Region of Interest based Predictive Algorithm for Subretinal Hemorrhage Detection using Faster R-CNN

Sai Ganesh N
VIT University - Chennai Campus
Suchetha M (suchetha.m@vit.ac.in)
VIT University - Chennai Campus
Rajiv Raman
Sankara Nethralaya
Edwin Dhas D
VIT University - Chennai Campus

Research Article

Keywords: Macular edema, Optical Coherence Tomography, Convolutional Neural Network, Region of Interest, Subretinal Fluid, Subretinal hemorrhage

Posted Date: April 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-246885/v1

Version of Record: A version of this preprint was published at Soft Computing on August 18th, 2021. See the published version at https://doi.org/10.1007/s00500-021-06098-1.

Noname manuscript No.
(will be inserted by the editor)

Region of Interest based Predictive Algorithm for Subretinal Hemorrhage Detection using Faster R-CNN

Sai Ganesh N · Suchetha M · Rajiv Raman · Edwin Dhas D

Received: date / Accepted: date

Abstract Macular Edema (ME) is an essential sort of macular issues caused due to the storing of fluid underneath the macula. Age-related Macular Degeneration (AMD) and Diabetic Macular Edema (DME) are the two customary visual contaminations that can lead to fragmentary or complete vision loss. This paper proposes a deep learning-based predictive algorithm that can be used to detect the presence of a Subretinal hemorrhage. This method initially detects the presence of Subretinal hemorrhage, and it then segments the Region of Interest (ROI) by a semantic segmentation process. The segmented ROI is applied to a predictive algorithm which is derived from the Fast Region Convolutional Neural Network (Fast R-CNN) algorithm, that can categorize the Subretinal hemorrhage as responsive or non-responsive. The dataset, provided by a medical institution, comprised of Optical Coherence Tomography (OCT) images of both pre and post-treatment images, was used for training the proposed Faster Region Convolutional Neural Network (Faster R-CNN). We also used the Kaggle dataset for performance comparison with the traditional methods that are derived from the Convolutional Neural Network (CNN) algorithm. The evaluation results using the Kaggle dataset and the hospital images provide an average sensitivity, selectivity, and accuracy of 85.3%, 89.64%, and 93.48% respectively. Further, the proposed method provides a time complexity in testing as $2.64s$, which is less than the traditional schemes like CNN, R-CNN, and Fast R-CNN.

Keywords Macular edema · Optical Coherence Tomography · Convolutional Neural Network · Region of Interest · Subretinal Fluid · Subretinal hemorrhage

Sai Ganesh N
Division of Healthcare Advancement, Innovation and Research, VIT Chennai, India
E-mail: sganesh154@gmail.com

Suchetha M
Division of Healthcare Advancement, Innovation and Research, VIT Chennai, India
E-mail: suchetha.m@vit.ac.in

Rajiv Raman
Shri Bhagwan Mahavir Vitreo-Retinal Services, Sankara Nethralaya, India
E-mail: rajivpgraman@gmail.com

Edwin Dhas D
Division of Healthcare Advancement, Innovation and Research, VIT Chennai, India
E-mail: edwindhas.dhason@vit.ac.in

1 Introduction

The macula of the human eye is an oval-framed zone that lies close to the point of convergence of the retina, which has a small pit named fovea. The fovea has a huge grouping of cones which is answerable for sharp and shaded vision. The macular issue is the bundle of ailments that destroy the macula, leading to clouded vision or vision loss. OCT is a contactless imaging procedure that has wide applications in ophthalmology. OCT imaging uses variable-sweep rates that create cross- sectional images of the visual tissues, including the retina [1]. A survey [2] shows that about 15% of people above 60 years old and 0.4% of people between 50-60 years old are affected by AMD. OCT images help to diagnose retinal varieties from the standard [3] retinal images with DME and AMD. The macula is the touchier piece of the eye and liable for clear vision. The growth of the Muller cells of the macula leads to Macular edema that forms a liquid amassing below the macula. This makes it grow and therefore, it has thickly pressed cones that are liable for a moment of subtleties in vision. At the

point when the macula is thickened, these cones can't work appropriately, therefore, the vision gets affected in doing the assignments such as reading, driving, or utilizing computers. A portion of the causes incorporates diabetes, maturing, waterfall medical procedure, tranquilize reactions, intrinsic ailments [4] are the few important reasons for macular edema.

The principal indication of macular edema is an obscured vision where the focal piece of the vision gets hazy, while the fringe vision is unaffected. It can prompt clear issues for the victim since the middle piece of the vision is required for practically fundamental undertakings like driving, reading, or utilizing computers [5]. The macula is a focal piece of the retina, and it is the basis for acceptable vision. DME showed liquid blisters inside the retina, and retinal widening is brought about liquid spillage to harmed macular veins. OCT images permit touchy discovery, appraisal of the liquid blisters, and retinal bulging [6]. Ophthalmologists assess the seriousness of DME by utilizing retinal bulging maps by implication to the intra-retinal liquid/blister regions [7]. The macula constitution could be a result of various issues, including the AMD and DME. AMD [8] is an eye illness bringing about obscured vision, vulnerable sides, or even no vision in the focal point of the eye field. It was the 4^{th} most normal, reason for visual deficiency [9] in the year 2013. In the US, around 13% of every upcoming instance for visual impairment (individuals matured between 25-60 years) every year is because of diabetic retinopathy [10]. The most widely recognized diabetic reason for vision loss in various social orders [11] is DME. During the beginning stage of retinopathy, the DME might impact the central vision. However, in diabetic victims (especially in type 2), DME is the highly progressive vision-trading off factor [12].

The three major cataracts includes nuclear waterfall, cortical waterfall, and Posterior Capsular waterfall. The Nuclear waterfall is shaped in the focal point that creates yellow formation in the eye and conviction a foggy picture. The Cortical waterfall is surrounded close to the edge of the point of convergence and commonly exists in mature age individuals. While the Posterior Capsular waterfall is a serious sort of waterfall which can harm the backside of the focal point [13]. Vein division is the most significant and advance for the identification of changes in retinal vascular structures. Therefore, various schemes have been developed to improve the results [14]. In [15], the authors used Recurrent Residual Convolutional Neural Network (RRCNN) and Recurrent Convolutional Neural Network (RCNN), considering U-Net models for the division of the retina vessel.

The remaining part of the paper is arranged as follows. Section II shows the related works that are associated with the detection of DME, Section III shows the framework of the CNN algorithm with diabetic macular edema; Section IV shows the proposed predictive algorithm using R-CNN and Faster R-CNN. Section V shows the experimental results and conversation of the proposed technique, and Section VI concludes the paper.

2 Related work

The authors Abdolreza Rashno et al. [1], discussed a completely mechanized calculation to portion liquid related (liquid-filled) and pimple region in OCT eye images of subjects with DME. The OCT images are sectioned utilizing a new neutrosophic change where the image is changed into three sets: D (valid), G (vague), and F (bogus). The work calculation shows a 7% improvement in the shakers coefficient and a 6% improvement in accuracy on the Duke dataset. Reza Rasti et al. [2], proposed a new CAD framework-dependent algorithm using Multi-Scale Convolutional mixture of Expert (MCME) group symbol to distinguish typical vision, and two essential kinds of macular pathology, such as dry AMD and DME. The work of MCME uses an information-related neural system that utilizes fast training of image that includes CNN to different scale subimages. Two distinctive OCT datasets of ME from Heidelberg gadgets were utilized for the assessment of the strategy. For examination reason, the authors have played out a broad scope of arrangement to contrast the outcomes and the good designs of the MCME strategy.

Uma Punniyamoorthy et al. [3], examined image handling strategies to identify the center circle, exudates, and the nearness of ME. Their strategy provides a sensitivity of 96.02%, the selectivity of 97.33%, and accuracy of 96.33% for the exudates location and a sensitivity of 97.35%, the selectivity of 98%, and accuracy of 98.70% for macular edema identification. The exhibition correlation for different strategies uncovers that the technique could be utilized as a facing procedure for diabetic retinopathy. Girish et al. [4], discussed a Fully convolutional network (FCN) for merchant autonomous IRC division. In this, a strategy was introduced which neutralizes the image commotion fluctuations and process the convolutional network on OCT images for the Optima pimple division opposite dataset (with four diverse merchant explicit images, to be specific, Cirrus, Nidek, Spectralis, and Topcon).

Xiaoming Liu et al. [5], examined another fully convolutional deep learning strategy that considers OCT layers and liquid locales in retinal OCT C-filters. In

this, approach a semi-regulated technique that uses the unlabeled information through an antagonistic learning procedure. The division strategy incorporates a division approach and discriminator arrangement, where these strategies are used with U-Net convolutional design. The target capacity division organizes a combined misfortune work, including two class cross-entropy misfortune named as cover misfortune and antagonistic misfortune. The authors have evaluated the performance of their algorithm in the Duke DME dataset and the S-One dataset, and it is found that the algorithm is most powerful than other best in class strategies for layers and liquid division in the OCT images. Yue Zhang et al. [6], introduced a multimodal information system in vehicles, that focuses on the after effects of two modalities, namely images, and speeds. Images that are handled in the vehicle discovery module give visual data regarding the highlights of vehicles, though speed calculation can additionally assess the potential areas of the objective vehicles. This scheme diminishes the number of up-and-comers being looked at while minimizing time utilization and computational expense. This uses shading Faster R-CNN, whose source of information is both the surface and shade of the vehicles, where the speed is estimated by the Kalman channel.

Wolf-Dieter Vogl et al. [7], examined two unique information-driven AI approaches working in a high-dimensional element place. The distinguish spatial fleeting marks dependent on retinal thickness which highlights the longitudinal ghastly space OCT imaging information and foresee singular patient results utilizing these quantitative qualities. The authors have used SC-OCT images of 94 patients with Branch retinal vein occlusion (BRVO) and 158 patients with focal eye vein impediment Central retinal vein occlusion (CRVO). Adeel et al. [8] introduced a computerized framework to detect the ME from fundus images. Also, they have presented another mechanized framework for the nitty-gritty evaluating the seriousness of illness utilizing information on exudates and macula. Another arrangement of highlights is utilized alongside a base separation classifier for precise limitation of the fovea, which is significant for the reviewing of ME. In this work, a framework that utilizes distinctive cross breed highlights and bolsters machines for localization of exudates are used. The point by point evaluation of ME with clinically noteworthy ME or non-clinically critical ME is finished by utilizing confined fovea and fragmented exudates.

Wenqi Wu et al. [9], discussed a neural network approach that uses Faster R-CNN. This system could increase the accuracy of face recognition, where the speed remains the same as that of faster R-CNN. Choosing the ROI from the homogeneously top components will perform different errands of RPN. Deep learning techniques can be applied in medical fields for the Diabetic retinopathy detection in fundus images using CNN and Support Vector Machine (SVM) algorithms. In the detection of Subretinal hemorrhage in OCT images, 5 machines are used to retrieve OCT images of DME patients, Cirrus 500, Cirrus 500 Angiography, Spectralis Heidelberg, swept-source, and swept-source angiography. Cirrus 500 and Cirrus 500 Angiography are from the Zeiss manufacturer instrument used to detect diabetic retinopathy. To detect the depth of Subretinal hemorrhage the swept-source-Topcon is used, where a swept-source angiography is an advanced form when compared to the previous machines.

