https://doi.org/10.31436/iiumej.v26i1.3194

RETINOPATHY DISEASE DETECTION AND CLASSIFICATION USING COORDINATE ATTENTION MODULE BASED CONVOLUTIONAL NEURAL NETWORK WITH LEAKY RECTIFIED LINEAR UNIT

PRAVIN BALASO CHOPADE[1*], PRABHAKAR N. KOTA[1], BHAGVAT D. JADHAV[2], PRAVIN MAROTRAO GHATE[2], SHRIRAM SADASHIV KULKARNI[3]

[1]*Dept. of Electronics &Telecommunications, M.E.S. College of Engineering, Pune, India*
[2]*Dept. of Electronics & Telecommunications, JSPM's Rajarshi Shahu College of Engineering, Pune, India*
[3]*Dept. of Information Technology, Sinhgad Academy of Engineering, Pune, India*

**Corresponding author: pbchopade@mescoepune.org*

(Received: 1 February 2024; Accepted: 30 October 2024; Published online: 10 January 2025)

ABSTRACT: The detection of Diabetic Retinopathy (DR) is an emergent research topic in recent decades, where DR is a primary cause of vision loss in humans. The existing techniques have limitations such as neuron death issues, vanishing gradient, and output offset. To overcome these issues, this paper proposes a Deep Learning (DL)-based technique for early and accurate DR detection. The Coordinate Attention Module (CAM) based Convolutional Neural Network (CNN) with Leaky Rectified Linear Unit (LReLU) is proposed for early and accurate detection of DR. The MESSIDOR dataset is preprocessed through the median filter to eliminate noise, and Contrast-Limited Adaptive Histogram Equalization (CLAHE) is utilized to increase the contrast level in an input image. The preprocessed images are given to Mayfly Optimization Algorithm-based Region Growing (MOARG) for image segmentation. Then, the features are extracted using ResNet50 and SqueezeNet, which extract deep learning features. The extracted features are given to CAM-based CNN with LReLU to detect DR, which overcomes the dead issues of neurons and minimizes the probability of inactive neurons. The proposed model achieves better results on the MESSIDOR datasets on the metrics of accuracy, precision, recall, specificity, f1-score, and Area Under Curve (AUC) values of about 99.72%, 99.46%, 99.25%, 99.61%, 99.37% and 99.14%, correspondingly, proving to be superior to the existing method, Capsule Network and Hybrid Adaptive DL based DR (HADL-DR).

***ABSTRAK*:** Pengesanan Retinopati Diabetik (DR) merupakan topik penyelidikan yang semakin mendapat perhatian dalam dekad-dekad kebelakangan ini, di mana DR merupakan punca utama kehilangan penglihatan pada manusia. Teknik sedia ada mempunyai beberapa kekangan seperti isu kematian neuron, vanishing gradient, dan output offset. Untuk mengatasi isu-isu ini, kertas ini mencadangkan teknik berasaskan Pembelajaran Mendalam (DL) untuk pengesanan awal dan tepat bagi DR. Modul Coordinate Attention Module (CAM) berasaskan Convolutional Neural Network (CNN) dengan Leaky Rectified Linear Unit (LReLU) dicadangkan untuk pengesanan awal dan tepat bagi DR. Dataset MESSIDOR diproses melalui penapis median yang digunakan untuk menghapuskan hingar, dan Contrast-Limited Adaptive Histogram Equalization (CLAHE) digunakan untuk meningkatkan tahap kontras pada imej input. Imej yang telah diproses diberikan kepada Algoritma Pengoptimuman Mayfly berasaskan Region Growing (MOARG) untuk segmentasi imej. Kemudian, ciri-ciri diekstrak menggunakan ResNet50 dan SqueezeNet yang mengekstrak ciri-ciri pembelajaran mendalam. Ciri-ciri yang diekstrak ini diberikan kepada CNN berasaskan CAM dengan LReLU untuk pengesanan DR, yang mengatasi isu kematian neuron dan meminimumkan

kebarangkalian neuron tidak aktif. Model yang dicadangkan mencapai keputusan yang lebih baik pada dataset MESSIDOR berdasarkan metrik ketepatan, ketepatan, panggilan semula, kekhususan, skor f1, dan nilai Kawasan di Bawah Lengkung (AUC) iaitu sekitar 99.72%, 99.46%, 99.25%, 99.61%, 99.37% dan 99.14%, masing-masing, membuktikan keunggulannya berbanding kaedah sedia ada, Capsule Network dan Hybrid Adaptive DL berasaskan DR (HADL-DR).

KEYWORDS: *Area Under Curve, Convolutional Neural Network, Coordinate Attention Module, Leaky Rectified Linear Unit, Mayfly Optimization Algorithm.*

1. INTRODUCTION

Retinopathy refers to a group of eye conditions that affect the retina and the light-sensitive tissue at the back of the eye [1]. Diabetic Retinopathy (DR) is the most common type of disease that occurs in people with diabetes, leading to vision impairment or blindness if not identified and detected at an early stage [2]. DR detection utilizes various medical image techniques to identify abnormalities in the retina [3]. Blurriness, dark areas of vision, floaters, difficulty perceiving colors, and vision loss are early symptoms of DR [4]. According to the World Health Organization (WHO), the prevalence of DR-related is approximately 4.8% among 37 million cases worldwide [5]. Therefore, individuals with DR require regular retinal screening to detect DR at an early stage, manage its progression, and prevent blindness [6]. Statistical analysis indicates that DR affects individuals over 18 and is increasing daily [7]. DR risk factor is separated into modifiable risks, such as smoking, blood pressure, and glucose levels, and non-modifiable risks, like age, genetic predisposition, and disease duration [8].

The severity of DR is classified into 5 stages: no DR, mild Non-Proliferative (NP) DR, moderate NPDR, severe NPDR, and proliferative DR based on the extent of damage to the retinal blood vessels [9]. In the No DR stage, the retina appears normal during an eye examination. In mild NPDR, there are small areas of swelling in blood vessels of the retina, but they do not cause vision issues, which is Van's early stage of DR [10] [11]. In moderate NPDR, severe changes occur in retinal blood vessels that contain blockages in vessels and decreased blood supply to the retina [12]. In severe NPDR, there is a substantial reduction in blood supply to the retina, which results in the growth of new blood vessels that compensate for the lack of blood flow [13]. Proliferative DR occurs when new, abnormal blood vessels start growing on the retina's surface, leading to retinal detachment and vision loss [14]. Several researchers have utilized DL techniques for the detection of DR, such as CNN and long short-term memory (LSTM), because they provide high accuracy [15]. Most of the research has utilized image patching and deep network architectures for DR detection [16]. The neuron death issue occurs when the ReLU activation function, which is observed in various neural networks, is applied to the MESSIDOR dataset. The traditional approach with ReLU leads to inactive neurons when gradients turn to zero during training, impacting the model's learning ability. The vanishing gradient issue occurs mainly in deeper networks where the activation functions compress gradients to smaller values, making it difficult for the network to propagate meaningful data to previous layers. The output offset rises when the predicted score deviates from its correct targets, occurring in networks that fail to converge optimally. The main contributions of the research are given below:

- The MOARG is used for segmentation, utilizing the global search capability of MOA to adapt intricate shapes in retinal images, such as lesions, abnormal regions, and blood vessels. This ensures that this model is capable and precise enough to deal with difficult variations in retinal structures, which is significant for accurate DR detection.

https://doi.org/10.31436/iiumej.v26i1.3194

- The CAM assists in visualizing the input image regions, enabling the model to highlight significant retinal regions that denote DR, thereby enhancing interpretability and performance.
- The LReLU function addresses the issue of dead neurons in traditional ReLU by introducing a small negative slope for negative input scores. It enables neurons to stay active even while receiving negative inputs, reducing inactive neuron issues and ensuring the model learns efficiently, enhancing its overall performance in DR detection.

