Towards developing a multiple modality fusion technique for automatic detection of Glaucoma

Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science
in
Electronics and Communication Engineering
by Research

by

Divya Jyothi Gaddipati
201331069
divyajyothi.gaddipati@research.iiit.ac.in

International Institute of Information Technology
Hyderabad - 500 032, INDIA
February 2020

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled "**Towards developing a multiple modality fusion technique for automatic detection of Glaucoma**" by **Divya Jyothi Gaddipati**, has been carried out under my supervision and is not submitted elsewhere for a degree.

Date	Adviser: Dr. Jayanthi Sivaswamy

Dedicated to
my brother, Jagadeesh
and my guide, Dr. Jayanthi Sivaswamy

Acknowledgments

Firstly, I would like to take the privilege to convey my gratitude to my advisor Dr. Jayanthi Sivaswamy for guiding me through every step of the way for the completion of the thesis. I am eternally grateful to her for giving me more than a couple of second-chances and believing in me. Her constant guidance, enthusiasm and support have inspired me to keep going. I would also like to thank Dr Krishnadas from Aravind Eye Hospital, Dr. T. Mansoori from Anand Eye Hospital, Dr. K.A. Vermeer, Dr. H.G. Lemij, and J.E.A. Majoor from the Rotterdam Eye Hospital, Netherlands for providing relevant clinical knowledge and access to their study data which have been of immense help for the completion of the thesis.

I owe special gratitude to a few people, whom I wish to take this opportunity to thank. First of all, I would like to thank the entire fraternity of the MIP research group at CVIT, especially Arunava Chakravarty, Sukesh Adiga, Jahnavi Gamalapati, A K Pujitha, Chetan Srinidhi who have helped me in the initial days to get a head start and continued throughout providing their invaluable advice and guidance whenever needed. Along with them and my fellow labmates - Anurag Deshmukh, Abhinav Dhere, Karthik Gopinath, Samrudhdhi Rangrej, Aabhas Majumdar, Alphin J, I enjoyed our regular group meetings with ma'am, which have been in a league of their own and a major learning experience. I would like to especially thank my friend and lab partner for the first two years of my Honors project, Jahnavi Chowdary with whom I started my journey in CVIT and without whom it wouldn't have been the same. I am extremely grateful for being given an opportunity to work with this delightful lot of highly talented people, it has really been an absolute pleasure and unforgettable experience.

As said by John Lennon, "I get by with a little help from my friends", this acknowledgement wouldn't feel complete without the special mention of the many friends I made during this journey towards my degree. Shruti Gullapuram, Moneish Kumar, Pravallika Etoori, Sangeetha Iyer, Aditya Srinivas, Sireesha Galla, Lasya Venneti, Nikitha Vallurupalli, Soujanya Ponnapalli, SVK Rohit and last but not the least, the Wing Extension, and many others whose names might not be here, have been incredible friends over the years regardless of my eccentricities and made IIIT feel like home to me. I would like to single out Sowmith Manepalli, Vamsee Chamakura and Manasa Lingamallu who have been very generous, in one way or another, with their help, inspiration and encouragement.

Finally, I would like to express my gratitude to my parents and brother for their unfailing support and understanding over the years which made the road to completing my research less bumpy. I am forever indebted to them for being so wonderful and never stopped believing in me even when I claimed my research would be completed "soon" for almost a year.

Abstract

Glaucoma is a major eye disease which when untreated, can gradually lead to irreversible loss in vision. The underlying causes are a loss of retinal nerve fibres (resulting in a thinning of the layer and enlargement of the optic cup) and peripapillary atrophy. Since these occur without any sign of symptoms in the initial stages, it is difficult to diagnose the disease in the early stages. Hence, the development of Computer-aided diagnostics (CAD) systems for early detection and treatment of the disease has attracted the attention of many medical experts and researchers alike.

Optical Coherence Tomography (OCT) and Fundus photography are two widely used retinal imaging techniques for obtaining the structural information of the eye which helps to analyze and detect the diseases. Existing automated systems rely largely on fundus images for assessment of glaucoma due to their fast acquisition and cost. OCT images provide vital and unambiguous information for understanding the changes occurring in the retina, specifically related to the retinal nerve fiber layer and the optic nerve head which are essential for disease assessment. However, the high cost of OCT is a deterrent for deployment in screening at large scale. Hence, the focus of this thesis is to investigate the potential of integrating the two retinal imaging techniques which provide complementary information of the eye for developing automatic glaucoma screening system.

