shown in Table 6 above. In the analysis of the SqueezeNet model, the accuracy rate, sensitivity rate, precision rate, specificity rate, F1-score rate, and AUC rate were all determined to be 0.9792, 0.9694, 0.9744, 0.9710, and 0.9798, respectively. The DenseNet model then achieved accuracy rates of 0.9615, sensitivity rates of 0.9475, precision rates of 0.945, specificity rates of 0.9622, F1-score rates of 0.9339, and finally, AUC rates of 0.9815. The EfficientNet model then obtained accuracy rates of 0.9570, sensitivity rates of 0.9308, precision rates of 0.9375, specificity rates of 0.9439, F1-score rates of 0.9254, and finally AUC rates of 0.9748. The CAM-RHSO model then achieved an accuracy rate of 0.989, a sensitivity rate of 0.9801, a precision rate of 0.9915, a specificity rate of 0.9715, an F1-score rate of 0.9869, and finally an AUC rate of 0.9905, in that order.

TABLE 6. Results of classifier method using E-ophtha-EX datasets.

Model	Sensitivity	Precision	Specificity	F1-Score	Accuracy	AUC
SqueezeNet	0.9694	0.9744	0.969	0.9710	0.9792	0.9798
DenseNet	0.9475	0.945	0.9622	0.9339	0.9615	0.9825
EfficientNet	0.9308	0.9375	0.9439	0.9254	0.9570	0.9748
CAM-RHSO	0.9801	0.9915	0.9715	0.9869	0.989	0.9905

TABLE 7. Results of classifier method using DIARETDB1 datasets.

Model	Sensitivity	Precision	Specificity	F1-Score	Accuracy	AUC
SqueezeNet	0.9481	0.9512	0.9435	0.9299	0.9459	0.959
DenseNet	0.943	0.923	0.9266	0.903	0.9375	0.964
EfficientNet	0.9305	0.92	0.9234	0.9181	0.9315	0.961
CAM-RHSO	0.953	0.9715	0.987	0.965	0.96	0.98

The results of the classifier method using the DIARETDB1 datasets are shown in Table 7 above. According to the analysis of the SqueezeNet model, the accuracy rate was 0.9459, the sensitivity rate was 0.9481, the precision rate was 0.9512, the specificity rate was 0.9435, the F1-score rate was 0.9299, and the AUC rate was 0.969. Next, the DenseNet model achieved accuracy rates of 0.9375, sensitivity rates of 0.943, precision rates of 0.923, specificity rates of 0.9266, F1-score rates of 0.903, and finally, AUC rates of 0.964. The EfficientNet model then achieved an accuracy rate of 0.9315, a sensitivity rate of 0.9305, a precision rate of 0.92, a specificity rate of 0.9234, an F1-score rate of 0.9181, and finally an AUC rate of 0.961, all in accordance with the data. The CAM-RHSO model then achieved respective accuracy rates of 0.96, sensitivity rates of 0.953, precision rates of 0.9715, specificity rates of 0.987, F1-score rates of 0.965, and AUC rates of 0.98.

The results of the classifier method using the DIARETDB1 datasets are shown in Table 8 above. The SqueezeNet model's evaluation resulted in accuracy rates of 0.9352, sensitivity rates of 0.9312, precision rates of 0.9232, specificity rates of 0.9099, F1-score rates of 0.9122, and finally, AUC rates of 0.949. The DenseNet perfect then achieved respective accuracy rates of 0.9205, sensitivity rates of 0.9074, precision

TABLE 8. Results of classifier method using IDRiD datasets.

