

A HYBRID CAD APPROACH USING VGG-16, RAS-UNET, AND CNN-MGCA FOR MICROANEURYSM DETECTION IN DIABETIC RETINOPATHY

S. Steffia and D. Murugan

Department of Computer Science and Engineering, Manonmaniam Sundaranar University, India

Abstract

. Detecting red lesions in color retinal fundus images is essential for preventing vision loss and blindness in people with diabetic retinopathy (DR). Among these lesions, microaneurysms (MAs) are the earliest and most common indicators of DR, making their identification particularly important for effective large-scale screening programs. However, accurately spotting MAs is challenging due to low contrast and varying image quality across different imaging conditions. To overcome these challenges, computer-aided diagnostic (CAD) systems powered by deep learning have shown immense potential for supporting timely and precise diagnosis. In this study, we propose a comprehensive CAD framework that combines advanced deep learning models to improve both detection and classification of retinal abnormalities. Our method begins by enhancing image quality—reducing noise, improving clarity, and standardizing image size to ensure consistent input for downstream analysis. We then differentiate between healthy and DR-affected retinas using a VGG-16 network enhanced with a Spatial Pyramid Pooling (SPP) layer to extract rich and meaningful features. These features are then fed into an Extreme Gradient Boosting (XGBoost) classifier, which separates normal from diseased cases. Next, to locate potential microaneurysms, we employ a Residual U-Net architecture with atrous depthwise separable convolutions (RAS-UNet). This model consists of an encoder, an atrous convolution module, and a decoder. The atrous module combines cascaded and parallel operations to capture features at multiple scales, enabling more reliable detection of MAs of different sizes. Finally, we refine the results by passing candidate regions through a Convolutional Neural Network with MGCA (CNN-MGCA) to distinguish true microaneurysms from false positives. We evaluated our system using a range of performance metrics, including accuracy, AUC, sensitivity, specificity, positive predictive value (PPV), F1-score, and FROC analysis. Overall, our experimental results demonstrate that the proposed approach outperforms existing methods reported in the literature, offering a promising tool for large-scale automated diabetic retinopathy screening and early intervention.

Keywords:

RAS-UNet, CNN-MGCA, Spatial Pyramid Pooling (SPP), Extreme Gradient Boosting (XGBoost)

1. INTRODUCTION

Diabetic Retinopathy (DR) is a prevalent ocular ailment observed in individuals with diabetes, often resulting in damage to the blood vessels in the retina. Symptoms of DR, including Microaneurysms (MAs), Hard Exudates (HE), and Hemorrhages (HM), are observable through color fundus retinal imaging, as indicated by various scientific studies [1]. The progression of vision impairment due to DR usually occurs without preliminary symptoms from the patients. However, by identifying and treating DR in its early stages, one can prevent the loss of vision.

Exudates, identifiable by the presence of white or pale yellow patches in the retina, occur as a consequence of protein loss in the smaller retinal veins. Hemorrhages, exhibiting irregular red

patterns with varying edges, arise from the seepage of blood from delicate and slender blood vessels in the retina. Microaneurysms (MAs), minute red circular markings on the surface of the retina, represent some earliest signs of DR in human eye [2]. These spots typically range in diameter from 10 μ m to 100 μ m [3] and are small outgrowths of capillaries, sometimes forming owing to the seepage from minuscule retinal blood vessels. MAs persist as the sole lesions present in the earliest stages of the disease and persist until DR develops [4]. Therefore, the early identification of MAs plays a crucial role in detecting and treating DR.

DR is a common eye disorder found in peoples with diabetes, frequently prompting harm to the blood vessels within the retina. Symptoms of DR, including MAs, HE, AND HM, are observable through color fundus retinal imaging as indicated by various scientific studies [1]. The progression of vision impairment due to DR usually occurs without preliminary symptoms from the patients. However, timely identification and intervention for DR can mitigate the risk of vision impairment. Exudates, characterized by white or pale yellow patches in the retina, result from the loss of proteins in small retinal veins. Hemorrhages, resembling irregular red spots with non-uniform borders, lead to leakage from fragile and thin blood vessels in the retina. Fig.1 displays examples of these specific lesions. Arrows are used to indicate the locations of the MAs within the image

Ophthalmologists typically use fundoscopy or specialized fundus imaging, captured with specific cameras, for detecting MAs. However, manual screening poses several challenges. The scarcity of ophthalmologists in remote areas, the labor-intensive and time-consuming nature of manual screening, and the potential for errors make a compelling case for a CAD system. This system plays a vital role in precisely identifying and categorizing microaneurysms in fundus images. Distinguishing them from elements such as small circular spots at vessel intersections, noise, and distortions poses a significant challenge. Additionally, irregularly shaped MAs, clustered MAs, those near the image's periphery, or within the macula are also challenging to detect. However, the advancement of automated technology has introduced various ML and DL methods for DR detection, showing notable efficacy and performance.

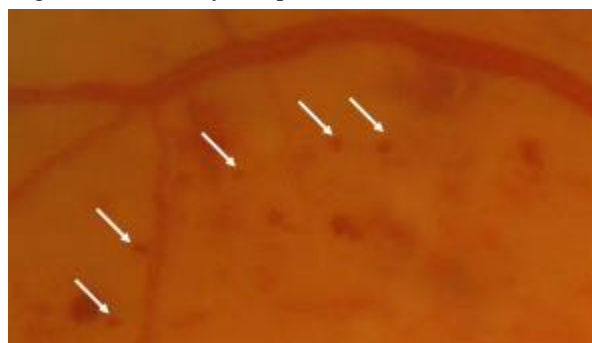


Fig.1 Sample of microaneurysms (MAs).

Employing automated methods for Diabetic Retinopathy (DR) identification not only extends the benefits of early diagnosis to a larger populace but also brings about cost savings, time efficiency, and enhanced diagnostic accuracy. Recent studies underscore the considerable utility of CAD methods in medical image processing [5]. Within this domain, sophisticated ML-driven methodology has been emerged, primarily focusing on the automatic segmentation and classification of retina images. These methods work on processing images derived from retinal scans, identifying regions affected by the disease, and determining the disease stage. This approach streamlines the task for ophthalmologists, enabling them to concentrate directly on targeted regions and implement treatments to combat the condition.

Machine learning methods are employed to discern and leverage retina-specific characteristics, including processes like optic disk identification, vessel augmentation, and lesion partitioning from the original input image. Following this, image categorization techniques, including KNN, SVM, and Naive Bayes [6]–[8], are applied to categorize the images. However, these conventional computational approaches depend on particular manually designed features to discern structures within the images. Consequently, they may lack the capability to grasp the intricate structure of patterns and often struggle to model the underlying structure of abnormalities. This limitation restricts their applicability in clinical contexts.

In recent times, there have been significant advancements in medical image processing methods, particularly in the domain of DL methods, notably Convolutional Neural Networks (CNN) [9]–[15]. A multitude of studies has emerged focusing on the classification of DR grades. For instance, K. Xu et al. [16] implemented a CNN-driven approach for categorizing images into normal and DR classifications into normal and DR categories. Their approach involved preprocessing steps such as data augmentation, image resizing, and normalization. Their model architecture comprised 8 convolutional stages, 4 max-pooling stages, 2 fully connected stages, and a Softmax function in the final stage, resulting in binary classification. Similarly, Li et al. [13] employed a Deep CNN (DCNN) for the classification of DR images. They employed fractional max-pooling to enhance the extraction of discriminative elements from the input data, followed by the classification of these elements using a Support Vector Machine. In another study, R. Pires et al. [17] introduced a sixteen-stage CNN model for categorizing DR images into referable and non-referable classes. To mitigate overfitting during the training process, they integrated dropout and L2 regularization methods.

The main aim is to establish an extensive CAD system tailored for widespread DR screening in color fundus images. This work focuses on developing a fully automated algorithm that combines both autonomous and guided learning approaches to accurately identify crimson abnormalities, particularly Micro aneurysms (MAs), within ocular images, aiming for heightened responsiveness, PPV, and specificity while minimizing spurious outcomes. The proposed CAD system integrates Vgg-16 with an SPP layer to discern texture characteristics from preprocessed fundus images. These output characteristics aid in distinguishing between healthy and DR cases. Additionally, the system employs RAS-UNet DL to visually segment abnormalities within DR case

images. Furthermore, the CNN-MGCA model is utilized for diagnosing micro aneurysms (MAs) from non-MA candidates. In the validation phase, different performance measures are discussed to validate the effectiveness of the introduced CAD mechanism. The important contributions of this research work are

Development of a fully automated method that combines unsupervised and supervised learning techniques to achieve accurate detection of red lesions, with a specific focus on Microaneurysms (MAs), within retinal images

The proposed CAD system integrates Vgg-16 with an SPP layer to extract essential texture features from preprocessed fundus images. These extracted features play a crucial role in distinguishing between healthy and DR cases.

The system employs the innovative RAS-UNet deep learning model for the visual segmentation of lesions within DR case images.

The novel CNN-MGCA deep learning model is employed to specifically diagnose microaneurysms (MAs) from non-MA candidates, aiming to achieve high sensitivity and specificity while minimizing false responses.

The proposed CAD system was validated by comparing it with other systems. Additionally, 6 different performance metrics were employed for evaluation...

The subsequent sections of the article are structured to furnish a thorough comprehension of the research. In Section II, the background and relevant literature reviews are presented to establish the context and highlight existing knowledge in the field. Following this, Section III offers an overview of the methodology employed in the analysis, outlining the approach and methods used in the research. The specifics of the conducted experiments are discussed in Section IV. Finally, Section VI encapsulates the study with a summary.

2. RELATED WORKS

Recent researchers have focused on the classification of DR stages. In this particular section, an overview of recent research endeavours is presented, delving into their respective proposed methodologies for the classification of DR stages.