Several algorithms [16–18] are proposed that can segment the Retinal vessels that use a smoothly regularized network and deeply supervised network. The authors Parhi et al., [19] proposed a fluid/cyst segmentation algorithm, where they proposed a quantitative assessment after segmenting the Subretinal hemorrhage. DWT and DCT [20] can also be used to grade the DME by extracting the DWT and DCT features. The commonly used algorithms for the discovery of DME in the clinical field are derived from algorithms such as CNN, SVM, and Fully convolutional neural networks. The challenges present in the identification of DME utilizing CNN, SVM, and Fully convolutional neural network calculations require significant investment and it needs a huge number of training and testing pictures since its computational complexity is high. Taimur Hassana et al. [25], proposed a merchant-free profound convolutional neural system and structure tensor diagram search-based segmentation system (CNN-STGS) to remove and break down the liquid pathology and retinal layers alongside 3-D retinal profiling. Navid Mohaghegh et al. [26], proposed a graphical macular interface framework (GMIS) for the specific, fast, and quantitative examination of visual contortion (VC) for patients having macular disorders. Zagoa et al. [27], introduced an effective algorithm that has two convolutional neural networks that can be used to choose the training patches so that the images with complex training images must be given special attention during the training phase. Qiu et al. [28], introduced a self-regulated iterative refinement learning (SIRL) strategy that has a pipeline design to build the exhibition of volumetric image classification in macular optical coherence tomography (OCT). Ajaz et al. [29], discussed the relationship between the geometrical vascular parameters estimated from the fluorescein angiography (FA) and optical coherence tomography (OCT) of the eyes with macular edema (ME).Novel deep learning architectures [30–34]

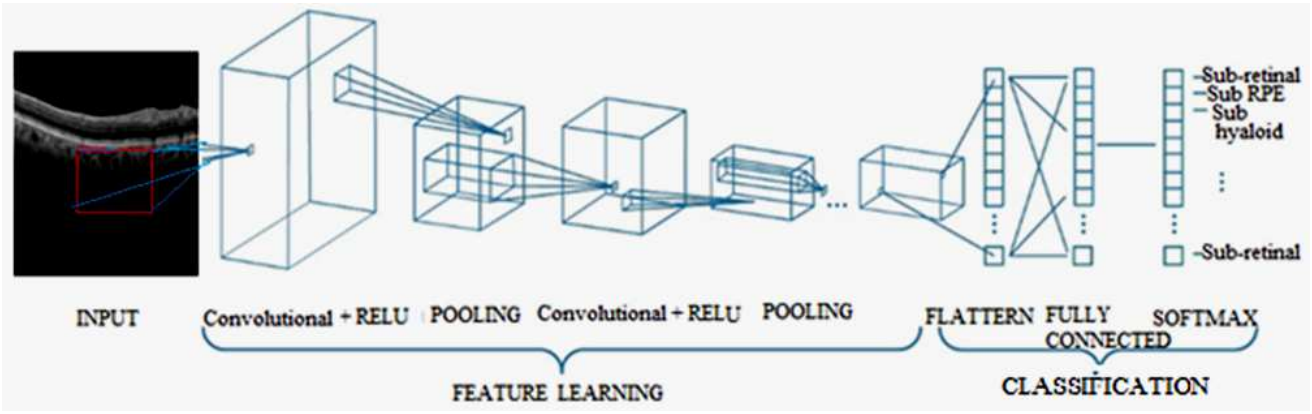

Fig. 1 Architecture of CNN

are also used in diverse applications which provides better classification results.

3 The Framework of Convolutional Neural Network

A convolutional neural network model (CNN or ConvNet), is one of the categories of the deep learning algorithm, where the model figures out how to perform grouping errands legitimately from pictures, video content, or sound. CNN's are especially valuable for discovering designs in images to perceive items, appearances, and scenes. They gain features straightforwardly from picture information, utilizing examples to order, and taking out the requirement for manual element extraction. A CNN can have several layers that each figure out how to identify various highlights of an image. Channels are applied to each input image at various regions, and the yield of each convolved image is utilized as the contribution to the following layer. A CNN consists of an input layer, Feature Learning (Convolutional + RELU, Pooling), and Classification layer (Flatten, Fully Connected, SoftMax) as depicted in fig 1.

The layers of CNN will perform feature learning and classification processes where the layers include convolution, ReLU, or activation and pooling. (i) Convolution: It gets the input image using the convolutional channels, where each channel initiates certain highlights from the input image. (ii) Rectified linear unit (ReLU): It takes into consideration about faster and progressively viable preparation by mapping negative qualities to zero and keeping up the positive qualities. The actuated highlights obtained in this stage are conveyed to the following layer. (iii) Pooling: Pooling reworks the yield by performing non-linear down assessing, diminishing the number of boundaries that the framework needs to learn. In the wake of learning features in various layers, the design of CNN movements to grouping. The Fully Connected layer is a completely related layer that yields a vector of K estimations, where the number of classes that the framework will have the choice to anticipate is K. This vector contains the probabilities for each class of any picture being characterized. The last layer of the CNN configuration uses a portrayal layer, for instance, SoftMax that gives the characterization and the non-linearity is communicated as (1),

$$f(x) = max(x, 0) \tag{1}$$

Also, the function $f(x)$ can be expressed interms of $tanh$ function as expressed in (1),

$$f(x) = tanh(x) \tag{2}$$

The function $f(x)$ can be expressed interms of logistic sigmoid function as

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

In the convolution neural network, the convolution of two continuous functions P and Q can be calculated using (4),

$$(P * Q)_x = \int_{\infty}^{\infty} P(t)Q(x - t)dt \tag{4}$$

In discrete form, the convolution of two functions can be expressed by replacing the time t to discrete values whose index is n as,

$$(P * Q)_x = \sum_{n=-\infty}^{\infty} P(n)Q(x - n) \tag{5}$$

Consider an input I and a filtering function H, then the convolution operation between I and H is expressed for a two-dimensional image as,

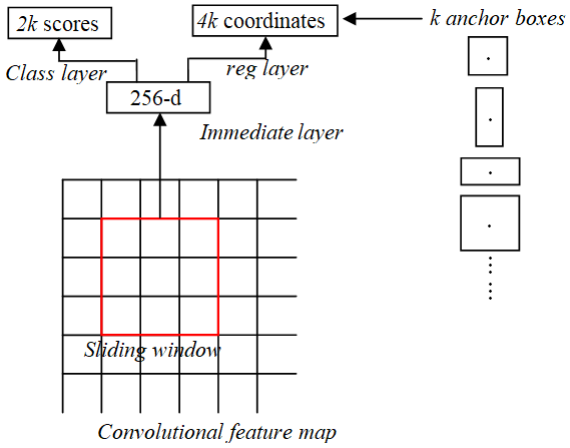

Fig. 2 Sliding window in RPN

$$(H * I)_{x,y} = \sum_{m=-a_1}^{a_1} \sum_{n=-b_1}^{b_1} H(m,n)I(x-m, y-n) \quad (6)$$

The filter H in matrix form is represented as,

$$H = \begin{bmatrix} H(-a_1,-b_1) & \dots & \dots & \dots & H(-a_1,b_1) \\ . & . & . & . & . \\ . & . & H(0,0) & . & . \\ . & . & . & . & . \\ H(a_1,-b_1) & \dots & \dots & \dots & H(a_1,b_1) \end{bmatrix} \quad (7)$$

The multi-task loss function is expressed as,

$$l = \propto_1 l_c + \propto_2 l_b + \propto_3 l_m \quad (8)$$

Where l_c, l_m and l_b are the classification, mask, and bounding box losses respectively. Let $p(0), p(1), ..p(k)$ be the probabilistic distribution, where the classification result is R, then the class loss is expressed as,

$$l_c = -log(p(r)) \quad (9)$$

If the size of ROI is $M \times N$, the mask loss l_m is expressed as

$$l_m = -\frac{1}{m \times n} \sum_{x=1}^{M} \sum_{y=1}^{N} [s(x,y) log s^k(x,y) + (1 - s(x,y)) log(1 - \hat{s}(x,y))] \quad (10)$$

Here the binary mask is $s(x,y)$ and the ground truth mask class is k and the predicted class cell is $s(x,y)$. If the ROI has the same number of rows and columns then the mask loss l_m is expressed as

$$l_m = -\frac{1}{m \times n} \sum_{x=1}^{M} \sum_{y=1}^{M} [s(x,y) log s^k(x,y) + (1 - s(x,y)) log(1 - \hat{s}(x,y))] \quad (11)$$

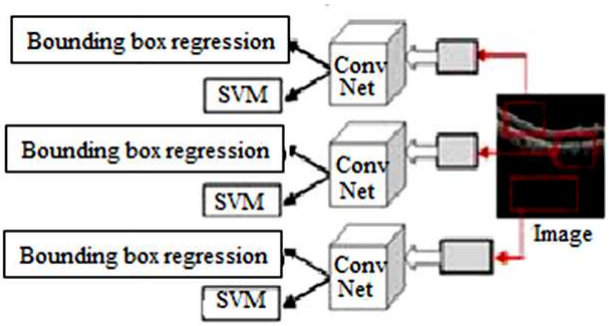

Fig. 3 R-CNN architecture

The proposed deep learning-based predictive algorithm is the advanced version of the CNN algorithm. The next section shows the proposed algorithm.

4 R-CNN and Faster R-CNN

4.1 Regional Proposal Network (RPN)

RPN has a classifier and regressor, which uses the concept of a sliding window, as shown in fig 2. The ZF model which is an expansion of AlexNet uses $256-d$ measurement and Visual Geometry Group (VGG-16) from the oxford model that uses $512-d$ measurement, where d represents dimensions. The scale and angle proportion are two significant parameters, and the RPN commonly uses 3 scales and 3 angle proportions. Thus, the aggregate of the nine proportions is feasible for every pixel.

The loss function for RPN is estimated as,

$$L(p_i, t_i) = \frac{1}{N_{cls}}(p_i, p_i^*) + \frac{\lambda}{N_{reg}} \sum p_i^* L_{reg}(t_i, t_i^*) \quad (12)$$

Here I represent the index of anchor, p_i is the predicted probability, if p_i^* is 1 then it is for positive anchors and if p_i^* is 0 then it is for negative anchors, N_{cls} represents the number of anchors in minibatch (512), N_{cls} represents the loss function of the enhanced Region proposal network, to represents the predicted bounding box with a vector of 4 parameterized coordinates and $*$ represents

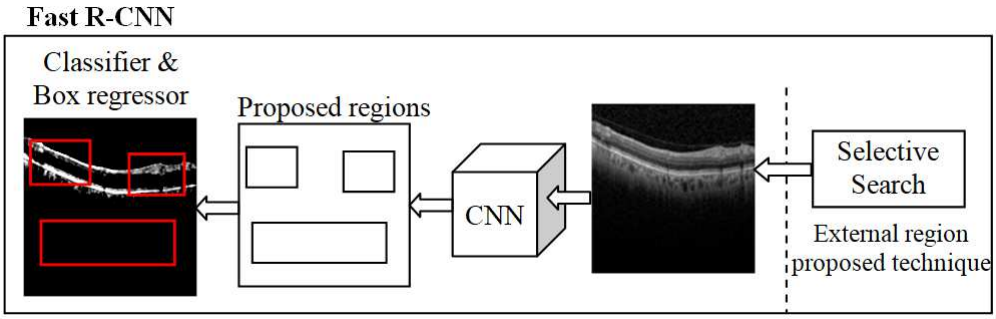

Fig. 4 Fast R-CNN architecture

the bounding box of ground truth. The loss function for RPN log loss of more than two classes.

$$L_{reg}(t_i, t_i^*) = R(t_i - t_i^*) \tag{13}$$

p^* is the regression term in the loss function and L_{reg} is the loss function for locating the target box. In this paper, we assume λ as 10. The challenges in R-CNN includes (i) It still sets aside a tremendous measure of effort to prepare that need to group 2000 regions for each image. (ii) It can't be completed steadily which utilizes 43 seconds for one test image. (iii) The specific pursuit calculation is fixed where learning is not performed during the stage. Fig 3 depicts the architecture of R-CNN.

