

RAG-FW: A hybrid convolutional framework for the automated extraction of retinal lesions and lesion-influenced grading of human retinal pathology

Taimur Hassan ¹, Muhammad Usman Akram ¹, Naoufel Werghi ¹, and Muhammad Noman Nazir ¹

¹Affiliation not available

October 30, 2023

Abstract

[Submitted in IEEE J-BHI]

Retinopathy refers to any damage in the retina that causes visual impairments or even blindness. Identification of retinal lesions plays a vital role in accurately grading retinopathy and for its effective treatment. Optical coherence tomography (OCT) imaging is the most popular non-invasive technique used for the retinal examination due to its ability to screen abnormalities in early stages. Many researchers have presented studies on OCT based retinal image analysis over the past. However, to our best knowledge, there is no framework yet available which can extract retinal lesions from multi-vendor OCT scans and utilize them for the intuitive grading of the human retina. To cater this lack, we propose a deep retinal analysis and grading framework (RAG-FW). RAG-FW is a hybrid convolutional framework which extracts retinal lesions such as intra-retinal fluid, sub-retinal fluid, hard exudates, drusen and chorioretinal abnormalities (including fibrotic scars and choroidal neovascular membranes) from multi-vendor OCT scans. Furthermore, it utilizes them for the lesion-influenced grading of retinopathy as per the clinical standards. RAG-FW has been trained using 113,261 retinal OCT scans from which 112,261 scans were used for training and 1,000 scans were used for the validation purposes. Furthermore, it has been rigorously tested on 43,613 scans from five highly complex publicly available datasets where it achieved the mean intersection-over-union score of 0.8055 for extracting the retinal lesions and the F₁ score of 99.52% for correctly classifying the retinopathy cases. The source code of RAG-FW is available at <http://biomisa.org/index.php/downloads/>.

RAG-FW: A hybrid convolutional framework for the automated extraction of retinal lesions and lesion-influenced grading of human retinal pathology

Taimur Hassan, Muhammad Usman Akram, *Member, IEEE*, Naoufel Werghi, *Senior Member, IEEE*, and Muhammad Noman Nazir

Abstract— Retinopathy refers to any damage in the retina that causes visual impairments or even blindness. Identification of retinal lesions plays a vital role in accurately grading retinopathy and for its effective treatment. Optical coherence tomography (OCT) imaging is the most popular non-invasive technique used for the retinal examination due to its ability to screen abnormalities in early stages. Many researchers have presented studies on OCT based retinal image analysis over the past. However, to our best knowledge, there is no framework yet available which can extract retinal lesions from multi-vendor OCT scans and utilize them for the intuitive grading of the human retina. To cater this lack, we propose a deep retinal analysis and grading framework (RAG-FW). RAG-FW is a hybrid convolutional framework which extracts retinal lesions such as intra-retinal fluid, sub-retinal fluid, hard exudates, drusen and chorioretinal abnormalities (including fibrotic scars and choroidal neovascular membranes) from multi-vendor OCT scans. Furthermore, it utilizes them for the lesion-influenced grading of retinopathy as per the clinical standards. RAG-FW has been trained using 113,261 retinal OCT scans from which 112,261 scans were used for training and 1,000 scans were used for the validation purposes. Furthermore, it has been rigorously tested on 43,613 scans from five highly complex publicly available datasets where it achieved the mean intersection-over-union score of 0.8055 for extracting the retinal lesions and the F_1 score of 99.52% for correctly classifying the retinopathy cases. The source code of RAG-FW is available at <http://biomisa.org/index.php/downloads/>.

Index Terms—Retinopathy, Deep Learning, Ophthalmology, Optical Coherence Tomography (OCT)

This work is supported by Center for Cyber-Physical Systems (C2PS). T. Hassan and N. Werghi are with Center for Cyber-Physical Systems (C2PS), EECS Department, Khalifa University, Abu Dhabi UAE (e-mail: enqr.taimoorhassan@gmail.com, naoufel.werghi@ku.ac.ae). T. Hassan and M. U. Akram are with Department of Computer and Software Engineering, National University of Sciences and Technology (NUST), Islamabad Pakistan (e-mail: usmakram@gmail.com). M. N. Nazir is with Department of Ophthalmology, Armed Forces Institute of Ophthalmology (AFIO), Rawalpindi Pakistan (e-mail: noman.nazir@gmail.com)

I. INTRODUCTION

Human beings perceive vision information through eyes [1]. A human eye is made up of three layers [2] where retina is the innermost layer responsible for producing vision. Retina is composed of two regions i.e. the macular region and the peripheral region. Macular region or macula is responsible for producing central vision whereas the side vision is produced by the peripheral retina. Light focused through biconvex lens falls on the retina which stimulates the rods and cones cells. Afterwards, the vision information is transferred to the brain for interpretation through the optic nerve [2]. Many retinal complications are caused due to diabetes where most of the times the retinal blood vessels get ruptured due to excessive hyperglycemia, resulting in the leakage of blood and other fluid deposits within the retinal structures. Such medical condition is known as diabetic retinopathy (DR) and it is the leading cause of blindness all over the world [3]. Furthermore, the fluid leakages within macula due to DR causes macular thickening or the macular edema (ME). ME can also occur due to cataract surgeries, macular degeneration, uveitis, retinal veins or arteries blockage, in a non-diabetic subject [4]. Moreover, retinal fluid leakages are also observed in the subjects affected from central serous retinopathy (CSR) and macular degeneration. CSR is a medical condition, mostly associated with youngsters, due to stress and it causes serous detachment underneath retina [5]. Macular degeneration, often known as age-related macular degeneration (AMD) is another retinal complication that mostly occurs in old people due to the formation of drusen. AMD is typically graded into two stages i.e. the dry or non-exudative AMD and the wet or exudative AMD. Drusen are formed in an early stage of AMD which produces distorted and blurred vision, and leads towards geographic atrophy (GA). The further progression of the disease causes wet macular degeneration in which abnormal blood vessels from choroid intercepts the retina and produces chorioretinal abnormalities such as scars and choroidal neovascular membranes (CNVM), leading to choroidal neovascularization (CNV). Although AMD alone does not cause blindness, it can cause severe visual impairments if not treated timely.

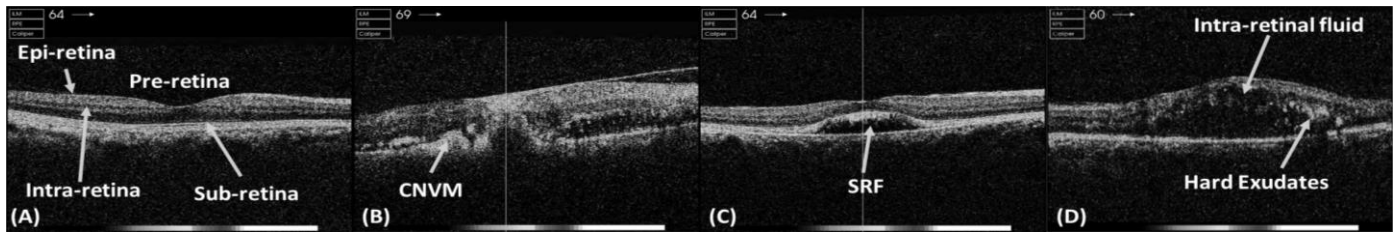


Figure 1: Retinal OCT scan of: (A) healthy subject with highlighted retinal regions, (B) wet AMD affected subjects with CNVM, (C) acute CSR affected subject with SRF, (D) DME affected subject with IRF and HE.

There are many retinal examination techniques which can effectively show the abnormal retinal symptoms in a diabetic and non-diabetic subject. The most widely used techniques are fundus photography (FP) [6], fundus fluorescein angiography (FFA) and optical coherence tomography (OCT). Fundus and OCT imaging are most commonly used by the clinicians due to their non-invasive nature and their ability to give objective visualization of retinal pathology. However, OCT imaging is preferred by the clinicians over fundus photography due to its ability to present early retinal abnormalities [7]. Figure 1 shows the retinal OCT brightness scans (B-scans) of healthy and diseased retinal pathologies where intra-retinal fluid (IRF), sub-retinal fluid (SRF), hard exudates (HE) and CNVM are highlighted.

II. RELATED WORK

Many studies have been presented in the past related to OCT based retinal image analysis. These studies are either focused on the clinical evaluation of OCT imaging for visualizing retinal diseases or are focused on the utilization of OCT imaging for the extraction of retinal layers, retinal lesions and automated retinal diagnosis.

A. Clinical Evaluations:

OCT imaging has been widely adapted by clinicians for screening retinal subjects and many times they have highlighted the significance of OCT imaging for evaluating the human retina. Clinically, OCT imagery acts as an initial examination procedure for screening diabetic macular edema (DME) subjects [8] or for the DME severity analysis [9] as it presents the early and objective visualization of DME pathology [10-16]. In addition to this, researchers have differentiated pseudophakic cystoid macular edema (P-CME) and DME cases solely based on OCT imagery [17]. In [18], an OCT based DME screening system has been proposed which is tested on 182 eyes of 107 patients and it was concluded that OCT based examination is very robust in picking retinal thickness variations. Furthermore, OCT imaging is utilized by the clinicians for studying non-diabetic pathologies as well. In [19], a comparison of AMD and CNV using OCT imaging, FFA and FP has been made where it was concluded that FP is a good tool for analyzing drusen and other pigmentary changes, but OCT images can pick even the slightest early variations of CNV and AMD.

B. Retinal Layers and Lesions Segmentation:

Extraction of retinal layers and retinal lesions is one of the key steps in analyzing the retinal pathology. Manual extraction of retinal layers and lesions is a very time consuming and

cumbersome process. Therefore, many researchers have presented an automated and semi-automated methods to overcome this [20-27]. Some researchers have also presented fluid reflectivity analysis from Enface OCT and OCT-Angiography (OCT-A) [28]. Toth et al. [29] presented a study to analyze the distribution of OCT features and sub-retinal lesion thicknesses in a macular scar or atrophic pathology. Moreover, an automated framework is presented in [30] for the extraction of retinal layers from optic nerve head (ONH) scans. Rashno et al. [31] developed a framework based on neutrosophic sets and graph algorithms to automatically segment intra-retinal fluids from OCT scans of DME subjects. Recently, researchers have utilized deep learning for the extraction of retinal layers and retinal fluids. Deep learning methods have several advantages over conventional methods such as more robustness and reliability for the extraction of retinal information [32]. Lee et al. proposed a modification of UNet architecture for the detection of retinal fluid from ME affected OCT images [33]. Roy et al. [34] developed a retinal layers and fluid detection framework named as (ReLayNet) and they evaluated their proposed framework on publicly available Duke dataset (which is named as Duke-II in this paper). Schlegl et al. [35] developed a CNN based model to detect and quantify SRF and IRF, and they achieved the accuracy of 0.92 and 0.94, respectively. They evaluated their proposed method on 1,200 OCT volumes of retinal vein occlusion (RVO), DME and AMD subjects where the scans were acquired through Cirrus and Spectralis machines [35]. Seebock et al. [36] presented a framework based on Bayesian UNet for detecting retinal anomalies in DME, AMD, dry GA and RVO pathologies. Identification of IRF from the AMD and DME affected scans was performed by [37] as well for which they extracted 312 distinct features and classified IRF through linear discriminant classifier (LDC), support vector machines (SVM) and Parzen window. G. N. Girish et al. [38] developed a framework based on fully convolutional network for the automated segmentation of intra-retinal cysts from multi-vendor retinal OCT scans. Fang et al. [39] presented lesion-aware CNN model that pays attention to lesions while classifying the macular diseases. They developed a lesion detection network which generates a soft attention map which is further utilized by the classification network to accelerate the performance of macular diagnosis based on OCT imagery. Deep learning methods have shown remarkable performance for extracting retinal information. But, they require a lot of annotated data and training time as well to yield better results.

