

Article

Deep Ensemble Learning Based Objective Grading of Macular Edema by Extracting Clinically Significant Findings from Fused Retinal Imaging Modalities

Bilal Hassan ^{1,†}, Taimur Hassan ^{2,†}, Bo Li ^{3,*}, Ramsha Ahmed ⁴ and Omar Hassan ⁵

¹ School of Automation Science and Electrical Engineering, Beihang University (BUAA), Beijing 100191, China

² Department of Electrical Engineering, Bahria University (BU), Islamabad 44000, Pakistan

³ School of Computer Science and Engineering, Beihang University (BUAA), Beijing 100191, China

⁴ School of Computer and Communication Engineering, University of Science & Technology Beijing (USTB), Beijing 100083, China

⁵ Department of Electrical and Computer Engineering, Sir Syed CASE Institute of Technology (SSCIT), Islamabad 44000, Pakistan

* Correspondence: libo@buaa.edu.cn

† These authors contributed equally in this research.

Received: 26 May 2019; Accepted: 26 June 2019; Published: 5 July 2019



Abstract: Macular edema (ME) is a retinal condition in which central vision of a patient is affected. ME leads to accumulation of fluid in the surrounding macular region resulting in a swollen macula. Optical coherence tomography (OCT) and the fundus photography are the two widely used retinal examination techniques that can effectively detect ME. Many researchers have utilized retinal fundus and OCT imaging for detecting ME. However, to the best of our knowledge, no work is found in the literature that fuses the findings from both retinal imaging modalities for the effective and more reliable diagnosis of ME. In this paper, we proposed an automated framework for the classification of ME and healthy eyes using retinal fundus and OCT scans. The proposed framework is based on deep ensemble learning where the input fundus and OCT scans are recognized through the deep convolutional neural network (CNN) and are processed accordingly. The processed scans are further passed to the second layer of the deep CNN model, which extracts the required feature descriptors from both images. The extracted descriptors are then concatenated together and are passed to the supervised hybrid classifier made through the ensemble of the artificial neural networks, support vector machines and naïve Bayes. The proposed framework has been trained on 73,791 retinal scans and is validated on 5100 scans of publicly available Zhang dataset and Rabbani dataset. The proposed framework achieved the accuracy of 94.33% for diagnosing ME and healthy subjects and achieved the mean dice coefficient of 0.9019 ± 0.04 for accurately extracting the retinal fluids, 0.7069 ± 0.11 for accurately extracting hard exudates and 0.8203 ± 0.03 for accurately extracting retinal blood vessels against the clinical markings.

Keywords: biomedical image processing; image analysis; image classification; machine intelligence; machine vision; optical coherence tomography; fundus photography

1. Introduction

Visual impairments severely degrade the quality of life and have an adverse effect on people suffering from other chronic health issues. Currently, blindness is considered as a major health problem worldwide. According to the Global Burden of Disease (GBD) in their 2017 report (released on 18 November 2018), loss of vision is categorized as the third leading form of impairments in humans and 48.2 million people are suffering from eye diseases all over the world. In addition to

this, 39.6 million people have severe visual impairments whereas 279 million people and 969 million people have moderate to low visual impairments, respectively [1,2]. Moreover, most of the visual impairments that were reported are due to retinopathy.

The prime cause of retinopathy is diabetes mellitus (DM). DM is caused due to the destruction of pancreatic beta cells (β -cell) affecting the glucose metabolism of the candidate subject. DM is graded into two types. Type I DM specifies deficiency of insulin whereas Type II DM is associated with insulin resistance [3–5]. Apart from this, DM also affects other vital organs of the human body including eyes, kidney, heart, etc. [6]. Macula produces the central vision and it is the most critical part of retina. Any damage to macula results in the loss of central vision. The retinal diseases that affect the central vision of a person are collectively known as maculopathy. The most common form of maculopathy is ME, which is caused by the leakage of extracellular fluid from hyper-permeable capillaries in the macula of the retina. ME is clinically graded into different stages depending upon the affected area of macular thickening. However, early detection and laser photocoagulation can prevent sudden blindness in most of the cases. Moreover, many retinal complications are often treatable and according to the initiative of “VISION 2020: The Right to Sight”, different measures are being taken to eradicate avoidable blindness by the year 2020 [7]. At the same time, it is equally important to equip ophthalmologists with state-of-the-art retinal computer aided diagnostic systems for efficient detection and grading of retinopathy.

The two non-invasive imaging modalities that are clinically in practice for retinal examination, are OCT and fundus imagery [8]. OCT captures the tissue reflection through light coherency. For retinal examination, a beam is bombarded on the fundus of retina yielding a cross-sectional axial scan (A-scan) [9,10]. The A-scans are joined together to produce a brightness scan (B-scan). Since OCT captures the cross-sectional retina, so early progression of retinopathy can be easily visualized. The early identification of retinopathy effectively leads towards the better treatment. Retinal OCT imagery has revolutionized the clinical examination and eye treatment [11,12]. Figure 1a shows the basic OCT scan acquisition schematics, which is based on Michelson interferometer (MI). In MI, a monochromatic coherent light source is used to penetrate the human eye to produce a cross sectional retinal scan. Beam splitter at the center splits the light source into two separate beams where one beam is directed towards a reference mirror and the other travels to the subject’s eye. These two beams upon reflection get recombined into a single beam producing axial scan at the detector.

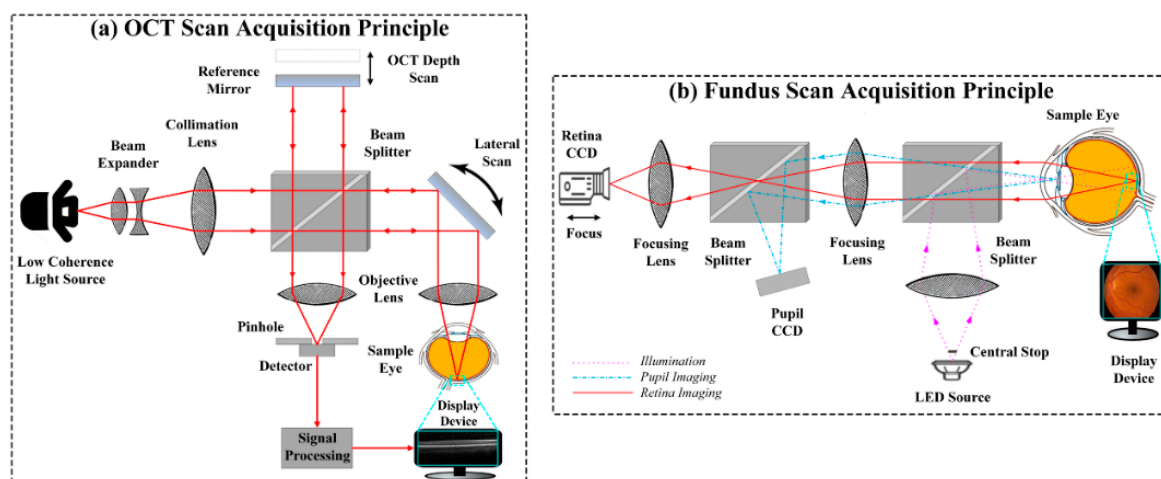


Figure 1. Working principle of retinal imaging modalities. (a) Optical coherence tomography (OCT) scan acquisition principle and (b) the fundus scan acquisition principle.

On the other hand, fundus photography also captures the central and peripheral retinal regions [13]. Figure 1b shows the acquisition principle of fundus imagery where a specialized microscope attached to a charge coupled device (CCD) camera for taking fundus photography. Fundus scans should ideally

be taken in dim conditions. In certain circumstances, it becomes vital to consider all of the retinal examination techniques to fully analyze the pathological conditions of the human retina. The optical principle of the fundus camera is the same as of ophthalmoscopy, which acquires about two to five times enlarged inverted fundus scan [14,15]. The light passes through the series of biconvex lenses, which are used for focusing light to pass through the central aperture forming an annulus. After that, light passes through the cornea and falls on the fundus and hence the fundus scan appears on the display device, which can then be saved. The advantages of fundus photography are: it does not require pupil dilation, it is easy to use, it does not require a skilled user and it captures the images that can easily be examined by specialists at any time anywhere. However, apart from the high cost of equipment and non-portability, a major limitation of fundus photography is that it obtains a 2D representation of 3D semi-transparent retinal tissues projected onto the imaging plane, which is catered through OCT imagery. Figure 2 shows ME visualization in both OCT and fundus scans.

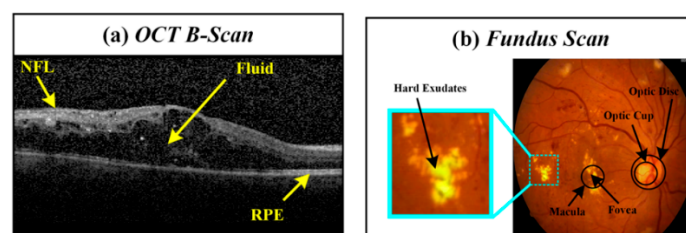


Figure 2. Appearance of macular edema symptoms in (a) an OCT B-scan and (b) fundus scan.

2. Related Work

In the past, many researchers have conducted clinical studies on analyzing ME using fundus and OCT scans [16–18] and concluded that OCT imaging provides better visualization of ME in comparison to fundus photography, especially in early stages where symptoms of ME are not relatively prominent. In addition to this, many studies have been conducted on devising automated algorithms for detecting ME from fundus or OCT scans individually. Most of the methods that use fundus images for the automated detection of ME are based on component segmentation, lesion detection and extraction of hard exudates (HE). Since in digital fundus scans, the contrast between HE and other retinal structures is relatively high, the most common approaches for detecting HE include marker-controlled watershed transformation [19], particle swarm optimization (PSO) based algorithm [20] and by means of local standard variation in a sliding window, morphological closing of the luminance channel and watershed transform [21]. However, illumination variations, which arise because of the changes in tissue pigmentation and imaging conditions, greatly affect these methods. Additionally, the methods based on extracting edge and color features are also proposed over the past for the segmentation of HE [22–25]. In general, such algorithms produce unsatisfactory results without including complex pre and post processing steps.

Different researchers have developed automated frameworks for the extraction of retinal layers and retinal fluids for analyzing ME affected pathologies [26–29]. Kernel regression and graph theory dynamic programming (KR + GTDP) [30] and software development life cycle (SDLC) [31] frameworks are also developed for segmenting retinal layers and retinal fluids in ME affected OCT scans. Srinivasan et al. [32] proposed a maculopathy detection framework using histogram of oriented gradients. Apart from this, deep learning frameworks [33–35] are also proposed recently for the automated extraction of retinal information from maculopathy affected OCT scans.

However, to the best of our knowledge, no method has been proposed in the past that fuses multiple retinal imaging modalities for objective evaluation of ME pathology. In this paper, we proposed a deep ensemble learning based framework that gives the objective grading ME pathology. The main contributions of our papers were as follows:

1. A novel method was presented in this paper that extracted the ME pathological symptoms from retinal fundus and OCT scans.
2. Instead of extracting handcrafted features, the proposed framework employed a deep convolutional neural network (CNN) model that gives the most relevant and useful features from retinal fundus and OCT scans for the objective evaluation of ME pathology irrespective of the scan acquisition machinery.
3. Many frameworks that have been proposed in the past were tested on a single dataset or on scans acquired through single OCT machinery. However, the proposed framework could give objective grading of ME pathology irrespective of OCT acquisition machinery and was rigorously tested on scans from different publicly available datasets.
4. The proposed framework employed an ensemble of artificial neural networks (ANN), support vector machines (SVM) and naïve Bayes (NB) for the in-depth grading of ME using both fundus and OCT retinal imaging modalities.
5. The proposed framework is adaptive and gives more weight to the clinical findings such as foveal swelling, fluid filled spaces and hard exudates while evaluating ME. This is achieved by fine-tuning the proposed CNN model on observing the critical ME symptoms from both fundus and OCT imagery.