4.2 Fast R-CNN

Fast region-based convolutional neural network (Fast R-CNN) improves the training and testing speed just as expanding the discovery exactness. Visual Geometry Group from oxford (VGG-16)is an architecture that is faster than R-CNN, faster at test time feature map on the PASCAL Visual Object Classes Challenge 2012 (PASCAL VOC 2012). Contrasted with Spatial pyramid pooling in deep convolutional networks for visual recognition (SPPNet) [4], the fast R-CNN trains VGG-16 three times faster, tests ten times faster, and is increasingly exact. Instead, the convolution activity does once per image, and an element map is produced. Fast-RCNN was fabricated for faster item discovery, covering up for the disadvantages of R-CNN. In any case, rather than taking care of the region proposal to CNN, we fetch the information picture to CNN to create a convolutional feature map. It perceives the region of recommendation and turns them into squares and by using an ROI pooling layer. From the ROI feature vector, used a SoftMax layer to anticipate the class of the proposed regions as depicted in Fig 4.

4.3 Proposed Predictive algorithm using R-CNN and Faster R-CNN

In the proposed methodology as depicted in Fig 5, OCT images, for responsive and non-responsive patients images (before treatment and after treatment images), are used as the input image. The Subretinal hemorrhage, present in the OCT image is then detected, which indicates that the patient is suffering from DME. Subretinal hemorrhage is segmented using a predictive algorithm, to detect whether it is responsive or non-responsive hemorrhage. Therefore, the stages include in the proposed method are (i) Acquiring of OCT image (ii) Detection of Subretinal hemorrhage (iii) Semantic Segmentation, and (iv) Predictive Algorithm. The intensity of an abnormal region of OCT image differs from the normal region due to the presence of Subretinal hemorrhage, Subhyaloid hemorrhage, or Sub RPE. If the accumulation is in the middle layer, topmost layer, and bottom layer, it is called Subretinal hemorrhage, Subhyaloid hemorrhage, and Sub RPE respectively. Subretinal hemorrhageis formed due to the formation of serous fluid(clear or lipid-rich exudates) in the Subretinal space, i.e., the fluid is formed due to the absence of retinal breaks, traction, or tears between the Retinal Pigment Epithelium (RPE) and the neurosensory retina (NSR). Subretinal hemorrhageindicates the breakdown of the normal anatomical structure of the retina and its relevant tissues, i.e., the RPE, braches the choroid and membrane. Subhyaloid hemorrhage is rare and usually contained in a self-created space between the posterior hyaloid and retina. We here describe a case of high altitude Subhyaloid hemorrhage and associated OCT findings.

Sub RPE is present in the dense clusters of the macular region, choriocapillaris, and the outer retinal layers. The OCT images can detect diseases in the reflectivity of the RPE detachment and retinal thickness. RPE is caused due to the reduction of atrophic retinal tissues, so that the ability to attenuate the light

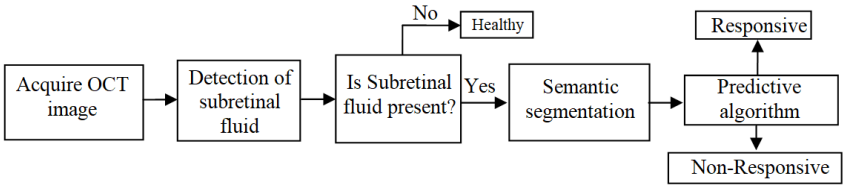

Fig. 5 Block diagram of a proposed method

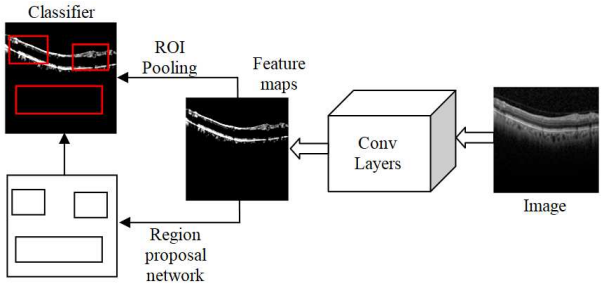

Fig. 6 The Architecture of faster R-CNN

reduces, which further reduces the retinal thickness. Retinal maps are used to estimate the volume an dto identify the extent of the atrophy which highlights the areas with the greatest atrophy,

This paper focuses on the detection of Subretinal hemorrhage. Therefore, semantic segmentation aims to segment the exact region that contains the Subretinal hemorrhage. After the detection of a Subretinal hemorrhage, a predictive algorithm is used to categorize, whether the image is responsive and non-responsive. The proposed predictive algorithm is derived from the Faster RCNN, which also uses the concepts present in the Regional proposal network, R-CNN, and Fast R-CNN.Both of the above algorithms (R-CNN and Fast R-CNN) use a specific pursuit to discover the region proposition. Specific hunt is a moderate and tedious procedure influencing the presentation of the system. Fig 6 depicts the proposed faster R-CNN.

Like fast R-CNN, the proposed Faster R-CNN uses a convolutional layer. Rather than using specific computation on the component escort for recognizing the region recommendations, a different system is used to anticipate the region proposition. The anticipated region proposition is reshaped, using an RoI pooling layer. It isused to characterize images inside the work area and foresee counterbalance esteems bounding boxes. The total of the past item identification computation uses regions to restrict the article inside the image. The framework doesn't consider the total image, instead, it considers the portions of the picture that have high possibilities of the existing item. The proposed training phase of Faster R-CNN includes the steps shown below. Step (i):Initialize the ImageNet pre-trained model to train the RPN. Step (ii):Train a different discovery organized by Fast R-CNN utilizing proposition created by step (i) RPN, instated by ImageNet pre-prepared model. Step (iii):Fix the Conv layer, adjust one of a kind layers to RPN, introduced by identifier organized in step (ii). Step (iv):Fix the Conv layer, adjust FC-layers of Fast R-CNN. The proposed faster R-CNN has the accompanying boundaries. The weight values are initialized as $N = ([0, 0.01])^2$.The learning update scheme uses weight decay as 0.0005 and momentum update as 0.9. The Loss function of the proposed faster R-CNN architecture is expressed as,

$$L = \frac{1}{N}\sum_{i} L_i + \lambda R(w) + \lambda \sum_{k}\sum_{l} W_{k,l}^2 \tag{14}$$

The next section shows the experimental results of the proposed work.

5 Experimental Results and Discussion

To validate the proposed scheme performance, we use the images obtained from the Kaggle dataset [24] and the images obtained from Sankara Nethralaya hospital. Fig 7 shows some sample test images obtained from the Kaggle dataset and Sankara Nethralaya hospital. The OCT images have been collected from the Kaggle dataset, which contains 968 images, and we have used a maximum of 40 images for training. Also, we have used the images of the Sankara Nethralaya hospital that contains the 400 images. The 40 training images are selected randomly, and the remaining images are used as test images. We have used MATLAB 2018a with a dual-core processor to measure the time complexity of the proposed algorithm.

Fig 8 and 9 show some of the semantic segmentation results obtained from the Kaggle dataset and images from the hospital respectively, where the blue color shows the segmented region.

The performance of the proposed scheme was evaluated usingparameters such as Selectivity S_l, Sensitivity

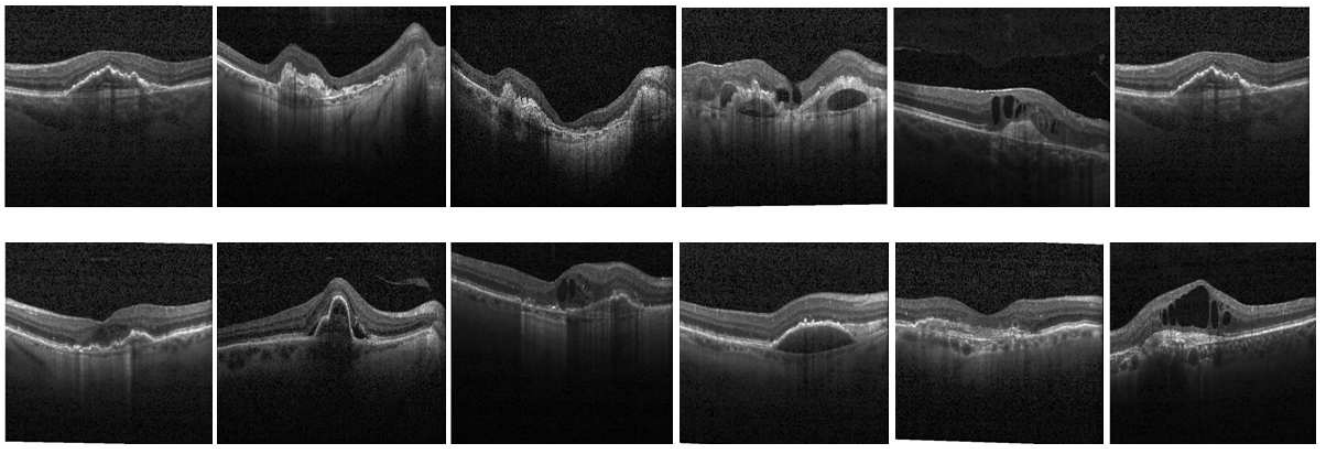

Fig. 7 Sample images from Kaggle dataset and Sankara Nethralaya hospital

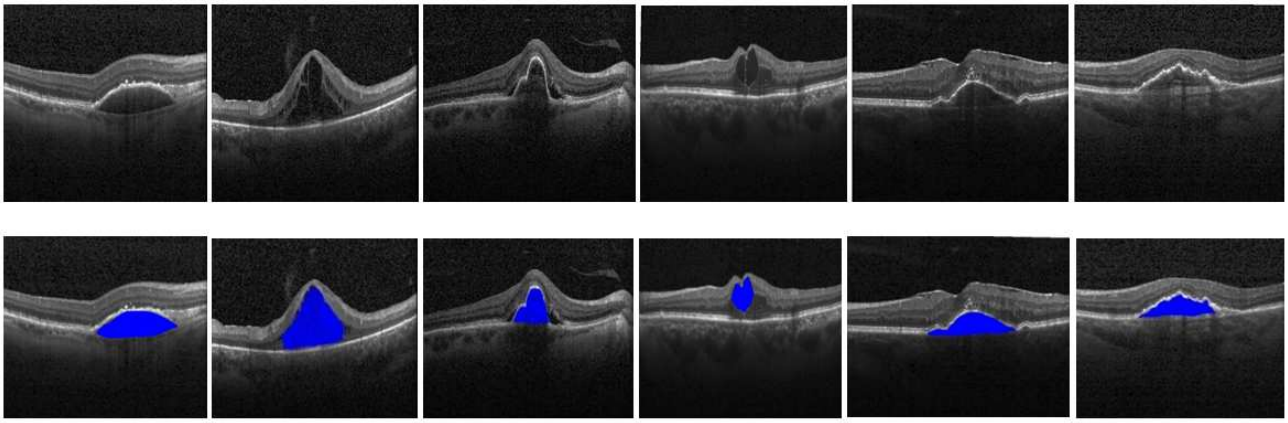

Fig. 8 Semantic segmentation results in Kaggle dataset (RowI: Input test images, Row II: Segmented output)