The remaining parts of the paper are arranged as follows: Section 2 provides a literature review, Section 3 explains the proposed methodology in detail, Section 4 provides the results and discussion, and Section 5 gives the conclusion.

2. LITERATURE REVIEW

This section describes some recent literature on retinopathy detection using Machine Learning (ML) and deep learning (DL) techniques.

Parthiban and Kamarasan [17] introduced an Intelligent Coyote Optimization Algorithm with DL-based DR Detection (ICOA-DLDRD) on fundus images. The model contained Gabor Filtering (GF)-based noise elimination and optimum region growing for segmentation. Lastly, COA with Deep Extreme Learning Machine (DELM) utilized cross-validation that assisted invisible data for DR. The developed model easily handled high-dimensional data, making it suitable for solving complex problems. However, this model was prone to overfitting issues, particularly when dealing with noisy data.

Palaniswamy and Vellingiri [18] developed an Internet of Things DL-DR Diagnosis (IoTDL-DRD) through Retinal Fundus Images (RFI). Initially, the RFI was preprocessed to eliminate noise and enhance contrast. Then, the optimum region growing segmentation was utilized to detect the lesion region in an image. Next, the DenseNet-based feature extraction and LSTM-based classification were utilized for efficient DR detection. The developed model was capable of handling long-term dependencies and sequential input data. However, it struggled with learning certain features and the inability to handle temporal dependencies, making the model complex to train.

Kalyani et al. [19] presented a Capsule Network (CapsNet) for DR detection and classification. The features were extracted from the fundus image using the convolutional and primary capsule layers. Subsequently, the class capsule and SoftMax layer were utilized to assign the image to a precise class, where the created CapsNet accurately identified the issues at every stage. The developed model outperformed individual transfer learning and numerous ensemble approaches. Nevertheless, it could not manage complex data due to its vanishing gradient problems.

Islam et al. [20] suggested Supervised Contrastive Learning (SCL) for DR detection and classification of severity levels from images. A dual-stage training method with a loss function was developed in SCL to recognize DR and classify its severity levels. The CLAHE was used to enhance image quality, and Xception CNN was developed as an encoder. The developed model learned discriminative and meaningful features in the input data, which made the model more effective. However, the model was sensitive to noisy data and attempted to identify the optimal hyperplane.

Hemanth and Alagarsamy [21] implemented a HADL-DR through optimal feature extraction and classification. Initially, an enhanced Multichannel-based Generative Adversarial

https://doi.org/10.31436/iiumej.v26i1.3194

Network (MGAN) with semi-maintenance was employed to detect blood vessel segmentation. Scale Invariant Feature Transform (SIFT) was used for feature extraction, and the best solution was selected through an enhanced Sequential Approximation Optimization (SAO). Then, the hybrid Recurrent Neural Network with LSTM (RNN-LSTM) was used to classify DR. The developed model was used to prevent the network from overfitting issues, enhancing the training data. However, it is computationally exhaustive and struggles with noisy data.

Das and Saha [22] developed a CNN-tuned Genetic Algorithm (GA) for DR detection and classification. The developed CNN contained a series of convolutional and pooling layers that were utilized for feature extraction. Then, the Support Vector Machine (SVM) was used for classification. By using GA, hyperparameters like the number of kernels, pooling, and convolutional layers were determined. According to the dimensions of an input image, minimum and maximum kernel sizes needed to be generated. It efficiently captured spatial hierarchies and enabled the detection of significant structures at different levels, thereby leading to overfitting issues when dealing with limited data.

Zhang et al. [23] presented DR grading through the Deep Graph Correlation Network (DGCN) without Manual Annotations. The DGCN comprised the Graph Convolutional Network (GCN) to progress automatic DR grading without annotations. The DGCN included GCN for exploiting intrinsic correlations from autonomous images learned through CNN. The presented model designed three loss functions of pseudo-contrastive, transformation-invariant constraint, and graph-center to optimize the DGCN. The developed model efficiently learned long short-term spatiotemporal dependencies but failed to capture the spatial correlations.

Abbood et al. [24] introduced a hybrid image enhancement technique for DR diagnosis through ResNet50. The introduced model included two stages: cropping the image to eliminate irrelevant content and applying Gaussian blurring and shape crop for noise elimination and enhancing contrast. The ResNet50 was used to extract a feature set to aid the classification process. The introduced model was evaluated using the MESSIDOR and EyePACS datasets, which provided effective solutions in terms of accuracy. Nonetheless, it encountered issues related to high running time and computational costs.

3. PROPOSED METHODOLOGY

This research proposes a CAM-based CNN with LReLU for early and accurate DR detection. The MESSIDOR dataset used comprises 1800 fundus images with five classes. The preprocessing is done using a median filter to eliminate noise, and CLAHE is used to enhance the contrast level in an input image. The preprocessed images are fed to MOARG for image segmentation. Then, the features are extracted using ResNet50 and SqueezeNet, which extract deep learning features. The extracted features are given to CAM-based CNN with LReLU for detection and classification. It overcomes neuron death issues by allowing negative values for certain neurons and reducing the likelihood of inactive neurons. Figure 1 illustrates the process of the proposed methodology.

https://doi.org/10.31436/iiumej.v26i1.3194

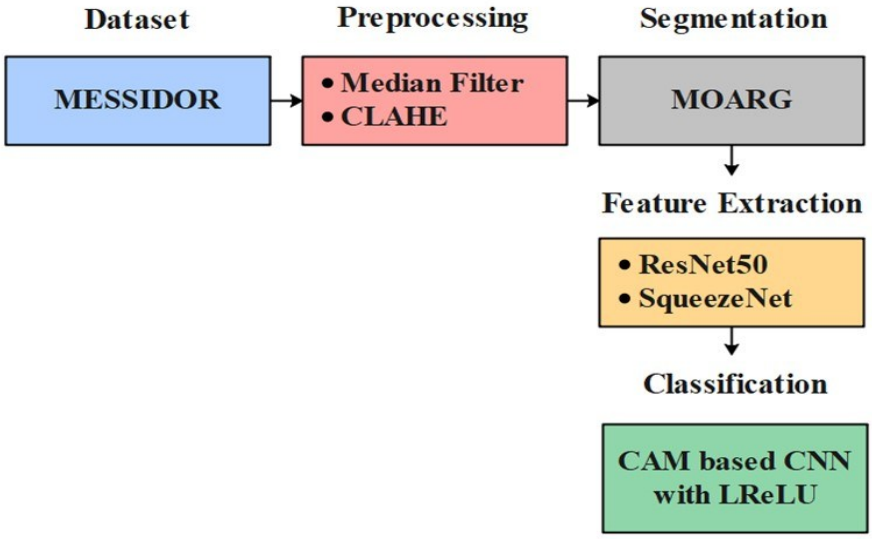

Figure 1. Process of the proposed methodology

3.1. Dataset

The dataset utilized in this manuscript is the MESSIDOR dataset [25], which consists of 1800 fundus images categorized into five different classes. Each class contains 360 images, and all images have been cropped and resized to 300×300. The classes of the dataset are listed below, and sample images are shown in Figure 2.