Model	Precision	Specificity	Sensitivity	F1-Score	AUC	Accuracy
SqueezeNet	0.9232	0.9099	0.9312	0.9122	0.949	0.9352
DenseNet	0.9225	0.9222	0.9074	0.917	0.947	0.9205
EfficientNet	0.9175	0.919	0.9008	0.91	0.944	0.913
CAM-RHSO	0.954	0.9796	0.9638	0.942	0.965	0.9750

rates of 0.9225, specificity rates of 0.9222, F1-score rates of 0.917, and AUC rates of 0.947. The EfficientNet perfect then achieved accuracy rates of 0.913, sensitivity rates of 0.9008, precision rates of 0.9175, specificity rates of 0.919, and finally AUC rates of 0.944. Next, the CAM-RHSO model achieved accuracy rates of 0.9750, sensitivity rates of 0.9638, precision rates of 0.954, specificity rates of 0.9796, F1-score rates of 0.942, and finally, AUC rates of 0.965.

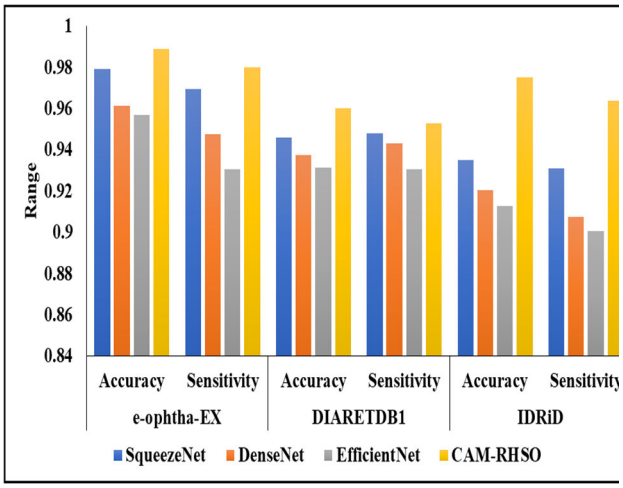

FIGURE 10. Analysis of various models.

The validation analysis of the proposed classifier for the entire dataset is shown in Table 9 above. In the analysis of the Capsule Network [23] approach, the first-class accuracy was 94.12, the precision rate was 0.94, the recall degree was 0.86, and the F1-score was finally 0.90. The second-class accuracy was 95.34, the recall rate was 0.95, and the F1-score was finally 0.91. The first-class accuracy was 89.04, the precision proportion was 0.61, the recall rate was 0.87, and the F1-score was finally 0. Approaching the first class accuracy of 95.80, the precision rate of 0.96, the recall rate of 0.90, and the F1-score of 0.93, respectively, the fourth class accuracy of 95.67, the precision rate of 0.96, the recall rate of 0.88, and the F1-score of 0.92, respectively, the first class accuracy of 90.94, the precision rate of 0.67, the recall rate of 0.91, and the F1-score of 0.77, respectively, and finally

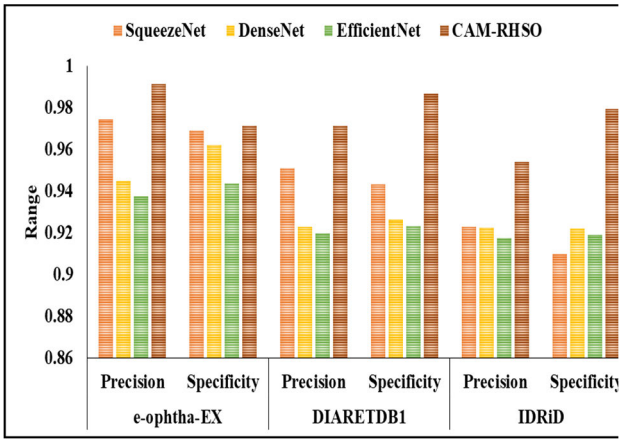

FIGURE 11. Graphical representation of the proposed model with existing models.