Srinivasan et al. [18] suggested the MSA methodology into the ResNetGB model to enhance the classification of DR grades. The initial step involves utilizing the encoder network to convert the retinal fundus (RF) image into a higher-stages interpretational space, amalgamating mid and high-stages characteristics to improve representation. Following this, the introduction of a Multi-Scale Feature Pyramid (MSFP) facilitates the capture of retinal patterns across various regions. The MSA strategy is subsequently applied to this elevated interpretation, resulting in a more robust and efficient classification process.

Gadekallu et al. [19], an integration of a deep neural network (DNN) model with PCA and the firefly procedure is employed for the identification of DR, leveraging a dataset sourced from the UCI machine learning database. To optimize the repository and refine the model's performance, irrelevant attributes present in the publicly available data are systematically eliminated through PCA, facilitating efficient feature extraction. Moreover, the firefly procedure is applied for additional dimensionality reduction. The outcome decreased dataset is subsequently input

into the DNN, contributing to an improved and more proficient identification of DR.

Yi et al. [20] employs a multi-step strategy, including data preprocessing, the introduction of an RA-EfficientNet network architecture, and the development of specialized classifiers. This network incorporates a Residual Attention (RA) block, enhancing the model's ability to discern subtle variations in lesions. These efforts collectively contribute to advancing the accuracy and effectiveness of diabetic retinopathy classification, providing valuable insights for medical practitioners

Yadav et al. [21] Introduce machine learning categorization tailored for the efficient detection of microaneurysms (MAs) without imposing significant computational demands. The methodology involves the segmentation of retinal images through histogram-based techniques. Subsequently, features are discerned from these partitioned pictures, encompassing statistical, structure and gray-level co-occurrence matrix (GLCM) methods. Post feature extraction, a range of classification algorithms, including Logistic Regression, KNN, NB, Kernel SVM, DT, SVM and RF, are implemented to discern between MAs and non-MAs.

Mateen et al. [22] Underscores the significance of precise previous identification of microaneurysms (MAs) in the treatment of DR to prevent irreversible blindness. The presented approach utilizes a integrated characteristic representation technique, incorporating pre-learned CNN models, specifically Inception-v3 and VGG-19. This methodology's effectiveness was assessed using two publicly accessible repository namely "E-Ophtha" and "DIARETDB1," achieving significant categorisation accuracy of 95% and 93%, respectively.

Raudonis et al. [23] introduces a novel model for the automated identification of MAs in color retinal fundus images. The approach involves three major stages: image decomposition into smaller patches, leveraging segmentation models for inference, and reconstructing the partitioning map from the output patches. The partitioning method employs a combination of three unique deep networks: UNet++, ResNet34-UNet and U-Net. The assessment of the method's performance involves calculating Dice scores and Intersection over Union values.

Liu et al. [24] introduced an innovative approach for detecting MAs and hard exudates (HEs) in diabetic retinopathy (DR) using a deep symmetric CNN. The adoption of a symmetric convolutional shape aims to enhance the efficiency of feature extraction. Furthermore, the technique tackles the issue of imbalanced positive/ negative samples by augmenting the size of the network, thereby reducing the potential for overfitting.

Hervella et al. [25], The deep ResNetGB model integrates the MSA scheme to improve the identification of DR grades. The process initiates with an encoder network that embeds retinal fundus (RF) images into a high-level interpretational space, amalgamating high and Middle-level characteristics to improve representation. Subsequently, where the MSFP technique is introduced to capture retinal patterns across various regions, and the MSA scheme is employed to this high-level explanation. The complete MSAResNetGB model is then trained using cross-entropy loss to classify patients into their respective DR grades.

Das et al. [26] presents an automated diagnostic approach for diabetic retinopathy utilizing deep learning(DL) techniques. The training data is collected from an openly available retina dataset. A new CNN model, denoted as "AD2Net," is developed, drawing on the strengths of Res2Net andDenseNet, and integrating an attention mechanism to enhance overall model performance. The input images undergo processing through this network to learn distinctive features at various disease stages. Consequently, the network classifies retinal fundus images into 5 disease stages: Normal, Moderate DR, Proliferative DR, Severe DR, and Mild DR, based on their severity.

Harshitha et al.[[27], an automated system for detecting DR is introduced, which efficiently organizes images into distinct levels based on disease-related characteristics. The proposed methodology utilizes a "Convolutional Neural Network (CNN)," where input images undergo convolution through a weighted matrix to extract essential image coordinates while preserving the original information (referred to as "Feature extraction"). Initially, a comparison of various CNN models is conducted to select the most suitable network model for the primary clustering task, with the goal of achieving optimal results.

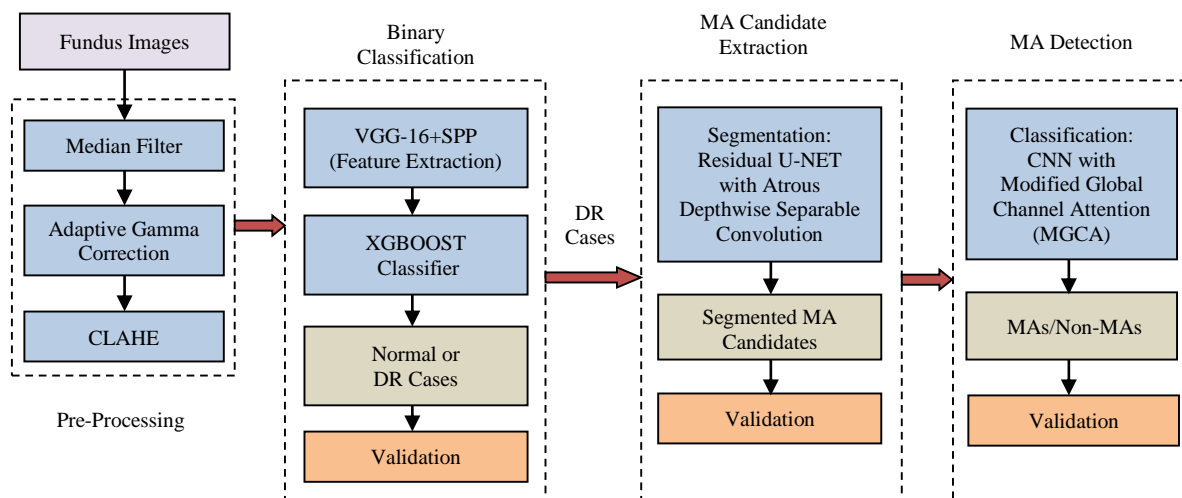


Fig.2. Block diagram outlining the introduced work

3. METHODOLOGY

The Fig.2 illustrates the proposed framework for MAs detection. The method for the proposed computer-aided diagnosis (CAD) system, which aims to identify diabetic retinopathy (DR) and recognize microaneurysms (MAs) in retinal images using deep learning techniques. It begins with data collection and preprocessing, focusing on noise reduction, quality enhancement, and standardizing image sizes. Using a pre-trained VGG-16 with a SPP layer, features are extracted and fed into a lightweight extreme Gradient Boosting (XGBoost) model for DR classification. A Residual U-NET incorporating atrous depthwise separable convolution (RAS-UNet) is employed to identify potential MAs by extracting multi-scale features. Subsequently, a Convolutional Neural Network with modified global channel attention (CNN-MGCA) is developed to classify MAs.

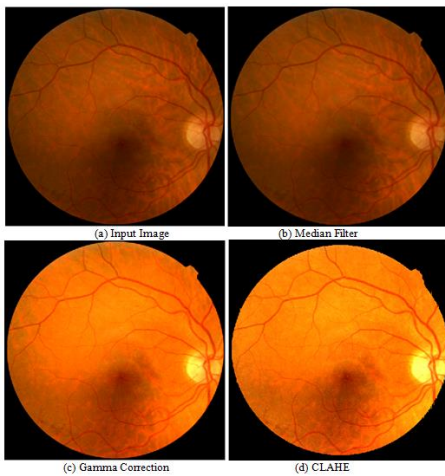


Fig.3. Output of preprocessing step

3.1 PREPROCESSING

Retina fundus images, often obtained from various clinical settings and diverse imaging devices, display significant intensity variations due to their diverse sources. To enhance the efficiency of the model training, the pre-processing stage was executed in this method. Initially, the input retinal images were resized to dimensions of 224×224 , maintaining their aspect ratio through the utilization of bicubic interpolation. Subsequently, a median filter is applied to alleviate noise, commonly encountered during image acquisition. Adaptive gamma correction is then utilized to address issues related to uneven lighting within the images, adjusting excessive or inadequate gray levels while augmenting contrast. Moreover, the Contrast-Limited Adaptive Histogram Equalization (CLAHE) technique is utilized, involving the conversion of the RGB to the Y-Cr-Cb color space. This approach employs AHE in brightness to improve the effects of uneven grayscale values, specifically in luminosity, ensuring a more standardized image quality for further analysis (See Fig.3). Upon completing the preprocessing phase, the enhanced RGB image is used for extraction of characteristics and binary categorization, utilizing the preloaded VGG-16 model in combination with SPP and lightweight XGBoost techniques.