0: No DR

1: Mild Non-Proliferative DR (NPDR)

2: Moderate NPDR

3: Severe NPDR

4: Proliferative DR

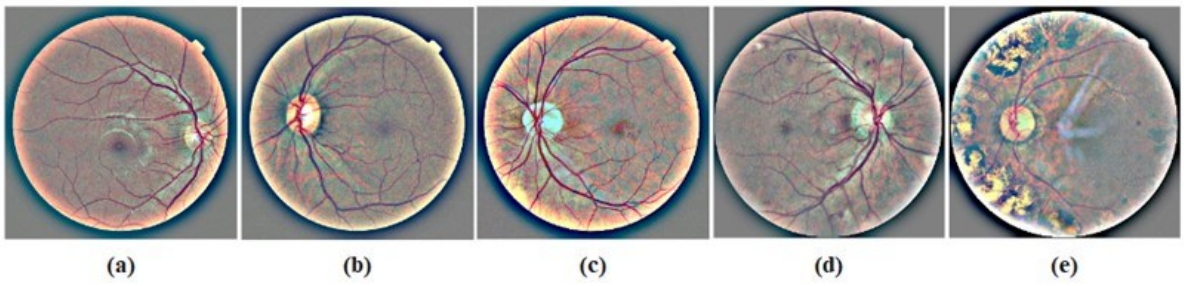

Figure 2. Dataset sample images (a) No DR (b) Mild NPDR (c) Moderate NPDR (d) Severe NPDR (e) Proliferative DR

3.2. Preprocessing

Preprocessing is a crucial phase in image processing because it improves the quality of the images from the MESSIDOR dataset. The median filter and CLAHE are applied to preprocess the MESSIDOR dataset. The median filter is applied to remove noise by replacing every pixel value with median intensity values in the adjacent pixels. The CLAHE is utilized to enhance the visibility of the retinal feature to improve the important regions for segmentation and feature extraction.

https://doi.org/10.31436/iiumej.v26i1.3194

3.2.1. Median Filter

The MESSIDOR dataset is preprocessed using a median filter, which removes noise in the MESSIDOR dataset. It is processed through shifting image pixels, adapting each score by the median value of the nearest pixel. The pixel is calculated by separating the entire pixel values from the nearest patterns in an arithmetic way and adapting the pixel measured as the average pixel value. It eliminates noise without minimizing image sharpness, and the mathematical formula is presented in Eq. (1).

$$\hat{f}(x,y) = median\{g(s,t)\},\ where\ (s,t) \in S_{xy} \tag{1}$$

where, S_{xy} denotes the coordinate groups in a rectangular image window, which has a center at (x,y). The $\hat{f}(x,y)$ is the reestablished image and $g(s,t)$ is estimated and corrupted region in S_{xy}. A median filter removes noise from an input image, which replaces each pixel with a median value in its local adjacent area.

3.2.2. CLAHE

The CLAHE is utilized to increase the contrast level, and the histogram of the resultant image is estimated based on the local window placed on every pixel in an adjacent region. Then, it is fixed through the limit value, and the region is mapped. However, reducing the number of histogram bars is unable to create a substantial variance, which results in pixels in the intermediate histogram and diminishing contrast among tissues—the output image after CLAHE is presented in Eq. (2).

$$CL = fZ_m(x,y) | \forall Z_m(x,y) \in Z_m \tag{2}$$

where, (x,y) represents the coordination of image pixels, fZ_m represents a transformation function based on density. By utilizing CLAHE, the contrast level is enhanced in an input image, helping to avoid saturation and clipping issues, which improves the image visibility without significantly amplifying noise. The preprocessed images are given as input to the segmentation process. The preprocessed (CLAHE) sample images are presented in Figure 3.

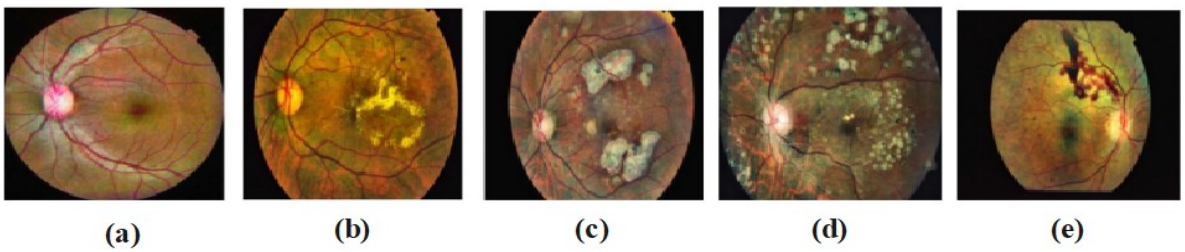

Figure 3. Preprocessed (CLAHE) sample images (a) No DR (b) Mild NPDR (c) Moderate NPDR (d) Severe NPDR (e) Proliferative DR

3.3. Segmentation

The preprocessed images are given as input to the Mayfly Optimization Algorithm-based Region Growing (MOARG) for image segmentation. Segmentation is identifying the particular region or structures in a fundus image. The optimization helps adjust the model's parameters to align with image features, which reduces communication and computational complexity. The RG is a pixel-based segmentation comprising similarity restrictions of intensity and texture, which are utilized to integrate pixels into the region. Primarily, the groups of pixels are incorporated with the assistance of iterations, and then the seed pixel is selected with regions. The group is carried out by incorporating equivalent, adjacent pixels to enhance the region size.

https://doi.org/10.31436/iiumej.v26i1.3194

The RG is completed if the adjacent pixel does not satisfy the similarity restriction and selects the other seed pixels. This process is repeated unless every pixel in an image corresponds to specific regions. The threshold and seed point selection are considered in this paper due to the increased segmentation accuracy. The MOARG is used for segmentation, utilizing the global search capability of the MOA to adapt complex shapes in the retinal images, including lesions, abnormal regions, and blood vessels. It ensures that this process is proficient and precise in dealing with complex variations in retinal structures essential for accurate DR detection. The process of region growth is presented as follows:

1. Provide the abnormal image as input.
2. Consider t as an optimum threshold produced through MOA.
3. Consider t as a seed point for the RG approach.
4. Include 4 adjacent pixels.
5. Evaluate the distance among adjacent pixels and the average region intensity.
6. Compute RG if $d \leq t$ on 4 neighboring pixels and save a coordinate point of new pixels.
7. Novel region mean is stored and returned to step 2 and RG is repeated until each pixel is integrated.