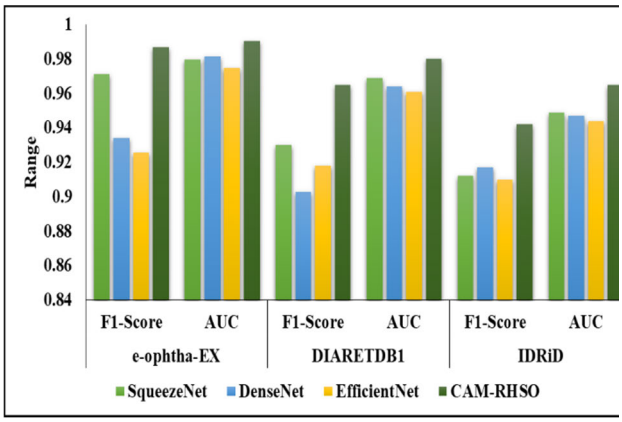

FIGURE 12. Comparative analysis based on F1-score and AUC.

the AlexNet+GoogleNet with SVM [36] approach the 1st class accuracy of 96.75 and the precision rate as 0.99 and also the recall degree as 0.91 and finally the F1-score as 0.95 respectively then the 3rd class accuracy of 95.56 and the precision rate as 0.84 and also the recall frequency as 0.99 and finally the F1-score as 0.90 respectively then the 1st class accuracy of 96.53 and the precision rate as 0.94 and also the recall rate as 0.91 and finally the F1-score as 0.93 respectively then the 4th class accuracy of 98.48 and the precision rate as 0.98 and also the recall rate as 0.96 and finally the F1-score as 0.97 SbCNF [37] approach the 1st class accuracy of 86.57 and the precision rate as 1.00 and also the recall rate as 0.68 and finally the F1-score as 0.81 respectively then the 1st class accuracy of 97.08 and the precision rate as 0.94 and also the recall proportion as 0.94 and finally the F1-score as 0.94 respectively then the 1st class accuracy of 3 87.77 and the precision degree as 0.47 and also the recall rate as 1.00 and finally the F1-score as 0.64 respectively then the 1st class accuracy of 98.32 0.94 and also the recall rate as 0.99 0.96 respectively. Finally, using our suggested methodology, the first-class accuracy was 98.72 percent, the F1 score was 0.96, and the precision rate was 0.98 percent. The second-class accuracy was 99.75 percent, the precision rate

TABLE 9. Validation analysis of proposed classifier for whole dataset.

Classifier	Classes	Accuracy	Precision	Recall	F1 Score
Capsule Network [23]	1	94.12	0.94	0.86	0.90
	2	95.34	0.95	0.88	0.91
	3	89.04	0.61	0.87	0.72
	4	96.16	0.96	0.89	0.92
CNN [24]	1	95.80	0.96	0.90	0.93
	2	95.67	0.96	0.88	0.92
	3	90.94	0.67	0.91	0.77
	4	96.38	0.96	0.91	0.93
AlexNet+GoogleNet with SVM [36]	1	96.75	0.99	0.91	0.95
	2	95.56	0.84	0.99	0.90
	3	96.53	0.94	0.91	0.93
	4	98.48	0.98	0.96	0.97
SbCNF [37]	1	86.57	1.00	0.68	0.81
	2	97.08	0.94	0.94	0.94
	3	87.77	0.47	1.00	0.64
	4	98.32	0.94	0.99	0.96
Proposed	1	98.72	0.99	0.96	0.98
	2	99.75	1.00	0.99	0.99
	3	98.09	0.93	0.99	0.96
	4	99.40	0.99	0.98	0.99

was 1.00 percent, the recall rate was 0.99 out of a hundred, and the F1 score was 0.99 percent. The third-class accuracy was 98.09 percent, the precision degree was 0.93 percent, the recall rate was 0.99 percent, and the F1 score was 0.96 percent.