C. Retinal Diagnosis:

Apart from extracting retinal layers and retinal lesions, many

researchers have developed computer aided diagnostic (CAD) solutions to screen retinal patients. The motivation for developing these CAD systems is to mass screen retinal subjects across the globe especially in rural areas where people cannot get the consultation of expert ophthalmologists. Furthermore, these CAD systems filters the retinal subjects so that only critical patients are referred to the doctor for immediate checkup. Some researchers have produced CAD systems to screen normal and abnormal retinal pathologies [40-41] while others have proposed methods which can classify between different retinal abnormalities [42-51]. In addition to this, a novel deep learning framework, namely RAC-CNN, is proposed in [52] for the classification of rods and cones from adaptive optics scanning light ophthalmoscope (AOSLO). This framework has been validated on healthy as well as achromatopsia subjects and the results were highly correlated with the time-consuming manual grading process [52]. Kermany et al. [53] proposed a deep retinal screening system that was tested on 1,000 OCT scans and achieved the accuracy of 96.6%. They also made their dataset publicly available (we named it Zhang dataset in this article). Rong et al. [54] presented a surrogate-assisted and CNN based classification of retinal OCT scans on their local as well as on Duke-III dataset [50] where they achieved the area under the curve (AUC) of 0.9856. Apart from this, we have also presented different frameworks over the past [55-64] to extract retinal layers, retinal fluids as well as for maculopathy diagnosis using retinal OCT images. However, to the best of our knowledge, there is no framework that can extract retinal lesions such as IRF, SRF, HE, drusen and chorioretinal abnormalities (CA) like fibrotic scars and CNVM from multi-vendor OCT scans and use them for the accurate grading of ME, AMD and CSR pathologies as per the clinical standards. Here, it should be noted that classification and grading are not the same thing. Grading is the next step after classification to measure the severity (stage) of the disease and it is based on the defined set of clinical standards. Retinal lesions play a significant role in accurately classifying retinal diseases [39]. In [35], retinal fluid recognition has been carried out for the AMD, DME and RVO pathologies. Moreover, the significance of extracting fluid biomarkers for retinal diagnosis is also highlighted in [36] and they identified IRF for AMD and DME cases. Similarly, the intra-retinal cysts from retinal OCT images have been automatically identified in [38]. However, none of them have utilized the extracted lesions from multi-vendor OCT scans for grading retinal diseases as per the clinical standards. Also, the challenge of segmenting HE, drusen and CA from multi-vendor OCT scans has not been addressed yet which is crucial for grading non-exudative and exudative AMD. In this paper, we present a retinal analysis and grading framework (RAG-FW), that not only extracts the retinal lesions, but also use them to grade the diagnosed pathologies. The major highlights of this paper are listed below:

- This paper presents RAG-FW, a hybrid convolutional framework for the automated extraction of retinal lesions

and lesion-influenced grading of retinopathy from multi-vendor OCT scans.

- RAG-FW has been made generalized so it provides retinal lesions extraction and retinopathy grading irrespective of the acquisition machinery, scan artifacts or the retinal pathology.
- RAG-FW has been rigorously tested on 43,613 multi-vendor OCT scans for extracting and identifying retinal lesions such as IRF, SRF, HE, drusen and chorioretinal abnormalities (like fibrotic scars and CNVM), along with the classification and grading of retinopathy.
- RAG-FW outperformed state-of-the-art solutions by achieving 14.15% improvements in extracting retinal fluids on Duke-II [25] and 2.02%, and 1.24% improvement in classifying retinopathy on Zhang and BIOMISA datasets.
- In addition to this, the transferability of RAG-FW for lesions extraction and retinopathy classification has been thoroughly tested through extensive cross-dataset validation where RAG-FW achieved the best mean intersection-over-union (IoU) score of 0.8019 for extracting retinal lesions and the best F_1 score of 99.53% for classifying AMD. The detailed evaluations on this transferability are shown in Table IX and X.

The rest of the paper is organized as follows. Section III presents the detailed architecture of RAG-FW, section IV describes the experimental setup followed by the results in section V. Section VI presents a detailed discussion about RAG-FW and section VII concludes the paper.

III. PROPOSED IMPLEMENTATION

RAG-FW employs hybrid convolutional network (RAG-Net) which contains dedicated units for lesions segmentation and retinopathy classification. Since both of these tasks are interlinked, requiring similar features, so a single CNN encoder is shared among them for feature extraction. Moreover, they have been trained jointly in which the weights adjusted by the segmentation unit for its convergence are utilized by the classification unit (which only fine-tunes them). The block diagram of the proposed framework is shown in Figure 2. First of all, the input scan is preprocessed through structure tensors [65] which retains the retina from the candidate scan and removes the background information, vendor annotations and acquisition artifacts etc. Afterwards, the scan is passed to the segmentation unit which extracts the potential lesions from it. Also, the classification unit screens the preprocessed scan against retinopathy which is further graded based upon the extracted lesions following the defined set of clinical rules as shown in Table II. This process is repeated for all the scans within the OCT volume and then the eye or volume level grading is generated by checking the longest scan connectivity. RAG-FW has been tested on 43,613 scans from different publicly available datasets acquired through different OCT machinery. The detailed discussion on each module is presented below:

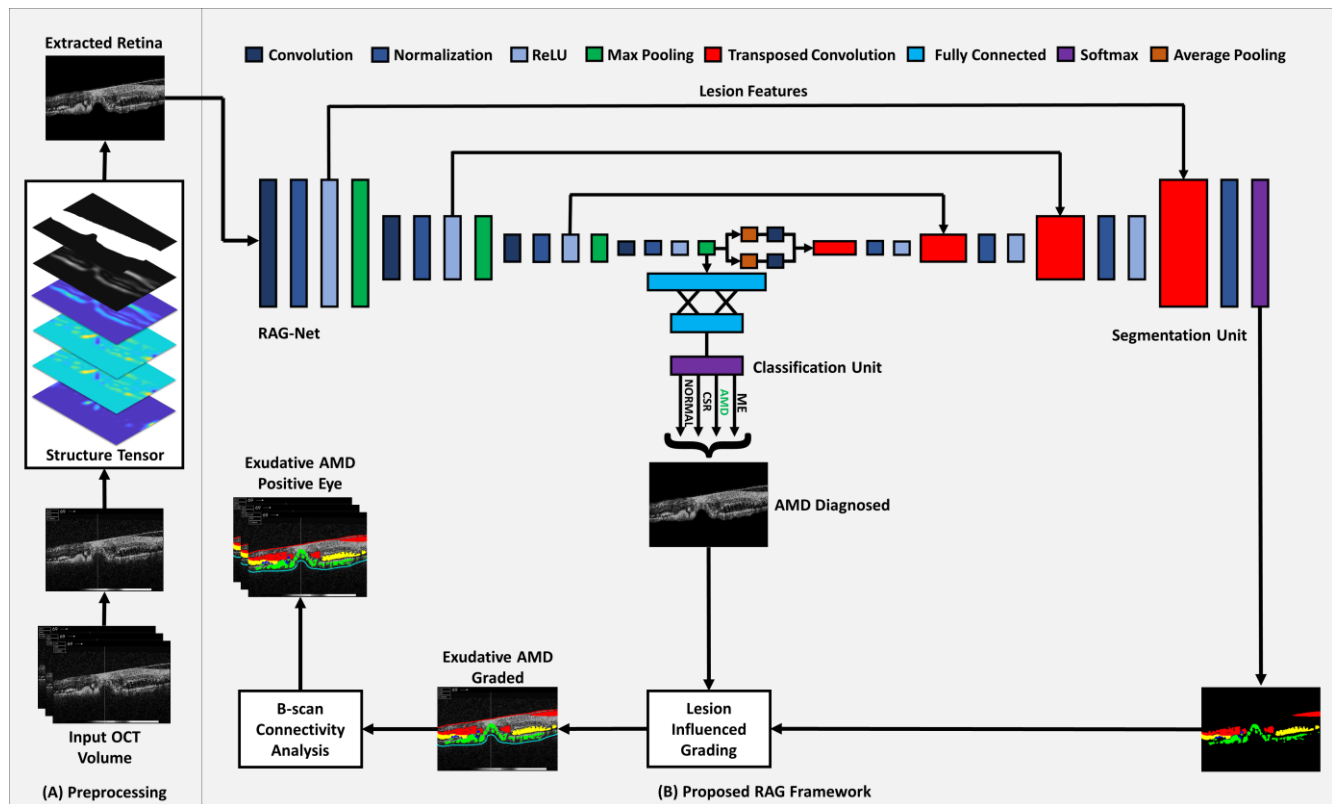


Figure 2: Block diagram of the proposed RAG framework (A) preprocessing stage to extract retina from the candidate scan, (B) proposed RAG framework that takes preprocessed scan as an input and uses RAG-Net for extracting retinal lesions and classifying retinopathy. Furthermore, RAG-FW uses the extracted lesion information for the lesion-influenced grading of retinopathy.

A. Preprocessing:

The OCT scan goes first into a preprocessing stage. The prime purpose of preprocessing stage is to isolate retina from the candidate scan which considerably improves the performance of CNN architectures in extracting lesions especially from the noisy and degraded scans. The preprocessing within the proposed framework is performed through structure tensors [65]. Structure tensors highlight the predominant orientations in the image gradients and tells the degree to which they are coherent within the specified neighborhood of a point. Let $S_T(z_1, z_2)$ be the structure tensor at pixel-level derived from the image gradients $I_X(z_1, z_2)$ and $I_Y(z_1, z_2)$ as expressed below:

$$S_T(z_1, z_2) = \begin{bmatrix} T_{XX}(z_1, z_2) & T_{XY}(z_1, z_2) \\ T_{YX}(z_1, z_2) & T_{YY}(z_1, z_2) \end{bmatrix} \quad (1)$$

and each tensor is computed through:

$$T_{XX}(z_1, z_2) = \sum_{r_1=-k}^k \sum_{r_2=-k}^k w(r_1, r_2) (I_X(q_1, q_2))^2 \quad (2)$$

$$T_{XY}(z_1, z_2) = \sum_{r_1=-k}^k \sum_{r_2=-k}^k w(r_1, r_2) (I_X(q_1, q_2) I_Y(q_1, q_2)) \quad (3)$$

$$T_{YX}(z_1, z_2) = \sum_{r_1=-k}^k \sum_{r_2=-k}^k w(r_1, r_2) (I_Y(q_1, q_2) I_X(q_1, q_2)) \quad (4)$$

$$T_{YY}(z_1, z_2) = \sum_{r_1=-k}^k \sum_{r_2=-k}^k w(r_1, r_2) (I_Y(q_1, q_2))^2 \quad (5)$$

where $q_1 = (z_1 - r_1)$, $q_2 = (z_2 - r_2)$, $w(r_1, r_2)$ denotes the

Gaussian window of finite length M that is used to remove noisy outliers while convolving image gradients and $k = \lfloor \frac{M}{2} \rfloor$ ($\lfloor \cdot \rfloor$ denotes floor operation). For each pixel within the candidate scan, we get a 2×2 symmetric structure tensor matrix as shown in Eq. (1) and repeating the same process for every pixel, we obtain four tensors representing image transitions at orthogonal orientations [65]. Afterwards, the tensor which represents the maximum retinal information is automatically selected through eigendecomposition [57]. Moreover, to isolate retina, the retinal mask is generated using the inner limiting membrane (ILM) and choroidal boundary (which are extracted from the selected tensor by iteratively searching for the first and last transition in each column of the scan). The retinal mask is then multiplied with the input scan to isolate retina as shown in Figure 3. The extracted retinal scan is then passed to the RAG-FW for further analysis.

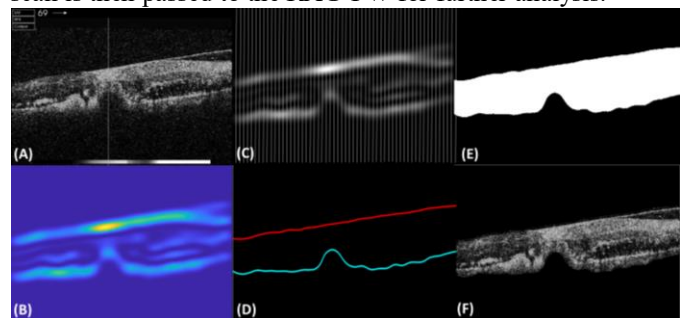


Figure 3: (A) OCT scan, (B) tensor with the maximum retinal information, (C) column-wise iterative search to pick ILM and choroidal boundary, (D) extracted retinal boundaries, (E) retinal mask, (F) isolated retina.

and used a joint training strategy in which the segmentation unit is trained first, and the classification unit utilize its weights and fine-tune its layers accordingly. This greatly reduces the overheads of managing two separate models and their training and reproducibility issues. The architectural details and hyper-parameters of RAG-Net is presented in Table II from which it can be observed that the RAG-Net architecture has 62,352,188 parameters in total where 62,240,828 parameters are learnable and 111,360 parameters are non-learnable. It should be noted here that the number of layers and their sizes have been finalized after rigorous experimentations on different publicly available datasets. Apart from this, both units employ cross-entropy loss Ce_L that measures the degree of dissimilarity between the actual output distribution and the predicted output distribution. Ce_L is computed through Eq. (6):

$$Ce_L = - \sum_{i=0}^{B_S} \sum_{j=1}^M T_{i,j} \log(P_{i,j}) \quad (6)$$

where B_S is the number of samples, M represents the number of classes, $T_{i,j}$ is the indicator that whether sample i belongs to class j or not, $P_{i,j}$ denotes the predicted probability of the i^{th} sample for the j^{th} class.