The MOA is a meta-heuristic optimization algorithm [26] for resolving difficult nonlinear issues that are motivated by the behavior of mayflies. The MOA location in 2D search space is defined by $a = (a_1, \dots, a_d)^T$ and $b = (b_1, \dots, b_d)^T$ respectively, and the velocity of $v = (v_1, \dots, v_d)^T$ is assigned to each mayfly. In each repetition, it maintains the global position ($gbest$) and best personal ($pbest$). The mayfly search is based on two populations, considering their position and speed with respect to specific rules. The male individual's speed and position are mathematically presented in Eq. (3) and (4).

$$x_i^{t+1} = x_i^t + v_i^{t+1} \tag{3}$$

$$v_i^{t+1} = v_i^t + a_1 e^{-\beta r_p^2}(pbest_i - x_i^t) + a_2 e^{-\beta r_g^2}(gbest - x_i^t) \tag{4}$$

where, x_i^t and v_i^t are the respective position and speed of ith the mayfly at kth iteration, $pbest_i$ is the best position of ith mayfly reached, and a_2 are coefficients of positive attractions and r_p is the Cartesian distance between mayfly and individuals' optimum positions. The female flies to the male for breeding to facilitate the mating approach, while the female is attracted to the male. The female mayflies and position are described in Eq. (5), (6), and (7).

$$v_i^{t+1} = v_i^t + a_2 e^{-\beta r_{mf}^2}(x_i^t - y_i^t),\ f(y_i) > f(x_i) \tag{5}$$

$$v_i^{t+1} = v_i^t = fl \times r,\ f(y_i) \leq f(x_i) \tag{6}$$

$$y_i^{t+1} = y_i^t + v_i^{t+1} \tag{7}$$

where, v_i^t and y_i^t are the speed and position of ith female mayfly at kth iteration, r_{mf} is the speed of a female from a male mayfly, and fl is a coefficient for random movement. Based on the above rule, the crossover is influenced by two population locations. The crossover is illustrated in Eq. (8) and (9).

$$offspring\ 1 = L \times male + (1 - L) \times female \tag{8}$$

$$offspring\ 2 = L \times female + (1 - L) \times male \tag{9}$$

https://doi.org/10.31436/iiumej.v26i1.3194

where, $offspring$ is an individual's offspring mayfly, L is a random value in a particular range, $male$ and $female$ is an individual male and female mayfly. Based on population size, if the new generation obtains the best fitness value, it will replace the respective variable of the previous generation. Otherwise, the previous generation will remain unchanged. At this stage, the objective function $f(x)$ is utilized to estimate and select the target population. Integrating MOA with RG explores the global search spaces that adapt to complex shapes in the segmentation process.

3.4. Feature Extraction

After segmentation, features are extracted using ResNet50 and SqueezeNet. By using this technique, the MESSIDOR dataset is much more informative. The ResNet50 extracts high-level features from the segmented regions to capture the significant patterns based on the abnormal blood vessels and lesions. The SqueezeNet focuses on low-level features, such as textures and edges in early layers, and higher-level features like patterns in deeper layers. Using both ResNet50 and SqueezeNet provides the corresponding set of deep features, thereby enhancing the overall feature extraction performance.

3.4.1. ResNet50-based Feature Extraction

ResNet50 has a residual block to resolve CNN's vanishing gradient and degradation issues. It can produce better classification results, and the residual function is given in Eq. (10).

$$y = F(x, W) + x \tag{10}$$

Here, x, y, and W are residual block input, output, and weight. It has different residual blocks in the different convolution layer filter sizes.

3.4.2. SqueezeNet-based Feature Extraction

SqueezeNet comprises fire modules containing squeeze and expand layers 1 convolutional (conv) layer, 8 fire modules, and 3 max-pooling layers. This layer has 1×1 conv and expand layers comprising an integration of 1×1 and 3×3 conv. In the expanded layer, the feature map of conv outputs is integrated into the channel of the fire module. The number of conv filters in the squeeze and expand layer is presented in Eq. (11).

$$X < Y_1 + Y_2 \tag{11}$$

where X is the number of 1×1 conv filters in the squeeze layer, and Y_1 and Y_2 are the number of 1×1 and 3×3 conv filters in the expand layer. Figure 4 represents the architecture of feature extraction methods. In the framework of detecting and classifying retinopathy diseases, SqueezeNet's approach captures significant low-level features like edges and textures to determine the fine-grained information in retinal images. This low-level information contains microaneurysms and vascular structures, which provide accurate diagnoses. While ResNet50 effectively captures high-level features, its deeper layers generate finer information because of the abstraction process. Even ResNet50 also extracts low-level features, but SqueezeNet is applied to obtain comparable accuracy with fewer parameters and smaller model sizes, making extraction effective. This hierarchical approach enables SqueezeNet to manage the fine-grained information, whereas ResNet50 increases the representation by focusing on high-level patterns. This approach ensures that no primary low-level characteristics are missed, enhancing classification performance in retinopathy detection. Therefore, combining each network utilizes a complementary benefit for a further process. The ResNet and SqueezeNet is combined by concatenating the extracted feature maps (2048 from ResNet50 and 1024 from SqueezeNet) into a unified feature vector. The final layer of ResNet50/SqueezeNet is Global

https://doi.org/10.31436/iiumej.v26i1.3194

Average Pooling (GAP). Overall, the output extracted no. of features is $N \times 3072$, which are fed as input to the classification process, where N is represented as the number of samples.

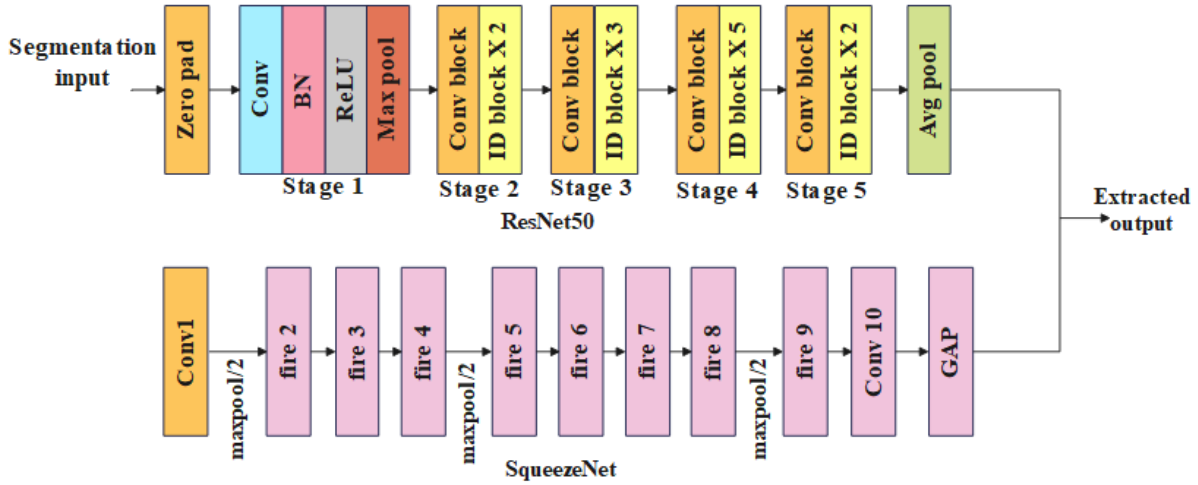

Figure 4. Architecture of feature extraction methods

3.5. Classification

The extracted features are classified through CAM-based CNN with LReLU, which is utilized to improve the model's capability to learn spatial relationships and focus on relevant features in an image. The CAM helps visualize the input image regions to highlight the important retinal regions, which further denotes that the DR helps enhance performance and interpretability. The LReLU addresses the issue of dead neurons in traditional ReLU by introducing a small negative slope for negative input scores. It allows neurons to stay active even when receiving negative inputs. It reduces inactive neuron issues and ensures the model learns efficiently, enhancing overall performance in DR detection. CNN is utilized to classify and identify large amounts of data. The advantage of CNN is that it automatically extracts and learns features from the training images, although existing techniques require manual feature extraction. The CNN contains 3 different conv, pooling, and Fully Connected (FC) layers. The conv is the initial and main element of CNN, extracting features from the input and containing fewer arrays named filters, which are used throughout the input and produce outputs of the feature maps. Different conv kernels are applied to extract various features. Numerous conv layers are based on the input image size. Then, pooling is used to reduce the dimension of the conv feature map. It produces down-sampling operations by reducing feature maps, which reduces computational complexity. Different pooling operations, such as max, min, and average pooling, are employed. The feature map outputs of the conv or pooling layer are altered to a 1D vector in which every input is combined through each result with weights and is named a dense layer. Lastly, it has FC layers; the last FC layer comprises an output number equivalent to many classes. The CNN approach employs the LeakyReLU activation function in all layers after convolution. The LeakyReLU's slope parameter is represented as α which manages the negative slope and has a default value of 0.01.