VI. CONCLUSION AND FUTURE SCOPE

Diabetic retinopathy therapy relies heavily on early diagnosis. The pace of this procedure is keeping pace with the development of relevant technologies. The severity of the fundus pictures was classified using A.I. models in this study. The research suggests an innovative two-stage DR detection approach, with the first stage including OD and BV segmentation and the second stage involving DR categorization using transfer learning. During preprocessing, we extracted the green channel, resized everything uniformly, applied a top-bottom hat transformation, and segmented the OD and BV. Next, publicly accessible datasets are used to train CAMNet for DR-perfect. These datasets are Messidor-2 and DIARETDB0. Results from the evaluation of the projected model on the dataset show promising clinical relevance. The experimental research demonstrates that the suggested perfect outperformed state-of-the-art representations by a wide margin, with an accuracy of 96% to 98% across three datasets. The study indicated that the classification accuracy of proliferative DR pictures was enhanced by adding data from automated image segmentation. Other imaging problems, especially those with sparse training data, might benefit from this method of segmentation-assisted classification. The potential for the employment of many algorithms to complement each other to enhance all sorts of deep learning issues justifies further investigation beyond imaging. Explore opportunities for fine-tuning and optimizing the IC2T model and the Rock Hyrax Swarm-Based Coordination Attention Mechanism. Investigate different hyperparameters, architectures, or training strategies to improve performance further.

REFERENCES

[1] D. J. Hemanth, O. Deperlioglu, and U. Kose, "An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network," *Neural Comput. Appl.*, vol. 32, no. 3, pp. 707–721, Feb. 2020.

[2] A. Sugeno, Y. Ishikawa, T. Ohshima, and R. Muramatsu, "Simple methods for the lesion detection and severity grading of diabetic retinopathy by image processing and transfer learning," *Comput. Biol. Med.*, vol. 137, Oct. 2021, Art. no. 104795.

[3] A. Jain, A. Jalui, J. Jasani, Y. Lahoti, and R. Karani, "Deep learning for detection and severity classification of diabetic retinopathy," in *Proc. 1st Int. Conf. Innov. Inf. Commun. Technol. (ICIICT)*, Apr. 2019, pp. 1–6.

[4] S. Das, K. Kharbanda, and R. Raman, "Deep learning architecture based on segmented fundus image features for classification of diabetic retinopathy," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102600.

[5] D. S. David and A. A. Jose, "Retinal image classification system for diagnosis of diabetic retinopathy using SDCMethods," *Artech J. Eff. Res. Eng. Technol.*, vol. 1, pp. 87–93, Feb. 2020.

[6] T. Ratanapakorn, A. Daengphoonphol, N. Eua-Anant, and Y. Yospaiboon, "Digital image processing software for diagnosing diabetic retinopathy from fundus photograph," *Clin. Ophthalmol.*, vol. 13, pp. 641–648, Apr. 2019.

[7] D. S. David, "Retinal image classification system for diagnosis of diabetic retinopathy using morphological edge detection and feature extraction techniques," *Artech J. Eff. Res. Eng. Technol.*, vol. 1, pp. 28–33, Oct. 2020.

[8] L. Math and R. Fatima, "Adaptive machine learning classification for diabetic retinopathy," *Multimedia Tools Appl.*, vol. 80, no. 4, pp. 5173–5186, Feb. 2021.

[9] A. Sharma, S. Shinde, I. I. Shaikh, M. Vyas, and S. Rani, "Machine learning approach for detection of diabetic retinopathy with improved preprocessing," in *Proc. Int. Conf. Comput., Commun., Intell. Syst. (ICCCIS)*, Feb. 2021, pp. 517–522.

[10] K. Shankar, Y. Zhang, Y. Liu, L. Wu, and C.-H. Chen, "Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification," *IEEE Access*, vol. 8, pp. 118164–118173, 2020.

[11] S. S. Rahim, V. Palade, and A. Holzinger, "Image processing and machine learning techniques for diabetic retinopathy detection: A review," in *Artificial Intelligence and Machine Learning for Digital Pathology*. Springer, 2020, pp. 136–154.

[12] S. Patel, "Diabetic retinopathy detection and classification using pretrained convolutional neural networks," *Int. J. Emerg. Technol.*, vol. 11, no. 3, pp. 1082–1087, 2020.

[13] A. Shanthini, G. Manogaran, G. Vadivu, K. Kottilingam, P. Nithyakani, and C. Fancy, "Threshold segmentation based multi-layer analysis for detecting diabetic retinopathy using convolution neural network," *J. Ambient Intell. Humanized Comput.*, vol. 2021, pp. 1–15, Mar. 2021.