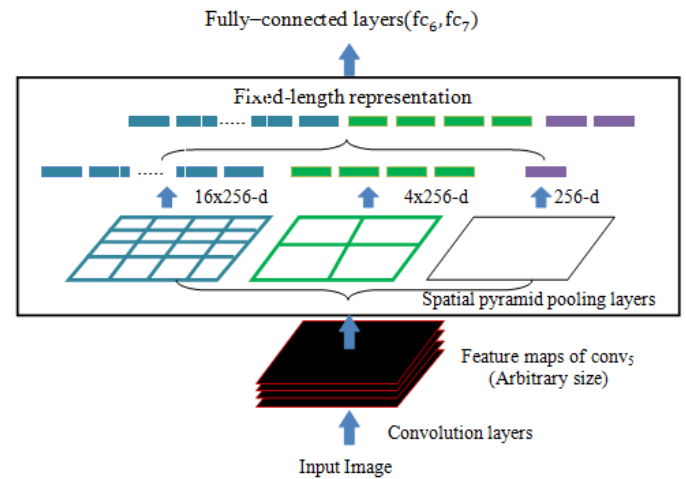


Fig.4. A network architecture incorporating a SPP layer. Here, 256 corresponds to the number of filters in the conv5 layer, which functions as the ultimate convolutional layer

3.2 FEATURE EXTRACTION

The VGG-16 model, a CNN preloaded on a large image dataset recognition task, is employed as the foundation. To adapt it for DR classification, a Spatial pyramid pooling (SPP) layer is integrated after the final convolutional layer. This modification defines the network to handle input retinal images of various sizes, addressing both arbitrary aspect ratios and scales. The SPP layer divides the elements maps into spatial bins and performs max pooling within each bin. This process generates fixed-dimensional vectors, regardless of the input image size, providing a robust representation of spatial details. The count of bins in the SPP layer is determined by the design choice, influencing the granularity of spatial information captured.

During the feature extraction phase in the proposed framework, VGG-16 with a SPP was employed to capture elements from the green channel of all processed datasets. The proposed model benefits from the integration of VGG16 [40] and SPP [41] through a stacking technique. The VGG16 model, as referenced in [40], is configured to accept an RGB image with dimensions of 224×224 as its input. The image undergoes a sequence of convolutional (conv) layers employing 3×3 receptive field filters, followed by a block containing 3 fully connected layers. These convolutional layers possess the ability to handle inputs of different sizes by sliding the input with a set of kernels, generating an output feature map represented as $V \in R^{(a \times d)}$.

$$V = f_{\text{VGG16}}(I) \quad (1)$$

where $f_{\text{VGG16}}(\cdot)$ represents the VGG16 network, as detailed in [40], performing a sequence of convolutions and pooling to compute the elements map while retaining the responses of the receptive fields in spatial domains. However, the fully connected layers within VGG16 require a fixed-length vector, imposing a limitation on the model for a specific input size. This necessity for a fixed size input during both training and testing phases is primarily owing to the utilization of these dense layers within the VGG16 architecture.

The SPP layer is designed to transform the variable-sized feature map into a fixed-size dimensional vector. Fig.4 illustrates

spatial pyramid pooling layer. Within each spatial bin, we aggregate the activations of every filter, consistently employing maximum pooling in our approach. The results from the spatial pyramid pooling step yield vectors of dN dimensions, where N represents the count of bins. These fixed-dimensional vectors serve as inputs to the dense layers.

The SPP achieves this transformation by applying max-pooling to the input feature map using windows and strides of arbitrary sizes. The pooling process is conducted over pyramid levels with $p \times p$ bins, treating each level as a pooling stage. The window size is denoted by W_s , and the stride by S .

$$W_s = \left\lceil \frac{a}{p} \right\rceil; \quad S = \left\lfloor \frac{a}{p} \right\rfloor \quad (2)$$

The features from all levels of the pyramids are combined through concatenation, resulting in a fixed-sized vector u with dimensions 1024. This flexible approach allows the SPP layer to effectively capture and summarize features across different spatial scales and locations within the input feature map. Consideration of these scales is crucial for enhancing the accuracy of deep networks. The Fig.5 depicts the detailed architecture of the VGG16-SPP model

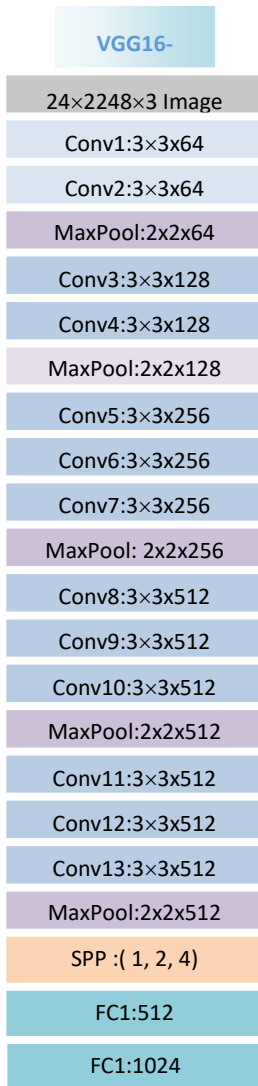


Fig.5. Proposed VGG16-SPP model architecture

The modified VGG-16 with the SPP layer is employed as a feature extractor. Input images, regardless of their sizes, undergo feature extraction through the network. The output from the SPP layer is a compact, informative representation of spatial features, encapsulating details at various scales. The features extracted by the VGG-16 with SPP are then used as input for an Extreme Gradient Boosting (XGBoost) model. XGBoost stands out as a robust and effective machine learning algorithm, particularly adept at handling classification tasks.

3.3 DR CLASSIFICATION

XGBoost stands out as a robust and effective machine learning algorithm that excels in classification tasks. It leverages the ensemble learning technique, combining the outputs of multiple weak learners (decision trees) to create a robust and accurate classifier. The XGBoost model is trained on a labeled dataset of retinal images, where each image is associated with a Diabetic Retinopathy severity class. During training, XGBoost learns to make decisions based on the spatial features provided by the pre-trained VGG-16 with SPP.

The XGBoost utilizes gradient boosting and decision trees to capture the correlation between input features and the target variable. Through an iterative process, it constructs an ensemble of decision trees, utilizing gradient descent to minimize the loss function. In the context of accessing features from VGG16-SPP with p samples and n features, the streamlined XGBoost is depicted below.

$$f(n) = \sum_{k=1}^K f_m(y) \quad (3)$$

The predictions for new sample labels are generated by $f(n)$, where $f_m(n)$ represents the cumulative output of M regression trees from 1 to m . Each individual regression tree is trained to minimize L , a loss function, which quantifies the variance between anticipated and real values. The loss function is adaptable and can be expressed as any differentiable function, like mean squared error (MSE) applied to regression tasks or binary output for classification scenarios.

XGBoost identifies the regression trees $\{f_m\}$ with m ranging from 1 to M by minimizing the specified objective function

$$l(\{f_m\}_m) = \frac{1}{M} \sum_{i=1}^n L(X_i, F(n_i)) + \sum_{k=1}^k \Omega(f_m) \quad (4)$$

Within this function, $\Omega(f_m)$ serves as a regularization term designed to penalize intricate trees, mitigating the risk of overfitting. The regularization term is commonly crafted to discourage the emergence of overly complex trees, playing a pivotal role in enhancing the model's capacity for effective generalization.

$$\Omega(f_{mk}) = \gamma T + \frac{1}{2} \mu \sum_{j=1}^T W_j^2 \quad (5)$$

where T represents the leaf count of the tree, w_j denotes the weight of the j^{th} leaf, and γ and λ are hyperparameters governing the strength of regularization. The training process for the trees involves the XGBoost algorithm, which employs a gradient-boosting strategy. At every iteration t a new tree f_t is incorporated into the ensemble by adjusting it to the negative gradient of the current predictions' loss function. This iterative approach enables

the model to iteratively refine its predictions and enhance its overall performance.

$$\frac{l(X_i, f_{i-1}(n_i))}{f_{i-1}(n_i)} \quad (6)$$

The weight controlling the learning rate is subsequently implemented to the ensemble, concomitant with the incorporation of the new tree. This step contributes to the gradual refinement of the model, allowing for a balanced and controlled learning process.

$$F_i(x) = F_{i-1}(m) + \theta \times f_i(n) \quad (7)$$

The proposed lightweight algorithm iterates through this procedure until the loss function converges or a predefined max number of trees K is attained. The ultimate prediction is calculated as the total of projections derived from each individual tree within the ensemble. This iterative process ensures the model's convergence and produces a robust combined prediction from the ensemble.

By integrating a pre-trained VGG-16 with SPP and an XGBoost model, this approach leverages the strengths of DL framework for extracting features and the robustness of gradient boosting for classification, resulting in a powerful system for Diabetic Retinopathy diagnosis.

3.4 RESIDUAL U-NET INCORPORATING ATROUS DEPTHWISE SEPARABLE CONVOLUTION (RAS-UNET) FOR SEGMENTATION

Atrous convolution: Atrous convolution serves as a potent technique, enabling precise control over feature resolution within deep CNN. It provides the capability to tailor the FOV, allowing the extraction of multi-scale information. This operation, an extension of the conventional convolution, permits explicit adjustments to the FOV. In the context of 2D signals, the application of atrous convolution (conv) on the input element map

y at every location j on the output element map x involves convolving with a filter w :

$$X[j] = \sum_k y[j + r \cdot k]W[k] \quad (8)$$

The rater in atrous convolution establishes the stride at which the input signal is sampled. It is essential to note that standard convolution represents a specific case where $r=1$ [42]. The adaptability of the filter's field-of-view (FOV) is achieved through the modification of the rate value.