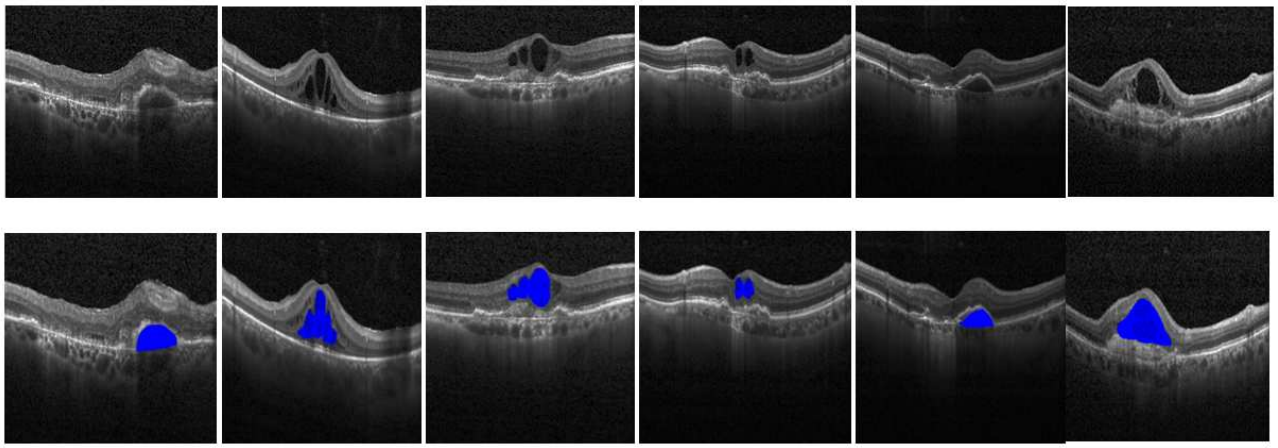

Fig. 9 Semantic segmentation results in SankaraNethralaya hospital images (Row I: Input test images, Row II: Segmented output)

Table 1 Comparison of the number of training, iteration, time complexity, Accuracy, and time of testing with the traditional methods

Algorithm	No. of training images (N_{train})	No of iteration (N_{iter})	Time complexity (T_{train}) (s)	Accuracy (%)	Time of testing (T_{test}) (s)
CNN	15	9	4.725	73.1	7.81
	20	12	6.3	76.4	
	25	16	8.4	79.12	
	30	20	10.5	83.56	
	35	24	12.6	86.72	
	40	27	14.175	88.5	
R-CNN	15	7	2.1875	74.32	5.42
	20	9	2.8125	78.64	
	25	12	3.75	81.45	
	30	15	4.6875	85.92	
	35	18	5.625	88.62	
	40	22	6.875	90.32	
Fast R-CNN	15	6	1.86	78.12	3.17
	20	8	2.48	81.23	
	25	10	3.1	84.32	
	30	13	4.03	87.52	
	35	16	4.96	89.32	
	40	20	6.2	91.24	
Proposed Faster R-CNN	15	4	1.1648	84.45	2.64
	20	6	1.7472	86.32	
	25	8	2.3296	88.54	
	30	11	3.2032	91.74	
	35	13	3.7856	93.92	
	40	16	4.6592	93.98	

S_e, and accuracy expressed as (15), (16), and (17) respectively.

$$S_l = \frac{T_n}{T_n + F_p} \tag{15}$$

$$S_e = \frac{T_p}{T_p + F_n} \tag{16}$$

$$Accuracy = \frac{T_n + T_p}{T_n + F_p + T_p + F_n} \tag{17}$$

The performance of the proposed scheme was compared with the traditional methods such as CNN [21], R-CNN [22], Fast R-CNN [23] algorithms. The experimental results were evaluated by using the different number of training image N_{trains} $15, 20, 25, 30, 35$, and 40. In all the schemes, as the number of training images increases,the number of iteration increases, However, the proposed method requires only 16 iteration which is less than the traditional methods for $N_{train} = 40$. Since the number of iterations also increments concerning the number of train pictures, the time complexity also increases concerning the number of train images. The proposed scheme consumes $1.1648s$ to train 15 images, and $4.6592s$ to train 40 images. The number of iterations needed to train one image is estimated by,

$$N_{iter}/image = \frac{N_{i,1}N_{i,2} + ...N_{i,L}}{N_{t,1}N_{t,2} + ...N_{t,L}} \tag{18}$$

For the proposed method, the number of iterations required to train one image is $N_{iter}/image = 0.3515$, while CNN, R-CNN, and Fast R-CNN provide $N_{iter}/image$ as $0.6545, 0.5030$, and 0.4424. The proposed scheme provides the least value of 0.3515 when compared to CNN, R-CNN, and Fast R-CNN approach. Similarly, the time required to train one image is estimated using the relation,

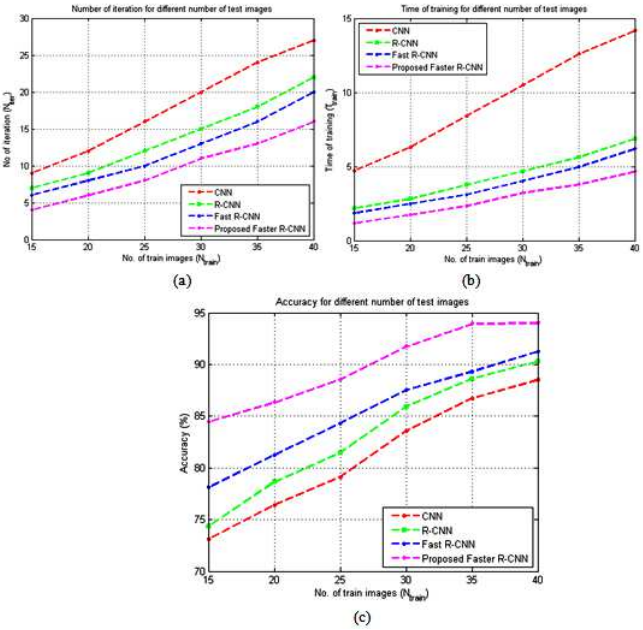

Fig. 10 Performance comparison of the proposed method with other methods (a) Number of iterations to complete training (b) Time complexity (in seconds) (c) Accuracy

$$T_{iter}/image = \frac{T_{i,1}T_{i,2} + ...T_{i,L}}{N_{t,1}N_{t,2} + ...N_{t,L}} \tag{19}$$

For the proposed method, the time to complete training for one image is $T_{iter}/image = 0.1023s$, while CNN, R-CNN, and Fast R-CNN provide $T_{iter}/image$ as $0.3436s$, $0.1571s$, and $0.1371s$. The proposed method provides a low time to train one image when compared to CNN, R-CNN, and Fast R-CNN approach. The time to execute one iteration can be calculated as,

$$T_{iter}/iteration = \frac{T_{i,1}T_{i,2} + ...T_{i,L}}{N_{i,1}N_{i,2} + ...N_{i,L}} \tag{20}$$

The proposed method provides $T_{iter}/iteration = 0.2912s$, while CNN, R-CNN, and Fast-CNN provide $T_{iter}/iteration$ as $0.525s, 0.3125s$,and $0.31s$ respectively. While comparing the metrics such as $N_{iter}/image, T_{iter}/image$, and $T_{iter}/iteration$ the time complexity of the proposed method is lesser than the traditional approaches.

Fig 10 (a), (b), and (c) depicts the graphical comparison of the number of iterations, time of training, and accuracy, respectively, for the different number of training images. The number of cycles, time of training, and accuracy increases as the number of training images increments. The number of cycles, time of training was less than the traditional methods for any n number of

Table 2 Variation of Selectivity, Sensitivity, and accuracy for different number of trained images

No. of train images (N_{train})	Kaggle dataset			Hospital dataset		
	Selectivity (%)	Sensitivity (%)	Accuracy (%)	Selectivity (%)	Sensitivity (%)	Accuracy(%)
15	78.12	79.34	84.45	76.764	77.912	83.464
20	81.23	81.59	86.32	79.874	80.162	85.334
25	82.52	83.64	88.54	81.164	82.212	87.554
30	83.92	85.62	91.74	82.564	84.192	90.754
35	84.56	88.32	93.92	83.204	86.892	92.934
40	85.98	90.36	93.98	84.624	88.932	92.994

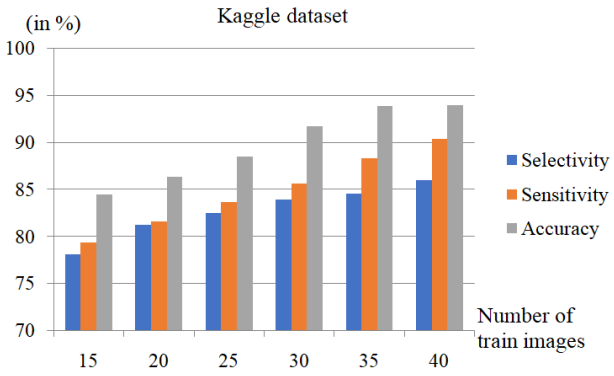

Fig. 11 Performance comparison for different number of test images using the Kaggle dataset

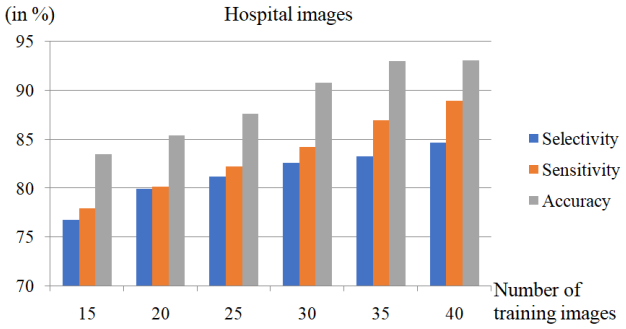

Fig. 12 Performance comparison using the images from the hospital for a different number of test images

training pictures. The accuracy of the proposed method likewise increments as the number of training pictures increases. The proposed method provides an accuracy of 93.98% when the number of train images is 40, while the accuracy is 84.45% when the number of train images is 15. For the training images of 40, the methods CNN, R-CNN, and Fast R-CNN provides an accuracy of 88.5%, 90.32%, and 91.24%, respectively. The proposed method provides a time complexity in testing as $2.64s$, which is lesser than the other schemes. The time complexity in testing for the methods CNN, R-CNN, Fast RCNN is estimated as $7.81s$, $5.42s$, and $3.17s$, respectively.

The comparison of Selectivity, Sensitivity, and Accuracy for the different number of training images with the Kaggle dataset is depicted in fig 11. The Selectivity and Sensitivity also increase as the number of train images increases, similar to accuracy. For the Kaggle dataset, the Sensitivity,Selectivity, and Accuracy of the proposed scheme were estimated as 90.36%, 85.98%, and 93.98% respectively forN_{train} = 40, while it is 79.34%, 78.12%, and 84.45% respectively for N_{train} = 15 .