[14] M. K. Behera and S. Chakravarty, "Diabetic retinopathy image classification using support vector machine," in *Proc. Int. Conf. Comput. Sci., Eng. Appl. (ICCSEA)*, Mar. 2020, pp. 1–4.

[15] R. Afrin and P. C. Shill, "Automatic lesions detection and classification of diabetic retinopathy using fuzzy logic," in *Proc. Int. Conf. Robot., Elect. Signal Process. Techn. (ICREST)*, Jan. 2019, pp. 527–532.

[16] N. Sikder, M. Masud, A. K. Bairagi, A. S. M. Arif, A.-A. Nahid, and H. A. Alhumyani, "Severity classification of diabetic retinopathy using an ensemble learning algorithm through analyzing retinal images," *Symmetry*, vol. 13, no. 4, p. 670, Apr. 2021.

[17] R. Cordero-Martínez, D. Sánchez, and P. Melin, "Comparison of image pre-processing for classifying diabetic retinopathy using convolutional neural networks," in *Proc. Int. Conf. Hybrid Intell. Syst.* Cham, Switzerland: Springer, 2022, pp. 194–204.

[18] Y. Zhou, B. Wang, L. Huang, S. Cui, and L. Shao, "A benchmark for studying diabetic retinopathy: Segmentation, grading, and transferability," *IEEE Trans. Med. Imag.*, vol. 40, no. 3, pp. 818–828, Mar. 2021.

[19] T. Shanthi and R. S. Sabeenian, "Modified AlexNet architecture for classification of diabetic retinopathy images," *Comput. Electr. Eng.*, vol. 76, pp. 56–64, Jun. 2019.

[20] Y. S. Boral and S. S. Thorat, "Classification of diabetic retinopathy based on hybrid neural network," in *Proc. 5th Int. Conf. Comput. Methodologies Commun. (ICCMC)*, Apr. 2021, pp. 1354–1358.

[21] E. Z. Aziza, L. M. E. Amine, M. Mohamed, and B. Abdelhafid, "Decision tree CART algorithm for diabetic retinopathy classification," in *Proc. 6th Int. Conf. Image Signal Process. their Appl. (ISPA)*, Nov. 2019, pp. 1–5.

[22] I. Qureshi, J. Ma, and Q. Abbas, "Recent development on detection methods for the diagnosis of diabetic retinopathy," *Symmetry*, vol. 11, no. 6, p. 749, Jun. 2019.

[23] G. Kalyani, B. Janakiramaiah, A. Karuna, and L. V. N. Prasad, "Diabetic retinopathy detection and classification using capsule networks," *Complex Intell. Syst.*, vol. 9, no. 3, pp. 2651–2664, Jun. 2023.

[24] E. Özbay, "An active deep learning method for diabetic retinopathy detection in segmented fundus images using artificial bee colony algorithm," *Artif. Intell. Rev.*, vol. 56, no. 4, pp. 3291–3318, Apr. 2023.

[25] P. Uppamma and S. Bhattacharya, "Diabetic retinopathy detection: A blockchain and African vulture optimization algorithm-based deep learning framework," *Electronics*, vol. 12, no. 3, p. 742, Feb. 2023.

[26] K. Ohri and M. Kumar, "Supervised fine-tuned approach for automated detection of diabetic retinopathy," *Multimedia Tools Appl.*, vol. 2023, pp. 1–22, Jul. 2023.

[27] K. Parthiban and M. Kamarasan, "Diabetic retinopathy detection and grading of retinal fundus images using coyote optimization algorithm with deep learning," *Multimedia Tools Appl.*, vol. 82, no. 12, pp. 18947–18966, May 2023.

[28] B. N. Bansode, A. S. Dildar, and G. S. Sable, "Deep CNN-based feature extraction with optimised LSTM for enhanced diabetic retinopathy detection," *Comput. Methods Biomech. Biomed. Eng., Imag. Visualizat.*, vol. 11, no. 3, pp. 960–975, May 2023.