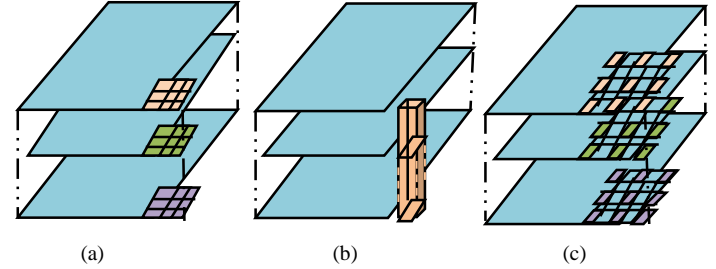


Fig.6. 3x3 DSC decomposes a regular convolution into (a) DC (b) PC (c) AC

The depthwise separable convolution (DC): This method that breaks down a regular convolution into a depthwise convolution, followed by a pointwise convolution (1×1 convolution), significantly reducing computational complexity. The depthwise convolution (DC) performs spatial convolutions separately for each input channel, and the pointwise convolution (PC) combines the outputs from the depthwise convolution. Atrous convolution (AC) is integrated within the depthwise convolution, illustrated in Fig.6. In this study, we term the resultant convolution as atrous separable convolution. Importantly, atrous separable convolution markedly reduces the computational demands of the proposed model while maintaining performance that is comparable or even improved.

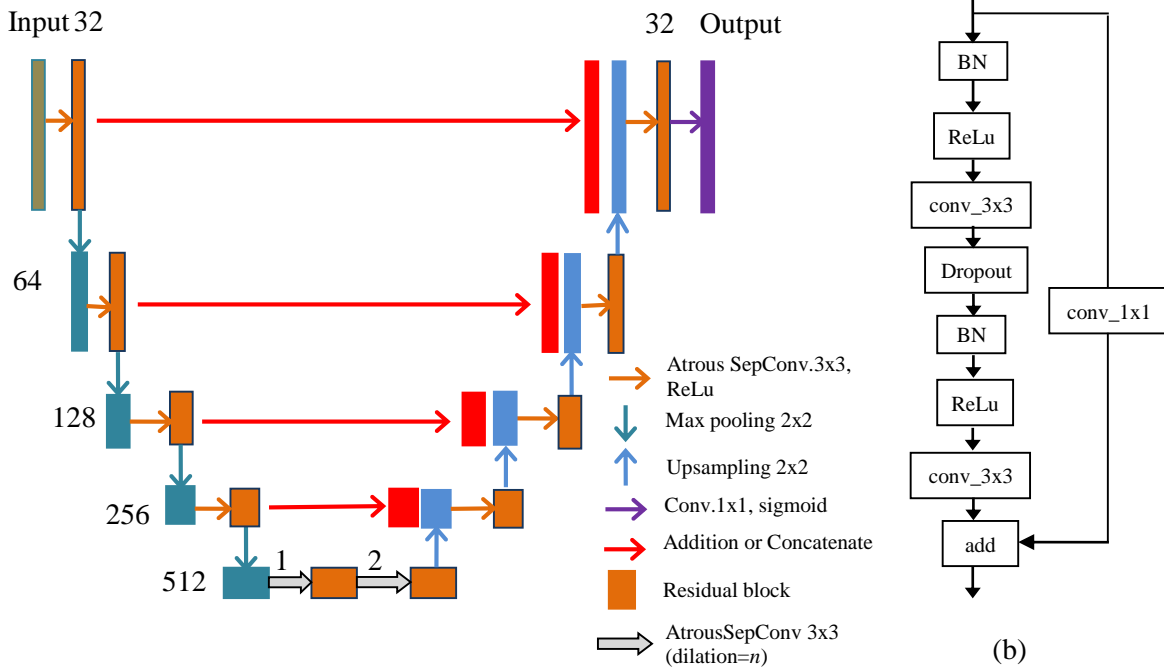


Fig.7. (a) RAS-UNet Architecture. (b) Residual block.

3.5 RESIDUAL U-NET

The Residual U-Net (RU-Net) is derived by integrating a residual block into the adapted U-Net structure, featuring a single downsampling path and 3upsampling paths. By incorporating residual units and skip connections, it simplifies training, facilitates information propagation, and allows for the design of a more efficient network with fewer attributes while attaining outstanding achievement in semantic classification tasks.

U-Net [43], introduced for biomedical image segmentation, comprises a downsampling and symmetrical upsampling path. Downsampling extracts features, while upsampling facilitates image expansion. Output from the upsampling path is fused with corresponding downsampling layers, combining additional dimensions and compensating for information loss. Inspired by this, a customized UNet with residual blocks ([48], [49], [45]) enhances segmentation tasks, leveraging insights from the specialized residual network [44]. Empirical evidence supports improved performance with the integration of residual blocks in U-Net methods.

Interested by this, we enhance U-Net by substituting its standard convolutional layers with residual blocks, resulting in a ResU-Net[45]. The ResU-Net encompasses a network is structured into 3 units: downsampling block, bridge block, and upsampling block. The downsampling block transforms the input fundus image into streamlined features, while the upsampling module reconstructs these features for pixel-wise categorization (semantic segmentation). The bridge component connects the encoding (downsampling) and decoding(upsampling) modules. All three segments utilize residual units, featuring 2 (3×3) convolution blocks and an identity mapping. Every convolution block consists of a Batch Normalization (BN) layer, a ReLU activation layer, and a convolutional layer. The encoding (downsampling) module, composed of three residual units, utilizes a stride of two in the initial convolution block for downsampling, foregoing pooling. Correspondingly, the decoding (upsampling) module, also with three residual units, involves upsampling of feature maps from lower levels and concatenation with corresponding encoding (downsampling) module feature maps. In the final stage of the decoding path, a 1×1 convolution and a sigmoid activation layer are utilized to convert multi-channel feature maps into the desired segmentation. The network comprises a total of 15 convolutional layers, presenting a more streamlined architecture in contrast to the 23 layers found in U-Net.

The residual unit, depicted in Fig.7(b), encompasses a shortcut connection and several stacked layers, including conv layers, ReLU layers, and BN layers [46]. Additionally, to prevent overfitting, a dropout [47] layer is inserted between the two convolutional layers within the residual unit. The structure of the residual unit can be succinctly expressed as follows:

$$Y_{\text{drop}}(i) = P(F(M[N(Y(i))])) \quad (9)$$

$$X(i+1) = F(M[N(Y_{\text{drop}}(i))]) + I(Y(i)) \quad (10)$$

Let $Y(i)$ and $Y(i+1)$ represent the input and output of the i^{th} residual block, respectively. The output of the dropout layer within the i^{th} residual block is denoted as $Y_{\text{drop}}(i)$. $N(\cdot)$ represents the BN function, $M(\bullet)$ is the activation function, $P(\cdot)$ denotes the

dropout operation, and $F(\cdot)$ corresponds to the convolutional operation. Specifically, the function $F(\cdot)$ represents an identity mapping.

The illustrated RAS-UNet in Fig.7(a) is designed with 10 residual modules and utilizes a sigmoid activation function in its final layer to produce a classification possibility distribution map. In contrast to ResU-Net, our method incorporates a dropout layer among the 2 atrous depthwise separable convolutional layers within the residual module, successfully addressing the issue of overfitting. The proposed architecture comprises four main components: an encoding module, a decoding module, a residual block, and an atrous convolution block. The encoding module, inspired by a convolutional network, involves iteratively applying 2 3×3 atrousSepConvs. Following each atrousSepConv operation, a 2×2 max-pooling operation is employed for downsampling, and the Rectified Linear Unit (ReLU) serves as the activation function. BN is incorporated for improved training. Within the decoding module, each step encompasses three pivotal processes: initially, an up-sampling operation doubles the size of each feature map; secondly, the element maps concatenate with corresponding ones from the encoding block; thirdly, two 3×3 atrousSepConv operations are executed preceding and succeeding the up-sampling operation, each is followed by a ReLU activation. The outputs from the up-sampling paths are amalgamated using predefined weight values, and this fused output is subsequently input into a residual module, a fully connected layer, and a sigmoid function to derive the final output. Each residual block in the up-sampling path is introduced with a concatenation layer that merges the feature channels. Specifically, this concatenation layer combines the output of the residual block from the preceding up-sampling path with the output of the previous step in the same up-sampling path. This strategic feature fusion across different up-sampling paths enhances the extraction of local features, contributing to a more precise performance for RAS-UNet. In a formal expression, the output of the a^{th} residual block in the final up-sampling can be represented as

$$Y_{(u_3,a)} = K\left(Y_{(d,b)}, Y_{(l_2,b)}, I\left(Y_{(l_3,b)} + 1\right)\right) \quad (11)$$

where, the index b denotes the b^{th} layer, where d and l signify the downsampling and upsampling paths, respectively. Additionally, 11, 12, and 13 correspond to the three upsampling paths. The function $I(\cdot)$ represents an upsampling operation, $K(\cdot)$ signifies residual learning, and Y denotes the output of a layer. Within the downsampling path, an atrous convolution operation is applied, followed by a max-pooling operation in U-Net, aiming to preserve intricate details. Following downsampling, the quantity of feature maps expands to 512.

To enhance network parameters and computational efficiency, we substituted the traditional convolution operation in the RU-Net structure with SepConv. Throughout the training process, we utilized Cross Entropy as the loss function, defined as:

$$L = -\frac{1}{N} \sum_1^N [x_a \log p_a + (1-x_a) \log(1-p_a)] \quad (12)$$

This adjustment is executed to diminish configuration settings and execution cost, thereby boosting the overall efficiency of the RAS-UNet architecture..

3.6 CNN-MGCA (MODIFIED GLOBAL CHANNEL ATTENTION) FORMA DETECTION

The attention mechanism dynamically adjusts higher-order abstract properties uncovered by the model, enhancing efficiency in computer vision applications. The widely used Squeeze-and-Excitation (SE) structure, developed by Hu et al. [50] and integrated into architectures such as MobileNet-v3 [51] and EfficientNet [52], offers flexibility and notable performance improvements, especially for Convolutional Neural Networks (CNN). While the SE-based inter-channel attention mechanism considers feature matrix channel correlations, the fully connected layers within the SE structure substantially increase model parameters. Although the bottleneck structure in SE reduces parameters through dimensionality reduction, it results in the loss of some feature information, introducing limitations in overall model performance.