Rest of the paper is organized as: Section 3 reports dataset details used in this study, Section 4 explains the proposed methodology, results are presented in Sections 5 and 6 describes the detailed discussion about the proposed framework. Section 7 concludes the paper and highlights the future directions.

3. Datasets

The proposed framework has been tested on retinal fundus and OCT B-scans from multiple publicly available Rabbani and Zhang datasets. Zhang's dataset only consisted of OCT scans of various retinal pathologies while Rabbani's datasets had scans of fundus, fluorescein angiography (FA) and OCT retinal imaging modalities. We excluded the retinal pathologies other than healthy and ME in these datasets. The detailed description of the datasets that were used for training and evaluation purposes is listed in Table 1. All the scans within the datasets were marked by the expert clinicians and we used them as a ground truth in evaluating the performance of the proposed framework.

Table 1. Details of the dataset used for training and testing purposes.

Dataset ^a	Imaging Modality ^b		Retinal Pathology ^b		Scans Dimension(s)	
	OCT	Fundus/FA	Healthy	ME	OCT	Fundus/FA
1 [36]	2764	-	1628	1136	496 × 512	-
2 [37]	-	24	-	24	-	512 × 612 768 × 868
3 [38]	12,800	100	12,900	-	512 × 650	1612 × 1536
4 [39]	-	120	60	60	-	720 × 576
5 [40]	-	35	-	35	-	720 × 576
6 [41]	-	60	25	35	-	720 × 576
7 [35]	62,988	-	51,390	11,598	512 × 496	-
					512 × 512	
					768 × 496	
					1024 × 496	
					1536 × 496	
Total	78,552	339	66,003	12,888	-	-
Split	Training			Validation		
Modality	OCT	Fundus/FA			OCT	Fundus/FA
Total Scans	73,552	239			5000	100
Healthy	63,318	135			2500	50
ME	10,234	104			2500	50

^a We only considered OCT and fundus imaging modalities consisting of healthy and ME retinal pathologies from these datasets; ^b the count shows the total number of scans in these datasets (including all the B-scans in OCT volumes).

4. Proposed Methodology

The proposed framework fuses retinal fundus and OCT imagery for the automated recognition and classification of ME and healthy subjects. The block diagram of the proposed framework is shown in Figure 3 where it can be observed that the proposed framework consisted of five major stages:

- Retinal imaging modality recognition;
- Preprocessing retinal scans;
- Extraction of clinically significant ME pathological symptoms;
- CNN for feature extraction;
- Retinal diagnosis.

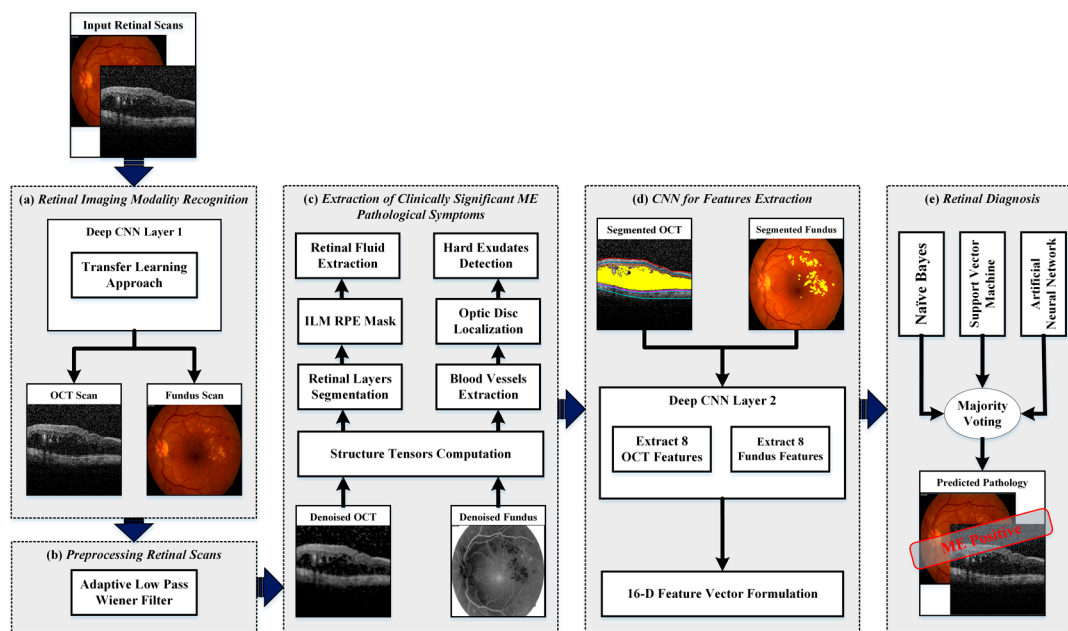


Figure 3. Proposed methodology for automated recognition and classification of healthy and macular edema (ME) affected eyes. (a) Retinal imaging modality recognition; (b) preprocessing retinal scans; (c) extraction of clinically significant ME pathological symptoms; (d) convolutional neural network (CNN) for feature extraction and (e) retinal diagnosis.

At first, the input retinal scans were categorized as fundus or OCT through the first layer of the deep CNN model. Afterwards, different acquisition artifacts and unwanted noise content from both type of imagery were removed through the preprocessing stage. After enhancing the scans, the information about retinal layers, retinal fluids and the hard exudate regions were automatically extracted through the set of coherent tensors, which highlights the clinically significant pathological features of ME retinal syndrome. The extracted retinal information was then mapped on the original scan from which the distinct features were extracted through deep CNN models. The extracted features from both fundus and OCT imagery were concatenated together to form a feature vector upon which the candidate subject was graded. The detailed description of each stage is presented in the subsequent subsections below.

4.1. Retinal Imaging Modality Recognition

The first stage of the proposed framework was related to the automated recognition of retinal fundus and OCT scans. For this purpose, we utilized the pre-trained AlexNet model [42]. AlexNet is a 25-layered CNN architecture that is trained on an ImageNet dataset. We modified the classification layer of the AlexNet network and retrained it on the local image modality recognition training dataset through transfer learning. The transfer learning phase is shown in Figure 4 and the detailed description

of the respective training dataset is presented in Table 1. The pretrained weights of the AlexNet model were very convergent for the recognition of retinal imaging modalities, which resulted in lesser training and fine-tuning time. The optimization during the training phase was performed through stochastic gradient descent (SGD) [43] where two 50% dropout layers were employed to reduce the overfitting. The main reason for employing AlexNet model instead of designing a CNN architecture from scratch is to achieve greater accuracy with the small amount of training dataset in a lesser time duration. Apart from this, the softmax function was used in the modified AlexNet architecture to compute final output probabilities. The softmax function is mathematically expressed in Equation (1) and the architectural description about AlexNet layers is presented in Table 2.

$$\sigma(X)_i = \frac{e^{x_i}}{\sum_{j=1}^N e^{x_j}} \quad (1)$$

where, $X = \{x_1, x_2, \dots, x_N\}$ is the input vector. After each convolution layer, the rectified linear units (ReLU) layer is employed that ensures that only the positive values retain in the feature map (because the negative values reflect the changes, which are dissimilar within the input and the convolutional kernel). After the ReLU layer, the max pooling layer has been added, which only keeps the maximum values within the neighborhood, which ultimately shrinks the resultant feature map.

Table 2. Function wise details of the AlexNet architecture.

Function	Layer	Description
Input Image	1	227 × 227 × 3 images
Convolution	2	9611 × 11 × 3 convolutions
	6	2565 × 5 × 48 convolutions
	10	3843 × 3 × 256 convolutions
	12	3843 × 3 × 192 convolutions
	14	2563 × 3 × 192 convolutions
ReLU	3	Assigns '0' to non-positive values
	7	
	11	
	13	
	15	
	18	
Max Pooling	5	3 × 3 max pooling
	9	
	16	
Dropout	19	50% dropout
	22	
Normalization	4	Five channels per element
	8	
Fully Connected	17	4096 fully connected layer
	20	
	23	
Softmax	24	Softmax activation function
Output	25	Two Classes (OCT, Fundus)

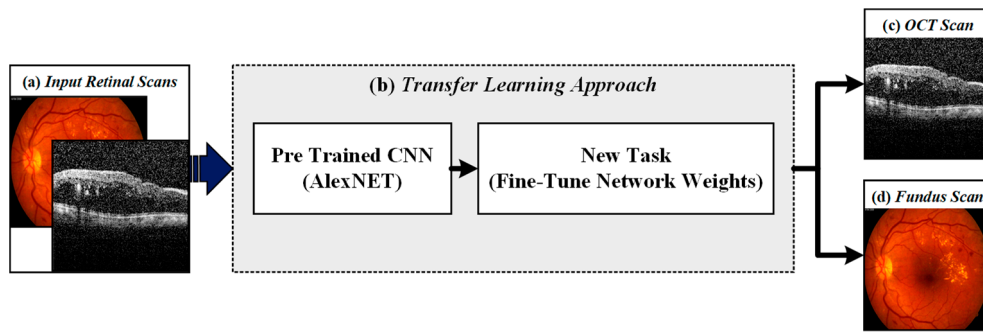


Figure 4. Retinal imaging modality recognition stage. (a) Input retinal scans; (b) transfer learning approach using AlexNet CNN architecture; (c) recognized OCT scan and (d) recognized fundus scan.

4.2. Preprocessing Retinal Scans

The acquisition of retinal scans is highly sensitive to the subject's head and eye movements and this often leads towards the scan degradation. Apart from this, the acquisition machines add different kind of scan annotations, which greatly affects the automated retinal analysis. In order to cater such noisy artifacts, a preprocessing stage was added, which removes the noisy contents effectively, while enhancing the retina. Since the annotations are mostly added in the top and bottom rows of the respective B-scan. They are automatically removed by setting the first and last fifty rows to zero. This threshold was empirically selected by analyzing the scans within all the datasets. Apart from this, the degraded scan areas as shown in Figure 5 were automatically removed by searching for the first and last highly sharp transitions for each column within the respective scan and then by setting the values in the identified noisy regions with the mean of background pixels.

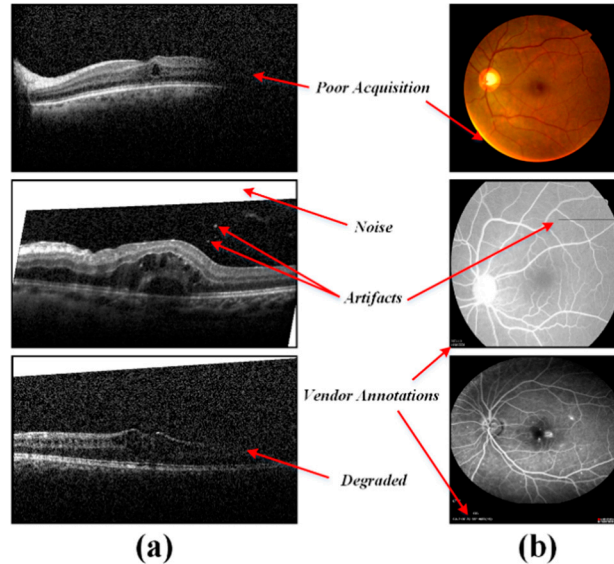


Figure 5. Degraded retinal scans. (a) OCT Scans and (b) fundus/ fluorescein angiography (FA) scans.