[29] S. Naik, D. Kamidi, S. Govathoti, R. Cheruku, and A. M. Reddy, "Efficient diabetic retinopathy detection using convolutional neural network and data augmentation," *Soft Comput.*, vol. 2023, pp. 1–22, Jun. 2023.

[30] M. S. A. Qaid, S. N. Basah, H. Yazid, M. H. M. Som, K. S. Basaruddin, and M. K. A. Hassan, "Performance analysis of diabetic retinopathy detection using fuzzy entropy multi-level thresholding," *Measurement*, vol. 216, Jul. 2023, Art. no. 112887.

[31] M. Nahiduzzaman, M. R. Islam, M. O. F. Goni, M. S. Anower, M. Ahsan, J. Haider, and M. Kowalski, "Diabetic retinopathy identification using parallel convolutional neural network based feature extractor and ELM classifier," *Expert Syst. Appl.*, vol. 217, May 2023, Art. no. 119557.

[32] G. Alwakid, W. Gouda, and M. Humayun, "Deep learning-based prediction of diabetic retinopathy using CLAHE and ESRGAN for enhancement," *Healthcare*, vol. 11, no. 6, p. 863, Mar. 2023.

[33] F. Mercaldo, M. Di Giammarco, A. Apicella, G. Di Iadarola, M. Cesarelli, F. Martinelli, and A. Santone, "Diabetic retinopathy detection and diagnosis by means of robust and explainable convolutional neural networks," *Neural Comput. Appl.*, vol. 35, no. 23, pp. 17429–17441, Aug. 2023.

[34] S. Krishnamoorthy, Y. Weifeng, J. Luo, and S. Kadry, "GO-DBN: Gannet optimized deep belief network based wavelet kernel ELM for detection of diabetic retinopathy," *Expert Syst. Appl.*, vol. 229, Nov. 2023, Art. no. 120408.

[35] C. Lin and Z. Jiang, "Development of preprocessing methods and revised EfficientNet for diabetic retinopathy detection," *Int. J. Imag. Syst. Technol.*, vol. 33, no. 4, pp. 1450–1466, Jul. 2023.

[36] B. Venkaiahppalaswamy, P. Prasad Reddy, and S. Batha, "Hybrid deep learning approaches for the detection of diabetic retinopathy using optimized wavelet based model," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, Art. no. 104146.

[37] K. Maithili, Y. M. Latha, A. Gangopadhyay, I. K. Varghese, A. Sundaram, C. Pandey, and A. Kiran, "Optimized CNN model for diabetic retinopathy detection and classification," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 7, pp. 317–331, 2023.

[38] EyePACS-1. (2015). *The World of Eyepacs*. [Online]. Available: http://www.eyepacs.com

[39] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton, and J.-C. Klein, "Feedback on a publicly distributed image database: The messidor database," *Image Anal. Stereology*, vol. 33, no. 3, p. 231, Aug. 2014.

[40] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, H. Uusitalo, H. Kälviäinen, and J. Pietilä, "DIARETDB0: Evaluation database and methodology for diabetic retinopathy algorithms," in *Machine Vision and Pattern Recognition Research Group*, vol. 73. Lappeenranta, Finland: Lappeenranta Univ. Technol., 2006, pp. 1–17.

[41] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unit: Unet-like pure transformer for medical image segmentation," in *Computer Vision—ECCV 2022*. Cham, Switzerland: Springer: Berlin, 2023, pp. 205–218.

[42] X. Chu, Z. Tian, Y. Wang, B. Zhang, H. Ren, X. Wei, H. Xia, and C. Shen, "Twins: Revisiting the design of spatial attention in vision transformers," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 9355–9366.

[43] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.

[44] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Munich, Germany, 2018, pp. 3–19.

[45] W. Wang, L. Yao, L. Chen, B. Lin, D. Cai, X. He, and W. Liu, "CrossFormer: A versatile vision transformer hinging on cross-scale attention," 2021, *arXiv:2108.00154*.