The Efficient Channel Attention (ECA) mechanism, as introduced by Wang et al. [53], achieves a nuanced equilibrium among model efficiency and complication when juxtaposed with the original Squeeze-and-Excitation (SE) architecture. It employs a method of adaptively adjusting convolution kernel sizes to proficiently uncover correlations among different channels within the feature matrix. Much like channel convolution [52], the ECA structure captures intrinsic correlations between channels in the feature map through a 1D convolution with a flexible kernel size in the CA mechanism. This strategy substantially streamlines model complexity in contrast to the SE framework. While the ECA framework adeptly manages parametric increments in the model, acting as a potent inter-channel attention mechanism, its 1D convolution constrains each channel's weight to be exclusively associated with a fixed number of neighboring channels. This limitation overlooks the correlations among global channel features within the feature map.

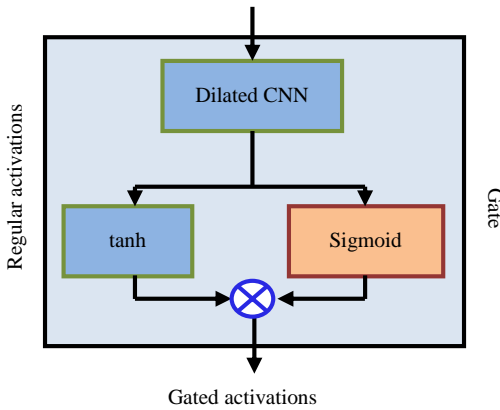


Fig.8. Gated convolution

3.7 GATED CONVOLUTION

Gated convolution is implemented by regulating the information flow within every layer through the application of the hyperbolic tangent nonlinearity to the output of dilated convolution, followed by attenuation using sigmoid gates (illustrated in Fig.8). While the sigmoid function is commonly acknowledged as an activation function, in this particular context, its purpose is to serve as a gating mechanism rather than fulfilling the role of a traditional activation function. The sigmoid

function's value range, spanning from 0 to 1, renders it well-suited as an attenuation factor for the activation of dilated convolution. This characteristic enables precise control over the information flow within each hidden layer.

In this mechanism, the hyperbolic tangent operation serves as a regular activation operation. Together, these two functions form what is known as gated activation. The learnable convolution filters, denoted as Z , are split into two parts: Z_f for the filter and Z_g for the gate. The final activation of layer k , p_k , is determined by the following equation:

$$p_k = \tanh(Z_f * X_k) \odot \sigma(Z_g * X_k) \quad (13)$$

where, W_f and W_g represent the filter and gate components of the learnable convolution filters, X_k is the input to layer k , \odot denotes element-wise multiplication, the hyperbolic tangent activation, denoted as \tanh , serves as the activation operation, while σ functions as a sigmoid activation, operating as a gating mechanism.

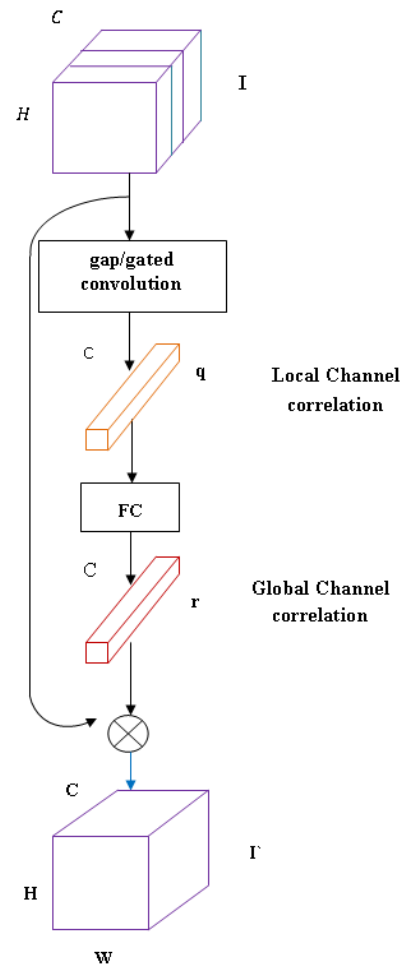


Fig.9. Schematic diagram of MGCA structure

In this study, developed a MGCA model addresses the limitation observed in the ECA structure, which focuses solely on local channel correlations. In contrast, MGCA considers correlations among all element map channels while keeping the count of model factors constant. This enhancement significantly refines the perceptual improvement of the model across diverse channels. The MGCA architecture is depicted in Fig.9.

Given the input feature matrix $I \in \mathbb{R}^{(H \times W \times C)}$, for the Modified Global Channel Attention (MGCA) structure, the elements of every channel are acquired through a gap operation. This operation requires computing the average value for all channels across the entire spatial dimensions. The resulting feature vector for each channel is then used to capture the global information.

$$p = \text{gap}(I) \tag{14}$$

The features capturing correlation among neighboring channels q are derived from the extracted features through a gated convolution operation. This operation is defined by Eq.(14)-Eq.(15). Here $p \in \mathbb{R}^{(C \times C)}$ represents the variable matrix of the gated convolution, and $\text{swish}(\cdot)$ is the activation function. The swish activation function is characterized by its non-linearity and smoothness, enhancing the model's ability to capture complex relationships within the data. The parameter matrix $M_1 \in \mathbb{R}^{(C \times C)}$ is determined based on the specific convolutional operation employed, contributing to the adaptability of the network in capturing local inter-channel correlations.

$$q = \text{swish}(M_1 p), \quad q \in \mathbb{R}^c \tag{15}$$

$$\text{swish}(x) = x \cdot \sigma(x) \tag{16}$$

To dynamically adjust the convolution kernel size based on different element maps, enhancing adaptability. In pursuit of greater spatial cross-channel, the relationship between the count of channels (C) in the element map and the convolution kernel size (k) is established as $C \propto k$. To formalize this relationship, the paper proposes a gated mapping represented by $C = \eta \cdot k + \beta$, where η and β are parameters learned during the network training process. Consequently, this adaptive design optimizes the convolutional kernel size k through gated convolution, allowing it to be tailored to the varying requirements of different feature maps and fostering improved capturing of local inter-channel correlations can be expressed as:

$$k = g_{\gamma, \beta}^{-1}(c) = \frac{\ln(c) - \beta}{\eta} \tag{17}$$

Consequently, the extraction of local inter-channel correlation features using a window length of k yields distinctive individual features. Each element can be represented as:

$$q_i = \text{swish} \left(\sum_{j=1}^k p_i(k) m_{(i,j)} \right), \quad i \in (1, c) \tag{18}$$

where, $p_i(k)$ signifies the k -channel features neighboring p_i .

After extracting the local inter-channel correlation feature $q \in \mathbb{R}^c$, obtaining the global channel correlation (GCC) feature r from q requires applying a subsequent global linear operation on q .

$$r = \sigma(M_2 q), \quad r \in \mathbb{R}^c \tag{20}$$

The matrix M_2 , representing the linear operation, is symbolized as $WM_2 \in \mathbb{R}^{(C \times C)}$, where it can be expressed as:

$$M_2 = \begin{pmatrix} m_{1,1} & \cdots & m_{1,c} \\ \vdots & \ddots & \vdots \\ m_{c,1} & \cdots & m_{c,c} \end{pmatrix} \tag{21}$$

Subsequently, the GCC $r = [r_1, r_2, \dots, r_c]$ is derived from the local inter-channel correlation q , where r_i can be represented as:

$$r_i = \sum_{j=1}^c m_{(i,j)} q_j \tag{22}$$

In the final step, the features of global channel correlation (denoted as r) undergo a weighting process with the input feature matrix I along the channel dimension. This process yields the feature matrix I' , which serves as the foundation for the global attention mechanism applied to the feature map. The expression for I' can be articulated as:

$$I' = I \otimes r, \quad I' \in \mathbb{R}^{H \times W \times C} \tag{23}$$

where \otimes represents a multiplication weighting operation applied across the feature map channel dimension.

The MGCA framework, introduced in this study, seizes global channel correlation information within the feature map via a dual-phase procedure. In the first phase, local inter-channel correlation is extracted using gated convolution, employing a constrained set of parameters. Subsequently, in the second phase, the acquired local inter-channel correlation is amalgamated to yield global channel correlation. This bifurcated operation adeptly addresses the issue posed by an extensive array of parameters linked to two fully connected operations, effectively averting overfitting concerns while extracting essential features of global channel correlation.

In the context of microaneurysm detection, after the feature extraction using the CNN-MGCA (Convolutional Neural Networks – Modified Global Channel Attention) architecture, a detection module is trained to differentiate between true microaneurysms and false positives

4. RESULTS AND DISCUSSION

4.1 MATERIAL

To assess and compare the efficacy of the suggested approach in identifying malicious activities in this study, four datasets were employed.

The Messidor dataset [28] consists of 1200 images in TIFF format, which are publicly accessible. These images are losslessly compressed and are come in 3 distinct resolutions: 1440×960, 2240×1488, or 2304×1536 pixels. The capture of these images utilized 8 bits per color plane.

The DIARETDB1 dataset, which is derived from the DIAbetic RETinopathyDataBase, Calibration Level 1 (DIARETDB1) [29], is a freely accessible and widely utilized dataset. It comprises 89 color fundus images in PNG format, each possessing a resolution of 1500×1152 pixels. Within this set, 84 images display non-proliferative indicators of Diabetic Retinopathy (DR) in the form of Microaneurysms (MAs), while 5 are classified as normal. This dataset provides annotations for both MAs and Exs. The images in DIARETDB1 are captured using a 50o Field of View (FOV) of the fundus camera, employing various settings intended to replicate real-life scenarios of image acquisition during mass screenings for DR.