The preprocessing stage further enhances retinal portions by increasing their variability with the background and also by removing the noisy outliers. This is accomplished through an adaptive low pass Wiener filter, which uses a localized neighborhood of a candidate pixel for denoising [34]. The response of the Wiener filter is expressed in Equations (2)–(4):

$$\hat{\uparrow} = \frac{1}{w_h w_v} \sum_{x_i \in w_h} \sum_{y_i \in w_v} O(x_i, y_i), \quad (2)$$

$$f^2 = \frac{1}{w_h w_v} \sum_{x_i \in w_h} \sum_{y_j \in w_v} O^2(x_i, y_j) - \hat{\mu}^2, \quad (3)$$

$$D(x_i, y_j) = \hat{\mu} + \frac{f^2 + a^2}{f^2} (O(x_i, y_j) - \hat{\mu}), \quad (4)$$

where, $O(x_i, y_j)$ and $D(x_i, y_j)$ represent the pixels of the original and denoised retinal scan respectively, w_h is the horizontal axis of denoising window while w_v is the vertical axis of denoising window. Local estimated mean and variance are represented by $\hat{\mu}$ and f^2 respectively and a^2 is the average of all estimated mean values.

4.3. Extraction of Clinically Significant ME Pathological Symptoms

ME is clinically graded into different categories as defined by the Early Treatment Diabetic Retinopathy Study (EDTRS). ME due to the presence of hard exudates and retinal fluids within the foveal diameter of 500 micrometers was considered to be clinically significant. ME outside this region was considered as non-clinically significant. Therefore, the accurate extraction of hard exudates, retinal fluid regions and the localization of fovea were very critical for effectively grading ME. Retinal fluids could be accurately observed through OCT scans while hard exudates were effectively visualized through fundus images. Therefore, the proposed framework, rather than relying on either fundus or OCT imagery, used both of them to effectively extract the retinal information for the reliable and objective grading of ME. In order to localize fovea, the proposed framework extracted the retinal layers from the OCT volume and measured the deepest inner limiting membrane (ILM) point within the foveal B-scan.

The extraction of retinal information from both types of imagery was performed through structure coherence matrix, also known as structure tensor. Structure tensor has gained tremendous popularity in medical image processing because it provides low-level feature analysis and it is very useful for detecting corners, edges and boundaries [44]. Structure tensor also known as Förstner interest operator, is a second moment matrix which computes the gradients of an image by using Gaussian derivative filters as expressed in Equations (5)–(8):

$$f\mathcal{T} = \begin{bmatrix} T_{\text{SS}}^2 & T_{\text{S}\dagger} \\ T_{\dagger\text{S}} & T_{\dagger\dagger}^2 \end{bmatrix}, \quad (5)$$

$$T_{\text{SS}}^2 = \sum_{x_i \in g_x} \sum_{y_j \in g_y} g(x_i, y_j) [\varphi_{\text{S}} D(x - x_i, y - y_j)]^2, \quad (6)$$

$$T_{\dagger\dagger}^2 = \sum_{x_i \in g_x} \sum_{y_j \in g_y} g(x_i, y_j) [\varphi_{\dagger} D(x - x_i, y - y_j)]^2, \quad (7)$$

$$T_{\text{S}\dagger} = T_{\dagger\text{S}} = \sum_{x_i \in g_x} \sum_{y_j \in g_y} g(x_i, y_j) [\varphi_{\text{S}\dagger} D(x - x_i, y - y_j)], \quad (8)$$

where, $f\mathcal{T}$ is the second order structure tensor matrix. T_{SS}^2 is the horizontally computed tensor, $T_{\dagger\dagger}^2$ is the vertically computed tensor and $T_{\text{S}\dagger}$, $T_{\dagger\text{S}}$ are the horizontal and vertical oriented tensors. φ_{S} , φ_{\dagger} and $\varphi_{\text{S}\dagger}$ are the partial derivative of denoised image within the pixel neighborhood with respect to x , y and both x , y orientations. $g(x, y)$ is the Gaussian window and $D(x, y)$ is the de-noised retinal scan. Figure 6 shows the structure tensor computational stage. Structure tensor uses a set of eigenvalues to measure the degree of coherency and the tensor with maximum coherency is automatically selected for extracting retinal information [34].

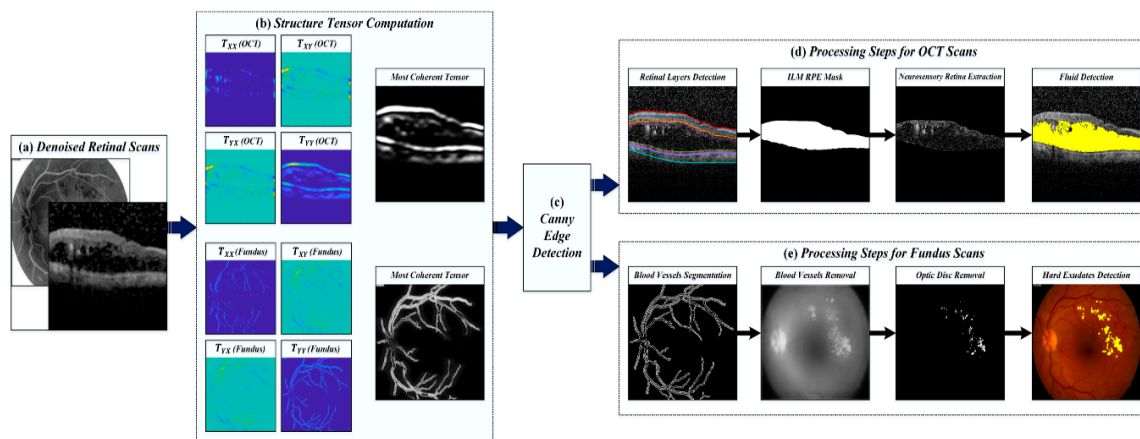


Figure 6. Extraction of clinically significant ME pathological symptoms. (a) Denoised retinal scans; (b) structure tensors computation stage; (c) detection of edges using the canny edge detection method; (d) processing steps for OCT scans and (e) processing steps for fundus scans.

After preprocessing the retinal fundus and OCT scans, the second moment matrix was automatically computed by the proposed framework for further analysis. $\int \mathcal{T}$ from the retinal fundus scan was computed for the extraction of blood vascular patterns. Afterwards, the optic disc region was automatically localized by analyzing the high intensity retinal regions. The extraction of blood vessels and localization of optic disc region was performed in order to improve the segmentation of hard exudates regions. Since blood vessels contain high frequency components, so, the tensors present their detailed visualization while suppressing all other contents as evident from Figure 6b. After computing $\int \mathcal{T}$ of the candidate fundus scan, the four coherent tensors were obtained. The best tensor (T_{MAX}) was then obtained by fusing T_{XX} and T_{YY} tensors, which together contained gives the maximum information about the blood vessels. Blood vessel segmentation in the proposed framework is quite robust as it can easily extract small blood capillaries as well, which are not even visible to the naked eye as shown in Figure 6e.

Structure coherence matrix of the retinal OCT scans is computed for the extraction of up to nine retinal layers [34]. Since most of the retinal layers are horizontally oriented so T_{YY} will be the most coherent tensor in $\int \mathcal{T}$ for extracting layers information as evident from Figure 6b. After extracting the nine retinal layers, ILM and the retinal pigment epithelium (RPE) layers were used to generate a retinal mask, which was then multiplied by the candidate OCT B-scan for the extraction of retinal fluids [34]. The extracted retinal information was then overlaid onto the respective scan for the extraction of clinically significant feature set by the proposed CNN model.

4.4. CNN for Feature Extraction

After extracting the hard exudates, retinal layers and retinal fluids, they were marked on the respective fundus and OCT scans for computing the distinct features to discriminate between healthy and ME affected subjects. These features were extracted through proposed CNN architecture. We designed a 14 layered structure tensor influenced CNN architecture containing one input layer, three convolution layers, three batch normalization layers, three ReLUs, two max pooling layers, one dropout layer with 50% threshold and one fully connected layer. The kernels within the convolution layers of the proposed CNN architecture contained weights that retain the structure tensor-based features while suppressing other content. This gave the significant variability between ME and healthy subjects. The proposed CNN model for feature extraction was designed from scratch and was trained on more than 0.07 million scans where the optimization was performed through SGD. In the proposed CNN model, the negative convolution sum values were removed through ReLU and the max pooling layer shrank the feature map to avoid unnecessary calculations. Since retinal fundus and OCT scans showed different clinically significant ME findings, therefore, the proposed CNN architecture extracted

distinct features from both imaging modalities (i.e., it extracted eight distinct features from retinal fundus scan and eight distinct features from OCT images), which were then concatenated together to generate a 16-D feature vector. These sixteen features were then used to grade healthy and ME subjects. The proposed CNN model shows promising results of feature extraction after getting trained on the dataset mentioned in Table 1. This was due to the robust extraction of retinal information, which were mapped on the retinal scans from which proposed that the CNN model generates the most meaningful and distinctive features as shown in Figure 7. The detailed configuration of the proposed CNN model for feature extraction is presented Table 3, while Table 4 contains the sixteen extracted features from some of the healthy and ME affected scans. Figure 7 shows detailed CNN model for features extraction from both imaging modalities.

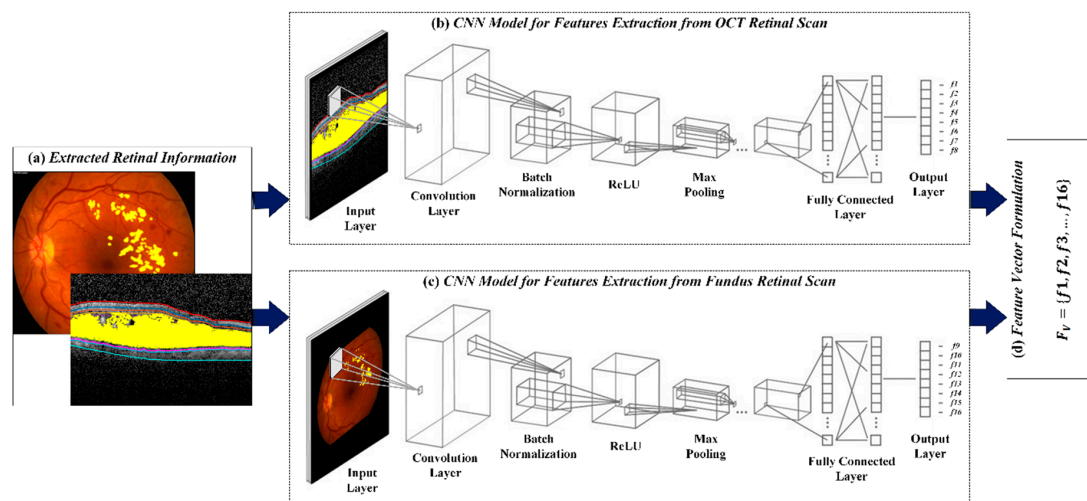


Figure 7. CNN for feature extraction. (a) Overlaid retinal scans with ME pathological symptoms; (b) CNN model for feature extraction from OCT retinal scans; (c) CNN model for feature extraction from retinal fundus scans and (d) 16-D feature vector containing eight OCT (f_1, f_2, \dots, f_8) and eight fundus features ($f_9, f_{10}, \dots, f_{16}$).

Table 3. Proposed structure tensor influenced CNN architecture.

1	Input layer	$227 \times 227 \times 3$ images with 'zerocenter' normalization
2	Convolution	$89 \times 9 \times 3$ convolutions with stride [1,1] and padding 'same'
3	Batch Normalization	Batch normalization with eight channels
4	ReLU	Rectified Linear Units
5	Max Pooling	2×2 max pooling with stride [2,2] and padding [0,0,0,0]
6	Convolution	$169 \times 9 \times 8$ convolutions with stride [1,1] and padding 'same'
7	Batch Normalization	Batch normalization with 16 channels
8	ReLU	Rectified Linear Units
9	Dropout	50% Dropout
10	Max Pooling	2×2 max pooling with stride [2,2] and padding [0,0,0,0]
11	Convolution	$329 \times 9 \times 16$ convolutions with stride [1,1] and padding 'same'
12	Batch Normalization	Batch normalization with 32 channels
13	ReLU	Rectified Linear Units
14	Fully Connected	Fully connected layer giving the significant eight features

Table 4. Selected fundus and OCT features from healthy and ME subjects.