[46] S. Saremi, S. Mirjalili, and A. Lewis, "Grasshopper optimisation algorithm: Theory and application," *Adv. Eng. Softw.*, vol. 105, pp. 30–47, Mar. 2017.

[47] X. S. Yang and X. He, "Bat algorithm: Literature review and applications," *Int. J. Bio-Inspired Comput.*, vol. 5, no. 3, p. 141, 2013.

[48] X. S. Yang, "A new metaheuristic bat-inspired algorithm," in *Nature Inspired Cooperative Strategies for Optimization* (Studies in Computational Intelligence). Berlin, Germany: Springer, 2010, pp. 65–74.

[49] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 1314–1324.

[50] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 1577–1586.

[51] B. Al-Khateeb, K. Ahmed, M. Mahmood, and D.-N. Le, "Rock hyraxes swarm optimization: A new nature-inspired metaheuristic optimization algorithm," *Comput., Mater. Continua*, vol. 68, no. 1, pp. 643–654, 2021.

[52] M. Sharma, S. Sharma, and G. Singh, "Performance analysis of statistical and supervised learning techniques in stock data mining," *Data*, vol. 3, no. 4, p. 54, Nov. 2018.

B. N. JAGADESH (Senior Member, IEEE) is currently an Associate Professor with the School of Computer Science and Engineering, VIT-AP University, Amaravathi, India. He has published more than 30 research articles in national and international journals with high reputation. His research interests include image processing, medical image analysis, multi-modal biometrics, and speech recognition.

M. GANESH KARTHIK received the M.Tech. degree from the Sree Vidyanikethan Engineering College (Autonomous), Tirupati, and the Ph.D. degree in CSE from the SRM Institute of Science and Technology, Kattankulathur Campus, Chennai. He is currently an Assistant Professor with the Department of CSE, GITAM School of Technology, GITAM (Deemed to be University), Bengaluru. His research interests include computer networks, the IoT security, and machine learning.

D. SIRI received the B.Tech. degree in information technology and the M.Tech. degree in computer science and engineering from Jawaharlal Nehru Technological University (JNTU), Hyderabad, and the Ph.D. degree from JJT University Rajasthan. She is currently an Assistant Professor with the Department of CSE, Gokaraju Rangaraju Institute of Engineering and Technology, Bachupally, Telangana, India. She published two articles in international journals and one paper in national and international conferences. Her research interests include machine learning, software engineering, and the IoT.

S. K. KHAJA SHAREEF received the B.Tech. degree in information technology from Kakatiya University, and the M.Tech. degree in information technology from Jawaharlal Nehru Technological University (JNTU) Hyderabad, Telangana. He is currently a Research Scholar with JNTU Hyderabad. He is also an Associate Professor with the Department of IT, MLR Institute of Technology, Hyderabad, Telangana. He is having published 25 publications in reputed journals and 16 patents. His research interests include information security, blockchain technology, and machine learning.

SRIHARI VARMA MANTENA received the M.Tech. degree in IT from Andhra University, and the Ph.D. degree in CSE from Acharya Nagarjuna University, Guntur. He has around ten years of teaching experience. He is currently an Assistant Professor with the Department of Computer Science and Engineering, SRKR Engineering College, Bhimavaram, India. He has published seven papers in refereed journals and conference proceedings. His research interests include data mining, computer networks and security, image processing, and the Internet of Things.

RAMESH VATAMBETI received the B.Tech. degree in computer science and engineering from Sri Venkateswara University, Tirupati, and the M.Tech. degree in IT and the Ph.D. degree in CSE from Sathyabama University, Chennai. He has around 18 years of teaching experience from reputed engineering institutions. He is currently a Professor with the School of Computer Science and Engineering, VIT-AP University, Amaravati, India. He has published more than 70 papers in refereed journals and conference proceedings. He is a reviewer for several refereed international journals and acted as the session chair and a technical committee member for several international conferences held in India and abroad. He has successfully guided three Ph.D. scholars. His research interests include computer networks, mobile ad-hoc and sensor networks, and machine learning.