3.5.1. Coordinate Attention Module

The attention mechanism in deep neural networks (DNN) focuses on relevant elements and removes irrelevant ones. The Coordinate Attention Module (CAM) is an efficient attention module that can be easily integrated into DNN [27]. It supports maximizing the feature extraction ability without incurring significant computational overhead. It uses an efficient

https://doi.org/10.31436/iiumej.v26i1.3194

technique for obtaining data location and channel for better feature extraction. Figure 5 presents the structure of CAM.

The CAM input is represented as $C \times H \times W$, where C is an input channel, H and W are the feature map's height and width. The CAM process is formulated below:

- The CAM festers 2D to dual 1D global pooling operations horizontally and perpendicularly. The 1D pooling operation outputs are $C \times 1 \times W$ and $C \times H \times 1$, correspondingly.
- Results of dual 1D pooling are integrated and managed to utilize 1×1 conv, batch norm, and non-linear layer. The output is $C/r \times 1 \times (W + H)$, here r is a compression ratio.
- The previous stage output is separated into dual sets ($C \times 1 \times W$ *and* $C \times H \times 1$), which are calculated to achieve weight in the horizontal and perpendicular ways.
- Lastly, input feature $C \times H \times W$ is reevaluated through the third stage results.

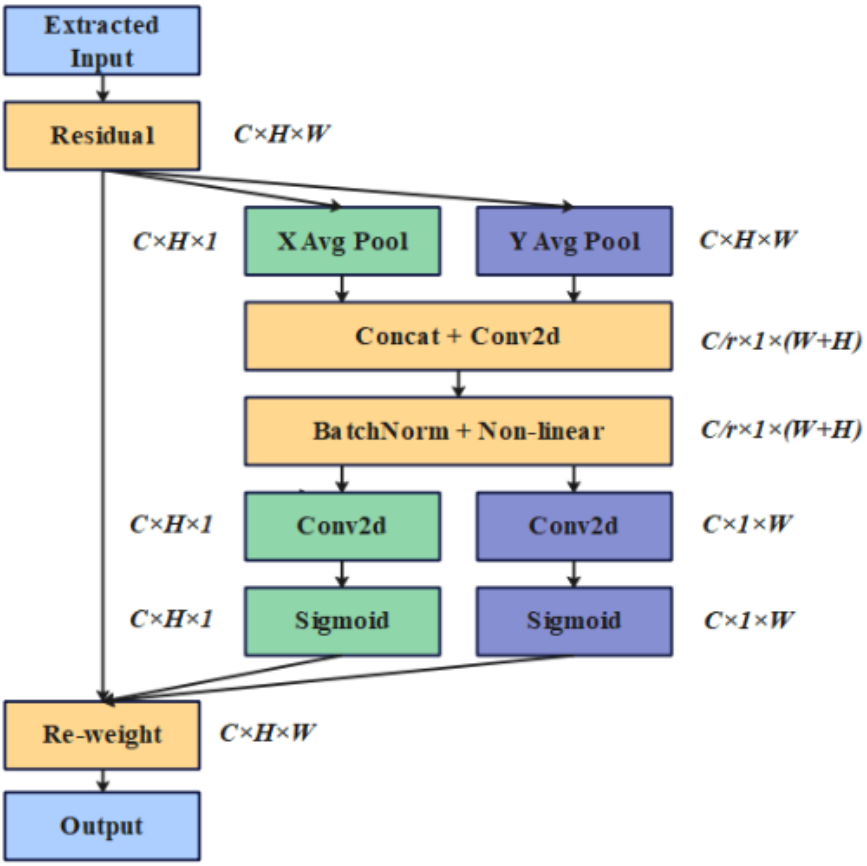

Figure 5. Structure of CAM

Assume the input of CAM is a vector presented in Eq. (12).

$$X = [x_1, x_2, \dots, x_c] \in R^{C \times H \times W} \tag{12}$$

It is calculated through dual 1D pooling in the horizontal and vertical ways. The channel c result at height h is presented in Eq. (13).

$$z_c^h(h) = \frac{1}{W} \sum_{0 \le i \le W} x_c(h, i) \tag{13}$$

https://doi.org/10.31436/iiumej.v26i1.3194

The channel c result at width w is presented in Eq. (14).

$$z_c^w(w) = \frac{1}{H}\sum_{0\le i\le H} x_c(j,w) \tag{14}$$

After producing feature maps, the concatenation operation is accomplished on a spatial dimension, and feature maps are integrated to 1×1 conv alteration function F_1 is presented in Eq. (15).

$$f = \delta\left(F_1(z^h, z^w)\right) \tag{15}$$

where, $[.,.]$ is an addition operation through geometric dimension, δ denotes non-linear activation function, $f \in R^{C/r\times(H+W)}$ presents the separating of feature map that encodes geometric data in vertical and horizontal directions and r is the reduction ratio to manage the block size. The f separates the geometric dimension into dual different tensors $f \in R^{C/r\times H}$ and $f \in R^{C/r\times W}$. The remaining two 1×1 conv alteration functions F_h and F_w are applied to separate transform f^h and f^w into tensors by the same input channel X, as is presented in Eq. (16) and (17).

$$g^h = \delta\left(F_h(f^h)\right) \tag{16}$$

$$g^w = \delta\left(F_w(f^w)\right) \tag{17}$$

The results of g^h and g^w in 2D are extended and utilized as attention weights. The result of CAM is presented in Eq. (18).

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{18}$$

where, y_c is the output of the CAM module, and the weights of g^h and g^w in 2D are fused with the input X to obtain the coordinate attention module output. Integrating CNN with CAM improves the model's capability to learn spatial relationships and focus on relevant features in an image.

3.5.2. Leaky Rectified Linear Unit

The drawbacks of ReLU and other activation functions lead to neuron death, vanishing gradient, and output offset. To overcome these issues, this paper proposes an LReLU that is a non-monotonic constant function that preserves an equal binary format. LReLU is used to reduce the dead neuron problem that occurs with ReLU, and LReLU also provides a smoother gradient flow for negative inputs. It contributes to increased model stability for retinopathy detection and classification in this research. The nature of activation functions is a bivariate function $b(x, g(x))$, where, x is an unfiltered pre-activation given as input to the final bivariate function, which is numerically presented in Eq. (19).

$$f(x) = \begin{cases} x & x \ge 0 \\ \alpha \times x \times \sigma(x) & x < 0 \end{cases} \tag{19}$$

where $\sigma(x)$ is a sigmoid function that is equivalent to $\frac{1}{1+e^{-x}}$ and α is a hyperparameter within the range of [0.1, 1]. As an output, the LReLU resolves the problems of non-zero gradients and saturation to speed up the learning procedure. The absence of a sigmoid function minimizes the power operations, which leads to faster processing. The Leaky approach makes the negative function area adaptable through a single input. The LReLU delivers a negative bump for small inputs that resolve the neuron death issues. The LReLU derivative function is presented in Eq. (20).

https://doi.org/10.31436/iiumej.v26i1.3194

$$f'(x) = \begin{cases} 1 & x \geq 0 \\ \alpha\left[\frac{1}{1+e^{-x}} + \frac{xe^{-x}}{(1+e^{-x})^2}\right] & x < 0 \end{cases} \quad (20)$$

where, α manages the depth of the negative dump for non-zero inputs, whereas the derivative is one in a positive region. The negative region is a combination of two functions, requiring a partial derivative to attain a gradient during backpropagation. By utilizing LReLU, the neuron death, vanishing gradient, and output offset are addressed, leading to early and accurate DR detection.