Features	Healthy				Macular Edema				
	Case 1	Case 2	Case 3	Mean ^a	Case 1	Case 2	Case 3	Mean ^a	
OCT	F1	2.28	1.53	1.69	1.51	1.97	-0.79	-2.62	1.17
	F2	2.13	4.35	1.58	2.18	0.58	0.63	1.29	0.86
	F3	-3.13	-1.92	-6.4	-3.57	-0.85	2.73	1.73	-1.56
	F4	4.83	0.84	1.48	2.87	-0.29	1.3	-3.38	0.19
	F5	0.59	0.28	0.44	0.29	1.9	1.35	0.76	0.68
	F6	4.82	1.87	7.59	4.46	2.73	3.29	2.59	2.93
	F7	1.03	0.37	2	0.71	-0.07	0.64	0.19	0.26
	F8	-0.65	0.94	0.58	-0.69	1.61	1.34	1.4	1.38
Fundus	F9	-1.85	0.61	-1.62	-0.95	0.3	2.3	0.1	0.27
	F10	2.67	1.4	2.51	2.93	1.28	1.42	1.17	1.09
	F11	-1.77	-7.83	-1.7	-1.14	-2.54	-2.11	-3.06	-2.71
	F12	1.29	0.18	0.01	0.26	1.44	1.13	1.46	1.43
	F13	1.61	-0.09	0.66	0.25	0.02	-2.49	0.81	-0.11
	F14	2.37	1.65	2.44	0.81	2.75	4.61	2.24	3.12
	F15	1.7	0.14	0.68	0.72	-1.04	1.46	-0.26	0.07
	F16	1.26	0.82	1.25	1.17	0.8	0.62	0.98	0.57

^a Mean value was computed using all the scans in a validation dataset.

4.5. Retinal Diagnosis

After extracting the sixteen clinically significant features from retinal fundus and OCT imagery, they were concatenated together and were utilized by the hybrid classification system for grading ME. The hybrid classification model in the proposed framework consisted of an ensemble of ANN, SVM and NB. The final decision was computed by measuring the majority votes of all three classification models. The description of each classification model is presented below.

4.5.1. Artificial Neural Networks

In this study, we used a feed forward artificial neural network classifier with one input layer, one output layer and two hidden layers. The input layer consisted of 16 nodes as per the extracted features. For hidden layers, we experimented with two to 40 nodes to find the optimum architecture (12 for the 1st hidden layer and nine for the 2nd hidden layer) of the neural network. A single output layer node gave the final classification probability. The sigmoid function was used for activation in each hidden layer whereas the final output layer contained softmax as the activation function. The weights during training were updated through gradient descent. Figure 8 shows the architecture of ANN used in the proposed study.

4.5.2. Support Vector Machines

We used a SVM classifier as well in the proposed classification model. SVM is among the most extensively used classifier [34], and in this research a non-linear decision boundary was computed through Gaussian radial basis function (RBF) and multilayer perceptron (MLP) hyperplanes for predicting ME and healthy subjects based on the extracted feature vector (F_V).

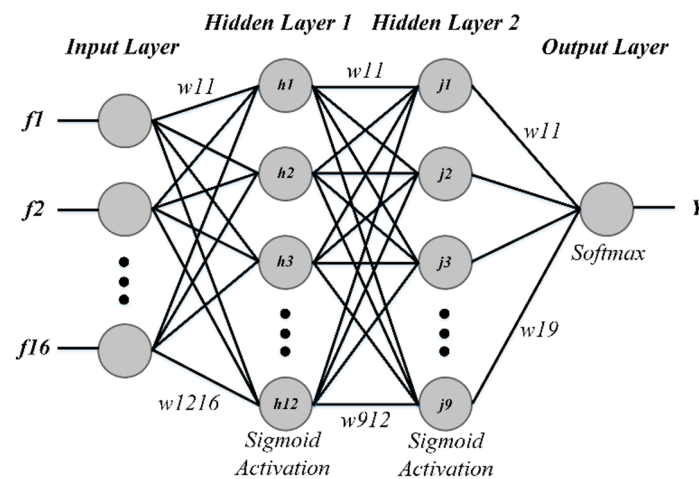


Figure 8. Architecture of an ANN model.

4.5.3. Naïve Bayes

NB is a probabilistic classifier, which makes a decision based on the maximum a posteriori (MAP) rule. In this study, we used the NB classifier to determine the probability of ME and healthy classes through a 16-D feature vector. The category with the maximum probability was then automatically chosen as a diagnosis for the respective feature vector. The probabilities were computed through Bayes Rule as expressed in Equations (9) and (10):

$$P(c_i|F_v) = P(F_v|c_i)P(c_i)/P(F_v), \quad (9)$$

$$Y = \arg_{c_i} \max[P(c_i|F_v)], \quad (10)$$

where, c_i represents the healthy and ME class, F_v is the 16-D test feature vector formulated during the feature extraction stage and Y represents the class assigned to the unlabeled scan, which has the largest probability given the F_v . F_v contains eight distinct features from the retinal fundus scan and eight distinct features from the OCT scan. We used Gaussian distribution to calculate the likelihood $P(F_v|c_i)$.

The detailed block diagram of classifiers training stage is shown in Figure 9. We used around 0.07 million retinal scans for training the hybrid classifier. Details of the training dataset are mentioned in Table 1. At first, sixteen distinct features were extracted from the labeled training scans to form a 16-D feature vector, which was then passed to all three classifiers separately and their decisions were finalized through majority voting. The performance of the proposed hybrid classifier during training was measured through K-fold cross validation as shown in Table 5 for different values of k. Once the classifiers achieved the desirable accuracy, they were used for retinal diagnosis of unlabeled scans during the classification stage as shown in Figure 9. Algorithm 1 summarizes the working flow of our proposed framework.

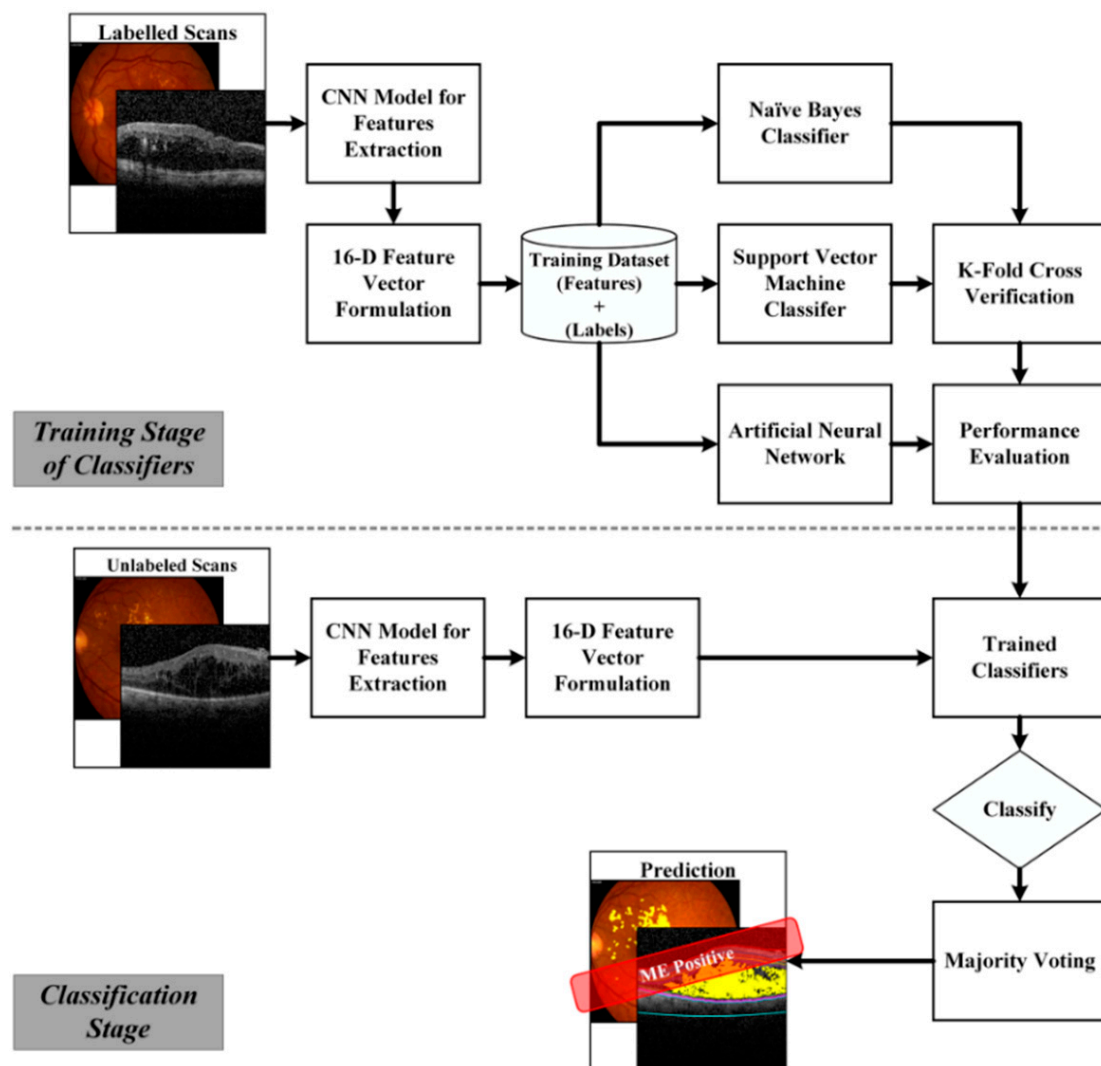


Figure 9. Block diagram of a hybrid classifier training and classification.

Table 5. Classifiers K-Fold cross validation performance.

K	Max Accuracy Achieved		
	ANN	SVM	NB
2	0.816	0.809	0.794
3	0.893	0.841	0.829
4	0.914	0.874	0.864
6	0.948	0.907	0.917
8	0.972	0.925	0.942
10	0.991	0.966	0.980
11	0.985	0.948	0.961
12	0.973	0.929	0.944

Algorithm 1: Proposed Framework**<START>**

- 1 Load the input retinal scan
- 2 Pass the input retinal scan to CNN Layer 1 to classify it as OCT or fundus scan
- 3 Preprocess the OCT/fundus scan using 2D adaptive low pass Wiener filter
- 4 **if** (input retinal scan == OCT)
 - then** Compute four tensors using 2D structure tensor approach
 - Select the most coherent tensor
 - Apply Canny edge detection
 - Segment nine retinal layers
 - Generate ILM-RPE mask
 - Extract neurosensory retina
 - Detect Fluid
 - Overlay the segmented retina layers and fluid on respective OCT scan
 - else** (input retinal scan == fundus)
 - then** Compute four tensors using 2D structure tensor approach
 - Select the most coherent tensor
 - Apply Canny edge detection
 - Segment and remove the blood vessels
 - Detect and remove the optic disc
 - Detect the hard exudates
 - Overlay the segmented blood vessels and hard exudates on respective fundus scan
- end**
- 5 Pass the retinal scan with overlaid retinal information to CNN Layer 2 for extraction of features
- 6 **if** (retinal scan == OCT)
 - then** Extract features (f_1, f_2, \dots, f_8)
 - else** (retinal scan == fundus)
 - then** Extract features $(f_9, f_{10}, \dots, f_{16})$
- end**
- 7 Formulate feature vector $F_v = \{f_1, f_2, \dots, f_{16}\}$
- 8 Pass the 16D F_v to hybrid ensemble machine learning classification system for ME grading
 - A = Decision (Artificial Neural Network)
 - B = Decision (Support Vector Machine)
 - C = Decision (NB)
 - Retinal Pathology = majority vote (A, B, C)

<END>

5. Results

We tested the proposed framework on an unlabeled dataset consisting of 5000 OCT B-scans out of which 2500 were of ME affected eyes and 2500 were of healthy eyes and 100 fundus scans with the same ratio of ME and healthy eyes. Since the feature vector is generated by concatenating the extracted features from both fundus and OCT scans, therefore, we individually computed the mean dice coefficient for measuring the performance of extracting hard exudates, blood vessels and retinal fluids.