5) RAG-FW for Retinopathy Grading:

After extracting the retinal lesions and classifying the candidate scan, it is graded accordingly. The grading strategy is based on set of defined clinical standards as mentioned in the Table II. According to Early Treatment Diabetic Retinopathy Study (ETDRS) [67], ME is graded as clinically significant (CSME) if 1) retinal thickening is observed at or within 500 μm from the center of macula or 2) hard exudates associated with retinal thickening are observed at or within 500 μm from the center of macula or 3) retinal thickening zone(s) of one-disc area (or more) are observed of which at least one part is within one-disc diameter from the center of macula. However, we note here that ETDRS developed this definition of CSME and non-CSME more than three decades ago (before the usage of OCT imagery for retinal examination) [68] and OCT images only show clinically significant ME cases even in early stages when they are needed to be dealt with [68-69]. So, instead of grading CSME and non-CSME, we have considered grading ME as DME or P-CME in our study as it has more clinical relevance. A study was presented in [17] in which the authors have concluded that OCT imagery can solely classify DME and P-CME. Using their study as one of the guidelines, we also made RAG-FW to automatically distinguish between DME and P-CME as shown in Table II. This is of course one perspective of differentiating DME and P-CME cases as the patient history of diabetes and findings from other modalities cannot be fully ignored. Moreover, according to Age-Related Eye Disease Study (AREDS), AMD is graded as dry if only drusen and retinal pigment epithelium (RPE) atrophic profiles are observed while it is graded as wet if the retinal fluids (where IRF are more critical as compared to SRF [71]) or other chorioretinal abnormalities such as fibrotic scars or CNVM are observed. It should be noted here

that since the presence of either CNVM or fibrotic scars indicate the wet form of AMD. So, rather than identifying them individually, we collectively recognized them as CA in the proposed study. Apart from this, CSR is also clinically graded as acute, acute-persistent or chronic [72]. Acute CSR occurs due to the accumulation of SRF (typically dome-shaped). Acute CSR becomes persistent when the SRF levels are not decreased within three months of their appearance. Some cases are resolved automatically while other turns into long-lasting chronic CSR (within three to six months of SRF appearance) where RPE atrophy, fibrosis and neovascular membrane (leading to CNV) can be observed [72]. It is worth noting here that the proposed RAG-FW can differentiate between DME and P-CME cases. Also, it can further grade AMD as exudative (wet) or non-exudative (dry) and CSR as acute or chronic by checking the presence of retinal lesions in in the classified pathology. For example, if the CSR classified subjects has CA along with SRF, then it is graded as chronic CSR and if it only has SRF then it is graded as acute CSR.

Table I: Architectural details and hyper-parameters of RAG-Net. Bold ones are only present in segmentation unit while the underlined ones are only present in the classification unit.

| Layers | Number of Layers* | Parameters |
|------------------------------------|--|------------|
| Convolution | 112 | 53,282,054 |
| Batch Normalization | 111 | 222,720 |
| ReLU | 102 | 0 |
| Pooling | 2 Max Pooling 4 Average Pooling | 0 |
| Addition | 32 | 0 |
| Input | 2 | 0 |
| Zero Padding | 2 | 0 |
| Lambda | 5 | 0 |
| Concatenation | 1 | 0 |
| Reshape | 1 | 0 |
| <u>Fully Connected and Flatten</u> | <u>2</u> | 8,847,370 |
| Softmax | 2 | 0 |
| <u>Classification</u> | <u>1</u> | 44 |
| Trainable Parameters | | 62,240,828 |
| Non-trainable Parameters | | 111,360 |
| Total Parameters | | 62,352,188 |

* detailed summary is also presented in the codebase package.

IV. EXPERIMENTAL SETUP

We conducted a series of experiments that aim to assess the performance of RAG-FW in 1) extracting and identifying retinal lesions, 2) classifying different retinal pathologies, 3) grading the retinopathy subjects as per clinical standards on publicly available datasets.

A. Data:

All the datasets which have been used for training and validating RAG-FW are presented in Table III. Duke and Zhang datasets are acquired through Spectralis, Heidelberg Inc. machine while BIOMISA dataset is acquired through Topcon 3D OCT 2000 series machine. Duke datasets contain normal as well as DME and AMD affected scans, BIOMISA

dataset contains scans with healthy, AMD, ME and CSR pathology while Zhang dataset contains scans of drusen (dry AMD), DME, CNV (wet AMD) and healthy subjects. Furthermore, all these datasets have been annotated by expert clinicians. To validate the performance of RAG-FW for grading retinopathy, we also performed blind testing in which newly acquired scans were passed to the expert clinician as well as to RAG-FW where the results of RAG-FW were cross-verified with respect to the clinician's findings.

Table II: Retinopathy grading strategy

| Diagnosed Pathology | Grading strategy as per clinical standards* |
|---------------------|--|
| AMD | Dry (Non-exudative) AMD: When only drusen are observed within the OCT scan. Wet (Exudative) AMD: When drusen and CA are discovered along with IRF, SRF or HE within the OCT scan. |
| ME | P-CME: When the fluid filled cysts are found within intra-retina or sub-retina causing retinal thickening without the presence of HE in the OCT scan. DME: When the retinal fluid (IRF or SRF) causing retinal thickening or HE are discovered within the OCT scan. |
| CSR | Acute: When SRF is observed within the candidate OCT scan. Chronic: When CA is also observed along with SRF within the OCT scan. |

*Note: These strategies are under full compliance with clinical standards defined by ETDRS and AREDS [70] but only considering findings from OCT imagery. It is emphasized here that the patient history and findings from the other modalities (such as FP or FFA etc.) cannot be ignored.

B. Training Details:

Both segmentation and classification unit in the proposed framework have been trained on scans from five publicly available datasets where the training and testing data split is honored as per each dataset standard. Apart from this, the training of segmentation and classification unit is conducted in a joint fashion where the segmentation unit is trained first for retinal lesions extraction. When the segmentation unit is trained, its weights (of the encoder end) are transferred to the classification unit where they are fine-tuned for classification purposes. This joint training mechanism greatly reduces the cost, overheads and reproducibility issues involved in training two separate models. Furthermore, the transferred weights allow faster convergence for the classification unit as evident from the accuracy and cross entropy loss in Figure 5. In the proposed framework, training for both segmentation and classification unit is conducted for 20 epochs using adaptive learning rate method (ADADELTA) [73] as an optimizer with a default learning rate of one and a decay factor of 0.95 [73]. Each epoch is composed of 512 iterations where the validation is performed after each epoch during the training phase. For validation, we took 1,000 scans from the training dataset (100 scans from Duke-I, 100 scans from Duke-II, 400 scans from Duke-III, 200 scans from BIOMISA and 200 scans from Zhang dataset) along with their ground truths. Moreover, both units have been implemented on Anaconda platform with Python 3.7.4 and Keras APIs on a machine with Intel Core i5 8400 processor operating at 2.8 GHz clock frequency, 16 GB RAM and NVIDIA RTX 2080 GPU. The source code is

available at: <http://biomisa.org/index.php/downloads/>.

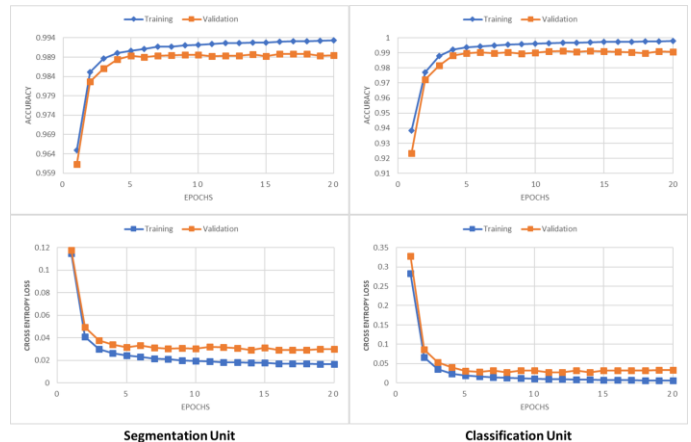


Figure 5: Training performance of RAG-Net segmentation and classification unit

C. Evaluation Metrics

The metrics which have been considered in this paper for the evaluation of RAG-FW are as follows:

1) Mean Dice Coefficient for Retinal Lesions Extraction

In order to measure the performance of RAG-FW for retinal lesion extraction and also to compare it with the existing state-of-the-art solutions, mean dice coefficient has been used. Mean dice coefficient is computed by taking an average of the dice coefficient (D_{Coff}) scores and D_{Coff} is computed through Eq. (6) [74]:

$$D_{Coff} = \frac{2(I_{LM} \cap G_T)}{I_{LM} + G_T} \quad (6)$$

where I_{LM} is the lesion map extracted from RAG-Net segmentation unit and G_T denotes the ground truth. D_{Coff} can also be computed through:

$$D_{Coff} = \frac{2T_p}{2T_p + F_p + F_N} \quad (7)$$

where T_p denotes the true positives (indicating pixel match between I_{LM} and G_T), F_p denotes the false positives (indicating number of incorrectly picked lesion pixels) and F_N denotes the false negatives (indicating number of missed lesion pixels).

2) Mean Intersection-over-Union for Lesions Extraction

To further indicate the responsiveness of RAG-FW in extracting different retinal lesions, we have used mean IoU. IoU is also known as Jaccard's Index and it also indicates how well the region of interest is extracted by comparing its similarity with the ground truth. IoU is computed through Eq. (8):

$$IoU = \frac{T_p}{T_p + F_p + F_N} \quad (8)$$

and IoU is related to D_{Coff} through Eq. (9):

$$IoU = \frac{D_{Coff}}{2 - D_{Coff}} \quad (9)$$

The mean IoU is then computed by taking the average of IoU scores.

Table III: Summary of Datasets Used

| Datasets | Pathologies | | | | | Total Scans | Subjects | Training Scans [#] | Testing Scans | Machinery |
|-----------------------|------------------|--------|--------|---------------|-------|-------------|---|-----------------------------|--------------------|--------------------|
| | Drusen (dry AMD) | ME | Normal | CNV (wet AMD) | CSR | | | | | |
| Duke – I [75] | 26,900 | - | 11,500 | - | - | 38,400 | 384 (269 AMD and 115 Normal) | 300 | 38,100 | Spectralis |
| Duke – II [25] | - | 610 | - | - | - | 610 | 10 (All DME) | 305 (subject 1-5) | 305 (subject 6-10) | Spectralis |
| Duke – III [50] | 723 | 1,101 | 1,407 | - | - | 3,231 | 45 (15 AMD, 15 DME and 15 Normal) | 3,048 | 183 | Spectralis |
| BIOMISA Dataset [76]* | 657 | 2,195 | 904 | 407 | 1,161 | 5,324 | 99 (27 AMD, 31 ME, 24 CSR and 17 Healthy) | 1,299 | 4,025 | Topcon 3D OCT 2000 |
| Zhang Dataset [53] | 8,866 | 11,598 | 51,390 | 37,455 | - | 109,309 | Not available | 108,309 | 1,000 | Spectralis |

* Number of eyes in BIOMISA dataset are as follows: AMD (Left: 19, Right: 9), ME (Left: 28, Right: 16), Normal (Left: 10, Right: 7), CSR (Left: 14, Right: 12). Also, BIOMISA dataset originally contains pseudo-colored OCT images. But, to maintain consistency across all datasets to achieve better segmentation, classification and grading performance, we only used the green channel of BIOMISA dataset scans as it contains the maximum retinal information.

[#] 1,000 out of 113,261 training scans are used for validation purposes.

3) Receiver Operator Characteristics (ROC) Curve for Retinal Lesions Extraction and Identification

To further validate the performance of RAG-FW and its robustness in extracting retinal lesions especially from the scans depicting complex retinal pathologies, we computed ROC curves where the classification threshold was varied from 0 to 1 in step of 0.001. In addition to this, we have computed ROC curves (using the same configurations) to measure the ability of the proposed framework for correctly classifying each pathology.

4) Confusion Matrices for Retinopathy Classification and Grading

The performance of the proposed framework for diagnosing and grading retinopathy is accessed by generating the confusion matrices and measuring the accuracy, sensitivity, specificity, precision and F_1 scores through:

$$\text{Accuracy} = A_C = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (10)$$

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

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

$$\text{Precision} = P_{PV} = \frac{T_P}{T_P + F_P} \quad (13)$$

$$F_1 = \frac{2(P_{PV} \times T_{PR})}{(P_{PV} + T_{PR})} \quad (14)$$

where T_N denotes the true negatives.

5) Blind Testing for Retinopathy Grading

In order to further validate the performance of RAG-FW for retinopathy grading, we performed blind testing in Armed Forces Institute of Ophthalmology (AFIO), Rawalpindi Pakistan, where the trained RAG-FW was fed with the newly acquired OCT scans. These scans were also given to the expert clinician for grading purposes in a separate room and the performance of RAG-FW in grading retinopathy was then cross-verified with the clinician results.

6) Qualitative Evaluations

We also presented different qualitative evaluations of RAG-FW to demonstrate its performance in extracting and recognizing retinal lesions as well as classifying and grading retinopathy.