The comparison of Selectivity, Sensitivity, and Accuracy for the different number of training images with the hospital dataset is depicted in fig 12. The Selectivity and Sensitivity also increase as the number of train images increases, similar to accuracy. For the hospital dataset, the Sensitivity, Selectivity, and Accuracy of the proposed scheme were estimated as 88.93%, 84.62%, and 92.99% respectively for N_{train} = 40, while it is 77.91%, 76.76%, and 83.46% respectively for N_{train} = 15.

The proposed faster R-CNNs loss function is estimated as,

$$L = \frac{1}{N}\sum_{i} L_i + \lambda R(w) + \lambda \sum_{k}\sum_{l} W_{k,l}^2 \tag{21}$$

$$L = \frac{1}{0.0012}\sum_{i} L_i + \lambda R(w) + \lambda \sum_{k}\sum_{l} W_{k,l}^2 \tag{22}$$

For $i = 0$, the loss function is expressed as

$$L = \frac{1}{0.0012} L_0 + \lambda R(w) + \lambda \sum_{k=9}\sum_{l=0.9} 0.005^2 \tag{23}$$

Here, W_k is weight decay, $K = 9$, and $l = 0.9$. Here $\lambda R(w) = 0$ because there is no regularization loss. The performance of the algorithm highly depends on the number of train images. The next section depicts the conclusion of the proposed scheme.

6 Conclusion

Diabetic macular edema is a common disease that occurs in most diabetic patients. It is the cumulation of aqueous from the center part of the retina called fovea near the optic disc. In this manner, this paper proposed a deep learning-based predictive algorithm for diabetic macular edema that uses a faster R-CNN approach. This scheme starts with detecting the presence of subretinal hemorrhage followed by segmentation. The semantic algorithm is used in the segmentation of subretinal hemorrhage, followed by the predictive algorithm, which is an extended approach of R-CNN. We have used datasets such as Kaggle and the images obtained from a hospital. The performance was evaluated using metrics such as accuracy, sensitivity, selectivity, and time complexity for training and testing. The proposed scheme provides an average selectivity,sensitivity, and accuracy of 89.64%, 85.3%, and 93.48%, respectively, when evaluated using the Kaggle dataset and the hospital images. The proposed method provides a time complexity in testing as $2.64s$. Also, the time to perform one iteration is $0.2912s$, which is less than the traditional schemes. The time and the number of iterations to perform on one image is estimated as $0.1023s$ and $0.3515s$, respectively, which is less than the traditional schemes namely CNN, R-CNN, and Fast R-CNN.

Declaration

The authors declare that they have no conflict of interest. The authors also declare that the Hospital dataset was obtained with prior permission from Shri Bhagwan Mahavir Vitreo-Retinal Services, Sankara Nethralaya, India.

Authorship contributions

Sai Ganesh N : Software implementation, Visualization
Suchetha M : Supervision, Conceptualization, Methodology, Data curation, Analysing, Investigation
Rajiv Raman : Investigation, Testing and validation
Edwin Dhas D : Visualization, Writing- Reviewing and Editing

References

1. Rashno, Abdolreza, et al. "Fully automated segmentation of fluid/cyst regions in optical coherence tomography images with diabetic macular edema using neutrosophic sets and graph algorithms." IEEE Transactions on Biomedical Engineering 65.5: 989-1001,(2017)
2. Rasti, Reza, et al. "Macular OCT classification using a multi-scale convolutional neural network ensemble." IEEE transactions on medical imaging 37.4: 1024-1034,(2017)
3. Punniyamoorthy, Uma, and Indumathi Pushpam. "Remote examination of exudates-impact of macular oedema." Healthcare technology letters 5.4: 118-123, (2018).
4. Girish, G. N., et al. "Segmentation of intra-retinal cysts from optical coherence tomography images using a fully convolutional neural network model." IEEE journal of biomedical and health informatics 23.1: 296-304,(2018).
5. Liu, Xiaoming, et al. "Semi-supervised automatic segmentation of layer and fluid region in retinal optical coherence tomography images using adversarial learning." IEEE Access 7 : 3046-3061, (2018).
6. Zhang, Yue, et al. "Vehicle tracking using surveillance with multimodal data fusion." IEEE Transactions on Intelligent Transportation Systems 19.7 : 2353-2361, (2018).
7. Vogl, Wolf-Dieter, et al. "Predicting macular edema recurrence from spatio-temporal signatures in optical coherence tomography images." IEEE transactions on medical imaging 36.9 : 1773-1783, (2017).
8. Syed, Adeel M., et al. "Fundus images-based detection and grading of macular edema using robust macula localization." IEEE Access 6 : 58784-58793, (2018).
9. Wu, Wenqi, et al. "Face detection with different scales based on faster R-CNN." IEEE transactions on cybernetics 49.11 : 4017-4028, (2018).
10. Wang, Depeng, and Liejun Wang. "On OCT image classification via deep learning." IEEE Photonics Journal 11.5 : 1-14, (2019).
11. Li, Hailiang, Yongqian Huang, and Zhijun Zhang. "An improved faster R-CNN for same object retrieval." IEEE Access 5 : 13665-13676, (2017).
12. Soomro, Toufique Ahmed, et al. "Deep learning models for retinal blood vessels segmentation: A review." IEEE Access 7 : 71696-71717, (2019).
13. Imran, Azhar, et al. "Comparative analysis of vessel segmentation techniques in retinal images." IEEE Access 7 : 114862-114887, (2019).
14. Biswal, Birendra, Thotakura Pooja, and N. Bala Subrahmanyam. "Robust retinal blood vessel segmentation using line detectors with multiple masks." IET Image Processing 12.3 : 389-399, (2017).
15. Jiang, Yun, et al. "Retinal vessels segmentation based on dilated multi-scale convolutional neural network." IEEE Access 7 : 76342-76352, (2019).
16. Lin, Yi, Honggang Zhang, and Guang Hu. "Automatic retinal vessel segmentation via deeply supervised and smoothly regularized network." IEEE Access 7): 57717-57724, (2018.
17. Kar, Sudeshna Sil, and Santi P. Maity. "Automatic detection of retinal lesions for screening of diabetic retinopathy." IEEE Transactions on Biomedical Engineering 65.3 : 608-618 (2017).
18. Xia, Haiying, et al. "Mapping functions driven robust retinal vessel segmentation via training patches." IEEE access 6 : 61973-61982, (2018).
19. Parhi, Keshab K., et al. "Automated fluid/cyst segmentation: A quantitative assessment of diabetic macular edema." Investigative Ophthalmology and Visual Science 58.8 : 4633-4633, (2017).
20. Acharya, U. Rajendra, et al. "Automated diabetic macular edema (DME) grading system using DWT, DCT features and maculopathy index." Computers in biology and medicine 84 : 59-68, (2017).
21. Mishra, Sapna S., Bappaditya Mandal, and Niladri B. Puhan. "Multi-level dual-attention based CNN for macular

optical coherence tomography classification." IEEE Signal Processing Letters 26.12 : 1793-1797, (2019).
22. Rasti, Reza, et al. "Macular OCT classification using a multi-scale convolutional neural network ensemble." IEEE transactions on medical imaging 37.4 : 1024-1034, (2017).
23. Qiao, Lifeng, Ying Zhu, and Hui Zhou. "Diabetic retinopathy detection using prognosis of microaneurysm and early diagnosis system for non-proliferative diabetic retinopathy based on deep learning algorithms." IEEE Access 8 : 104292-104302, (2020).
24. Kermany, Daniel S., et al. "Identifying medical diagnoses and treatable diseases by image-based deep learning." Cell 172.5 : 1122-1131, (2018).
25. Hassan, Taimur, et al. "Deep structure tensor graph search framework for automated extraction and characterization of retinal layers and fluid pathology in retinal SD-OCT scans." Computers in biology and medicine 105 : 112-124, (2019).
26. Mohaghegh, Navid, Ebrahim Ghafar-Zadeh, and Sebastian Magierowski. "NGRID: A novel platform for detection and progress assessment of visual distortion caused by macular disorders." Computers in biology and medicine 111 : 103340, (2019).
27. Zago, Gabriel Tozatto, et al. "Diabetic retinopathy detection using red lesion localization and convolutional neural networks." Computers in biology and medicine 116 : 103537, (2020).
28. Qiu, Jiaming, and Yankui Sun. "Self-supervised iterative refinement learning for macular OCT volumetric data classification." Computers in biology and medicine 111 : 103327, (2019).
29. Ajaz, Aqsa, et al. "Association between Optical Coherence Tomography and Fluorescein Angiography based retinal features in the diagnosis of Macular Edema." Computers in biology and medicine 116 : 103546, (2020).
30. Navaneeth, Bhaskar, and M. Suchetha. "PSO optimized 1-D CNN-SVM architecture for real-time detection and classification applications." Computers in biology and medicine 108 : 85-92, (2019).
31. Lekha, S., and M. Suchetha. "A novel 1-D convolution neural network with SVM architecture for real-time detection applications." IEEE Sensors Journal 18.2 : 724-731, (2017).
32. Bhaskar, Navaneeth, and Suchetha Manikandan. "A deep-learning-based system for automated sensing of chronic kidney disease." IEEE Sensors Letters 3.10 : 1-4, (2019).
33. Devarajan, D., Ramesh, S.M. and Gomathy, B. A metaheuristic segmentation framework for detection of retinal disorders from fundus images using a hybrid ant colony optimization. Soft Comput 24, 1334713356 (2020).
34. Wang, C., Zhao, Z. and Yu, Y. Fine retinal vessel segmentation by combining Nest U-net and patch-learning. Soft Comput (2021).