The dataset in [37] consisted of 24 diabetic macular edema eyes with seven diffuse pattern of fluid leakage, 10 focal pattern of leakage and seven mixed pattern of leakage. They also provided three different expert markings of hard exudates for all these cases, which we used in validating the performance of the proposed system in extracting hard exudates regions as shown in Table 6. It can be observed from Table 6 that the proposed framework achieved the overall mean dice coefficient of 0.7069 ± 0.11 in extracting hard exudates. Figure 10 shows the visual comparison of the proposed framework for extracting hard exudates with three different expert markings.

Since the annotations against blood vessels in fundus/FA scans and retinal fluids in OCT scans were not available in the datasets used in this study, we arranged these annotations through a local expert clinician for comparative analysis. We evaluated the efficiency of the proposed system for blood vessels extraction through mean dice coefficient computed against the manual markings done by a local clinician as shown in Table 7. We obtained the overall mean dice coefficient of 0.8589 ± 0.04 for blood vessels segmentation in the case of healthy eyes and 0.8012 ± 0.03 in the case of ME affected eyes. Whereas, for both retinal conditions we achieved the overall mean dice coefficient of 0.8203 ± 0.03 . These results validate the accuracy of proposed systems in blood vessels segmentation against various retinal pathologies, even in the presence of hard exudates, hemorrhages and micro-aneurysms in ME fundus/FA scans. It shows the effectiveness of the proposed method in detailed extraction of blood vessels. Figure 11 shows the extracted blood vessels by the proposed system in healthy and ME affected fundus/FA scans.

Similarly, we evaluated the performance of the proposed system for the extraction of retinal fluids through mean dice coefficient computed against the manual markings done by a local clinician as shown in Table 8. We obtained the overall mean dice coefficient of 0.9026 ± 0.03 for retinal fluid extraction on the Rabbani dataset [36] and 0.9012 ± 0.04 on the Zhang dataset [35].

Table 6. Mean dice coefficient for hard exudates segmentation against the expert markings [37].

Leakage Pattern	Scans	Against Grader 1	Against Grader 2	Against Grader 3
Diffuse	1	0.7529	0.8931	0.8372
	2	0.3573	0.5973	0.6698
	3	0.6838	0.6727	0.6744
	4	0.6718	0.7391	0.7739
Focal	1	0.5871	0.6397	0.6928
	2	0.2339	0.7288	0.694
	3	0.7169	0.8275	0.8631
	4	0.3887	0.6349	0.634
Mixed	1	0.6035	0.882	0.875
	2	0.5941	0.6691	0.7329
	3	0.6551	0.7938	0.8339
	4	0.5582	0.7661	0.7851
Mean \pm STD (All Dataset)		0.5726 ± 0.16	0.7669 ± 0.10	0.7813 ± 0.08
Mean \pm STD (Overall)		0.7069 ± 0.11		

Table 7. Mean dice coefficient for blood vessels segmentation against a local clinician’s annotations.

Scans	Rabbani Dataset 1 [41]		Rabbani Dataset 2 [37]
	Healthy	ME	ME
1	0.7817	0.7871	0.7914
2	0.8808	0.8315	0.8135
3	0.8747	0.8197	0.7861
4	0.7983	0.7890	0.8047
5	0.8669	0.8444	0.7971
6	0.8518	0.8071	0.8168
7	0.8811	0.8004	0.7896
8	0.8190	0.7927	0.8238
9	0.8468	0.8058	0.8275
10	0.8617	0.7871	0.7914
Mean ± STD (All Dataset)	0.8589 ± 0.04	0.8185 ± 0.03	0.7839 ± 0.02
Mean ± STD (Overall)	0.8203 ± 0.03		

Table 8. Mean dice coefficient for retinal fluids extraction against a local clinician’s annotations.

Scans	Rabbani Dataset [36]	Zhang Dataset [35]
1	0.9194	0.9152
2	0.8689	0.8560
3	0.9082	0.9351
4	0.8551	0.9145
5	0.8726	0.9243
6	0.9322	0.8796
7	0.9238	0.8986
8	0.8887	0.8731
9	0.9162	0.8766
10	0.8724	0.9259
Mean ± STD (All Dataset)	0.9026 ± 0.03	0.9012 ± 0.04
Mean ± STD (Overall)	0.9019 ± 0.04	

Whereas, for both the datasets we achieved the overall mean dice coefficient of 0.9019 ± 0.04 . These results show that the proposed method performed well in retinal fluid extraction irrespective of the datasets and the OCT acquisition equipment. Figure 12 shows the extracted retinal fluid by the proposed system in healthy and ME affected OCT scans.

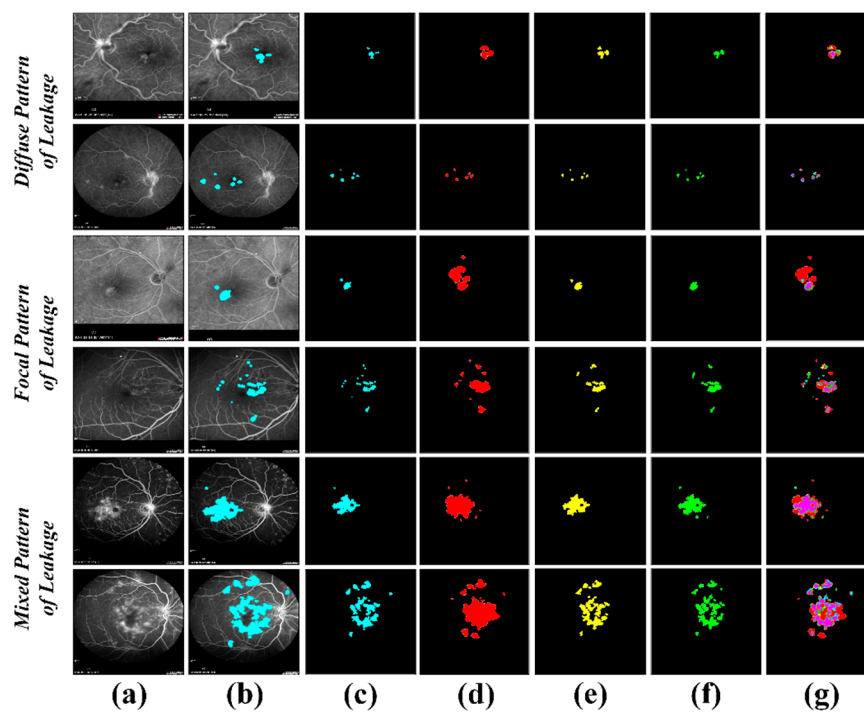


Figure 10. Segmented hard exudates region in the [37] dataset. (a) Original FA scans; (b) overlaid hard exudates region on original FA scans; (c) extracted hard exudates region through the proposed system; (d) Grader 1 markings of hard exudates region; (e) Grader 2 markings of hard exudates region; (f) Grader 3 markings of hard exudates region; (g) hard exudates extraction against expert markings—red, yellow and green represents the expert markings of Grader 1, 2 and 3, respectively, cyan represents the hard exudates region extracted through the proposed system and magenta represents the overlapped region of hard exudates.

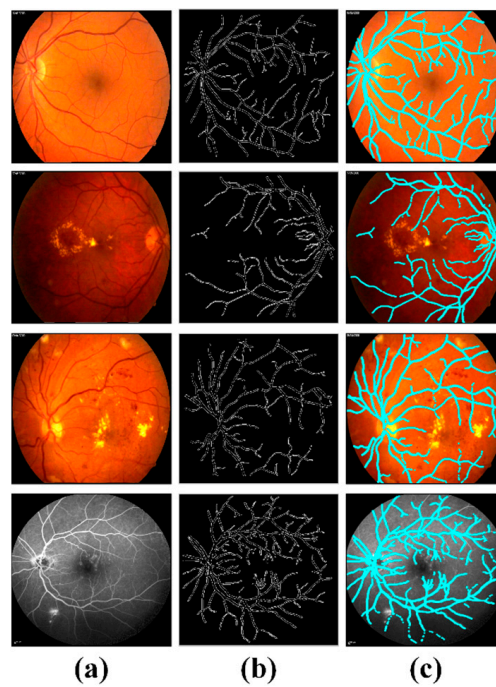


Figure 11. Extracted blood vessels through the proposed system in the [37,41] datasets. (a) Original fundus/FA scans; (b) extracted blood vessels through proposed system and (c) overlaid blood vessels on original fundus/FA scans.

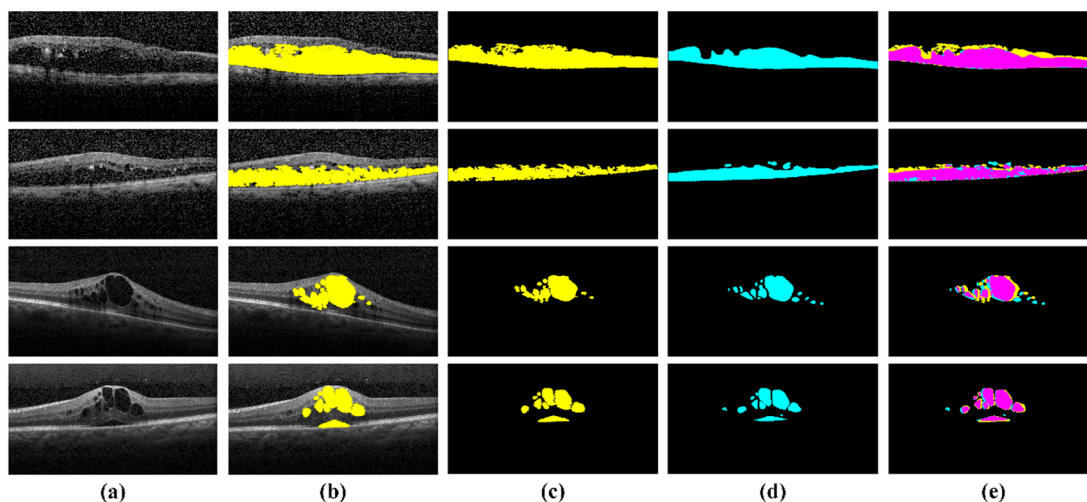


Figure 12. Extracted retinal fluid regions in the [35,36] datasets. (a) Original OCT scans; (b) overlaid retinal fluid regions on original OCT scans; (c) extracted retinal fluid regions through the proposed system; (d) local clinician markings of retinal fluid regions (e) retinal fluid extraction against expert markings—cyan represents the expert markings of local clinician, yellow represents the retinal fluid regions extracted through the proposed system and magenta represents the overlapped region of retinal fluids.