V. RESULTS

RAG-FW has been validated on five publicly available datasets for lesions segmentation and lesion-influenced grading of retinopathy. In addition to this, the performance of RAG-FW for retinal lesion extraction has been thoroughly compared with state-of-the-art segmentation frameworks such as PSPNet [66], SegNet [77], UNet [78], FCN (8s and 32s) [79-80] where preprocessing stage has been added in all the frameworks for fair comparison. To increase the readability, we first presented the evaluations of RAG-FW with respect to each dataset separately and then also on the combination of all the datasets as shown below:

A. Evaluations on Zhang Dataset

Zhang dataset is one of the largest OCT dataset till date. It contains a total of 109,309 OCT scans from dry AMD, wet AMD, DME and the healthy subjects (dry AMD is named as drusen and wet AMD is named as CNV in the dataset). The images in Zhang dataset are acquired through Spectralis, Heidelberg Inc. where the number of patients and eyes are not disclosed. The dataset also contains chest X-ray scans from healthy and pneumonia affected subjects but we have excluded those scans in our study because we are only focusing here on the human retina. Zhang dataset has been made publicly available by [53] primarily for the retinopathy classification and only contains ground truths to measure the retinopathy diagnostic performance. However, for more in-depth evaluation of RAG-FW for extracting retinal lesions, we got the Zhang dataset annotated from the expert clinicians in AFIO, Rawalpindi for each retinal lesion. On Zhang dataset,

first of all, we thoroughly compared RAG-FW with [53] for classifying retinopathy. The classification performance of RAG-FW and [53] are shown in Figure 6 through confusion matrix where it can be seen that RAG-FW achieved the accuracy of 98.6% (2.02% better than [53]) and F_1 score of 99.06% (0.46% better than [53]) for retinopathy classification on 1,000 test scans. For fair comparison, we have used the same training and testing configurations as described in [53] along with their original ground truths. RAG-FW achieved better results than [53] because the weights used by the proposed classification unit are accelerated by the segmentation unit to pay attention to the lesion features. So, rather than relying solely on image representation of each pathology, the classification unit screens retinopathy by considering the lesion regions as well.

| | | True Labels | | | |
|------------------|--------|-------------|-----|--------|--------|
| | | CNV | DME | DRUSEN | NORMAL |
| Predicted Labels | CNV | 247 | 1 | 9 | 0 |
| | DME | 2 | 249 | 0 | 0 |
| | DRUSEN | 1 | 0 | 241 | 1 |
| | NORMAL | 0 | 0 | 0 | 249 |

(A)

| | | True Labels | | | |
|------------------|--------|-------------|-----|--------|--------|
| | | CNV | DME | DRUSEN | NORMAL |
| Predicted Labels | CNV | 242 | 9 | 7 | 0 |
| | DME | 5 | 237 | 2 | 0 |
| | DRUSEN | 3 | 1 | 236 | 4 |
| | NORMAL | 0 | 3 | 5 | 246 |

(B)

Figure 6: Comparison of retinal diagnosis on Zhang dataset, (A) RAG-FW, (B) [53]

Afterwards, we compared the performance of proposed segmentation unit with state-of-the-art models for extracting different retinal lesions from Zhang dataset as shown in Table IV. Here, the metric of evaluation is mean IoU and it can be observed that the proposed framework outperformed other networks by achieving the mean IoU score of 0.7852 which is 4.58% better than PSPNet (the 2nd best) and 54.17% better than FCN-32. It can be seen that although RAG-FW stood 2nd best for extracting IRF and SRF but the gap between RAG-FW and the best performing networks is very minimal i.e. RAG-FW lags 0.32% from PSPNet in extracting IRF and 1.62% from SegNet in extracting SRF. However, if we compare the performance of RAG-FW for extracting CA and HE with the 2nd best network then we can see that RAG-FW extracted these lesions with a neat gap of 4.25% and 14.28%, respectively. Apart from this, Zhang dataset contains many complex OCT scans where the retinal lesion extraction is quite challenging as evident from Figure 7 (C), (D) and (E) in which multiple lesions such as IRF and CNVM can be observed. However, the proposed RAG-FW not only extracted the retinal lesions proficiently but was able to judge them correctly as compared to its competitors even when these lesions have the similar texture in the scan. This is because RAG-FW preserves the scan contextual information during lesion segmentation and adds the best features together during encoding process. More scans depicting the lesion extraction performance of RAG-FW and its comparison with other segmentation networks is presented in the codebase package.

B. Evaluations on Duke-I Dataset

Duke-I contains a total of 38,400 B-scans depicting dry AMD

and the normal pathology. These scans are acquired through using Spectralis, Heidelberg Inc. The Duke-I dataset is primarily designed to evaluate the performance of automated frameworks for the extraction of retinal layers. Therefore, it does not contain any annotations for the retinal lesions. In order to evaluate the lesion extraction performance of the proposed framework we got it annotated from AFIO, Rawalpindi. The lesion extraction performance of RAG-FW on Duke-I can also be observed in Table IV where it can be seen that RAG-FW achieved the mean IoU score of 0.8193 for correctly extracting and identifying drusen (the only lesion present in Duke-I). Comparing the performance of RAG-FW with other segmentation frameworks, we can see that RAG-FW achieved 3.18% better performance than PSPNet and 69.02% better performance than FCN-32. We have also utilized scans from Duke-I dataset along with other Duke datasets for evaluating the performance of RAG-FW for retinopathy classification and grading. The detailed discussion about this is presented in Section V (F).

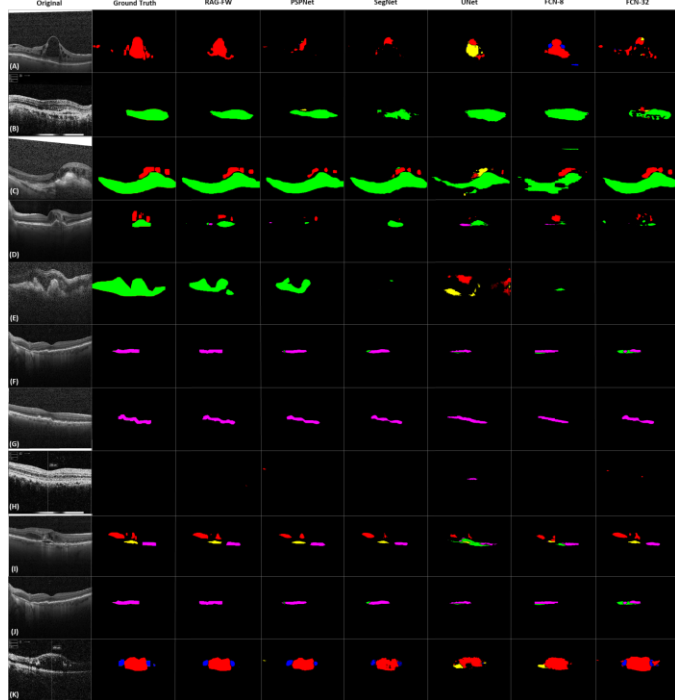


Figure 7: Visual comparison of RAG-FW with PSPNet, SegNet, UNet, FCN-8 and FCN-32 for lesions extraction from different retinal pathologies. Red color shows IRF, yellow color shows SRF, blue color shows HE, pink color shows drusen and green color shows extracted CA regions.

C. Evaluations on Duke-II Dataset

Duke-II is the second publicly available dataset from Vision and Image Processing (VIP) lab, Duke University, USA which we have considered in the evaluation of proposed framework. The dataset has been first introduced in [25] where the authors have evaluated their framework on it for retinal layers and fluid extraction. Duke-II contains 610 OCT scans from 10 DME subjects. These scans are acquired through Spectralis, Heidelberg Inc. We also want to highlight here that Duke-II is one of the dataset that has extensively used by the research community working in OCT based retinal image analysis.

Duke-II dataset has been primarily designed to evaluate the automated tools for the extraction of retinal layers and retinal fluid and contains ground truths which are marked by two clinicians (where the markings have significant variability with each other [25]). RAG-FW has been applied on Duke-II for the extraction and identification of retinal lesions as well as for the classification and grading of retinopathy. Table V shows the comparison of the proposed framework with [25], [31] and [55] in terms of mean D_{Coff} and mean IoU. For fairness, the comparison shown in Table V is based on original ground truths provided in the Duke-II dataset and markings from both clinicians have been considered. It can be seen from Table V that RAG-FW produced better fluid extraction results than [25], [31] and [55] on Duke-II and achieved the mean D_{Coff} and mean IoU score of 0.664 and 0.497, respectively. Here, we want to mention that [34] also extracted the retinal fluid from Duke-II dataset and achieved the mean D_{Coff} of 0.77. But, we haven't included this in Table V because [34] achieved this score by considering the markings of only one clinician. So, its comparison with RAG-FW would have been unfair since we computed the mean D_{Coff} score using the original markings from both clinicians.

Table IV: Performance comparison of RAG-FW in terms of IoU for extracting retinal lesions from all five publicly available datasets

| Lesion | RAG-FW | PSPNet | SegNet | UNet | FCN-8 | FCN-32 | |
|----------|----------|---------------|---------------|---------------|--------|--------|--------|
| Zhang | Drusen | 0.8153 | 0.7831 | 0.8143 | 0.7725 | 0.4256 | 0.3573 |
| | CA | 0.9385 | 0.8986 | 0.8732 | 0.8451 | 0.7238 | 0.6853 |
| | IRF | <u>0.8165</u> | 0.8192 | 0.8036 | 0.7824 | 0.4824 | 0.3572 |
| | SRF | <u>0.7342</u> | 0.7125 | 0.7463 | 0.7241 | 0.4164 | 0.3752 |
| | HE | 0.6217 | <u>0.5329</u> | 0.5079 | 0.5471 | 0.3064 | 0.0241 |
| | Mean IoU | 0.7852 | <u>0.7492</u> | 0.7490 | 0.7342 | 0.4709 | 0.3598 |
| Duke-I | Drusen | 0.8193 | <u>0.7932</u> | 0.7546 | 0.6913 | 0.3751 | 0.2538 |
| | Mean IoU | 0.8193 | <u>0.7932</u> | 0.7546 | 0.6913 | 0.3751 | 0.2538 |
| Duke-II | IRF | 0.8693 | 0.8371 | <u>0.8678</u> | 0.6718 | 0.3249 | 0.2378 |
| | SRF | 0.8439 | 0.8296 | <u>0.8387</u> | 0.6516 | 0.3182 | 0.2453 |
| | HE | 0.6347 | <u>0.5769</u> | 0.5019 | 0.5194 | 0.2157 | 0.0152 |
| | Mean IoU | 0.7826 | <u>0.7478</u> | 0.7361 | 0.6142 | 0.2862 | 0.1661 |
| Duke-III | Drusen | <u>0.8162</u> | 0.8249 | 0.8063 | 0.7632 | 0.4079 | 0.3083 |
| | IRF | 0.8253 | <u>0.8017</u> | 0.7953 | 0.7791 | 0.4531 | 0.3729 |
| | SRF | <u>0.7536</u> | 0.7421 | 0.7545 | 0.7493 | 0.4536 | 0.3412 |
| | HE | 0.6342 | <u>0.5673</u> | 0.4841 | 0.5218 | 0.3524 | 0.0453 |
| | Mean IoU | 0.7573 | <u>0.7340</u> | 0.7100 | 0.7033 | 0.4167 | 0.2669 |
| BIOMISA | Drusen | 0.8264 | 0.7753 | <u>0.8053</u> | 0.7832 | 0.4129 | 0.3143 |
| | CA | 0.9436 | 0.8874 | 0.8602 | 0.8368 | 0.7364 | 0.6647 |
| | IRF | 0.8352 | <u>0.8276</u> | 0.7915 | 0.8154 | 0.4251 | 0.3127 |
| | SRF | 0.7158 | 0.7093 | <u>0.7148</u> | 0.6931 | 0.3957 | 0.3416 |
| | HE | 0.6384 | <u>0.5571</u> | 0.5168 | 0.5762 | 0.3158 | 0.0185 |
| | Mean IoU | 0.7918 | <u>0.7513</u> | 0.7377 | 0.7409 | 0.4571 | 0.3303 |

Furthermore, we have compared the proposed framework with [27] and [36] as well for fluid extraction but since these frameworks are tested on their local in-house datasets. So, the comparison with [27] and [36] is indeed indirect. Also, it should be noted here that IoU is a more strict measure as compared to D_{Coff} and from Table V, it is evident that RAG-FW achieved 2.56% better fluid extraction results than [55] (the 2nd best) and 14.15% better results than [25] in terms of

mean D_{Coff} (or 3.82% better results than [55] and 19.91% better results than [25] in terms of mean IoU). Apart from this, we have compared the performance of RAG-FW with the popular segmentation networks as well for the extraction of IRF, SRF and HE as shown in Table IV. Here, the ground truths for IRF, SRF and HE have been obtained through AFIO, Rawalpindi and all the segmentation networks have been evaluated on these ground truths. It can be seen that the RAG-FW achieved the mean IoU score of 0.7826 which is 4.44% better than the 2nd best performing network (PSPNet) and 78.77% better than FCN-32.

Table V: Comparison of retinal fluid extraction on different datasets based on mean dice coefficient and mean IoU, Bold indicates the best performance while the second best performance is underlined.

| Metric | Dataset | RAG-FW | [25] | [31] | [36] | [27] | [55] |
|------------------------|---------|--------------|-------|-------|-------|--------|--------------|
| Mean D_{Coff} | Duke-II | 0.664 | 0.57 | 0.61 | - | - | <u>0.647</u> |
| | BIOMISA | 0.934 | - | - | - | - | <u>0.906</u> |
| | Local* | - | - | - | 0.789 | 0.9273 | - |
| Mean IoU | Duke-II | 0.497 | 0.398 | 0.438 | - | - | <u>0.478</u> |
| | BIOMISA | 0.876 | - | - | - | - | <u>0.828</u> |
| | Local | - | - | - | 0.651 | 0.864 | - |

* The comparison of RAG-FW with [27] and [36] is indirect because they have extracted the retinal fluid on their local datasets.