The e-optha database, [30] and [31], comprises two distinct subdatasets identified as e-optha-MA and e-optha-EX. These subsets are specifically tailored to represent the presence of MAs and EXs, individually. For the purposes of this study, we employed e-optha-MA, which includes 148 color fundus images

featuring microaneurysms or small hemorrhages, and 233 images without any lesions, all in JPEG format. The images are available in two dimensions: 2544×1696 and 1440×960. Expert ophthalmologists have meticulously annotated these images.

The IDRiD dataset [32] was utilized for classification and severity assessment. This dataset comprises 81 images in JPEG format, each accompanied by ground truth (GT) annotations for four categories of lesions: Hard EX, MAs, HMs, and soft EXs, provided in TIF format. The images are annotated at the pixel level and were split into 54 for training and 27 for testing. Within this set, EX and HM lesions are observed in 81 different images, with MA present in all 81 images.

4.2 EVALUATION METRIC

The evaluation metrics for binary classification, segmentation, and MA detection, including Accuracy (Acc), Sensitivity (Sen), Specificity (SPE), Positive predictive value (PPV), F1-score (F1) and Area under the curve (AUC), are crucial for assessing the performance of the proposed comprehensive CAD system.

Sensitivity (SEN) characterizes the rate of true positives (TP), while specificity (SPE) denotes the proportion of true negatives (TN). It's crucial to recognize that a method can exhibit accuracy without sensitivity or vice versa. Accuracy (ACC) is the ratio of true results (TP or TN) to the total number of images. False positive (FP) signifies the ratio of erroneously predicted positives, whereas false negative (FN) represents the ratio of incorrectly predicted negatives. Positive Predictive Value (PPV) gauges the correct positive predictions relative to the sum of correct and incorrect positive predictions. The F1-score computes the harmonic average of accuracy and SEN, offering a balanced measure. Lastly, the Area Under the Curve (AUC) approximates half the sum of SEN and SPE. These metrics collectively offer a thorough assessment of the model's achievement in binary classification, segmentation, and MA detection. The corresponding mathematical formulations are outlined in Eq.(24)-Eq.(29).

$$SEN = \frac{TP}{TP + FN} \quad (24)$$

$$SPE = \frac{TN}{TN + FP} \quad (25)$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (26)$$

$$PPV = \frac{TP}{TP + FP} \quad (27)$$

$$F_1 = \frac{2 \times PPV \times SEN}{PPV + SEN} \quad (28)$$

$$AUC \approx 0.5(SEN + SPE) \quad (29)$$

4.3 HEALTHY VS DR CLASSIFICATION RESULTS

In this segment, we utilized the feature vector obtained from the VGG16-SPP architecture for the purpose of binary classification through the application of XGBoost. The objective was to distinguish between healthy and Diabetic Retinopathy (DR) cases. The process involved training multiple models on distinct datasets, specifically IDRiD, MESSIDOR, e-ophtha, and

DIARETDB1, and subsequently selecting the best-trained model for predicting unknown labels.

The training procedure was conducted using XGBoost, a popular machine learning algorithm known for its effectiveness in classification tasks. For each dataset (IDRiD, MESSIDOR, e-ophtha, and DIARETDB1), a separate XGBoost model was trained, utilizing the respective feature vectors obtained from the VGG16-SPP architecture. To ensure robustness and avoid overfitting, a 5-fold cross-validation approach was employed during the training process.

The objective was to identify the model with the highest predictive accuracy, which could then be utilized to make predictions on previously unseen or unknown labels. This comprehensive approach ensures that the chosen model has undergone rigorous evaluation and selection based on multiple datasets, enhancing its generalization capabilities to different datasets and improving its overall reliability in classifying healthy and DR cases.

The Table.3 illustrates the outcomes of the training process, highlighting the performance metrics across various datasets. Notably, the binary dataset within the IDRiD dataset attains the highest rank, followed by the E-OPHTHA dataset in the second position. The DIARETDB1 dataset is positioned third, while the MESSIDOR dataset occupies the last position in the ranking.

Given the outcomes presented in Table.1, a specific XGBoost model was generated. This model underwent training and testing solely on the IDRiD dataset, aiming to predict labels for datasets with undisclosed classifications.

Specifically, images labeled as DR cases or label 1 were identified and progressed to the subsequent phase of analysis. This approach leverages the strengths of the XGBoost model trained on the IDRiD dataset to generalize predictions to other datasets with previously undisclosed labels.

Table.1. Average Performance Metrics for the Proposed Model (Vgg-16+SPP+XGBoost)

Dataset	SEN	SPE	ACC	PPV	F1	AUC
IDRiD	0.95	0.90	0.960	0.953	0.951	0.961
MESSIDOR	0.860	0.750	0.80	0.85	0.854	0.875
E-OPHTHA	0.90	0.90	0.90	0.91	0.929	0.920
DIARETDB1	0.94	0.75	0.865	0.90	0.920	0.950

4.4 SEGMENTATION RESULT

In this section, we showcase the outcomes of segmenting the Microaneurysm (MA) candidates using the RAS-UNet model. To perform the segmentation of MA lesions, the model was trained on the ground truths (GTs) from the IDRiD dataset, yielding weights that can be subsequently loaded for predictions on other datasets. Notably, training the RAS-UNet model with IDRiD data resulted in superior accuracy when compared to training on alternative datasets. Table 2 presents the performance metrics for the proposed RAS-UNet model across all four datasets. The results of the RAS-UNet segmentation on the four datasets are illustrated in Fig.10.

Table.2. Performance Metrics for the Proposed RAS-UNet Model

Dataset	SEN	SPE	ACC	PPV	F1	AUC
IDRiD	0.930	0.947	0.975	0.958	0.955	0.9780
MESSIDOR	0.880	0.905	0.912	0.891	0.88	0.915
E-OPHTHA	0.903	0.912	0.945	0.92	0.939	0.9521
DIARETDB1	0.892	0.903	0.932	0.91	0.915	0.9402

We performed a comparative analysis of average evaluation metrics for U-Net[33], ResU-Net[34], and our proposed RAS-UNet model. Table 3 presents a summary of the comparisons, showcasing the averages of diverse evaluation metrics for the RAS-UNet system we propose. In RAS-UNet model, attained overall averages of 91.6%, 90.1%, 94.6%, 94.5%, 92.2% and 91% for SPE, SEN, AUC, ACC, F1 and PPV individually.

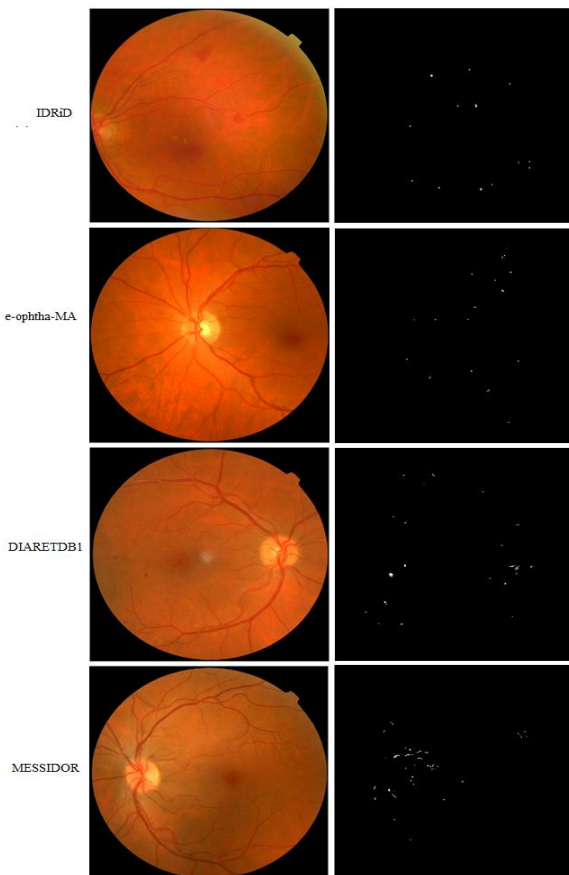


Fig.10. Segmentation outcomes of the RAS-UNet on four datasets

Table.3. Average Performance Measures Comparison

Dataset	SEN	SPE	ACC	PPV	F1	AUC
U-Net [33]	0.860	0.922	0.915	0.958	0.955	0.942
ResU-Net [34]	0.876	0.931	0.923	0.891	0.88	0.958
RAS-UNet	0.901	0.916	0.945	0.91	0.922	0.946

4.5 MA DETECTION RESULT

To evaluate the efficacy of the suggested Microaneurysm identification approach, we conducted a 5-fold cross-validation

independently for each dataset. In this approach, we randomly split each dataset into ten equally sized partitions. In each iteration, one partition served as the test data, and the remaining four were employed to train the classifier. This cross-validation procedure was reiterated five times, producing five distinct evaluation outcomes. These results were then averaged to produce a single estimation, providing a robust evaluation of the MA detection method's performance across different folds and ensuring reliable generalization to the entire dataset. The Table.4 displays the MA detection results across the utilized datasets.

Table.4 displays the average evaluation criteria of the proposed model (CNN-MGCA) on 4 datasets

Dataset	SEN	SPE	ACC	PPV	F1	AUC
IDRiD	0.97	0.968	0.98	0.973	0.974	0.987
MESSIDOR	0.92	0.925	0.952	0.942	0.948	0.965
E-OPHTHA	0.94	0.95	0.965	0.958	0.96	0.975
DIARETDB1	0.93	0.92	0.948	0.945	0.930	0.955

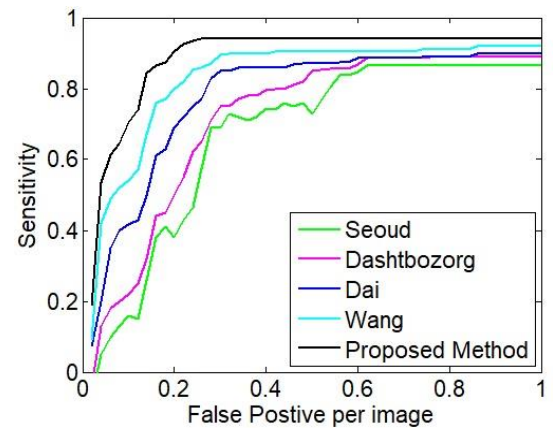


Fig.11. FROC curves for Microaneurysm (MA) detection approaches are compared in a comprehensive study. The suggested approach is evaluated alongside state-of-the-art approaches [36], [37], [39], [38] using the IDRiD dataset.