Figures

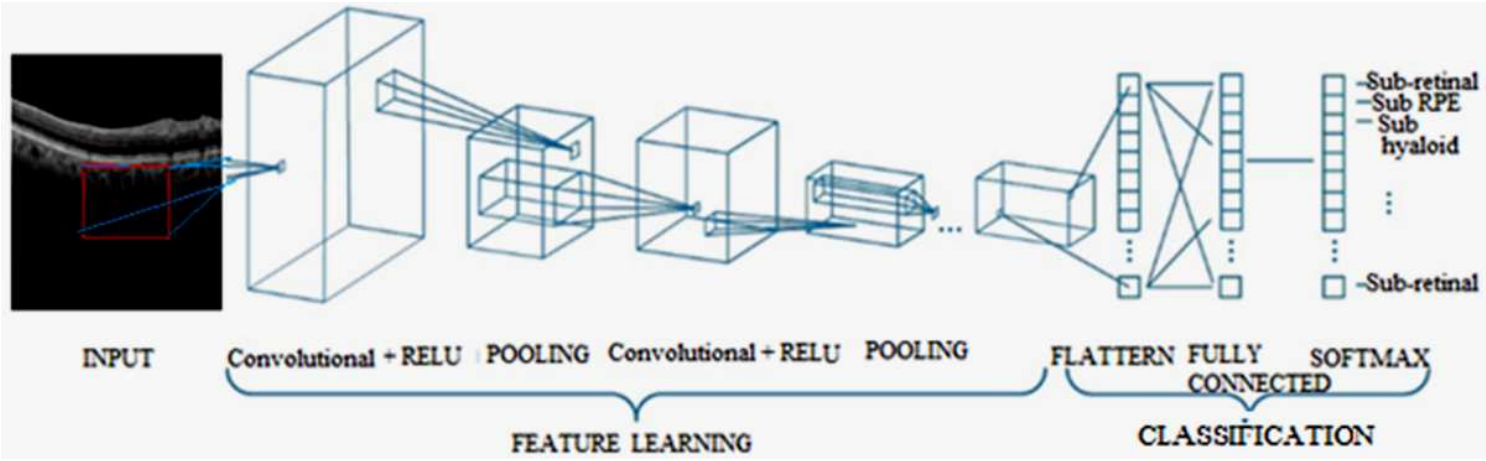

Figure 1

Architecture of CNN

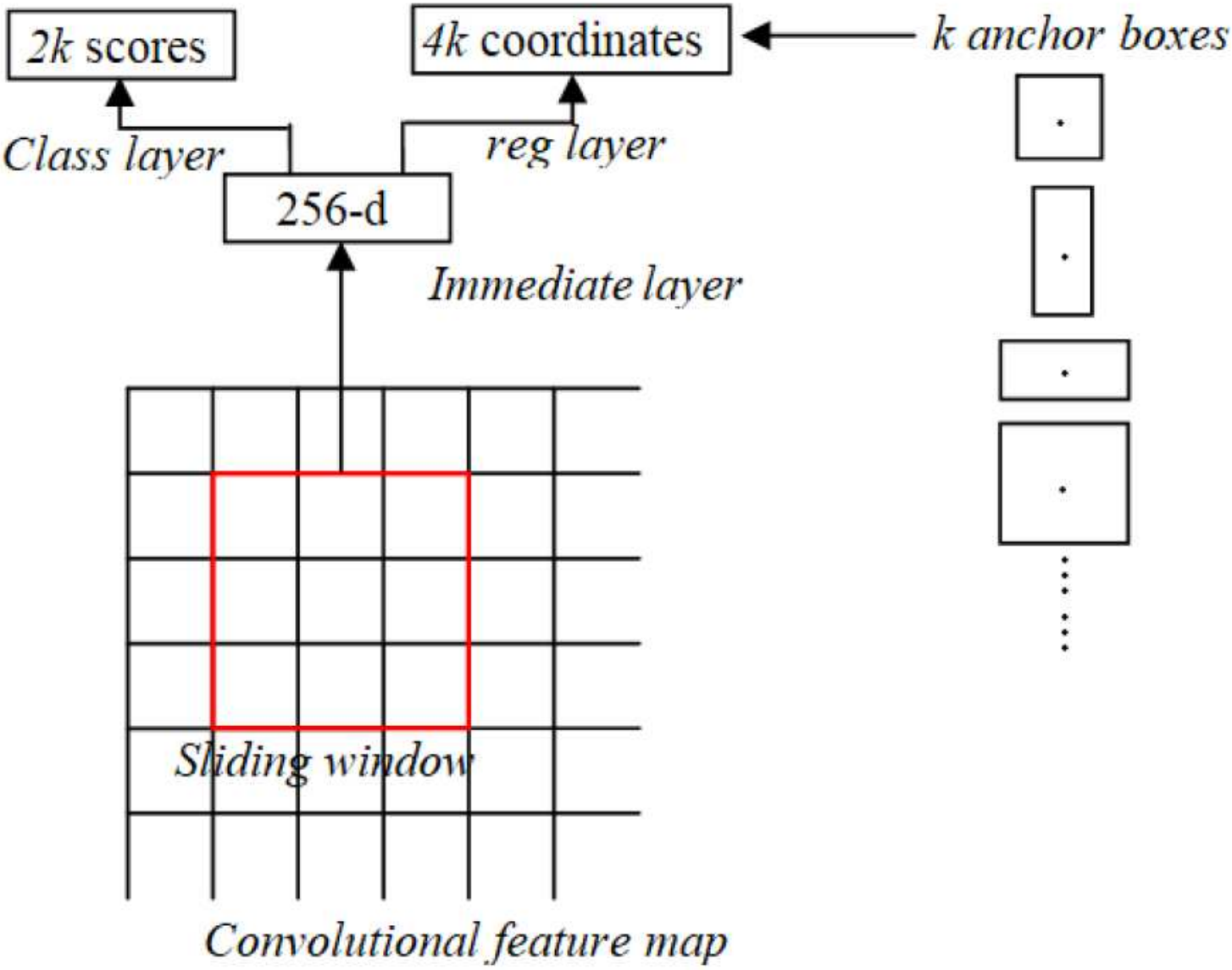

Figure 2

Sliding window in RPN

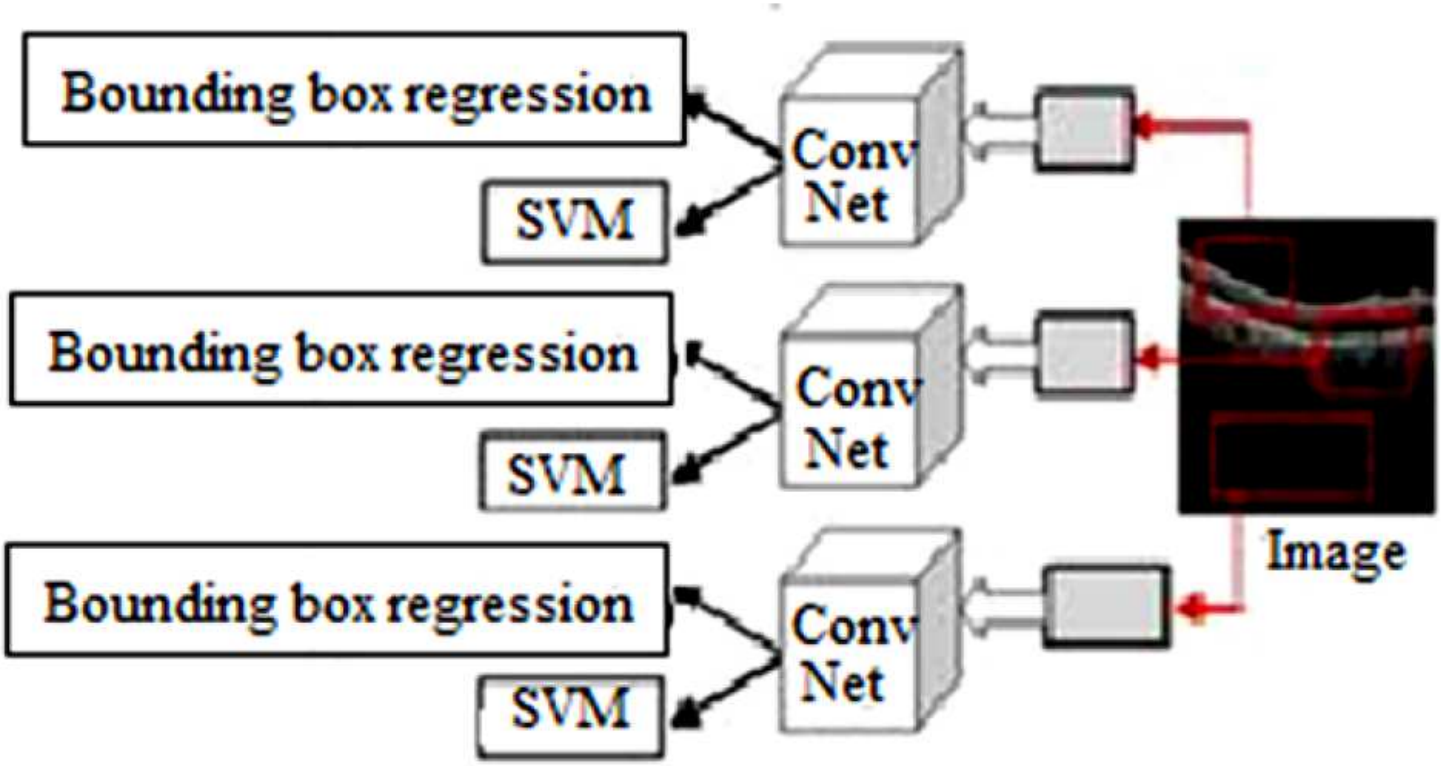

Figure 3

R-CNN architecture

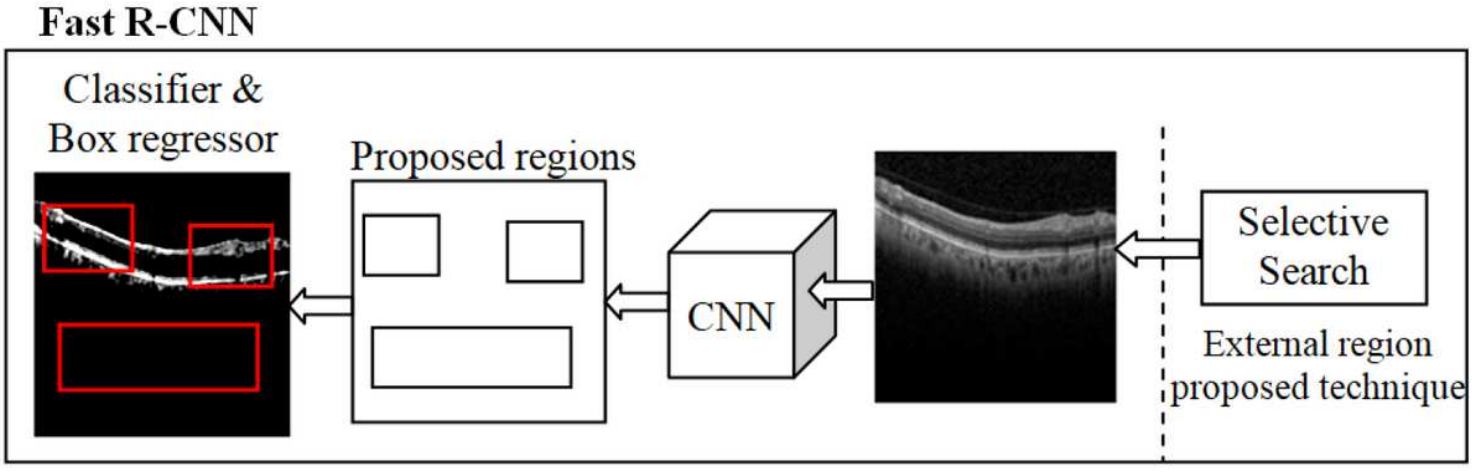

Figure 4

Fast R-CNN architecture

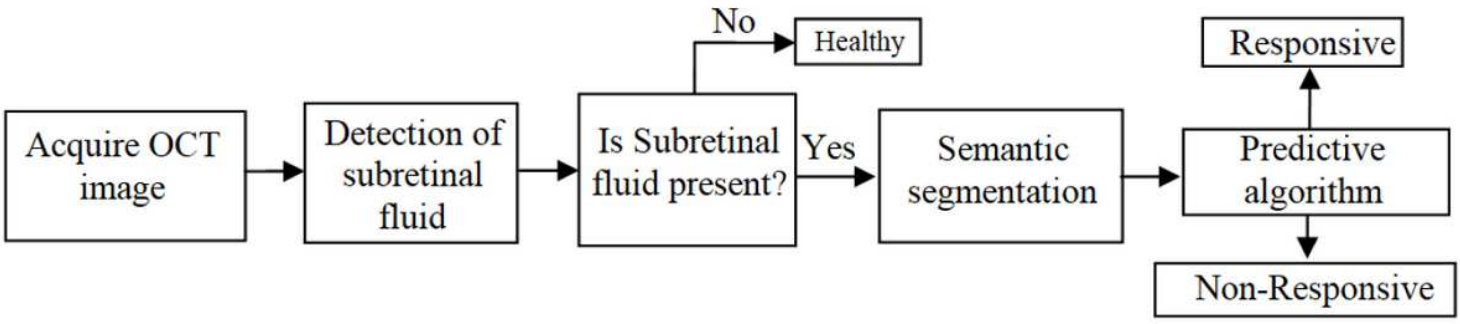

Figure 5

Block diagram of a proposed method

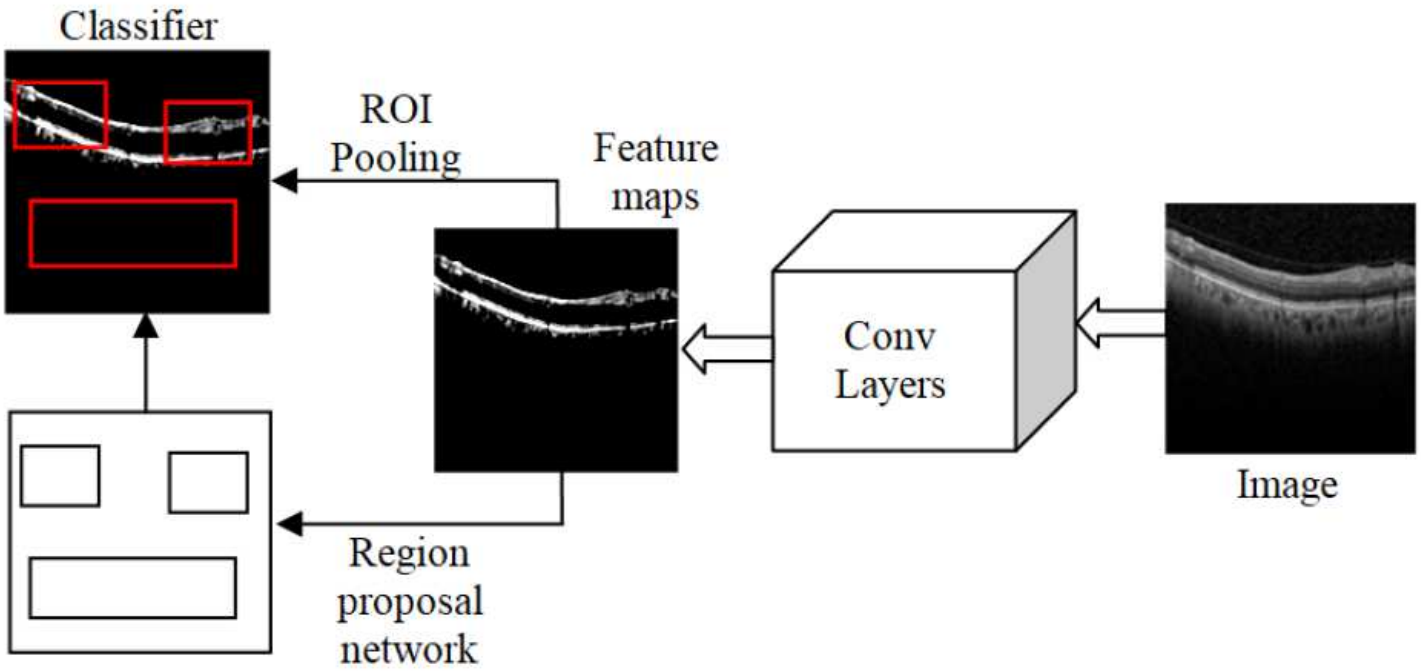

Figure 6