4. EXPERIMENTAL RESULT

The CAM-based CNN with LReLU is implemented in Python with the following configuration: Intel Core i7 processor, 16 GB RAM, and Windows 10 OS. The metrics, including accuracy, precision, recall, specificity, f1-score, AUC, Dice Score Coefficient (DSC), Intersection-Over-Union (IoU), Structural Similarity Index (SSIM), and Mean IoU (MIoU), are considered to calculate CAM-based CNN with LReLU, as shown mathematically in Eqs. (21)-(30).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (21)$$

$$Precision = \frac{TP}{TP+FP} \quad (22)$$

$$Recall = \frac{TP}{TP+FN} \quad (23)$$

$$Specificity = \frac{TN}{TN+FP} \quad (24)$$

$$F1-score = 2 \times \frac{Precision \times recall}{Precision+recall} \quad (25)$$

$$AUC = \frac{\sum R_i(I_I) - I_I(I_I+I)/2}{I_I+I_f} \quad (26)$$

$$DSC = \frac{2 \times TP}{(TP+FP)+(TP+FN)} \quad (27)$$

$$IoU = \frac{TP}{TP+FP+FN} \quad (28)$$

$$SSIM = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \quad (29)$$

$$MIOU = \frac{1}{k+1}\sum_{i=0}^{k} \frac{TP}{FN+FP+TP} \quad (30)$$

where, R_i is the rate of ith image, I_I and I_f is the number of positive and negative images, μ_x and μ_y are the average values of x and y, σ_x^2 and σ_y^2 are the variances of x and y, σ_{xy} is the covariance of x and y, TP, TN, FP and FN are the True Positive, True Negative, False Positive and False Negative values, correspondingly.

4.1. Quantitative and Qualitative Analysis

The quantitative and qualitative analysis of CAM-based CNN with LReLU is given in this section based on accuracy, precision, recall, specificity, F1-score, and AUC. The Long-Short Term Memory (LSTM) needs higher training data and is slowly trained on large datasets. The Recurrent Neural Network (RNN) cannot handle huge amounts of data due to the vanishing gradient issue. The proposed CAM-based CNN with LReLU overcomes the problem of neuron

https://doi.org/10.31436/iiumej.v26i1.3194

death by allowing negative values for some neurons and minimizing the probability of inactive neurons.

Table 1. Performance of segmentation

Method	DSC	IoU	SSIM	MIOU
WOA	0.9079	0.9062	0.9068	0.9071
GOA	0.9261	0.9258	0.9247	0.9253
MOA	0.9358	0.9347	0.9353	0.9341
MOARG	**0.9574**	**0.9566**	**0.9571**	**0.9559**

Figure 6. Performance of segmentation

Table 1 and Figure 6 present the performance of MOARG on the MESSIDOR dataset. The performance of the Whale Optimization Algorithm (WOA), Grasshopper Optimization Algorithm (GOA), and MOA are evaluated and compared with MOARG. The results indicate that MOARG achieves better results by utilizing DSC, IoU, SSIM, and MIOU, with respective values of 0.9574, 0.9566, 0.9571, and 0.9559, which is superior to the existing techniques.

Table 2. Performance of optimization in segmentation

Method	DSC	IoU	SSIM	MIOU
CSO	0.8964	0.8953	0.8949	0.8926
WOA	0.9079	0.9062	0.9068	0.9071
GOA	0.9261	0.9258	0.9247	0.9253
MOA	**0.9358**	**0.9347**	**0.9352**	**0.9341**

https://doi.org/10.31436/iiumej.v26i1.3194

Figure 7. Performance of optimization in segmentation

Table 2 and Figure 7 present the performance of MOA in segmentation on the MESSIDOR dataset. The performance of Cat Swarm Optimization (CSO), WOA, and GOA is measured and compared with MOA. The obtained results prove that MOA achieves superior outcomes by utilizing DSC, IoU, SSIM, and MIOU with respective values of 0.9358, 0.9347, 0.9352, and 0.9341, thereby yielding higher results compared to existing techniques.

Table 3. Performance of feature extraction

Method	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)	AUC (%)
ResNet18	91.82	91.66	91.74	91.61	91.55	91.47
ResNet50	93.68	93.54	93.65	93.59	93.46	93.38
SqueezeNet	94.73	94.61	94.68	94.62	94.58	94.51
ResNet50+ SqueezeNet	**96.56**	**96.48**	**96.51**	**96.43**	**96.37**	**96.33**

Figure 8. Performance of feature extraction

https://doi.org/10.31436/iiumej.v26i1.3194

Table 3 and Figure 8 display the outcomes of ResNet50+SqueezeNet on the MESSIDOR dataset. ResNet18, ResNet50, and SqueezeNet are measured and matched with ResNet50+SqueezeNet. The outcomes demonstrate the ResNet50+SqueezeNet's commendable performance in terms of accuracy, precision, recall, specificity, f1-score, and AUC, with values of 96.56%, 96.48%, 96.51%, 96.43%, 96.37%, and 96.33%, surpassing the previous techniques.

Table 4. Performance of activation functions

Method	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)	AUC (%)
ReLU	94.57	94.43	94.51	94.39	94.35	94.27
PReLU	95.83	95.62	95.78	95.56	95.43	95.38
SGReLU	97.76	97.48	97.63	97.51	97.25	97.29
LReLU	**98.69**	**98.51**	**98.55**	**98.42**	**98.36**	**98.31**

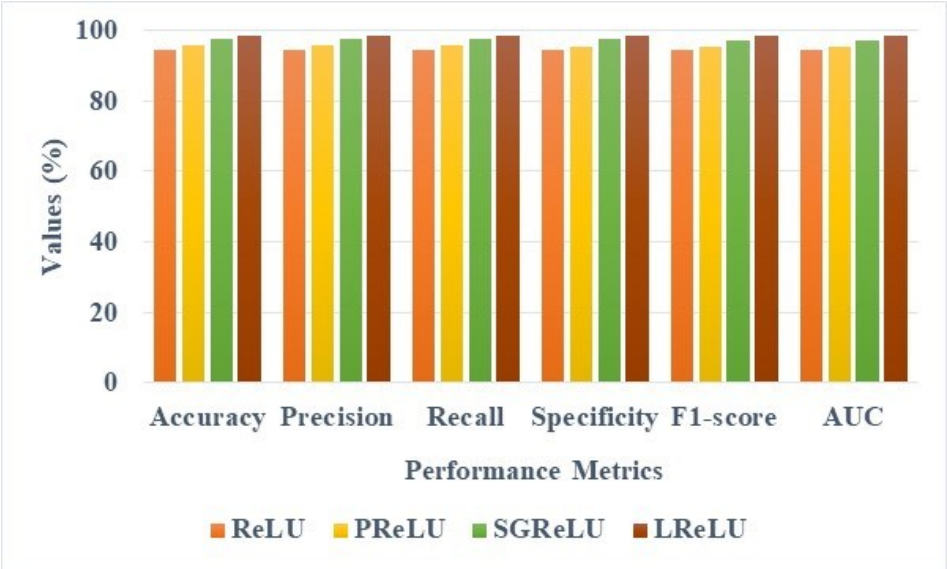

Figure 9. Performance of activation functions

Table 4 and Figure 9 showcase the performance of LReLU on the MESSIDOR dataset. ReLU, PReLU, and SGReLU are measured and matched against LReLU. The results indicate that LReLU outperforms with a commendable display of findings with 98.69% accuracy, 98.51% precision, 98.55% recall, 98.42% specificity, 98.36% f1-score, and 98.31% AUC value. This exemplifies the exceptional performance of the proposed method.