Moreover, we performed a classification of healthy and ME scans based on a 16-D feature vector extracted through the proposed CNN model. We passed the 16-D feature vector extracted from retinal fundus and OCT imagery to the proposed hybrid classifier for grading ME. The hybrid classification model in the proposed framework consisted of an ensemble of ANN, SVM and NB. The final decision was computed by measuring the majority votes of all three classification models. The hybrid classifier correctly classified 94.33% of all the unlabeled dataset scans, while individual performances of each classifier along with other methods reported in literature are listed in Table 9. We used sensitivity (SE), specificity (SP), positive predictive values (PPV), negative predictive values (NPV) and diagnostic accuracy (A) as the five measuring metrics to evaluate the hybrid classifier as expressed in Equations (11)–(15):

$$\text{Sensitivity} = \frac{T_P}{T_P + F_N'} \quad (11)$$

$$\text{Specificity} = \frac{T_N}{T_N + F_P'} \quad (12)$$

$$\text{PPV} = \frac{T_P}{T_P + F_P'} \quad (13)$$

$$\text{NPV} = \frac{T_N}{T_N + F_N'} \quad (14)$$

$$\text{Diagnostic Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (15)$$

T_P and T_N are the true positives and true negatives respectively, which specify the correctly classified (CC) cases. In this study, T_P indicates whether the input scan was macular edema and it was classified as macular edema too, while T_N represents the cases where actual input scan was of healthy eye and the classification also showed it as healthy. F_P and F_N stands for false positive and false negative, respectively, these are false classification indicators. F_P cases are those in which actual input scan was of healthy eye and classifier classified it as ME, while F_N is the reverse of F_P .

Table 9. Measure outcomes of the proposed ensemble hybrid classifier in comparison to other state-of-the-art techniques.

Methods	Validation Dataset		CC		TP	TN	FP	FN	SE	SP	PPV	NPV	A	
	OCT	Fundus	OCT	Fundus										
Proposed	ANN	5000 ^{R, Z}	100 ^R	4653	94	2457	2291	259	93	0.96	0.90	0.90	0.96	0.93
	SVM			4648	93	2407	2322	228	143	0.94	0.91	0.91	0.94	0.92
	NB			4559	92	2374	2289	261	176	0.93	0.90	0.90	0.93	0.91
	Hybrid			4716	95	2473	2338	212	77	0.97	0.92	0.92	0.97	0.94
[19]	X	20 ^{Ψ, Ψ}	X	-	-	-	-	-	0.943* 0.967 [∧]	1* 1 [∧]	0.92* 0.949 [∧]	-	-	
[20]	X	100 ^Φ	X	93	33	60	0	7	0.825	1	-	-	0.93	
[21]	X	30	X	-	-	13	2	-	0.928	-	0.924	-	-	
[23]	X	150 ^ζ	X	-	72	71	-	-	0.96	0.946	-	-	-	
[24]	X	400 ^Φ 104 ^ξ	X	-	-	-	-	-	0.95 1	0.9 0.74	-	-	-	
[25]	X	15 ^Ψ 15 ^C	X	-	-	-	-	-	0.978 0.907	0.99 0.994	0.833 0.74	-	-	
[26]	30 ^D	X	28	X	15	13	-	-	1	0.933	-	-	-	
[27]	19 ^C	X	-	X	-	-	-	-	0.91	0.96	-	-	-	
[28]	16	X	-	X	-	-	-	-	-	-	-	-	0.875	
[29]	90 ^B	X	88	X	60	28	-	-	1	0.933	-	-	0.977	
[31]	50 ^B	X	42	X	28	14	-	-	0.93	0.8	-	-	0.84	
[32]	45 ^D	X	43	X	30	13	-	-	1	0.866	-	-	-	
[34]	42281 ^D 4260 ^B	X	-	X	-	-	-	-	0.991	0.986	-	-	0.985	
[35]	500 ^Z	X	483	X	237	246	-	-	0.968	0.996	-	-	0.982	

^ΨDigital Retinal Images for Vessel Extraction (DRIVE) Dataset, ^ΨStructured Analysis of the Retina (STARE) Dataset, ^ΦMethods to Evaluate Segmentation and Indexing Techniques in the field of Retinal Ophthalmology (MESSIDOR) Dataset, ^ζBristol Eye Hospital Dataset, ^ξHamilton Eye Institute Macular Edema (HEI-DMED) Dataset, ^CCustom Dataset used by the authors, ^DDUKE Dataset, ^BBiomedical Image and Signal Analysis (BIOMISA) Dataset, ^ZZhang Dataset, ^RRabbani Dataset, *Results with fixed threshold, [∧]Results with variable threshold; ANN = Artificial Neural Network, SVM = Support Vector Machine, NB = Naïve Bayes Classifier, CC = Correctly Classified, TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative, SE = Sensitivity, SP = Specificity, PPV = Positive Predicted Values, NPV = Negative Predicted Values and A = Diagnostic Accuracy.

Figures 13 and 14 shows some healthy and ME OCT cases from both Rabbani [36] and Zhang [35] datasets, which are correctly processed by the proposed framework whereas Figure 15 shows some of the healthy and ME fundus scans, which were correctly processed by the proposed system.

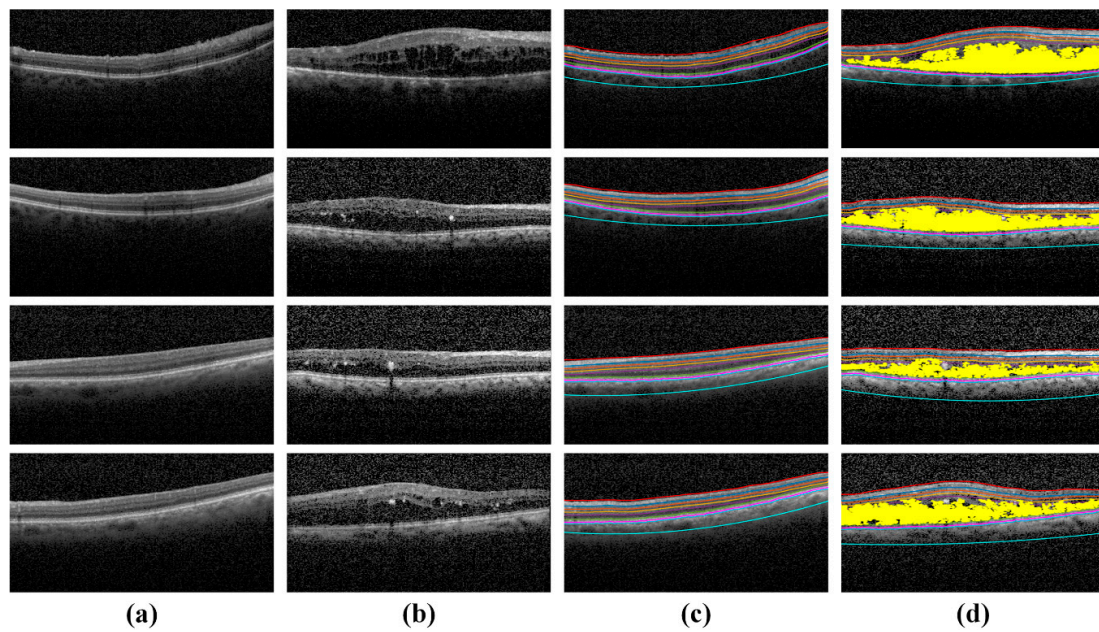


Figure 13. Healthy and ME OCT scans in the Rabbani dataset [36], which are processed by the proposed system. (a) Original healthy OCT B-scans; (b) original ME OCT B-scans; (c) classified as healthy by the proposed system and (d) classified as ME by the proposed system.

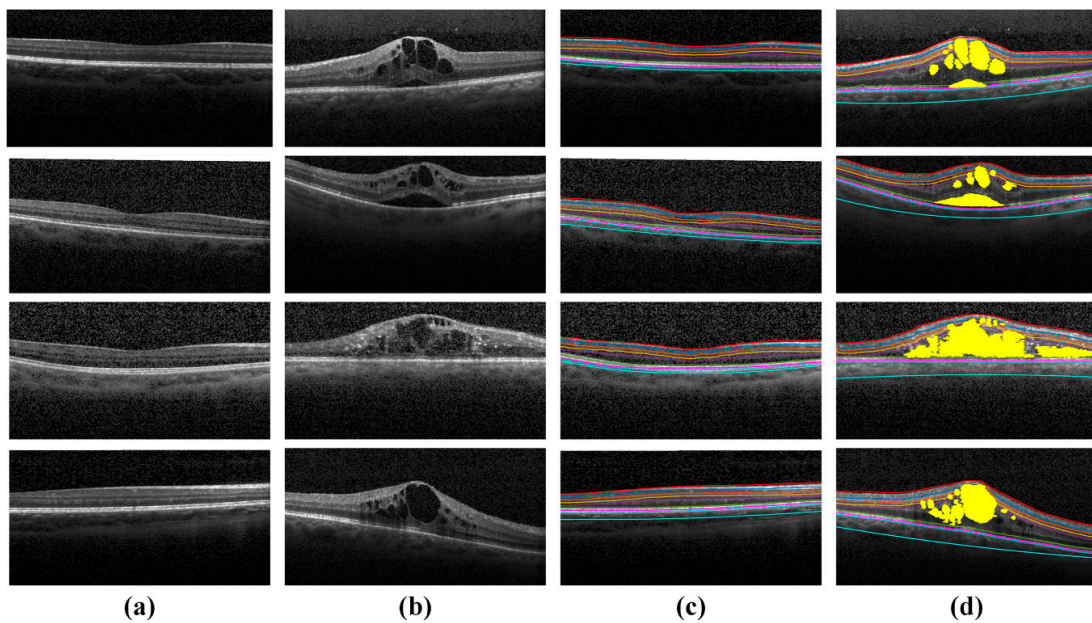


Figure 14. Healthy and ME OCT scans in the Zhang dataset [35], which are processed by the proposed system. (a) Original healthy OCT B-scans; (b) original ME OCT B-scans; (c) classified as healthy by the proposed system and (d) classified as ME by the proposed system.

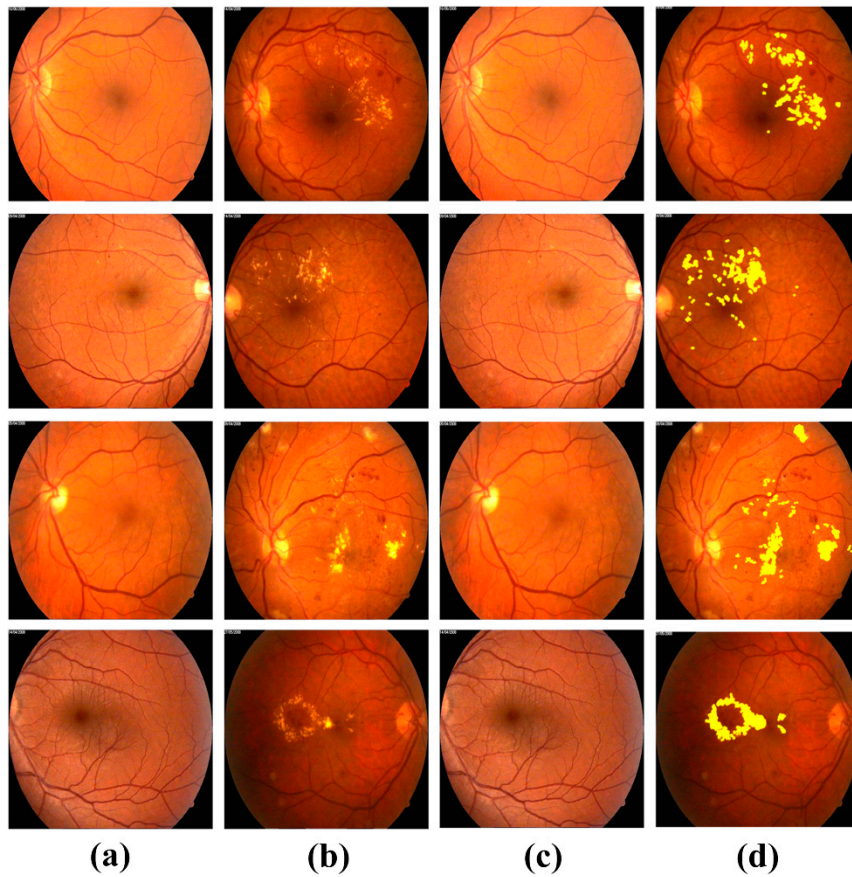


Figure 15. Healthy and ME fundus scans in the Rabbani Dataset [39,40], which are processed by the proposed system. (a) Original healthy fundus scans; (b) original ME fundus scans; (c) classified as healthy by the proposed system and (d) classified as ME by the proposed system.