The Architecture of faster R-CNN

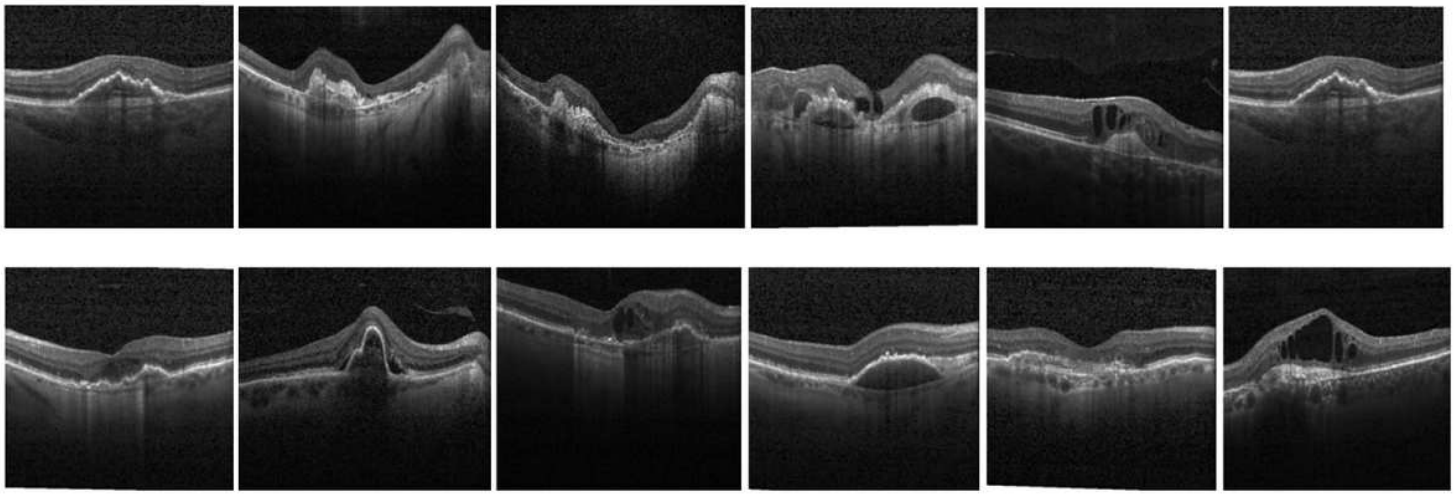

Figure 7

Sample images from Kaggle dataset and Sankara Nethralaya hospital

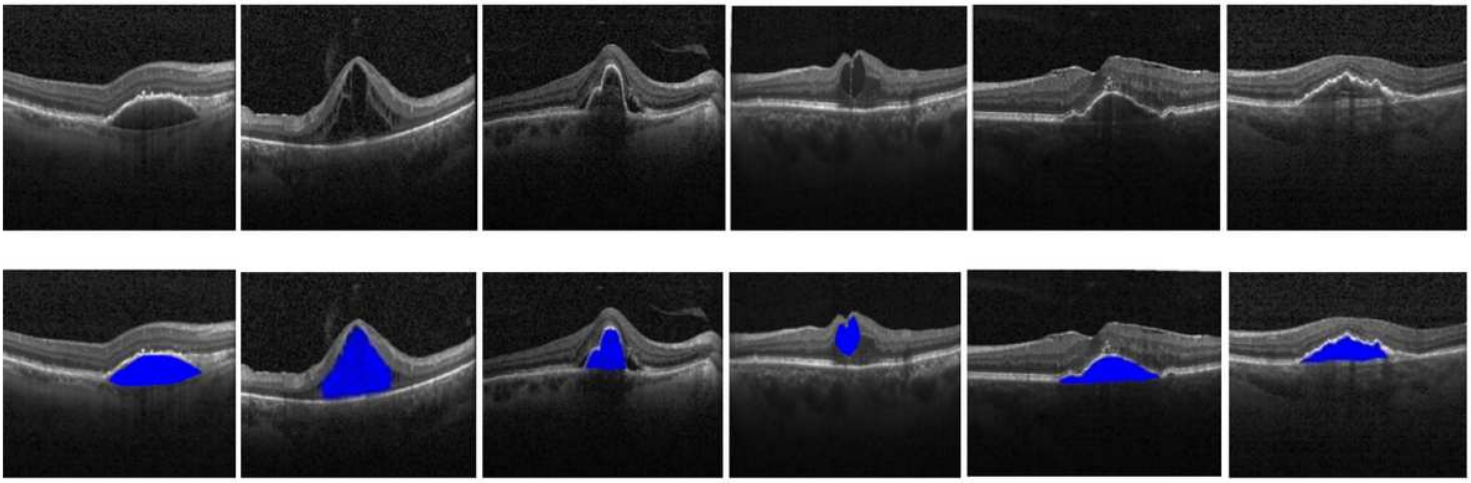

Figure 8

Semantic segmentation results in Kaggle dataset (RowI: Input test images, Row II: Segmented output)

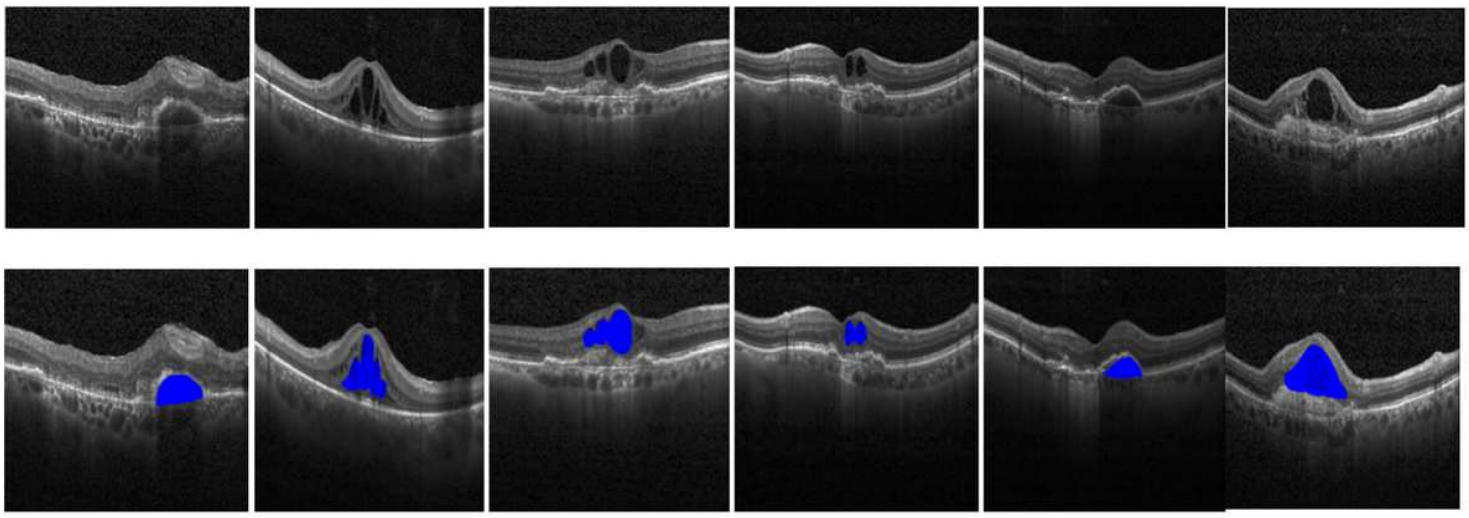

Figure 9

Semantic segmentation results in Sankara Nethralaya hospital images (Row I: Input test images, Row II: Segmented output)