D. Evaluations on Duke-III Dataset

Duke-III is the third publicly available dataset from VIP lab which we used for the evaluation of RAG-FW. The dataset is first introduced in [50] for the classification of AMD, DME and normal subjects. Duke-III contains OCT scans from 15 AMD subjects, 15 DME subjects and 15 healthy subjects. These scans are acquired through Spectralis, Heidelberg Inc. The classification performance of RAG-FW on Duke-III dataset is presented in Table VI. Here, it should be noted that the performance for all methods is measured subject-wise not the scan-wise as per the dataset standard where the proposed framework and [54] correctly classified all the 45 subjects and [50] classified 43 out of 45 subjects correctly. In addition to this, we have evaluated RAG-FW for the extraction of different retinal lesions as shown in Table IV. Here the ground truths for retinal lesions are obtained through AFIO, Rawalpindi Pakistan since Duke-III does not originally contain annotations for the retinal lesions. Results reported in Table IV shows that the best performance for extracting drusen and SRF in term of IoU is achieved by the PSPNet and SegNet, respectively. However, RAG-FW is able to achieve the best mean IoU score of 0.7573 for extracting retinal lesions which is 3.07% better than PSPNet and 64.75% better than FCN-32. Overall, RAG-FW achieved the best performance because it outperformed its competitors in extracting IRF by 2.85% and HE by 10.54% respectively.

E. Evaluations on BIOMISA Dataset

The last dataset on which we have evaluated RAG-FW is the BIOMISA dataset. Contrary to other datasets used in this paper, BIOMISA dataset contains scans acquired from Topcon 3D OCT 2000 machine. Furthermore, the scans within BIOMISA dataset reflect dry AMD, wet AMD, DME, P-

CME, acute CSR, chronic CSR and normal pathologies. Moreover, BIOMISA dataset can also be used for multimodal analysis as well as it contains both fundus and OCT scans for each subject. It should be noted here that BIOMISA dataset only contains pseudo-colored OCT images while other datasets contain original grayscale ones. To achieve better performance, we maintained the consistency with other datasets by using the green channel of BIOMISA OCT scans only as it contains maximum amount of retinal information. Moreover, BIOMISA dataset has been made publicly available in [76] and it is primarily designed to test the automated frameworks for the extraction of retinal layers, retinal lesions, classification and grading of retinopathy.

Table VI: Comparison of retinal diagnosis on different datasets. Bold indicates the better performance and the 2nd best results are underlined

| | Metric | RAG-FW* | [50] | [53] | [54] | [56] | [57] |
|----------|-----------------|---------------|---------------|--------------|-------------|---------------|---------------|
| Zhang | A _C | 98.60% | - | <u>96.6%</u> | - | - | - |
| | T _{PR} | 98.27% | - | <u>97.8%</u> | - | - | - |
| | T _{NR} | 99.60% | - | <u>97.4%</u> | - | - | - |
| | P _{PV} | 99.86% | - | <u>99.4%</u> | - | - | - |
| | F ₁ | 99.06% | - | <u>98.6%</u> | - | - | - |
| Duke-III | A _C | 100% | <u>95.55%</u> | - | 100% | - | - |
| | T _{PR} | 100% | 100% | - | 100% | - | - |
| | T _{NR} | 100% | <u>86.67%</u> | - | 100% | - | - |
| | P _{PV} | 100% | <u>93.75%</u> | - | 100% | - | - |
| | F ₁ | 100% | <u>96.77%</u> | - | 100% | - | - |
| BIOMISA | A _C | 99.01% | - | - | - | 96.07% | <u>97.78%</u> |
| | T _{PR} | <u>99.09%</u> | - | - | - | 96.08% | 100% |
| | T _{NR} | 98.67% | - | - | - | <u>96.04%</u> | 93.33% |
| | P _{PV} | 99.95% | - | - | - | <u>97.85%</u> | 96.77% |
| | F ₁ | 99.37% | - | - | - | 96.95% | <u>98.35%</u> |
| COMBINED | A _C | 99.32% | - | - | - | - | - |
| | T _{PR} | 99.40% | - | - | - | - | - |
| | T _{NR} | 99.12% | - | - | - | - | - |
| | P _{PV} | 99.65% | - | - | - | - | - |
| | F ₁ | 99.52% | - | - | - | - | - |

* support lesion-influenced grading

In the results depicted in Table IV, we can see that RAG-FW achieved the mean IoU score of 0.7918 which is 5.11% better than the PSPNet result. RAG-FW also achieved the best IoU scores for all the lesion classes as compared to its competitors with a significant lead in extracting CA and HE. Furthermore, we report a qualitative comparison in Figure 7 (B), (H) and (K) where it can be observed that RAG-FW produced better results than other networks. If we look at Figure 7 (K), we can see that RAG-FW achieved better extraction of IRF and HE as compared to its competitors without confusing between different lesion areas as done by PSPNet, UNet and FCN-8. We note, however, the best performance in Figure 7 (H) showing normal pathology is achieved by UNet and SegNet where RAG-FW confused a tiny region of hyper-reflectivity as IRF. Moreover, Table V presents the performance comparison of RAG-FW with [55] for retinal fluid extraction where it can be observed that RAG-FW achieved the mean

D_{Coff} of 0.934 and mean IoU of 0.876 which is 2.99% better than [55] in term of mean D_{Coff} (or 5.47% better than [55] in terms of mean IoU). Apart from this, Table VI shows the performance of RAG-FW for classifying retinopathy where it can be seen that the proposed framework achieved the accuracy and F_1 score of 99.01% and 99.37%, respectively for classifying retinopathy. Comparing the classification performance of RAG-FW with its competitors in Table VI, we can see that it produces 1.24% better results than [57] in terms of accuracy, 2.66% better results than [56] in terms of specificity, 2.10% better results than [56] in terms of precision and 1.02% better results than [57] in terms of F_1 score. However, it produces second best performance in terms of sensitivity where it lags only 0.91% than [57]. One of the key feature of RAG-FW is that it can grade retinopathy as well unlike its competitors. So, in order to measure the performance of RAG-FW for retinopathy grading, we performed blind testing as mentioned in Section IV (C). Table VII shows few blind testing results in which the clinician grading is also reported along with RAG-FW grading. Apart from this, we have added the fundus scan as well in Table VII to give better insight on the results. It can be observed from Table VII that RAG-FW has effectively graded the retinopathy especially for the P-CME pathology in 2nd row of Table VII, acute and chronic CSR in 5th and 6th row of Table VII, respectively. So, by looking at these results, we can see the potential of RAG-FW for mass screening retinal patients across the world (especially in remote areas), aiding ophthalmologists in overcoming blindness. The overall retinopathy grading results across all datasets is presented in Figure 12.

F. More Evaluations on Combined Dataset

Apart from evaluating RAG-FW on each dataset separately and comparing it with state-of-the-art solutions, we have further evaluated it on the combination of all datasets for extracting retinal lesions as well as classifying and grading retinopathy. The combined dataset contains 43,613 multi-vendor OCT scans for evaluation purposes. The performance of RAG-FW for extracting different retinal lesions on combined dataset is presented in Table VIII in terms of mean IoU where it can be observed that the proposed framework achieved the mean IoU score of 0.8055 ± 0.1009 . Comparing the performance of RAG-FW with other networks, we can see that RAG-FW has outperformed the 2nd best performance of PSPNet by 3.62%. Furthermore, to evaluate the performance of RAG-FW to see how much it confuses between different retinal lesions (especially the ones which can be easily misclassified like IRF and SRF or CA and drusen), we computed ROC curves as shown in Figure 8. It can be observed from Figure 8 that the minimum AUC score is achieved for recognizing IRF vs SRF i.e. 0.9758. But considering the fact that both IRF and SRF have highly similar image features, the performance of RAG-FW in distinguishing these regions is quite impressive. Furthermore, we can observe in Figure 8 that RAG-FW has efficiently discriminated between IRF/ SRF and HE regions by achieving the AUC score of 0.9893 and 0.9833, respectively. Also, the maximum AUC score is achieved for recognizing fluid and no-fluid

region i.e. 0.9938. Apart from this, Figure 9 presents some qualitative results of RAG-FW demonstrating its performance for extracting different retinal lesions from multi-vendor OCT scans. In Figure 9, it can also be seen that how effectively RAG-FW has extracted retinal lesions from some of the complex OCT scans demonstrating different retinal pathologies. For example, by observing Figure 9 (S), (T) and (U), we can see the performance of RAG-FW in extracting CA regions from Wet AMD affected scans.

Table VII: Blind retinopathy grading

| OCT B-scan | Fundus scan (for reader cross reference) | Clinician Grading | RAG-FW Grading |
|------------|---|----------------------|-------------------|
| | | DME | DME |
| | | P-CME | P-CME |
| | | Dry AMD | Dry AMD |
| | | Wet AMD | Wet AMD |
| | | Acute CSR | Acute CSR |
| | | Chronic CSR | Chronic CSR |
| | | DME | DME |
| | | DME | DME |

Similarly, by observing Figure 9 (V), (W) and (X), we can see how effectively RAG-FW has picked drusen from Dry AMD affected scans. Apart from this, the performance of RAG-FW for extracting IRF and SRF from Wet AMD, Acute CSR and DME affected scans can be seen in Figure 9 (G), (H), (I) and (J) and examples of extracted HE regions can be seen in Figure 9 (K) and (L). Moreover, the performance of RAG-FW for classifying retinopathy on combined dataset can be seen in a Figure 10 as well as in Table VI where it can be observed that the proposed framework achieved the accuracy and F_1 score of 99.32% and 99.52%, respectively for classifying retinopathy. To further validate RAG-FW for classifying

retinopathy, we computed ROC curves to see how much it confuses between different retinal pathologies. Figure 11 shows these curves where it can be observed that RAG-FW achieved the minimum AUC score of 0.9323 for diagnosing ME as compared to other pathologies while the best performance is achieved for diagnosing normal vs abnormal retinal pathologies i.e. the AUC score of 0.9417.

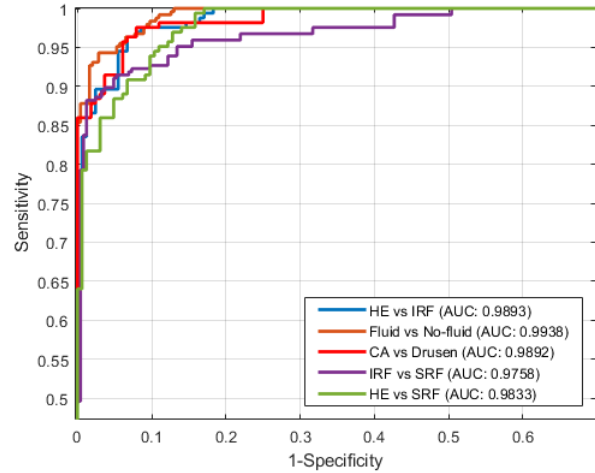


Figure 8: ROC curves showing performance of RAG-FW in distinguishing different retinal lesions from multi-vendor OCT scans (best visualized in color). It can be observed how well the lesions having similar image features are identified.

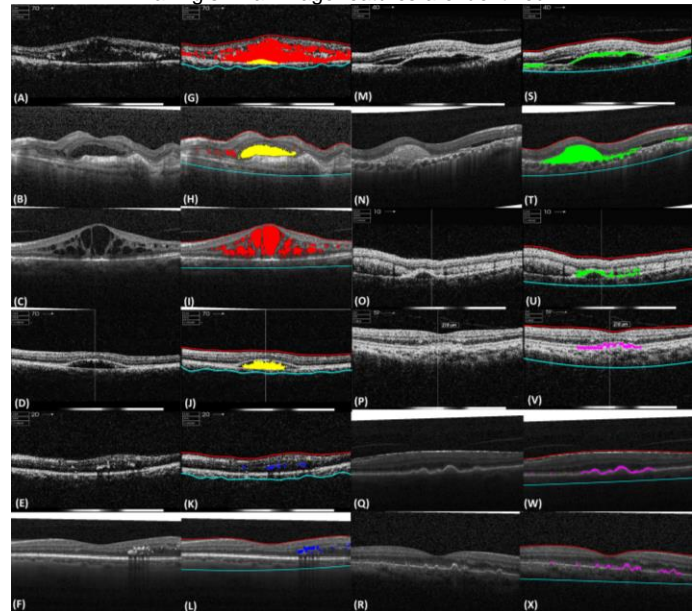


Figure 9: Qualitative results demonstrating the performance of RAG-FW for extracting IRF (red), SRF (yellow), HE (blue), drusen (magenta) and CA (green) regions from multi-vendor OCT scans. ILM and choroidal boundary is represented by red and cyan color, respectively. Apart from classifying retinopathy, RAG-FW can also grade it as per the clinical standards. The retinopathy grading performance of RAG-FW can be seen in Figure 12 (A) where it can be observed that RAG-FW achieved the accuracy of 98.70%. In addition to this, we have performed another experiment in which we have only used classification unit for the direct grading of retinopathy instead of classifying the scans first and grading them based upon the extracted lesions. The performance of RAG-FW for direct grading is shown in Figure 12 (B) where it can be observed that the classification

has confused a lot of DME and P-CME affected scans as well as Dry AMD and Wet AMD affected scans. This result evidences that the extraction of retinal lesions is crucial for the accurate retinopathy grading. It also concurs with the relevance of retinal lesions extraction for the retinopathy classification highlighted in [39]. In another series of experiments, we assessed the generalization capacity of the proposed framework through cross-datasets validation. Table IX reports a comparison of our proposed framework and other segmentation models. We can notice that the best performance is achieved by RAG-FW with a mean IoU score of 0.7936 for Duke → Zhang, 0.8019 for Zhang → Duke, 0.6871 for Duke → BIOMISA, 0.6835 for BIOMISA → Duke, 0.6687 for BIOMISA → Zhang and 0.6703 for Zhang → BIOMISA.