To appraise the efficiency of Microaneurysm (MA) detection, the assessment employed the free-response operating characteristic (FROC) curve [35]. This approach involves plotting sensitivity against the mean number of false positives per image (FPI). Sensitivity represents the percentage of Microaneurysms accurately identified by the technique, while FPI quantifies the occurrences of non- microaneurysms incorrectly classified as Microaneurysms.

After examining the FROC (Fig.11) curves, the proposed method showcases comparable or superior outcomes when juxtaposed with unsupervised techniques and specific supervised approaches documented in current literature. Fig.12 illustrates instances of Microaneurysm (MA) detection across the four datasets.

We performed a comparative analysis of the mean values associated with diverse evaluation metrics across studies conducted by Seoud et al. [5], Dashtbozorg et al. [23], Dai et al. [30], and Wang et al. [25]. The Table.5 presents a condensed overview of the average comparisons for various evaluation criteria within the proposed CNN-MGCA system.

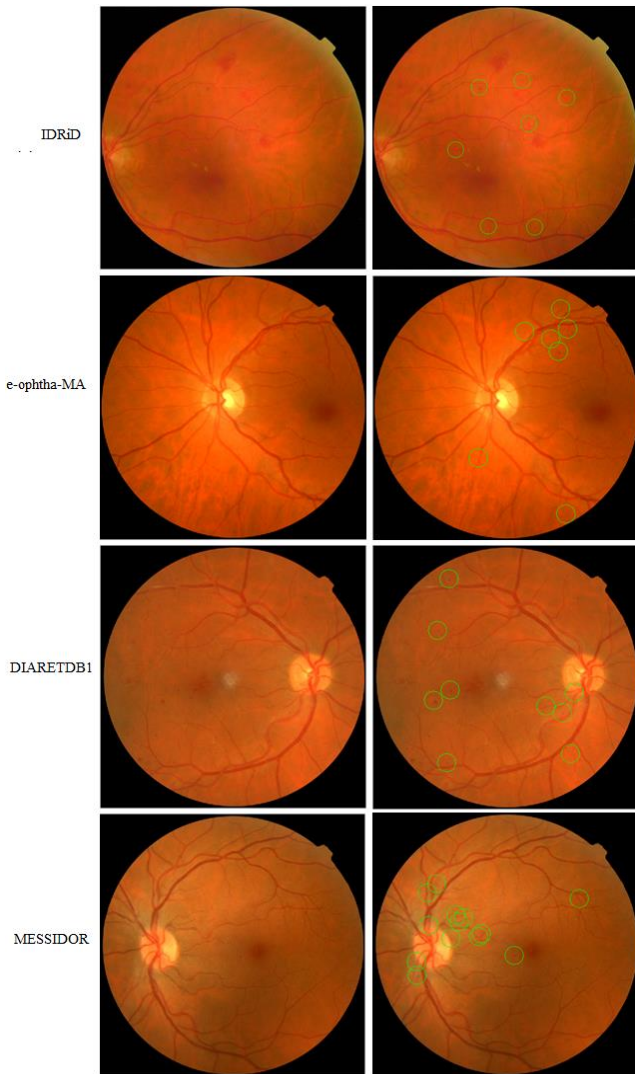


Fig.12. Examples of Microaneurysm (MA) detection on the four datasets

In CNN-MGCA classification, attained overall averages of 94.1%, 94%, 97%, 96.01%, 95.3%, and 95.4% for SPE, SEN, AUC, ACC, F1, and PPV individually.

Table.5. Averaging Performance Measures Across Various Comparisons

Dataset	SEN	SPE	ACC	PPV	F1	AUC
Seoud et al. [5]	0.84	0.825	0.87	0.85	0.855	0.875
Dashtbozorg et al. [23]	0.843	0.867	0.89	0.874	0.886	0.902
Daiet al. [30]	0.87	0.863	0.914	0.89	0.903	0.918
Wang et al. [25]	0.90	0.89	0.935	0.925	0.920	0.943
Proposed	0.94	0.941	0.961	0.954	0.953	0.970

5. CONCLUSION

This research introduces a CAD framework that leverages DL methodologies. The proposed CAD framework addresses key aspects of the diagnostic process. Firstly, it focuses on preprocessing tasks such as noise reduction, quality enhancement,

and standardizing retinal image sizes. Subsequently, the system distinguishes between healthy and diabetic retinopathy (DR) cases by leveraging Vgg-16 with a SPP layer to extract elements from retinal images. These features serve as input for the Extreme Gradient Boosting (XGBoost) algorithm, contributing to the classification task. To identify potential microaneurysms (MAs), the system employs a Residual U-NET incorporating atrous depthwise separable convolution (RAS-UNet) for MA candidate extraction. The RAS-UNet incorporates three crucial components: a downsampling module, an upsampling module, and an atrous convolution module, which integrates both cascade and parallel atrous convolution operations. This design enhances the model's capability to perceive and identify both small and large microaneurysms (MAs). The final step involves classification using a CNN-MGCA to distinguish microaneurysms from non-microaneurysm candidates. The effectiveness of the introduced approach is demonstrated through its comparative performance against state-of-the-art techniques. The capabilities of the developed method are showcased using the IDRiD, MESSIDOR, E-OPHTHA, and DIARETDB1 datasets. Furthermore, the evaluation results across these four public datasets highlight the robustness of the suggested microaneurysms identification method, underscoring its insensitivity to variations in imaging device attributes, graphic precision, and imaging modality.

REFERENCES

- [1] U.R. Acharya, M.R.K. Mookiah, J.E.W. Koh, J.H. Tan, S.V. Bhandary, A.K. Rao, H. Fujita, Y. Hagiwara C.K. Chua, and A. Laude, "Automated Screening System for Retinal Health using Bi-Dimensional Empirical Mode Decomposition and Integrated Index", *Computers in Biology and Medicine*, Vol. 75, pp. 54-62, 2016.
- [2] N. Salamat, M.M.S. Missen and A. Rashid, "Diabetic Retinopathy Techniques in Retinal Images: A Review", *Artificial Intelligence in Medicine*, Vol. 97, pp. 168-188, 2019.
- [3] C. Pereira, D. Veiga, J. Mahdjoub, Z. Guessoum, L. Gonçalves, M. Ferreira and J. Monteiro, "Using a Multi-Agent System Approach for Microaneurysm Detection in Fundus Images", *Artificial Intelligence in Medicine*, Vol. 60, No. 3, pp. 179-188, 2014.
- [4] W. Cao, N. Czarnek, J. Shan and L. Li, "Microaneurysm Detection using Principal Component Analysis and Machine Learning Methods", *IEEE Transactions on NanoBioscience*, Vol. 17, No. 3, pp. 191-198, 2018.
- [5] A. Maier, C. Syben, T. Lasser and C. Riess, "A Gentle Introduction to Deep Learning in Medical Image Processing", *Journal of Medical Physics*, Vol. 29, No. 2, pp. 86-101, 2019.
- [6] A. Pinz, S. Bernogger, P. Datlinger and A. Kruger, "Mapping the Human Retina", *IEEE Transactions on Medical Imaging*, Vol. 17, No. 4, pp. 606-619, 1998.
- [7] M. Al-Antary, M. Hassouna, Y. Arafah and R. Khalifah, "Automated Identification of Diabetic Retinopathy using Pixel-based Segmentation Approach", *Proceedings of International Conference on Watermarking Image Processing*, Vol. 24, pp. 16-20, 2020.