Figures 16 and 17 shows the training performance of AlexNet and the proposed CNN model for the modality recognition and feature extraction, respectively. The training of AlexNet was conducted for 40 epochs where each epoch was completed in 35 iterations. The training of the proposed CNN model for feature extraction was conducted for 30 epochs where each epoch contained 50 iterations. The performance of the AlexNet and proposed CNN model during training phase was measured through the accuracy and cross-entropy loss function as expressed in Equation (16):

$$C_L = \sum I_{F_V, w} \log(P_{F_V, w}) \quad (16)$$

where $I_{F_V, w}$ is an indicator that the w is the correct class for the feature vector F_V , $P_{F_V, w}$ is the probability computed for F_V that it belongs to class w and C_L is the cross-entropy loss. The summation in Equation (9) runs for the total number of classes.

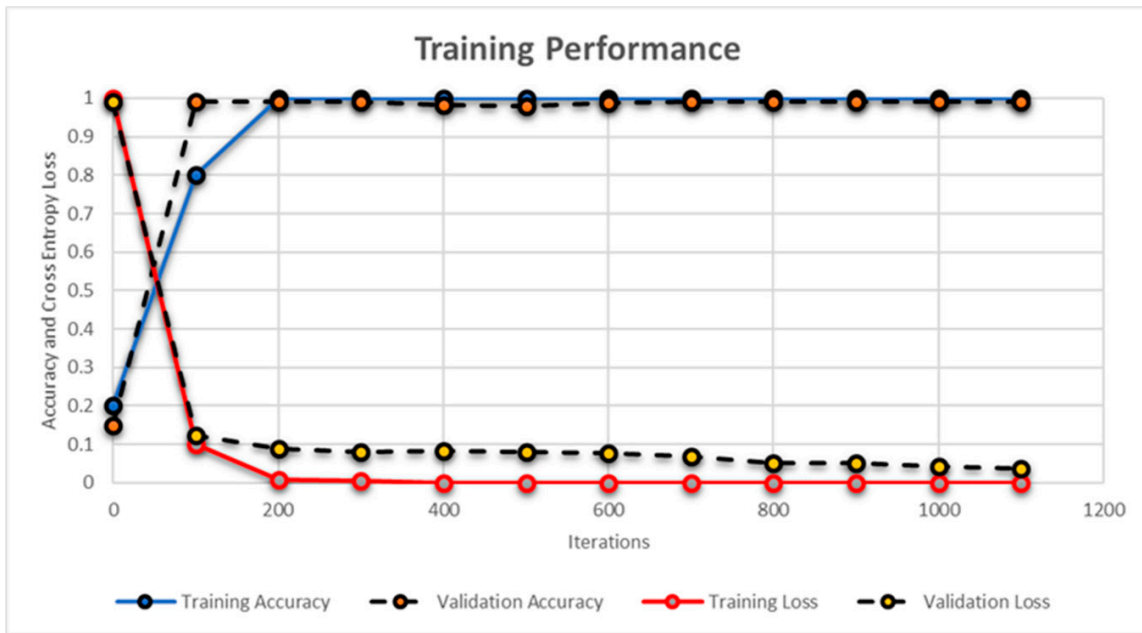


Figure 16. Training performance of the AlexNet model. Blue line shows the training accuracy, red line shows the training loss and the dashed black lines shows the validation accuracy and validation loss.

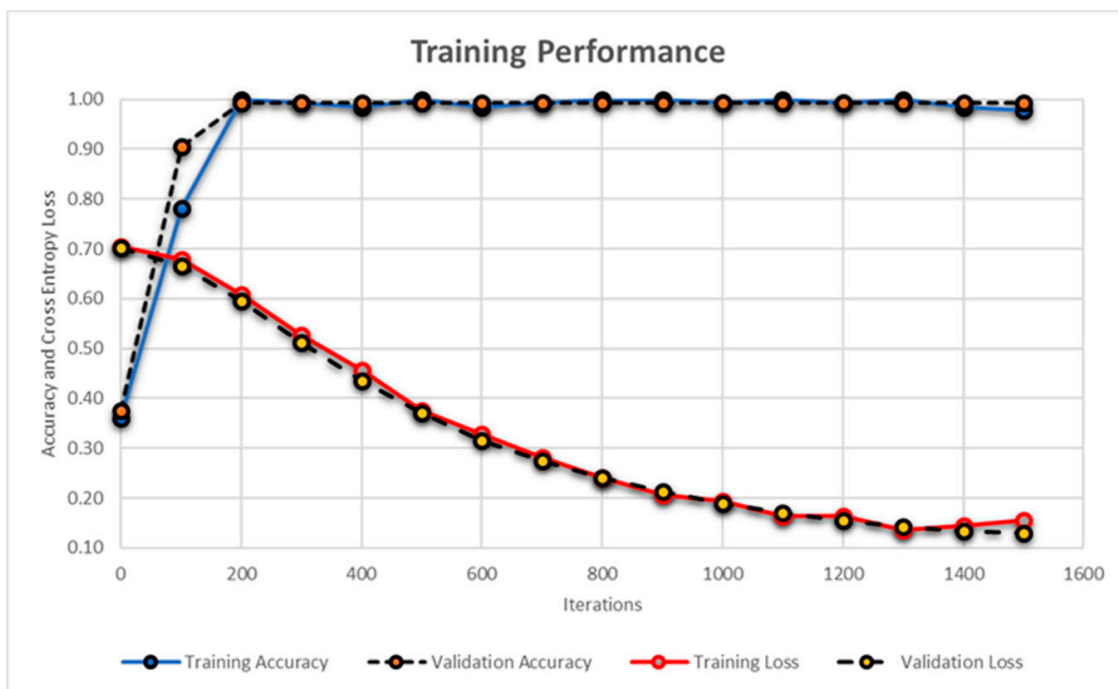


Figure 17. Training performance of the proposed CNN model. Blue line shows the training accuracy, red line shows the training loss and the dashed black lines shows the validation accuracy and validation loss.

Apart from this, for every 100 iterations, the validation was performed where validation performance was also measured through accuracy and cross-entropy loss function. The validation was performed in order to get the unbiased evaluation of the candidate model during the training phase as evident from Figures 16 and 17. Furthermore, we employed 50% dropout layers within each model to reduce overfitting on the dataset. The proposed CNN model achieved the accuracy of 99.23% in 1500 iterations during the training phase while the AlexNet model achieved the accuracy of 98.79%. These results were obtained through MATLAB R2018a and Table 10 shows the details of systems and

software along with the average time required for computing the results by each classifier. Although, the average time of hybrid classifier was a few seconds more than individual classifiers, the accuracy achieved by the proposed classification model was 94.33%.

Table 10. Details of the system and software used for conducting this research.

System		Software		Average Time for Single Classification (seconds)		
				Classifier	OCT	Fundus
Made	DELL			ANN	4.6	3.3
Processor	i7-4500U @ 2.4GHz	Windows 10 Pro	MATLAB	SVM	5.7	4.2
RAM	8GB DDR2	64-bit	R2018a	NB	3.2	1.8
Graphics	AMD HD 8670M			Hybrid	6.8	5.1

6. Discussion

A deep retinal diagnostic framework was proposed here that combines retinal fundus and OCT imagery for the extraction of clinically significant ME findings and uses the extracted information for the reliable and accurate grading of ME. According to EDTRS, ME was clinically graded based upon the locality of edema with respect to fovea i.e., if the retinal fluids or hard exudates are observed within the foveal diameter of 500 micrometers, then ME is graded as clinically significant otherwise it is graded as non-clinically significant. Clinically significant macular edema is more critical as compared to non-clinically significant macular edema as it produces retinal thickening near the fovea, which causes non-recoverable visual impairments (or even blindness). Retinal fundus and OCT imagery are the most common and non-invasive retinal examination techniques, which depicts the prominent symptoms of retinopathy. OCT imagery shows the early symptoms of retinopathy due to its ability to present retinal cross-sectional regions. Therefore, the retinal blood vessels leakages and retinal fluids accumulation can be easily visualized through OCT scans. However, accurate visualization of hard exudates from OCT imagery is a very cumbersome task, therefore, the retinal fundus scans are clinically used for this purpose. To the best of our knowledge, all the retinal diagnostic frameworks that have been proposed in the past for ME diagnosis are based on single retinal imaging modality, which do not completely depict the retinal abnormalities. The proposed framework is unique as it fuses the findings from both retinal fundus and OCT imagery for the effective, reliable and objective diagnosis as well as grading of ME subjects (especially those having a diabetic history). The proposed framework works in a way that it first recognizes the type of imagery through the pre-trained AlexNet CNN model. The retinal imaging modality recognition is one of the crucial steps of the proposed framework since both images do not contain any metadata that can depict their unique information or description. Therefore, in order to develop a generalized framework that can perform automated analysis and can automatically mass screen retinal patients, the respective imagery has to be automatically recognized first. After recognizing the retinal images, the proposed framework extracts the retinal layers and retinal fluids from the candidate OCT scans and it also extracts the hard exudate regions from the fundus scans. The extracted retinal information is then overlaid onto the respective scans and the annotated scans are then passed to the proposed CNN model, which extracts the eight distinct features from the annotated fundus scan and eight distinct features from the annotated fundus scans. These features are fused together to form a 16-D feature vector, which is passed to the proposed hybrid classifier formed through the ensemble of ANN, SVM and NB. One of the major aims of the proposed framework was to accurately diagnose and grade ME pathologies. Since ME is clinically graded into different categories depending upon the disease severity levels so in order to get reliable and accurate diagnosis, the hybrid classification was proposed that gives a decision based upon the majority votes obtained through all the three supervised classifiers. This increases the diagnostic performance of the proposed framework without compromising the time performance as evident from Table 10. Apart from this, the proposed framework was extensively tested on multiple publicly available datasets and was compared with

state-of-the-art solutions against different metrics and ground truths (provided by expert clinicians) as evident from the results section. Table 9 depicts the detailed diagnostic comparison with other existing solutions where it can be seen that the proposed framework was the only generic framework that was validated on multiple publicly available datasets containing both retinal fundus and OCT imagery and achieved the diagnostic accuracy of 94.33%.

7. Conclusions and Future Work

In this paper, we proposed a computer aided diagnostics method for segmentation of retinal pathological symptoms and classification of macular edema using two retinal imaging modalities (OCT and fundus imaging). The proposed framework was based on a hybrid classification model in which 16 unique features are extracted for distinguishing macular edema cases from healthy ones. The dataset used for conducting this study consisted of more than 78,891 retinal scans in total, out of which we used 73,791 scans for training purpose and 5100 for evaluation purpose. The proposed classification model correctly classified 4811 retinal scans, achieving 94.33% accuracy. The proposed system was quite robust in general, insensitive to OCT B-scans orientations and performed extremely well against the noisy and degraded scans as shown in Figure 5. Moreover, the proposed technique could be optimized for detecting other ocular diseases such as age-related macular degeneration (ARMD), idiopathic central serous chorioretinopathy (CSCR), Glaucoma, diabetic retinopathy, etc., as well as for segmenting other retinal layers. It could also be extended for the 3D modeling of the human retina.

Author Contributions: Conceptualization, validation, formal analysis, data curation, visualization and writing—review and editing, B.H., T.H., R.A., B.L., O.H.; methodology, writing—original draft preparation, B.H., T.H., R.A., O.H.; software, B.H., T.H., R.A.; supervision, project administration and funding acquisition, B.L.

Funding: This research and its APC was funded by National Key R&D Program of China, grant number 2017YFB0202601.