Table VIII: Performance comparison of retinal lesion extraction based on mean IoU. Bold indicates the best score while the second best is underlined. (Note that the following results are computed on the combination of all the datasets)

| Lesions | RAG-FW | PSP-Net | SegNet | UNet | FCN-8 | FCN-32 |
|---------|---------------|---------------|---------------|--------|--------|--------|
| SRF | <u>0.7682</u> | 0.7382 | 0.7711 | 0.7531 | 0.5641 | 0.4823 |
| IRF | 0.8373 | <u>0.8120</u> | 0.7910 | 0.7538 | 0.4921 | 0.4033 |
| CA | 0.9509 | <u>0.9143</u> | 0.9116 | 0.8947 | 0.7364 | 0.7107 |
| HE | 0.6416 | <u>0.5914</u> | 0.5177 | 0.5604 | 0.2307 | 0.0211 |
| Drusen | 0.8299 | <u>0.8257</u> | 0.7907 | 0.8155 | 0.4407 | 0.3543 |
| Mean | 0.8055 | <u>0.7763</u> | 0.7564 | 0.7555 | 0.4928 | 0.3943 |
| ± | ± | ± | ± | ± | ± | ± |
| STD | 0.1009 | <u>0.1080</u> | 0.1293 | 0.1104 | 0.1648 | 0.2230 |

| | | True Labels | | | |
|------------------|--------|--------------|-------------|------------|--------------|
| | | AMD | ME | CSR | NORMAL |
| Predicted Labels | AMD | 28072 | 27 | 15 | 47 |
| | ME | 43 | 2199 | 16 | 19 |
| | CSR | 18 | 19 | 786 | 42 |
| | NORMAL | 27 | 9 | 11 | 12263 |

Figure 10: Performance of RAG-FW for classifying retinopathy on combined dataset

Here, it should be noted that we have combined all the Duke datasets together for this experiment and named it as Duke to avoid redundant cross-dataset combinations.

It can also be noticed from Table IX that all models have shown their best performance for the pair of Duke and Zhang datasets. This is because Zhang dataset and Duke datasets are acquired through Spectralis machine so they have similar scan properties while BIOMISA dataset is acquired through Topcon machine. Table IX also shows the comparison between retinal lesions which were uncommon in both datasets pair. For example, Duke datasets does not contain Wet AMD (CNV) scans contrary to Zhang dataset. However, we didn't filter the scans containing CA regions in the experiment as CA is not present in Duke datasets. If these scans are filtered, then the performance of RAG-FW would have been even better. We have also conducted cross-dataset validations for the retinopathy classification. Here, we only excluded the evaluation for the CSR since the scans depicting

this disease was only present in the BIOMISA dataset. It can be observed in the results, depicted in Table X that the RAG-FW still preserves a high level of accuracy for the Duke and Zhang pairs as they have been acquired with same machine. We have though an exception at F_1 and P_{PV} scores (highlighted in bold) for the ME class when Duke datasets are used for testing due to the imbalanced ME class in Duke-II (only 305 samples). For the other pair of datasets, we notice a decrease in the performance which is expected due to the differences in the scanner specifications. Note that because of space limitation we could not report the confusion matrices of these validations, but they can be consulted in the code documentation <http://biomisa.org/index.php/downloads/>.

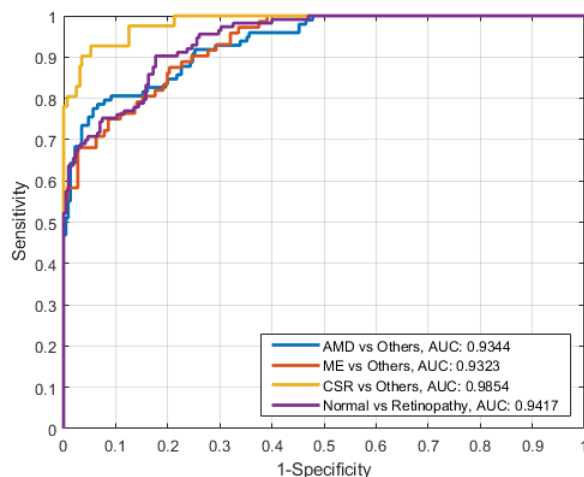


Figure 11: ROC performance curve of RAG-FW classification unit in classifying retinopathy

| | | True Labels | | | | | | | |
|------------------|-------------|-------------|------|---------|---------|-----------|-------------|--------|--|
| | | P-CME | DME | Wet AMD | Dry AMD | Acute CSR | Chronic CSR | Normal | |
| Predicted Labels | P-CME | 591 | 54 | 3 | 14 | 7 | 2 | 8 | |
| | DME | 17 | 1537 | 15 | 11 | 2 | 5 | 11 | |
| | Wet AMD | 2 | 25 | 531 | 151 | 6 | 9 | 16 | |
| | Dry AMD | 0 | 0 | 14 | 27376 | 0 | 0 | 31 | |
| | Acute CSR | 4 | 8 | 3 | 6 | 484 | 16 | 18 | |
| | Chronic CSR | 5 | 2 | 5 | 4 | 21 | 265 | 24 | |
| | Normal | 7 | 2 | 4 | 23 | 8 | 3 | 12263 | |

(A)

| | | True Labels | | | | | | | |
|------------------|-------------|-------------|-----|---------|---------|-----------|-------------|--------|--|
| | | P-CME | DME | Wet AMD | Dry AMD | Acute CSR | Chronic CSR | Normal | |
| Predicted Labels | P-CME | 271 | 535 | 46 | 37 | 12 | 7 | 12 | |
| | DME | 159 | 641 | 61 | 25 | 18 | 4 | 13 | |
| | Wet AMD | 141 | 353 | 357 | 13223 | 64 | 51 | 7 | |
| | Dry AMD | 2 | 3 | 22 | 14205 | 3 | 2 | 24 | |
| | Acute CSR | 14 | 37 | 4 | 19 | 211 | 103 | 29 | |
| | Chronic CSR | 31 | 54 | 79 | 28 | 203 | 122 | 35 | |
| | Normal | 8 | 5 | 6 | 48 | 17 | 11 | 12251 | |

(B)

Figure 12: Performance comparison of RAG-FW for retinopathy grading (A) lesion-influenced grading, (B) direct grading. The scans from all the datasets were considered for evaluation

Table IX: Performance comparison of lesion extraction when trained on one dataset and tested on another (Training Dataset → Testing Dataset) in terms of mean IoU. Bold highlight the best performance while the 2nd best performance is underlined

| Configuration | RAG-FW | PSPNet | SegNet | UNet | FCN-8 | FCN-32 |
|-----------------|---------------|---------------|--------|--------|--------|--------|
| Duke → Zhang | 0.7936 | <u>0.7254</u> | 0.6997 | 0.7098 | 0.3827 | 0.2531 |
| Zhang → Duke | 0.8019 | <u>0.7369</u> | 0.7029 | 0.7152 | 0.3853 | 0.2594 |
| Duke → BIOMISA | 0.6871 | <u>0.5893</u> | 0.5827 | 0.5408 | 0.2284 | 0.1432 |
| BIOMISA → Duke | 0.6835 | <u>0.5836</u> | 0.5763 | 0.5349 | 0.2231 | 0.1387 |
| BIOMISA → Zhang | 0.6687 | <u>0.5701</u> | 0.5576 | 0.5103 | 0.2098 | 0.1092 |

Zhang → BIOMISA **0.6703** 0.5715 0.5618 0.5138 0.2142 0.1146

Table X: Classification performance when trained on one dataset and tested on another (Training Dataset → Testing Dataset). Bold indicates the low scores due to class imbalance situation and red indicates the best classification scores across all pathologies

| Configuration | Metric | AMD | ME | Healthy |
|-----------------|-----------------|---------------|---------------|---------------|
| Duke → Zhang | T _{PR} | 0.9380 | 0.9360 | 0.9760 |
| | T _{NR} | 0.9560 | 0.9507 | 0.9373 |
| | P _{PV} | 0.9552 | 0.8635 | 0.8385 |
| | F ₁ | 0.9465 | 0.8983 | 0.9020 |
| Zhang → Duke | T _{PR} | 0.9960 | 0.9444 | 0.9891 |
| | T _{NR} | 0.9877 | 0.9939 | 0.9953 |
| | P _{PV} | 0.9946 | 0.6061 | 0.9887 |
| | F ₁ | 0.9953 | 0.7384 | 0.9889 |
| Duke → BIOMISA | T _{PR} | 0.7182 | 0.7915 | 0.6929 |
| | T _{NR} | 0.7612 | 0.7066 | 0.7664 |
| | P _{PV} | 0.5206 | 0.7363 | 0.4643 |
| | F ₁ | 0.6036 | 0.7629 | 0.5560 |
| BIOMISA → Duke | T _{PR} | 0.8768 | 0.9286 | 0.7986 |
| | T _{NR} | 0.8027 | 0.8535 | 0.8776 |
| | P _{PV} | 0.9101 | 0.0590 | 0.7322 |
| | F ₁ | 0.8932 | 0.1110 | 0.7639 |
| BIOMISA → Zhang | T _{PR} | 0.7240 | 0.7440 | 0.7800 |
| | T _{NR} | 0.7620 | 0.7427 | 0.7307 |
| | P _{PV} | 0.7526 | 0.4908 | 0.4912 |
| | F ₁ | 0.7380 | 0.5914 | 0.6028 |
| Zhang → BIOMISA | T _{PR} | 0.7134 | 0.7749 | 0.6902 |
| | T _{NR} | 0.7488 | 0.7027 | 0.7538 |
| | P _{PV} | 0.5063 | 0.7296 | 0.4504 |
| | F ₁ | 0.5923 | 0.7516 | 0.5451 |

We also conducted a series of experimentation to compare our contour-based scheme to edge-based methods. The results reported in Table XI confirms the superiority of the contour-based approach in preserving the geometrical properties of the lesion.

VI. DISCUSSION

This paper presents a hybrid convolutional framework named as RAG-FW. RAG-FW employs RAG-Net that contains a segmentation and a classification unit for the extraction of retinal lesions and lesion-influenced grading of retinal diseases. Retinal lesions play a vital role in analyzing and measuring the severity of retinopathy. To the best of our knowledge, this paper presents the first framework that not only recognizes the retinal lesions but also use them for the severity grading of human retina from multi-vendor OCT scans.

Table XI: Dice for contour and edge-based lesion extraction. Results are computed on the combination of all the datasets. Bold indicates the better performance

| Retinal Lesions | Edge-based Extraction | Contour-based Extraction |
|-----------------|-----------------------|--------------------------|
| SRF | 0.6764 | 0.8689 |
| IRF | 0.7951 | 0.9114 |
| CA | 0.8246 | 0.9748 |
| HE | 0.6179 | 0.7816 |
| DRUSEN | 0.7713 | 0.9070 |

Apart from this, RAG-FW has been extensively tested on 43,613 retinal OCT scans from different publicly available datasets and it is thoroughly compared with existing state-of-the-art solutions against different metrics. The training time

for the RAG-FW is around 5 hours and 10 minutes whereas it takes around 21 seconds on average for the lesions extraction and grading a complete OCT volume against retinopathy.