Table 5. Performance of classification

Method	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)	AUC (%)
LSTM	93.55	93.48	93.37	93.51	93.24	95.08
RNN	95.89	95.61	95.56	95.62	95.16	95.36
CNN	97.68	97.24	97.13	97.54	97.45	97.21
CAM based CNN with LReLU	**99.72**	**99.46**	**99.25**	**99.61**	**99.37**	**99.14**

https://doi.org/10.31436/iiumej.v26i1.3194

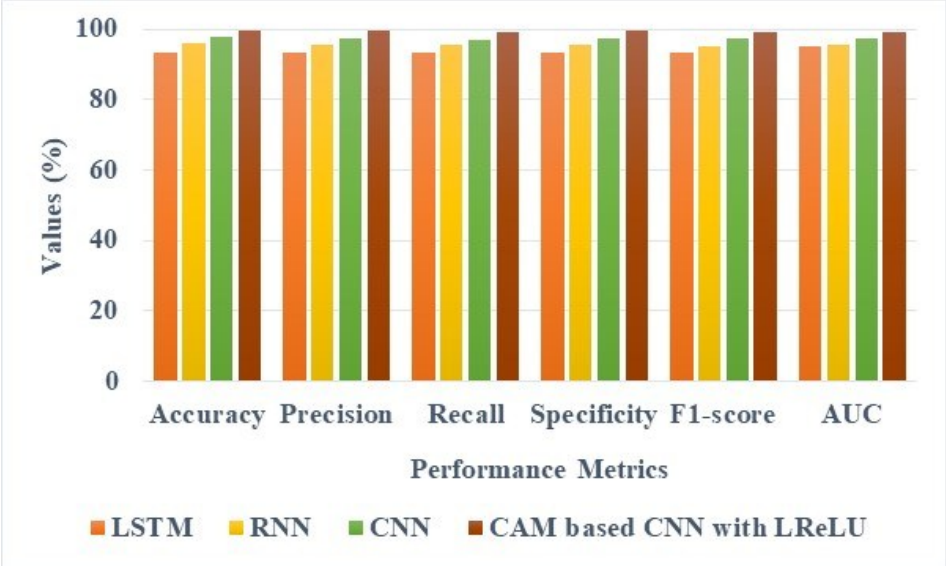

Figure 10. Performance of classification

Table 5 and Figure 10 reveal the performance of the proposed CAM-based CNN with LReLU on the MESSIDOR dataset. The LReLU activation function with LSTM, RNN, and CNN is utilized, and their performance is measured and compared with CAM-based CNN with LReLU. The obtained result speaks for the CAM-based CNN with LReLU's exemplary outcomes in terms of accuracy, precision, recall, specificity, f1-score, and AUC values of about 99.72%, 99.46%, 99.25%, 99.61%, 99.37%, and 99.14%, respectively, and in doing so, eclipses the previous methods.

4.2. Comparative Analysis

A comparative analysis of CAM-based CNN with LReLU performance based on six metrics is shown in Table 6. Existing techniques such as ICOA-DLDRD [17], IoTDL-DRD [18], CapsNet [19], SCL [20], and RNN-LSTM [21] are used to evaluate the classifier capability. The CAM-based CNN with LReLU is trained, tested, and validated on the MESSIDOR dataset. The obtained outcomes imply that CAM-based CNN with LReLU achieves superior results in terms of accuracy, precision, recall, specificity, f1-score, and AUC values of about 99.72%, 99.46%, 99.25%, 99.61%, 99.37%, and 99.14%, respectively, which takes over the other classifiers.

Table 6. Comparative Analysis Using the MESSIDOR Dataset

Method	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)	AUC (%)
ICOA-DLDRD [17]	99.65	99.18	99.01	99.76	N/A	N/A
IoTDL-DRD [18]	99.08	N/A	94.28	99.34	N/A	94.84
CapsNet [19]	97.65	97.54	94.18	N/A	95.85	N/A
SCL [20]	80.52	77.64	93.63	N/A	84.89	84.60
RNN-LSTM [21]	98.29	96.45	95.78	94.03	N/A	97.37
Proposed CAM-based CNN with LReLU	**99.72**	**99.46**	**99.25**	**99.61**	**99.37**	**99.14**

4.3. Discussion

This section discusses the advantages of CAM-based CNN with LReLU and the drawbacks of existing methods. The ICOA-DLDRD [17] requires highly labeled data, which consumes more time for training. The IoTDL-DRD [18] was not adapted to manage long-term dependencies and sequential input data. The CapsNet [19] could not manage complex data because of a vanishing gradient issue. The SCL [20] was sensitive to noisy data and tried to identify the optimum hyperplane. The RNN-LSTM [21] was computationally exhaustive and struggled with noisy data. The proposed CAM-based CNN with LReLU overcomes these existing models' limitations, including dead neurons, vanishing gradient, output offset issues, and the probability of inactive neurons. The MOARG enhances segmentation accuracy by exploring the global search space and adopting complex shapes, resulting in a more prominent segmentation performance than traditional algorithms, namely CSO, WOA, and GOA. The CAM-based CNN with LReLU model performs commendably, outperforming the other techniques with improved accuracy, precision, recall, specificity, F1-score, and AUC values of about 99.72%, 99.46%, 99.25%, 99.61%, 99.37%, and 99.14%. These values are superior to the existing models, namely ICOA-DLDRD [17], IoTDL-DRD [18], CapsNet [19], SCL [20], and RNN-LSTM [21], as shown in Table 6. This display of conduct deduces its capability to generate accurate and reliable DR detection and classification.

5. CONCLUSION

This research proposes a CAM-based CNN with LReLU for early and accurate detection of Diabetic Retinopathy. The MESSIDOR dataset is preprocessed through the median filter to remove noise, while CLAHE is deployed to increase the contrast level in an input image. The preprocessed images are given as input to the MOARG for image segmentation. Then, the features are extracted using ResNet50 and SqueezeNet, which extract deep learning features. The extracted features are provided to CAM-based CNN with LReLU for the detection and classification of Diabetic Retinopathy. It overcomes the issue of dead neurons by allowing negative values for some neurons and minimizing the probability of inactive neurons. The CAM-based CNN with LReLU obtains commendable findings on the MESSIDOR dataset in terms of metrics of respective accuracy, precision, recall, specificity, F1-score, and AUC values of about 99.72%, 99.46%, 99.25%, 99.61%, 99.37%, and 99.14%. This delivery of such outputs confirms the high-quality, superior capabilities of the suggested method over the existing methods. In the future, hyperparameter tuning will further enhance Diabetic Retinopathy performance.