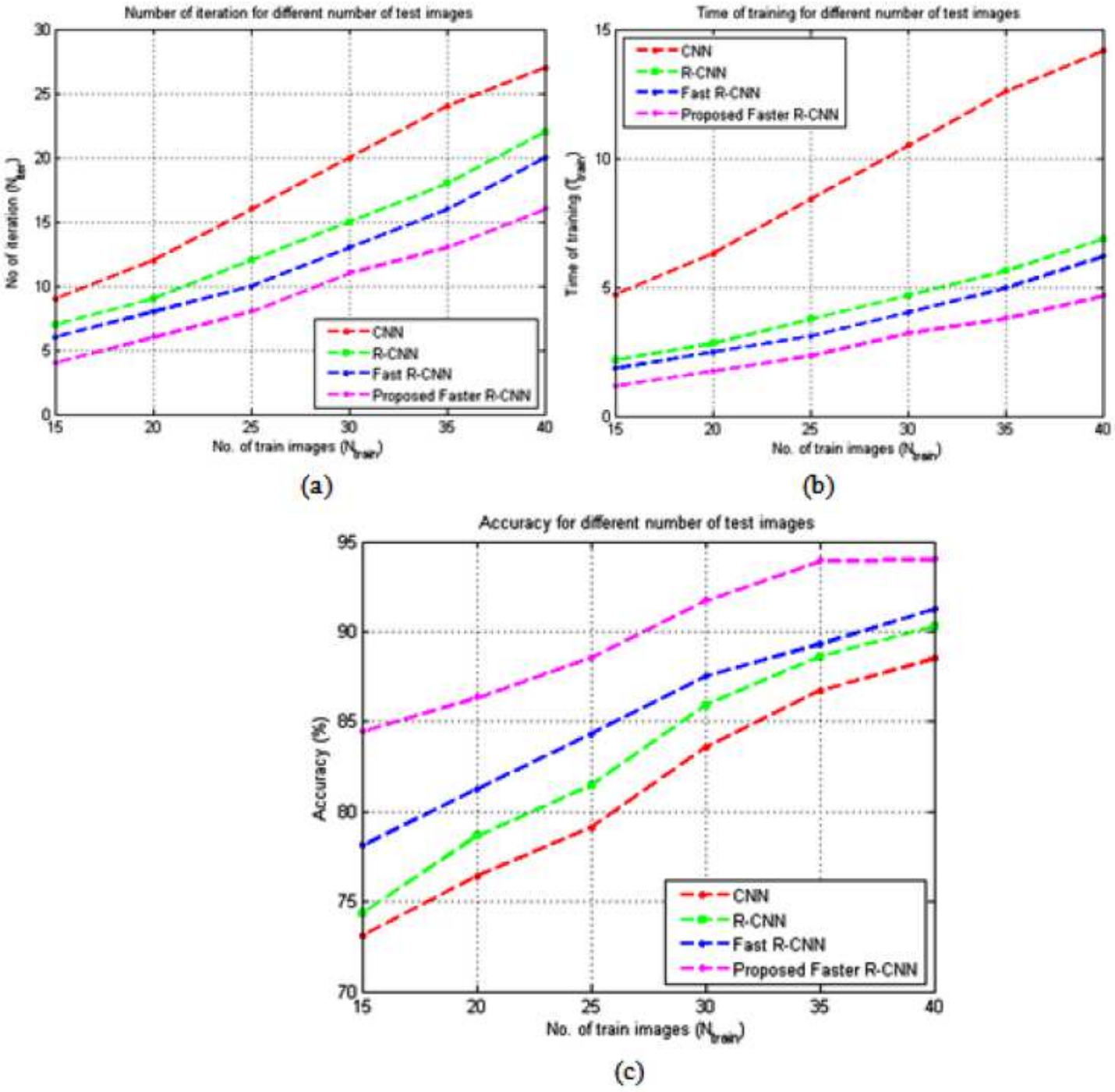

Figure 10

Performance comparison of the proposed method with other methods (a) Number of iterations to complete training (b) Time complexity (in seconds) (c) Accuracy

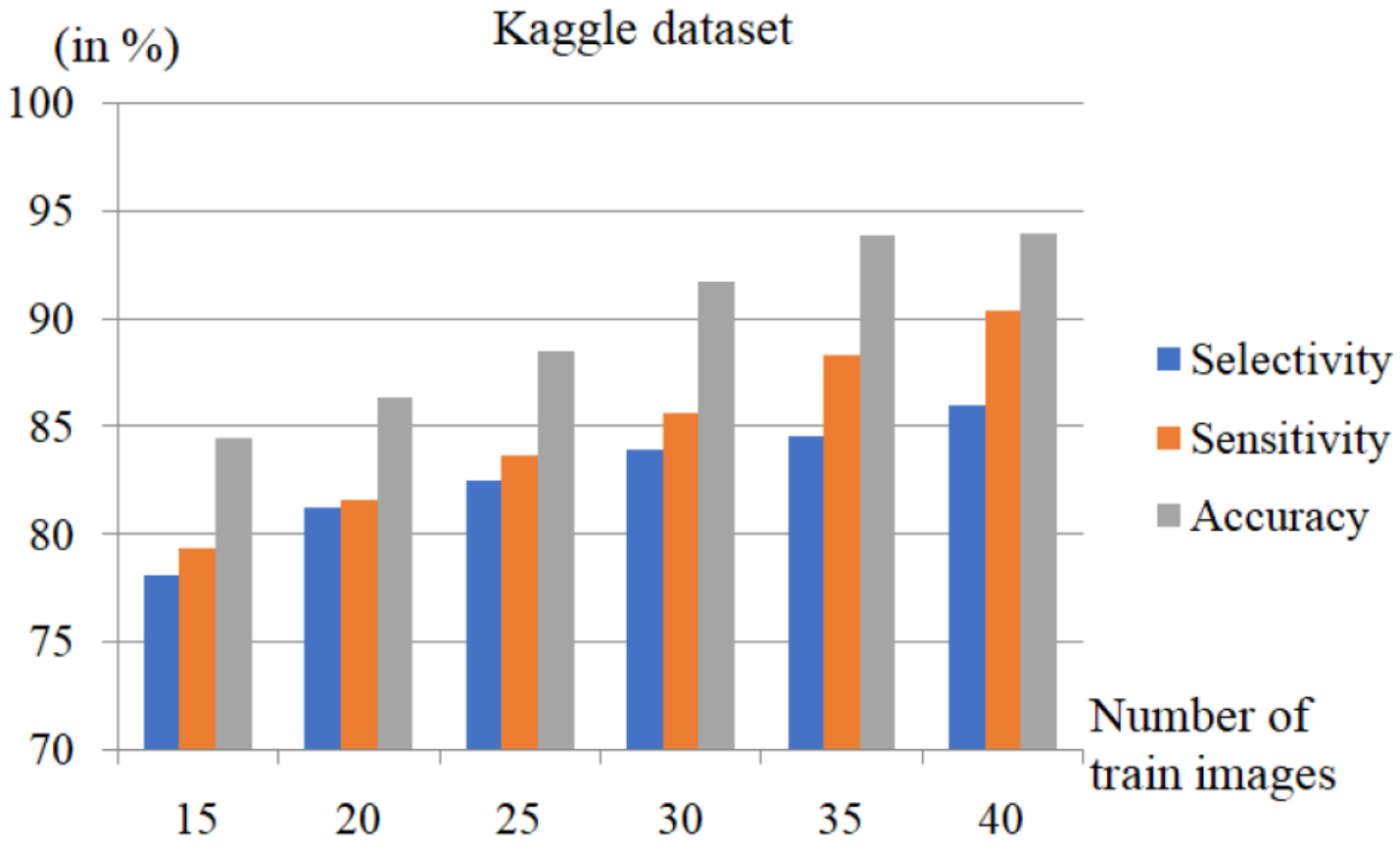

Figure 11

Performance comparison for different number of test images using the Kaggle dataset

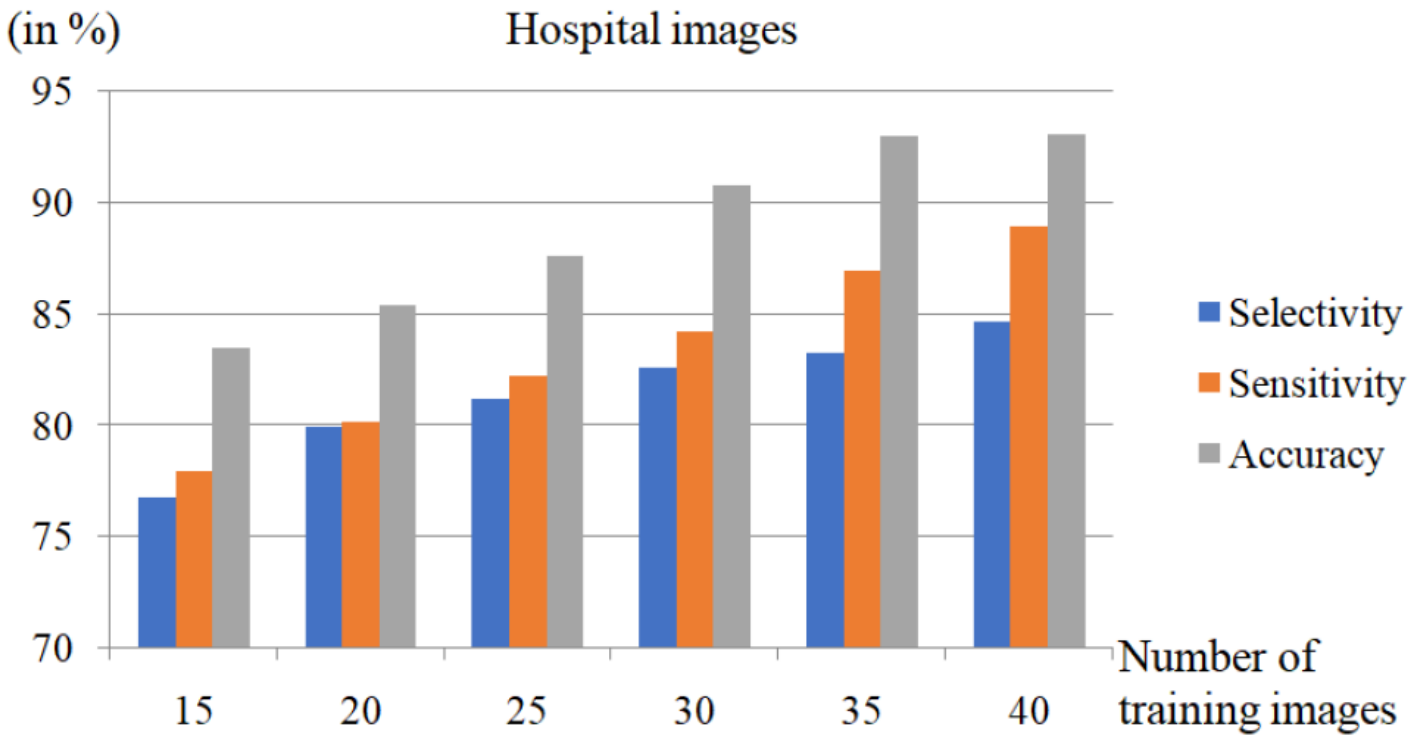

Figure 12

Performance comparison using the images from the hospital for a different number of test images