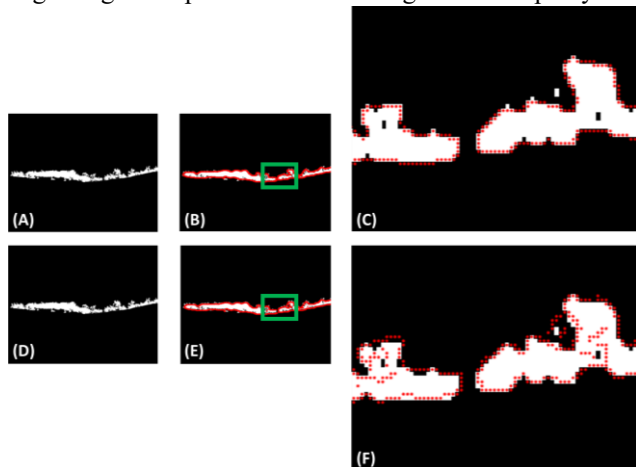


Figure 13: Qualitative comparison between the contour-based method (top) and edge-based method (bottom) for the retinal fluid region extraction. (A) original extracted fluid, (B) fluid geometry preserved through contouring, (C) zoomed version showing contours, (F) original extracted fluid, (E) fluid geometry preserved through edges, (F) zoomed version showing edges

Lesions such as IRF and SRF have also been recognized by [35], [37] and [38] but they haven't utilized these biomarkers for grading the candidate retina (also [37] and [38] identified IRF only). Authors in [39] designed lesion-aware CNN model for the accurate classification of retinopathy but they haven't tested it on multi-vendor OCT scans nor they have performed lesion-influenced grading. We have extensively tested the proposed framework on each dataset separately as well as on their combination. Furthermore, we have compared it with existing state-of-the-art solutions for retinal lesions extraction and retinopathy classification where the proposed framework significantly outperforms them in various metrics as discussed in the results section. Moreover, the preprocessing stage significantly improves the lesion extraction results for both proposed framework as well as for the pre-trained segmentation models as evident from Table XII. Since the B-scans are grayscale in nature (except for BIOMISA scans, which were intentionally processed as grayscale to make them compatible with other datasets), some of the background regions especially near vendor annotations looks quite similar to lesions areas such as fluid (especially in BIOMISA scans). Therefore, without preprocessing stage, all the segmentation models can easily misclassify those regions. Apart from this, after grading all the B-scans, the proposed framework uses a simple yet effective mechanism for generating the eye or volume level retinal classification. Since any disease specific retinal abnormalities are presented in a consecutive manner within the OCT volume. So, the eye-level diagnosis is generated by checking the longest grading connectivity across all the B-scans of the OCT volume as shown in Figure 14. Since the longest connectivity in Figure 14 is of CSR, so the complete volume (or an eye) will be graded as CSR affected.

Table XII: Mean IoU for lesion extraction with and without preprocessing stage for all segmentation frameworks

| Frameworks | With Preprocessing | Without Preprocessing |
|------------|--------------------|-----------------------|
| RAG-FW | 0.8055 | 0.7694 |
| PSPNet | 0.7763 | 0.7474 |
| SegNet | 0.7564 | 0.7357 |
| UNet | 0.7555 | 0.7342 |
| FCN-8 | 0.4928 | 0.4671 |
| FCN-32 | 0.3943 | 0.3521 |

We also molded RAG-FW to grade retinopathy directly instead of classifying it first and then grading it as shown in Figure 12. However, there have been several misclassifications between P-CME and DME, Dry AMD and Wet AMD, Acute CSR and Chronic CSR as it can be observed in the figure. This suggests that retinal lesions play a significant role in distinguishing between different disease stages. Although direct retinopathy grading results in the removal of the segmentation unit but its grading performance is far less as compared to the current approach. Also, the grading based on the segmentation unit only is also not feasible because the presence of retinal lesions in any pathology is non-mutually exclusive e.g. the presence of retinal fluid and hard exudates can be observed for DME pathologies as well as for the exudative AMD pathologies. So, if the candidate scan is not diagnosed (classified) before, it cannot be properly graded.

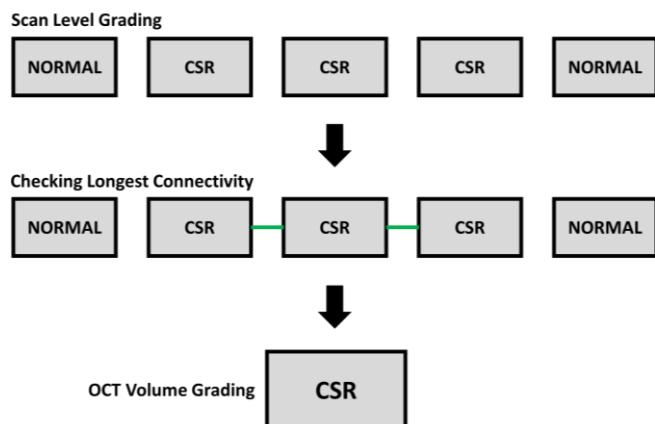


Figure 14: OCT volume grading by checking longest connectivity between consecutive B-scans

VII. CONCLUSION

This paper presents RAG-FW, which is a multi-vendor framework for the extraction and recognition of abnormal retinal lesions. Furthermore, the extracted lesions are utilized for the in-depth and intuitive grading of retinopathy. RAG-FW is a generic framework that can work irrespective of the acquisition machinery or the retinal pathology. It is also invariant to the scan noisy artifacts. RAG-FW has been tested on 43,613 retinal OCT scans from different publicly available datasets acquired using Spectralis, Heidelberg Inc. and Topcon 3D OCT 2000 machines, and it has been compared with many existing state-of-the-art solutions against various metrics. RAG-FW achieved the mean IoU score of 0.8055 for extracting retinal lesions and achieved the F_1 score of 99.52% for screening retinopathy. The proposed framework is currently limited in grading ME, CSR and AMD pathologies. However, it can be extended in future to cater more pathologies, such as glaucoma, and their severities as well.

Furthermore, incorporating findings from other retinal modalities such as FP can further strengthen the intuitive grading of retinopathy.

ACKNOWLEDGEMENTS

We would like to thank the associate editor and three anonymous reviewers for their valuable feedback and suggestions which has significantly helped us in improving quality of the paper. We would also like to thank Divam Gupta and Rounaq Jhunjhunu Wala for providing a single repository containing implementation of different deep segmentation models on GitHub.

REFERENCES

- [1] "Implant gives rats sixth sense for infrared light". Wired UK. 14 February 2013. Accessed 14 February 2013.
- [2] Richard S. Snell, Michael A. Lemp, "Clinical Anatomy of the Eye", 2nd edition, May 31st, 2013.
- [3] Ning Cheung, Paul Mitchell, Tien Ying Wong, "Diabetic retinopathy", *Lancet*, 376(9735):124–36. doi: 10.1016/S0140-6736(09)62124-3, 2010.
- [4] Grant M. Comers, "Cystoid Macular Edema", Kellog Eye Center, Accessed June 2016.
- [5] Saine, PJ. "Fundus Photography: What is a Fundus Camera?" Ophthalmic Photographers Society Accessed May 30, 2016.
- [6] Joel S. Schuman, "Introduction to Optical Coherence Tomography Technology".
- [7] Maria Wang, Inger Christine Munch, Pascal W. Hasler, Christian Prunte, Michael Larsen, "Central serous chorioretinopathy", *Acta Ophthalmologica*, 86 (2): 126–45, 2008.
- [8] Wei Zhang, Kaori Yamamoto and Sadao Hori, "Optical Coherence Tomography for assessment of diabetic macular edema", *Int J Ophthalmol*, Volume 1, December 18, 2008.
- [9] Shrestha A, Maharjan N, Shrestha A, Thapa R, Poudyal G, "Optical Coherence Tomographic assessment of macular thickness and morphological patterns in diabetic macular edema: Prognosis after modified grid photocoagulation", 4 (7): 128-133, *Nepal J Ophthalmol* 2012.
- [10] Rosanna Zacharias Hannouche, Marcos Pereira Ávila, "Detection of diabetic foveal edema with bio microscopy, fluorescein angiography and optical coherence tomography", *Arq Bras Oftalmol*, 71(5):759-63, 2008.
- [11] Desislava Koleva-Georgieva, "Optical Coherence Tomography Findings in Diabetic Macular Edema", February 24, 2012.
- [12] Bartosz L. Sikorski, Grazyna Malukiewicz, Joanna Stafiej, Hanna Lesiewska-Junk, and Dorota Raczynska, "The Diagnostic Function of OCT in Diabetic Retinopathy", *Mediators of Inflammation*, Volume 2013 (2013), Article ID 434560, 12 pages.
- [13] Virgili G, Menchini F, Murro V, Peluso E, Rosa F, Casazza G., "Optical coherence tomography (OCT) for detection of macular oedema in patients with diabetic retinopathy", *Cochrane Database Syst Rev*. 2011 Jul 6;(7): CD008081. doi: 10.1002/14651858.
- [14] George Trichonas, Peter K Kaiser, "Optical coherence tomography imaging of macular oedema", *Br J Ophthalmol* 2014, 98: ii24-ii29 doi:10.1136/bjophthalmol-2014-305305.
- [15] Adam Martidis, Jay S Duker, Paul B Greenberg, Adam H Rogers, Carmen A Puliafito, Elias Reichel, Caroline Bauman, "Intravitreal triamcinolone for refractory diabetic macular edema", *Elsevier journal of ophthalmology*, doi:10.1016/S0161-6420(02)00975-2, 24th April 2002.
- [16] Yasser M Helmy, Heba R Atta Allah, "Optical Coherence Tomography classification of diabetic cystoid macular edema", *Clinical Ophthalmology - Dove press*, August 27, 2013.
- [17] Marion R. Munk, Lee M. Jampol, Christian Simader, Wolfgang Huf, Tamara J. Mittermüller, Glenn J. Jaffe, Ursula Schmidt-Erfurth, "Differentiation of Diabetic Macular Edema From Pseudophakic Cystoid Macular Edema by Spectral-Domain Optical Coherence Tomography", *Investigative Ophthalmology and Visual Science*, October 2015.
- [18] Michael R. Hee, Carmen A. Puliafito, Jay S. Duker, Elias Reichel, BS Jeffrey G. Coker, BS Jason R. Wilkins, Joel S. Schuman, MS Eric A.