Acknowledgments: We are extremely thankful to Rabbani et al. and Zhang et al. for making their datasets publicly available, which enabled us to conduct this study. We are also thankful to local clinician for helping us in annotation of datasets.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **2018**, *392*, 1789–1858. [[CrossRef](#)]
2. Hassan, B.; Ahmed, R.; Li, B.; Noor, A.; Hassan, Z.U. A comprehensive study capturing vision loss burden in Pakistan (1990–2025): Findings from the Global Burden of Disease (GBD) 2017 study. *PLoS ONE* **2019**, *14*, e0216492. [[CrossRef](#)] [[PubMed](#)]
3. Harney, F. Diabetic retinopathy. *Medicine* **2006**, *34*, 95–98. [[CrossRef](#)]
4. Alghadyan, A.A. Diabetic retinopathy—An update. *Saudi J. Ophthalmol.* **2011**, *25*, 99–111. [[CrossRef](#)] [[PubMed](#)]
5. Acharya, U.R.; Vidya, K.S.; Ghista, D.N.; Lim, W.J.E.; Molinari, F.; Sankaranarayanan, M. Computer-aided diagnosis of diabetic subjects by heart rate variability signals using discrete wavelet transform method. *Knowl. Based Syst.* **2015**, *81*, 56–64. [[CrossRef](#)]
6. Verma, L.; Prakash, G.; Tewari, H.K. Diabetic retinopathy: Time for action. No complacency please! *Bull. World Health Organ.* **2002**, *80*, 419. [[PubMed](#)]
7. Mingguang, H.; Wanjiku, M.; Susan, L.; Paul, C. Global Efforts to Generate Evidence for Vision 2020. *Ophthalmic Epidemiol.* **2015**, *22*, 237–238.
8. Hassan, T.; Akram, M.U.; Hassan, B.; Nasim, A.; Bazaz, S.A. Review of OCT and fundus images for detection of Macular Edema. In Proceedings of the 2015 IEEE International Conference on Imaging Systems and Techniques (IST), Macau, China, 16–18 September 2015; pp. 1–4.

9. Huang, D.; Swanson, E.A.; Lin, C.P.; Schuman, J.S.; Stinson, W.G.; Chang, W.; Puliavito, C.A. Optical coherence tomography. *Science* **1991**, *254*, 1178–1181. [[CrossRef](#)]
10. Swanson, E.A.; Izatt, J.A.; Hee, M.R.; Huang, D.; Lin, C.P.; Schuman, J.S.; Fujimoto, J.G. In vivo retinal imaging by optical coherence tomography. *Opt. Lett.* **1993**, *18*, 1864–1866. [[CrossRef](#)]
11. De Carlo, T.E.; Romano, A.; Waheed, N.K.; Duker, J.S. A review of optical coherence tomography angiography (OCTA). *Int. J. Retin. Vitre.* **2015**, *1*, 5. [[CrossRef](#)]
12. Fercher, F.; Drexler, W.; Hitzenberger, C.K.; Lasser, T. Optical coherence tomography-principles and applications. *Rep. Prog. Phys.* **2003**, *66*, 239. [[CrossRef](#)]
13. Schmitz-Valckenberg, S.; Holz, F.G.; Bird, A.C.; Spaide, R.F. Fundus autofluorescence imaging: Review and perspectives. *Retina* **2008**, *28*, 385–409. [[CrossRef](#)] [[PubMed](#)]
14. Sepah, Y.J.; Akhtar, A.; Sadiq, M.A.; Hafeez, Y.; Nasir, H.; Perez, B.; Nguyen, Q.D. Fundus autofluorescence imaging: Fundamentals and clinical relevance. *Saudi J. Ophthalmol.* **2014**, *28*, 111–116. [[CrossRef](#)]
15. Chee, K.L.; Santiago, P.A.; Lingam, G.; Singh, M.S.; Naing, T.; Mangunkusumo, A.E.; Naser, M.N. Application of Ocular Fundus Photography and Angiography. In *Ophthalmological Imaging and Applications*; CRC Press: Boca Raton, FL, USA, 2014; pp. 135–156.
16. Virgili, G.; Menchini, F.; Dimastrogiovanni, A.F.; Rappizzi, E.; Menchini, U.; Bandello, F.; Chiodini, R.G. Optical coherence tomography versus stereoscopic fundus photography or biomicroscopy for diagnosing diabetic macular edema: A systematic review. *Investig. Ophthalmol. Vis. Sci.* **2007**, *48*, 4963–4973. [[CrossRef](#)] [[PubMed](#)]
17. Browning, J.; McOwen, M.D.; Bowen, R.M., Jr.; Tisha, L.O. Comparison of the clinical diagnosis of diabetic macular edema with diagnosis by optical coherence tomography. *Ophthalmology* **2004**, *111*, 712–715. [[CrossRef](#)] [[PubMed](#)]
18. Strøm, C.; Sander, B.; Larsen, N.; Larsen, M.; Lund-Andersen, H. Diabetic macular edema assessed with optical coherence tomography and stereo fundus photography. *Investig. Ophthalmol. Vis. Sci.* **2002**, *43*, 241–245.
19. Reza, A.W.; Eswaran, C.; Hati, S. Automatic tracing of optic disc and exudates from color fundus images using fixed and variable thresholds. *J. Med. Syst.* **2009**, *33*, 73–80. [[CrossRef](#)]
20. Sreejini, K.S.; Govindan, V.K. Automatic grading of severity of diabetic macular edema using color fundus images. In Proceedings of the 2013 Third International Conference on Advances in Computing and Communications (ICACC), Cochin, India, 29–31 August 2013; pp. 177–180.
21. Walter, T.; Klein, J.C.; Massin, P.; Erginay, A.A. Contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina. *IEEE Trans. Med. Imaging* **2002**, *21*, 1236–1243. [[CrossRef](#)]
22. Giancardo, L.; Meriaudeau, F.; Karnowski, T.P.; Li, Y.; Garg, S.; Tobin, K.W., Jr.; Chaum, E. Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. *Med. Image Anal.* **2012**, *16*, 216–226. [[CrossRef](#)]
23. Osareh, A.; Shadgar, B.; Markham, R. A computational-intelligence-based approach for detection of exudates in diabetic retinopathy images. *IEEE Trans. Inf. Technol. Biomed.* **2009**, *13*, 535–545. [[CrossRef](#)]
24. Deepak, K.S.; Sivaswamy, J. Automatic assessment of macular edema from color retinal images. *IEEE Trans. Med. Imaging* **2012**, *31*, 766–776. [[CrossRef](#)] [[PubMed](#)]
25. Yazid, H.; Arof, H.; Isa, H.M. Automated identification of exudates and optic disc based on inverse surface thresholding. *J. Med. Syst.* **2012**, *36*, 1997–2004. [[CrossRef](#)] [[PubMed](#)]
26. Hassan, B.; Raja, G. Fully Automated Assessment of Macular Edema using Optical Coherence Tomography (OCT) Images. In Proceedings of the 2016 International Conference on Intelligent Systems Engineering (ICISE), Islamabad, Pakistan, 15–17 January 2016; pp. 5–9.
27. Wilkins, G.R.; Houghton, O.M.; Oldenburg, A.L. Automated segmentation of intraretinal cystoid fluid in optical coherence tomography. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 1109–1114. [[CrossRef](#)] [[PubMed](#)]
28. Sugmk, J.; Kiattisin, S.; Leelasantham, A. Automated classification between age-related macular degeneration and diabetic macular edema in OCT image using image segmentation. In Proceedings of the 7th 2014 Biomedical Engineering International Conference (BMEiCON), Fukuoka, Japan, 26–28 November 2014; pp. 1–4.

29. Hassan, B.; Raja, G.; Hassan, T.; Akram, M.U. Structure tensor based automated detection of macular edema and central serous retinopathy using optical coherence tomography images. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **2016**, *33*, 455–463. [[CrossRef](#)] [[PubMed](#)]
30. Chiu, S.J.; Allingham, M.J.; Mettu, P.S.; Cousins, S.W.; Izatt, J.A.; Farsiu, S. Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. *Biomed. Opt. Express* **2015**, *6*, 1172–1194. [[CrossRef](#)] [[PubMed](#)]
31. Sahar, S.; Ayaz, S.; Akram, M.U.; Basit, I. A Case Study Approach: Iterative Prototyping Model Based Detection of Macular Edema in Retinal OCT Images. In Proceedings of the 27th International Conference on Software Engineering and Knowledge Engineering (SEKE), Pittsburgh, PA, USA, 6–8 July 2015; pp. 266–271.
32. Srinivasan, P.P.; Kim, L.A.; Mettu, P.S.; Cousins, S.W.; Comer, G.M.; Izatt, J.A.; Farsiu, S. Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomed. Opt. Express* **2014**, *5*, 3568–3577. [[CrossRef](#)] [[PubMed](#)]
33. Lee, C.S.; Tying, A.J.; Deruyter, N.P.; Wu, Y.; Rokem, A.; Lee, A.Y. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *Biomed. Opt. Express* **2017**, *8*, 3440–3448. [[CrossRef](#)] [[PubMed](#)]
34. Hassan, T.; Akram, M.U.; Masood, M.F.; Yasin, U. Deep structure tensor graph search framework for automated extraction and characterization of retinal layers and fluid pathology in retinal SD-OCT scans. *Comput. Biol. Med.* **2018**, *105*, 112–124. [[CrossRef](#)] [[PubMed](#)]
35. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* **2018**, *172*, 1122–1131. [[CrossRef](#)]
36. Rasti, R.; Rabbani, H.; Mehridehnavi, A.; Hajizadeh, F. Macular OCT classification using a multi-scale convolutional neural network ensemble. *IEEE Trans. Med. Imaging* **2017**, *37*, 1024–1034. [[CrossRef](#)]
37. Rabbani, H.; Allingham, M.J.; Mettu, P.S.; Cousins, S.W.; Farsiu, S. Fully automatic segmentation of fluorescein leakage in subjects with diabetic macular edema. *Investig. Ophthalmol. Vis. Sci.* **2015**, *56*, 1482–1492. [[CrossRef](#)] [[PubMed](#)]
38. Mahmudi, T.; Kafieh, R.; Rabbani, H.; Akhlagi, M. Comparison of macular OCTs in right and left eyes of normal people. In Proceedings of the Medical Imaging 2014: Biomedical Applications in Molecular, Structural, and Functional Imaging, San Diego, CA, USA, 17–21 August 2014; Volume 9038, p. 90381W.
39. Alipour, S.H.M.; Rabbani, H.; Akhlaghi, M.R. Diabetic retinopathy grading by digital curvelet transform. *Comput. Math. Methods Med.* **2012**, *2012*, 761901. [[CrossRef](#)]
40. Esmaeili, M.; Rabbani, H.; Dehnavi, A.M.; Dehghani, A. Automatic detection of exudates and optic disk in retinal images using curvelet transform. *IET Image Process.* **2012**, *6*, 1005–1013. [[CrossRef](#)]
41. Alipour, S.H.M.; Rabbani, H.; Akhlaghi, M.; Dehnavi, A.M.; Javanmard, S.H. Analysis of foveal avascular zone for grading of diabetic retinopathy severity based on curvelet transform. *Graefe's Arch. Clin. Exp. Ophthalmol.* **2012**, *250*, 1607–1614. [[CrossRef](#)] [[PubMed](#)]
42. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the 5th International Conference on Neural Information Processing Systems, Lake Tahoe, CA, USA, 3–6 December 2012.
43. Bengio, Y. Practical recommendations for gradient based training of deep architectures. In *Neural Networks: Tricks of the Trade*; Springer: Berlin, Germany, 2012; pp. 437–478.
44. Köthe, U. Edge and junction detection with an improved structure tensor. *Jt. Pattern Recognit. Symp.* **2003**, 25–32. [[CrossRef](#)]