- [8] S.A. Alryalat, M. Al-Antary, Y. Arafa, B. Azad, C. Boldyreff, T. Ghnaimat, N. Al-Antary, S. Alfegi, M. Elfalah and M. Abu-Ameerh, "Deep Learning Prediction of Response to Anti-VEGF among Diabetic Macular Edema Patients: Treatment Response Analyzer System (TRAS)", *Diagnostics*, Vol. 12, No. 2, pp. 1-8, 2022.
- [9] R. Azad, M. Asadi-Aghbolaghi, M. Fathy and S. Escalera, "Bi-Directional ConvLSTM U-Net with Densley Connected Convolutions", *Proceedings of International Conference on Computer Vision Workshop*, Vol. 87, pp. 1-10, 2019.
- [10] R. Azad, M. Asadi-Aghbolaghi, M. Fathy and S. Escalera, "Attention Deeplabv3+: Multi-Level Context Attention Mechanism for Skin Lesion Segmentation", *Proceedings of International Conference on Computer Vision*, Vol. 98, pp. 251-266, 2021.
- [11] Z. Zeng, Y. Xulei, Y. Qiyun, Y. Meng and Z. Le, "SeSe-Net: Selfsupervised Deep Learning for Segmentation", *Pattern Recognition Letters*, Vol. 128, pp. 23-29, 2019.
- [12] M. Bakator and D. Radosav, "Deep Learning and Medical Diagnosis: A Review of Literature", *Multimodal Technologies Interaction*, Vol. 2, No. 3, pp. 1-12, 2018.
- [13] N. Eftekhari, H.R. Pourreza, M. Masoudi, K. Ghiasi-Shirazi and E. Saeedi, "Microaneurysm Detection in Fundus Images using a Two-Step Convolutional Neural Network", *Biomedical Engineering OnLine*, Vol. 18, No. 1, pp. 1-16, 2019.
- [14] Y.H. Li, N.N. Yeh, S.J. Chen and Y.C. Chung, "Computer-Assisted Diagnosis for Diabetic Retinopathy based on Fundus Images using Deep Convolutional Neural Network", *Mobile Information Systems*, Vol. 2, No. 47, pp. 1-14, 2019.
- [15] Y. Hatanaka, K. Ogohara, W. Sunayama, M. Miyashita, C. Muramatsu and H. Fujita, "Automatic Microaneurysms Detection on Retinal Images using Deep Convolution Neural Network", *Proceedings of International Workshop on Advanced Image Technology*, Vol. 7, pp. 1-12, 2018.
- [16] G.S. Scotland, P. McNamee, A.D. Fleming, K.A. Goatman, S. Philip, G.J. Prescott, P.F. Sharp, G.J. Williams, W. Wykes, G.P. Leese and J.A. Olson, "Costs and Consequences of Automated Algorithms Versus Manual Grading for the Detection of Referable Diabetic Retinopathy", *British Journal of Ophthalmology*, Vol. 94, No. 6, pp. 712-719, 2010.
- [17] K. Xu, D. Feng and H. Mi, "Deep Convolutional Neural Network-based Early Automated Detection of Diabetic Retinopathy using Fundus Image", *Molecules*, Vol. 22, No. 12, pp. 1-7, 2017.
- [18] R. Pires, S. Avila, J. Wainer, E. Valle, M.D. Abramoff and A. Rocha, "A Data-Driven Approach to Referable Diabetic Retinopathy Detection", *Artificial Intelligence in Medicine*, Vol. 96, pp. 93-106, 2019.
- [19] V. Srinivasan and V. Rajagopal, "Multi-Scale Attention-Based Mechanism in Gradient Boosting Convolutional Neural Network for Diabetic Retinopathy Grade Classification", *International Journal of Intelligent Engineering and Systems*, Vol. 15, No. 4, pp. 489-498, 2022.
- [20] T.R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P.K.R. Maddikunta, I.H. Ra and M. Alazab, "Early Detection of Diabetic Retinopathy using PCA-Firefly based Deep Learning Model", *Electronics*, Vol. 9, No. 2, pp. 1-16, 2000.
- [21] S.L. Yi, X.L. Yang, T.W. Wang, F.R. She, X. Xiong and J.F. He, "Diabetic Retinopathy Diagnosis based on RA-Efficientnet", *Applied Sciences*, Vol. 11, No. 22, pp. 1-18, 2021.
- [22] D. Yadav, A.K. Karn, A. Giddalur, A. Dhiman, S. Sharma, Muskan and A.K. Yadav, "Microaneurysm Detection using Color Locus Detection Method", *Measurement*, Vol. 176, pp. 1-9, 2021.
- [23] M. Mateen, T.S. Malik, S. Hayat, M. Hameed, S. Sun and J. Wen, "Deep Learning Approach for Automatic Microaneurysms Detection", *Sensors*, Vol. 22, No. 2, pp. 1-9, 2022.
- [24] V. Raudonis, A. Kairys, R. Verkauskiene, J. Sokolovska, G. Petrovski, V.J. Balciuniene and V. Volke, "Automatic Detection of Microaneurysms in Fundus Images using an Ensemble-Based Segmentation Method", *Sensors*, Vol. 23, No. 7, pp. 1-14, 2023.
- [25] T. Liu, Y. Chen, H. Shen, R. Zhou, M. Zhang, T. Liu and J. Liu, "A Novel Diabetic Retinopathy Detection Approach based on Deep Symmetric Convolutional Neural Network", *IEEE Access*, Vol. 9, pp. 1-9, 2021.
- [26] A.S. Hervella, J. Rouco, J. Novo and M. Ortega, "Retinal Microaneurysms Detection using Adversarial Pre-Training with Unlabeled Multimodal Images", *Information Fusion*, Vol. 79, pp. 1-7, 2021.
- [27] S. Das, K.M.S. Kharbanda, R. Raman and D. Edwin Dhas, "Deep Learning Architecture based on Segmented Fundus Image Features for Classification of Diabetic Retinopathy", *Biomedical Signal Processing and Control*, Vol. 68, pp. 1-19, 2021.
- [28] C. Harshitha, A. Asha, J.L.S. Pushkala, R.N.S. Anogini and C. Karthikeyan, "Predicting the Stages of Diabetic Retinopathy using Deep Learning", *Proceedings of International Conference on Inventive Computation Technologies*, Vol. 11, pp. 1-13, 2021.
- [29] E. Decenciere, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordenez, P. Massin, A. Erginay, B. Charton and J.C. Klein, "Feedback on a Publicly Distributed Image Database: The Messidor Database", *Image Analysis and Stereology*, Vol. 33, No. 3, pp. 231-234, 2014.
- [30] "Diaretdb1-Standard Diabetic Retinopathy Database", Available at: <https://www.kaggle.com/datasets/nguyenhung1903/diaretdb1-standard-diabetic-retinopathy-database>, Accessed in 2019.
- [31] U. Imran and K.A. Almejalli, "Intelligent Automated Detection of Microaneurysms in Fundus Images using Feature-Set Tuning", *IEEE Access*, Vol. 8, pp. 65187-65296, 2020.
- [32] E. Decenciere, G. Cazuguel, X. Zhang, G. Thibault, J.C. Klein, F. Meyer and D. Elie, "TeleOphta: Machine Learning and Image Processing Methods for Teleophthalmology", *IRBM*, Vol. 34, No. 2, pp. 196-203, 2013.
- [33] P. Porwal, "Diabetic Retinopathy: Segmentation and Grading Challenge Workshop", *Proceedings of International Symposium on Biomedical Imaging*, Vol. 59, pp. 1-10, 2018.
- [34] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", *Proceedings of International Conference on*

- Medical Image Computing and Computer Assisted Intervention*, Vol. 87, pp. 234-241, 2015.
- [35] Z. Zhang, Q. Liu and Y. Wang, "Road Extraction by Deep Residual U-Net", *IEEE Geoscience and Remote Sensing Letters*, Vol. 15, No. 5, pp. 749-753, 2018.
- [36] P.C. Bunch, J.F. Hamilton, G.K. Sanderson and A.H. Simmons, "A Free Response Approach to the Measurement and Characterization of Radiographic Observer Performance", Vol. 8, pp. 124-135, 1977.
- [37] L. Seoud, T. Hurtut, J. Chelbi, F. Cheriet and J.M.P. Langlois, "Red Lesion Detection using Dynamic Shape Features for Diabetic Retinopathy Screening", *IEEE Transactions on Medical Imaging*, Vol. 35, No. 4, pp. 1116-1126, 2016.
- [38] B. Dashtbozorg, J. Zhang, F. Huang and B.M. Ter Haar Romeny, "Retinal Microaneurysms Detection using Local Convergence Index Features", *IEEE Transactions on Image Processing*, Vol. 27, No. 7, pp. 3300-3315, 2018.
- [39] S. Wang, H.L. Tang, L.I.A. Turk, Y. Hu, S. Sanei, G.M. Saleh and T. Peto, "Localizing Microaneurysms in Fundus Images through Singular Spectrum Analysis", *IEEE Transactions on Biomedical Engineering*, Vol. 64, No. 5, pp. 990-1002, 2017.
- [40] L. Dai, R. Fang, H. Li, X. Hou, B. Sheng, Q. Wu and W. Jia, "Clinical Report Guided Retinal Microaneurysm Detection with Multi-Sieving Deep Learning", *IEEE Transactions on Medical Imaging*, Vol. 37, No. 5, pp. 1149-1161, 2018.
- [41] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, Vol. 13, pp. 1-14, 2014.
- [42] K. He, X. Zhang, S. Ren and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 37, No. 9, pp. 1904-1916, 2015.
- [43] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A.L. Yuille, "Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution and Fully Connected CRFs", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 40, No. 4, pp. 834-848, 2017.
- [44] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention*, Vol. 67, pp. 234-241, 2015.
- [45] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, Vol. 6, pp. 770-778, 2016.
- [46] Z. Zhang, Q. Liu and Y. Wang, "Road Extraction by Deep Residual U-Net", *IEEE Geoscience and Remote Sensing Letters*, Vol. 15, No. 5, pp. 749-753, 2018.
- [47] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", *Proceedings of International Conference on Machine Learning*, Vol. 79, pp. 1-11, 2015.
- [48] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", *Journal of Machine Learning Research*, Vol. 15, No. 56, pp. 1929-1958, 2014.
- [49] Z. Feng, J. Yang, L. Yao, Y. Qiao, Q. Yu and X. Xu, "Deep Retinal Image Segmentation: A FCN-Based Architecture with Short and Long Skip Connections for Retinal Image Segmentation", *Neural Information Processing*, Vol. 54, pp. 713-722, 2017.
- [50] M.Z. Alom, M. Hasan, C. Yakopcic, T.M. Taha and V.K. Asari, "Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, Vol. 14, pp. 1-12, 2018.
- [51] J. Hu, L. Shen and G. Sun, "Squeeze-and-Excitation Networks", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, Vol. 45, pp. 7132-7141, 2018.
- [52] A. Howard, M. Sandler, G. Chu, L.C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q.V. Le and H. Adam, "Searching for MobileNetV3", *Proceedings of International Conference on Computer Vision*, Vol. 19, pp. 1-7, 2019.
- [53] M. Tan and Q.V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", *Proceedings of International Conference on Machine Learning*, Vol. 18, pp. 1-11, 2019.
- [54] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo and Q. Hu, "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, Vol. 8, pp. 1-12, 2020.