- Swanson, James G. Fujimoto, "Topography of diabetic macular edema with optical coherence tomography", Elsevier journal of ophthalmology, doi:10.1016/S0161-6420(98)93601-6, 14 March 2005.
- [19] Nils F. Mokwa, Tina Ristau, Pearse A. Keane, Bernd Kirchhoff, Srinivas R. Sadda and Sandra Liakopoulos, "Grading of Age-Related Macular Degeneration: Comparison between Color Fundus Photography, Fluorescein Angiography, and Spectral Domain Optical Coherence Tomography", *Journal of Ophthalmology*, Volume 2013 (2013), Article ID 85915, 6 pages.
- [20] Fernández, D.C., Salinas, H.M. and Puliafito, C.A., "Automated detection of retinal layer structures on optical coherence tomography images", *Optics Express*, 13(25), pp.10200-10216, 2005.
- [21] G. R. Wilkins, O. M. Houghton, A. L. Oldenburg, "Automated Segmentation of Intraretinal Cystoid Fluid in Optical Coherence Tomography", *IEEE Transactions on Biomedical Engineering*, 59(4), 1109-1114, (2012).
- [22] Q. Yang, C. A. Reisman, Z. Wang, Y. Fukuma, M. Hangai, N. Yoshimura, A. Tomidokoro, M. Araie, A.S. Raza, D. C. Hood and K. Chan, "Automated layer segmentation of macular OCT images using dual-scale gradient information". *Optics Express*, 18(20), pp.21293-21307, 2010.
- [23] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in SD-OCT images congruent with expert manual segmentation", *Optics Express*, 18(18), pp.19413-19428, 2010.
- [24] Y. Huang, R. P. Danis, J. W. Pak, S. Luo, J. White, X. Zhang, A. Narkar, and A. Domalpally, "Development of a semi-automatic segmentation method for retinal OCT images tested in patients with diabetic macular edema", *PLOS one*, 8(12), p.e82922, 2013.
- [25] Stephanie J. Chiu, Michael J. Allingham, Priyatham S. Mettu, Scott W. Cousins, Joseph A. Izatt, and Sina Farsiu, "Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema", *Biomedical Optics Express*, Vol. 6, No. 4, April 2015.
- [26] Kevin J. McHugh, Dian Li, Jay C. Wang, Leon Kwark, Jessica Loo, Venkata Macha, Sina Farsiu, Leo A. Kim, Magali Saint-Geniez, "Computational modeling of retinal hypoxia and photoreceptor degeneration in patients with age-related macular degeneration", *PLoS One*, June 2019.
- [27] Dehui Xiang, Geng Chen, Fei Shi, Weifang Zhu, Qinghui Liu, Songtao Yuan and Xinjian Chen, "Automatic Retinal Layer Segmentation of OCT Images With Central Serous Retinopathy", *IEEE Journal of Biomedical and Health Informatics*, Vol 23, No. 1, January 2019.
- [28] Roberta Farci, Alexandre Sellam, Florence Coscas, Gabriel J. Coscas, Giacomo Diaz, Pietro Emanuele Napoli, Eric Souied, Maria Silvana Galantuomo and Maurizio Fossarello, "Multimodal OCT Reflectivity Analysis of the Cystoid Spaces in Cystoid Macular Edema", *BioMed Research International*, March 2019.
- [29] Cynthia A. Toth, Vincent Tai, Maxwell Pistilli, Stephanie J. Chiu, Katrina P. Winter, Ebenezer Daniel, Juan E. Grunwald, Glenn J. Jaffe, Daniel F. Martin, Gui-shuang Ying, Sina Farsiu, Maureen G. Maguire, "Distribution of OCT Features within Areas of Macular Atrophy or Scar after 2 Years of Anti-VEGF Treatment for Neovascular AMD in CATT", *American Academy of Ophthalmology*, 2018.
- [30] D. Kaba, Y. Wang, C. Wang, X. Liu, H. Zhu, A. G. Salazar-Gonzalez, and Y. Li, "Retina layer segmentation using kernel graph cuts and continuous max-flow", *Optics Express*, Vol. 23, Issue 6, pp. 7366-7384, 2015.
- [31] Abdolreza Rashno, Dara D. Koozekanani, Paul M. Drayna, Behzad Nazari, Saeed Sadri, Hossein Rabbani, Keshab K. Parhi, "Fully-Automated Segmentation of Fluid/Cyst Regions in Optical Coherence Tomography Images with Diabetic Macular Edema using Neutrosophic Sets and Graph Algorithms," in *IEEE Transactions on Biomedical Engineering*, vol. PP, no. 99, pp. 1-1, 2017.
- [32] Leyuan Fang, David Cunefare, Chong Wang, Robyn H. Guymer, Shutao Li, and Sina Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search", *Biomedical Optics Express*, Vol. 8, No. 5, May 2017.
- [33] Cecilia S. Lee, Ariel J. Tying, Nicolaas P. Deruyter, Yue Wu, Ariel Rokem, and Aaron Y. Lee, "Deep-learning based, automated segmentation of macular edema in optical coherence tomography", *Biomedical Optics Express*, Vol. 8, No. 7, July 2017.
- [34] Abhijit Guha Roy, Sailesh Conjeti, Sri Phani Krishna Karri, Debdoot Sheet, Amin Katouzian, Christian Wachinger and Nassir Navab, "RelayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks", *Biomedical Optics Express*, Vol. 8, No. 8, 1 August 2017.
- [35] Thomas Schlegl, Sebastian M. Waldstein, Hrvoje Bogunovic, Franz Endstraßer, Amir Sadeghipour, Ana-Maria Philip, Dominika Podkowinski, Bianca S. Gerendas, Georg Langs, Ursula Schmidt-Erfurth, "Fully Automated Detection and Quantification of Macular Fluid in OCT Using Deep Learning", *Elsevier Ophthalmology Journal*, Vol. 125, No. 4, April 2018.
- [36] Philipp Seebock, Jose Ignacio Orlando, Thomas Schlegl, Sebastian M. Waldstein, Hrvoje Bogunovic, Sophie Klimscha, Georg Langs and Ursula Schmidt-Erfurth, "Exploiting Epistemic Uncertainty of Anatomy Segmentation for Anomaly Detection in Retinal OCT", *IEEE Transactions on Medical Imaging*, May 2019.
- [37] Plácido L. Vidal, Joaquim De Moura, Jorge Novo, Manuel G. Penedo And Marcos Ortega, "Intraretinal fluid identification via enhanced maps using optical coherence tomography images", *Biomedical Optics Express*, October 2018.
- [38] G. N. Girish, Bibhash Thakur, Sohini Roy Chowdhury, Abhishek R. Kothari, and Jeny Rajan, "Segmentation of Intra-Retinal Cysts From Optical Coherence Tomography Images Using a Fully Convolutional Neural Network Model", *IEEE Journal of Biomedical and Health Informatics*, Vol 23, No. 1, January 2019.
- [39] Leyuan Fang, Chong Wang, Shutao Li, Hossein Rabbani, Xiangdong Chen, Zhimin Liu, "Attention to Lesion: Lesion-Aware Convolutional Neural Network for Retinal Optical Coherence Tomography Image Classification", *IEEE Transactions on Medical Imaging*, August 2019.
- [40] Bilal Hassan, Gulistan Raja, "Fully automated assessment of Macular Edema using Optical Coherence Tomography (OCT) images", 2016 International Conference on Intelligent Systems Engineering (ICISE), 15th - 17th January 2016.
- [41] Ravi M. Kamble, Genevieve C. Y. Chan, Oscar Perdomo, Manesh Kokare, Fabio A. Gonzalez, Henning Muller, Fabrice Meriaudeau, "Automated Diabetic Macular Edema (DME) Analysis using Fine Tuning with Inception-Resnet-v2 on OCT Images", *IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, 2018.
- [42] Abhishek, A.M., Berendschot, T.T., Rao, S.V. and Dabir, S., "Segmentation and analysis of retinal layers (ILM & RPE) in optical coherence tomography images with edema", *IEEE Conference on Biomedical Engineering and Sciences (IECBES)*, pp. 204-209, December 2014.
- [43] Soichiro Kuwayama, Yuji Ayatsuka, Daisuke Yanagisano, Takaki Uta, Hideaki Usui, Aki Kato, Noriaki Takase, Yuichiro Ogura and Tsutomu Yasukawa, "Automated Detection of Macular Diseases by Optical Coherence Tomography and Artificial Intelligence Machine Learning of Optical Coherence Tomography Images", *Journal of Ophthalmology*, April 2019.
- [44] Filippo Arcadu, Fethallah Benmansour, Andreas Maunz, John Michon, Zdenka Haskova, Dana McClintock, Anthony P. Adamis, Jeffrey R. Willis and Marco Prunotto, "Deep Learning Predicts OCT Measures of Diabetic Macular Thickening From Color Fundus Photographs", *Retina*, January 2019.
- [45] L. Zhang, W. Zhu, F. Shi, H. Chen, X. Chen, "Automated Segmentation of Intra-retinal Cystoid Macular Edema For Retinal 3D OCT Images With Macular Hole" *International Symposium on Biomedical Imaging*, 12, 1494 - 1497, (2015).
- [46] J. Sugruk, S. Kiattisin, A. L.Lasantitham, "Automated classification between age-related macular degeneration and diabetic macular edema in OCT image using image segmentation", *IEEE Biomedical Engineering International Conference*, (2014).
- [47] Joaquim de Moura, Plácido L. Vidal, Jorge Novo and Marcos Ortega, "Automatic Identification of Diabetic Macular Edema Using a Transfer Learning-Based Approach", *2nd XoveTIC Conference*, A Coruña, Spain, 5-6 September 2019.
- [48] Feng Li, Hua Chen, Zheng Liu, Xuedian Zhang, Zhizheng Wu, "Fully automated detection of retinal disorders by image-based deep learning", *Graefe's Archive for Clinical and Experimental Ophthalmology*, January 2019.
- [49] Jaakko sahlsten, Joel Jaskari, Jyri Kivinen, Lauri turunen, esa Jaanio, Kustaa Hietala, Kimmo Kaski, "Deep Learning Fundus Image Analysis for Diabetic Retinopathy and Macular edema Grading", *Nature Scientific Reports*, July 2019.
- [50] Pratul P. Srinivasan, Leo A. Kim, Priyatham S. Mettu, Scott W. Cousins, Grant M. Comer, Joseph A. Izatt, and Sina Farsiu, "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images",

- Biomedical Optics Express, Vol. 5, No. 10 | DOI:10.1364/BOE.5.003568, 12 Sep 2014
- [51] Sertan Kaymak, Ali Serener, "Automated Age-Related Macular Degeneration and Diabetic Macular Edema Detection on OCT Images using Deep Learning", IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), 2018.
- [52] David Cunefare, Alison L. Huckenpahler, Emily J. Patterson, Alfredo Dubra, Joseph Carroll and Sina Farsiu, "RAC-CNN: multimodal deep learning based automatic detection and classification of rod and cone photoreceptors in adaptive optics scanning light ophthalmoscope images", Biomedical Optics Express, August 2019.
- [53] Kermany D, Goldbaum M, Cai W et al., "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning", Cell, 172(5):1122-1131. doi:10.1016/j.cell.2018.02.010, 2018.
- [54] Yibiao Rong, Dehui Xiang, Weifang Zhu, Kai Yu, Fei Shi, Zhun Fan and Xinjian Chen, "Surrogate-Assisted Retinal OCT Image Classification Based on Convolutional Neural Networks", IEEE Journal of Biomedical and Health Informatics, Vol 23, No. 1, January 2019.
- [55] Taimur Hassan, Muhammad Usman Akram, Muhammad Furqan Masood, Ubaidullah Yasin, "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, February 2019
- [56] Taimur Hassan, M. Usman Akram, Mahmood Akhtar, Shoab Ahmad Khan, Ubaidullah Yasin, "Multilayered Deep Structure Tensor Delaunay Triangulation and Morphing Based Automated Diagnosis and 3D Presentation of Human Macula", Journal of Medical Systems, September 2018.
- [57] Taimur Hassan, Muhammad Usman Akram, Arslan Shaukat, Sajid Gul Khawaja and Bilal Hassan, "Structure Tensor Graph Searches Based Fully Automated Grading and 3D Profiling of Maculopathy From Retinal OCT Images", IEEE Access, September 2018.
- [58] Bilal Hassan, Taimur Hassan, Bo Li, Ramsha Ahmed and Omar Hassan, "Deep Ensemble Learning Based Objective Grading of Macular Edema by Extracting Clinically Significant Findings from Fused Retinal Imaging Modalities", MDPI Sensors, July 2019.
- [59] Samina Khalid, M. Usman Akram, Taimur Hassan, Amina Jameel, Tehmina Khalil, "Automated Segmentation and Quantification of Drusen in Fundus and Optical Coherence Tomography Images for Detection of ARMD", Journal of Digital Imaging, December 2017.
- [60] Samina Khalid, M. Usman Akram, Taimur Hassan, Ammara Nasim and Amina Jameel, "Fully Automated Robust System to Detect Retinal Edema, Central Serous Chorioretinopathy, and Age Related Macular Degeneration from Optical Coherence Tomography Images", BioMed Research International, March 2017.
- [61] Taimur Hassan, M. Usman Akram, Bilal Hassan, Adeel M. Syed, and Shafaat Ahmed Bazaz, "Automated segmentation of subretinal layers for the detection of macular edema," Appl. Opt. 55, 454-461 (2016).
- [62] Bilal Hassan, Gulistan Raja, Taimur Hassan, and M. Usman Akram, "Structure tensor based automated detection of macular edema and central serous retinopathy using optical coherence tomography images," J. Opt. Soc. Am. A 33, 455-463, 2016.
- [63] Adeel M. Syed, Taimur Hassan, M. Usman Akram, Samra Naz, Shehzad Khalid, "Automated diagnosis of macular edema and central serous retinopathy through robust reconstruction of 3D retinal surfaces" Computer Methods and Programs in Biomedicine, 137: 1-10 (2016).
- [64] Taimur Hassan, Anam Usman, M Usman Akram, M Furqan Masood, Ubaidullah Yasin, "Deep Learning Based Automated Extraction of Intra-Retinal Layers for Analyzing Retinal Abnormalities", IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom), September 2018.
- [65] Peter Bakker, "Image structure analysis for seismic interpretation", Ph.D. Thesis, Delft University of Technology, 2002.
- [66] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, Jiaya Jia, "Pyramid Scene Parsing Network", IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [67] "Diabetic Macular Edema", EyeWiki, Accessed: November 4th, 2019.
- [68] Nidhi Relhan and Harry W. Flynn J, "The Early Treatment Diabetic Retinopathy Study historical review and relevance to today's management of diabetic macular edema", Current Opinion in Ophthalmology, Wolters Kluwer, May 2017.
- [69] Charlotte Strøm, Birgit Sander, Nicolai Larsen, Michael Larsen, Henrik Lund-Andersen, "Diabetic Macular Edema Assessed with Optical Coherence Tomography and Stereo Fundus Photography", Retina, Investigative Ophthalmology and Visual Science, January, 2012.
- [70] "Age-Related Macular Degeneration PPP - Updated 2015", American Academy of Ophthalmology, Accessed: November 4th, 2019.
- [71] Retina Specialist, "The Good and Bad of Retinal Fluid", Accessed October 27th, 2019.
- [72] Faruque Ghanchi, "Focus on CSR", The Royal College of Ophthalmologists, 2013.
- [73] Matthew D. Zeiler, "ADDELTA: An Adaptive Learning Rate Method", arXiv:1212.5701, December 2012.
- [74] Murguia, M., and Villasenor, J. L., Estimating the effect of the similarity coefficient and the cluster algorithm on biogeographic classifications. Ann. Bot. Fennici. 40:415-421, 2003.
- [75] S. Farsiu, S. J. Chiu, R.V. O'Connell, F.A. Folgar, E. Yuan, J.A. Izatt, and C.A. Toth, "Quantitative Classification of Eyes with and without Intermediate Age-related Macular Degeneration Using Optical Coherence Tomography", Ophthalmology, 121(1), 162-172 Jan. (2014).
- [76] Taimur Hassan, M. Usman Akram, M. Furqan Masood and Ubaidullah Yasin, "BIOMISA Retinal Image Database for Macular and Ocular Syndromes", 15th International Conference on Image Analysis and Recognition (ICIAR-2018), Portugal, June 2018.
- [77] Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, December 2017.
- [78] Olaf Ronneberger, Philipp Fischer, Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", Medical Image Computing and Computer Assisted Intervention (MICCAI), 2015.
- [79] Jonathan Long, Evan Shelhamer, Trevor Darrell, "Fully Convolutional Networks for Semantic Segmentation", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
- [80] Evan Shelhamer, Jonathan Long, Trevor Darrell, "Fully Convolutional Networks for Semantic Segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, April 2017.