REFERENCES

[1] Özbay E. (2023) An active deep learning method for diabetic retinopathy detection in segmented fundus images using artificial bee colony algorithm. Artif. Intell. Rev., 56:3291-3318. https://doi.org/10.1007/s10462-022-10231-3

[2] Raiaan MAK, Fatema K, Khan IU, Azam S, ur Rashid MR, Mukta MSH, Jonkman M, De Boer F. (2023) A Lightweight Robust Deep Learning Model Gained High Accuracy in Classifying a Wide Range of Diabetic Retinopathy Images. IEEE Access, 11:2169-3536. https://doi.org/10.1109/ACCESS.2023.3272228

[3] Usman TM, Saheed YK, Ignace D, Nsang A. (2023) Diabetic retinopathy detection using principal component analysis multi-label feature extraction and classification. Int. J. Cognit. Comput. Eng., 4:78-88. https://doi.org/10.1016/j.ijcce.2023.02.002

[4] Erciyas A, Barışçı N, Ünver HM, Polat H. (2023) Improving detection and classification of diabetic retinopathy using CUDA and Mask RCNN. Signal, Image Video Process., 17(4):1265-1273. https://doi.org/10.1007/s11760-022-02334-9

[5] Luo X, Wang W, Xu Y, Lai Z, Jin X, Zhang B, Zhang D. (2024) A deep convolutional neural network for diabetic retinopathy detection via mining local and long-range dependence. CAAI Trans. Intell. Technol., 9(1):153-166. https://doi.org/10.1049/cit2.12155
[6] Feng M, Wang J, Wen K, Sun J. (2023) Grading of Diabetic Retinopathy Images Based on Graph Neural Network. IEEE Access, 11:98391-98401. http://dx.doi.org/10.1109/ACCESS.2023.3312709
[7] Saranya P, Umamaheswari KM. (2024) Detection of exudates from retinal images for non-proliferative diabetic retinopathy detection using deep learning model. Multimedia Tools Appl., 83(17), pages 52253-52273. https://doi.org/10.1007/s11042-023-17462-8
[8] Fatima, Imran M, Ullah A, Arif M, Noor R. (2022) A unified technique for entropy enhancement based diabetic retinopathy detection using hybrid neural network. Comput. Biol. Med., 145:105424. https://doi.org/10.1016/j.compbiomed.2022.105424
[9] Nanda P, Duraipandian N. (2022) A Novel Optimizer in Deep Neural Network for Diabetic Retinopathy Classification. Computer Systems Science & Engineering, 43(3), 1099-1110. DOI: 10.32604/csse.2022.024695
[10] Kasim Ö. (2023) Ensemble classification based optimized transfer learning feature method for early stage diagnosis of diabetic retinopathy. J. Ambient Intell. Hum. Comput., 14(8):11337-11348. https://doi.org/10.1007/s12652-023-04648-z
[11] Rachapudi V, Rao KS, Rao TSM, Dileep P, Roy TLD. (2023) Diabetic retinopathy detection by optimized deep learning model. Multimedia Tools Appl., 82(18):27949-27971. https://doi.org/10.1007/s11042-023-14606-8
[12] Butt MM, Iskandar DNFA, Abdelhamid SE, Latif G, Alghazo R. (2022) Diabetic Retinopathy Detection from Fundus Images of the Eye Using Hybrid Deep Learning Features. Diagnostics, 12(7):1607. https://doi.org/10.3390/diagnostics12071607
[13] Mustafa H, Ali SF, Bilal M, Hanif MS. (2022) Multi-stream deep neural network for diabetic retinopathy severity classification under a boosting framework. IEEE Access, 10:113172-113183. https://doi.org/10.1109/ACCESS.2022.3217216
[14] Vinayaki VD, Kalaiselvi R. (2022) Multithreshold image segmentation technique using remora optimization algorithm for diabetic retinopathy detection from fundus images. Neural Process. Lett., 54:2363-2384. https://doi.org/10.1007/s11063-021-10734-0
[15] Murugappan M, Prakash NB, Jeya R, Mohanarathinam A, Hemalakshmi GR, Mahmud M. (2022) A novel few-shot classification framework for diabetic retinopathy detection and grading. Measurement, 200:111485. https://doi.org/10.1016/j.measurement.2022.111485
[16] Jagadesh BN, Karthik MG, Siri D, Shareef SKK, Mantena SV, Vatambeti R. (2023) Segmentation Using the IC2T Model and Classification of Diabetic Retinopathy Using the Rock Hyrax Swarm-Based Coordination Attention Mechanism. IEEE Access, 11:124441-124458. https://doi.org/10.1109/ACCESS.2023.3330436
[17] Parthiban K, Kamarasan M. (2023) Diabetic retinopathy detection and grading of retinal fundus images using coyote optimization algorithm with deep learning. Multimedia Tools Appl., 82(12):18947-18966. https://doi.org/10.1007/s11042-022-14234-8
[18] Palaniswamy T, Vellingiri M. (2023) Internet of Things and Deep Learning Enabled Diabetic Retinopathy Diagnosis Using Retinal Fundus Images. IEEE Access, 11:27590-27601. https://doi.org/10.1109/ACCESS.2023.3257988
[19] Kalyani G, Janakiramaiah B, Karuna A, Prasad LVN. (2023) Diabetic retinopathy detection and classification using capsule networks. Complex Intell. Syst., 9(3):2651-2664. https://doi.org/10.1007/s40747-021-00318-9
[20] Islam MR, Abdulrazak LF, Nahiduzzaman M, Goni MOF, Anower MS, Ahsan M, Haider J, Kowalski M. (2022) Applying supervised contrastive learning for the detection of diabetic retinopathy and its severity levels from fundus images. Comput. Biol. Med., 146:105602. https://doi.org/10.1016/j.compbiomed.2022.105602
[21] Hemanth SV, Alagarsamy S. (2023) Hybrid adaptive deep learning classifier for early detection of diabetic retinopathy using optimal feature extraction and classification. J. Diabetes Metab. Disord., 22(1):881-895. https://doi.org/10.1007/s40200-023-01220-6
[22] Das S, Saha SK. (2022) Diabetic retinopathy detection and classification using CNN tuned by

genetic algorithm. Multimedia Tools Appl., 81(6):8007-8020. https://doi.org/10.1007/s11042-021-11824-w

[23] Zhang G, Sun B, Chen Z, Gao Y, Zhang Z, Li K, Yang W. (2022) Diabetic retinopathy grading by deep graph correlation network on retinal images without manual annotations. Front. Med., 9:872214. https://doi.org/10.3389/fmed.2022.872214

[24] Abbood SH, Hamed HNA, Rahim MSM, Rehman A, Saba T, Bahaj SA. (2022) Hybrid retinal image enhancement algorithm for diabetic retinopathy diagnostic using deep learning model. IEEE Access, 10:73079-73086. https://doi.org/10.1109/ACCESS.2022.3189374

[25] Dataset link: https://www.kaggle.com/datasets/mohammadasimbluemoon/diabeticretinopathy-messidor-eyepac-preprocessed

[26] Hu A, Deng Z, Yang H, Zhang Y, Gao Y, Zhao D. (2021) An Optimal Geometry Configuration Algorithm of Hybrid Semi-Passive Location System Based on Mayfly Optimization Algorithm. Sensors, 21(22):7484. https://doi.org/10.3390/s21227484

[27] Xu T, Yuan Z. (2022) Convolution neural network with coordinate attention for the automatic detection of pulmonary tuberculosis images on chest x-rays. IEEE Access, 10:86710-86717. https://doi.org/10.1109/ACCESS.2022.3199419