Firstly, we propose a deep learning approach directly operating on 3D OCT volumes for glaucoma assessment which showed promising results, thus demonstrating the prominence of the highly discriminative features learnt from OCT for automated glaucoma detection. Next, we present a novel CAD solution wherein both OCT and fundus modality images are leveraged to learn a model that can perform a mapping of fundus to the OCT feature space. We show how this model can be subsequently used to detect glaucoma given an image from only one modality (fundus), thus enabling the automated screening operation to be executed on a large scale. The results show that fundus to OCT feature space mapping is an attractive option for glaucoma detection.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Glaucoma

Glaucoma is a major optic neuropathy, which is the second most common cause of blindness in the world. It is called as the ”silent thief of sight” as it gradually causes damage to the optic nerve without any sign of symptoms in its nascent stage and consequently leads to permanent loss in vision. This irreparable and asymptomatic nature of the disease necessitates early detection before significant loss has developed and thus curbing the risk of visual impairment and related morbidity.

Glaucoma originates from elevated intraocular pressure caused by a fluid produced by the eye. This increase in stress on the eye results in the damage of optic nerve, which is responsible for communication between brain and eye [1] as shown in Fig. 1.1. The diagnosis of glaucoma also depends on recognition of characteristic structural changes to the Optic Nerve Head (ONH) and the Peripapillary regions. As the disease doesn't show any symptoms, these structural changes of the retina are one of the vital sources for early detection. The ONH consists of two main structures: optic disc (OD) and optic cup (OC). Glaucoma leads to an enlargement of the optic cup with respect to the disc, often known as ”cupping”. Cupping is quantitatively assessed using the vertical cup-disc diameter ratio (CDR) and the thickness profile of the neuroretinal rim, the region between the OC and the OD. Another indicator is the narrowing of the rim. The effect of glaucoma in the peripapillary region include thinning of the retinal nerve fiber layer (RNFL) around the optic nerve often known as peripapillary atrophy (PPA).

Although, diagnosis of glaucoma by a medical expert remains as the gold standard, it is highly labor-intensive and subjected to interpretation errors. Recent advances in developing cost effective computer aided diagnostic (CAD) systems and a considerable increase in available data have enhanced automated detection of many ocular diseases, and also attracted attention of a multitude of researchers and medical experts alike. It not only assists the clinicians by providing objective opinion with relevant insights, but is also easily accessible by patients which in turn aids the early detection of the disease at a large scale. Thus, it increases efficiency and cost-effectiveness of the medical process. Although, the success of the automated screening systems is limited by their detection performance.

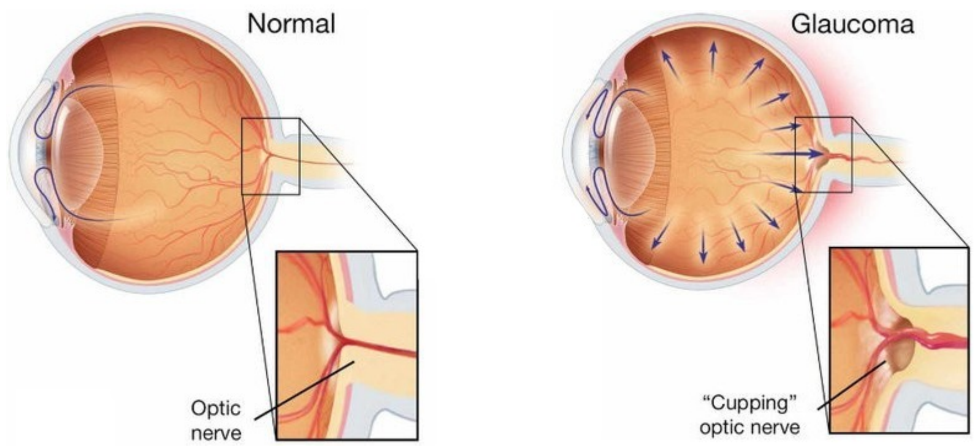


Figure 1.1: Normal and Glaucomatous eye

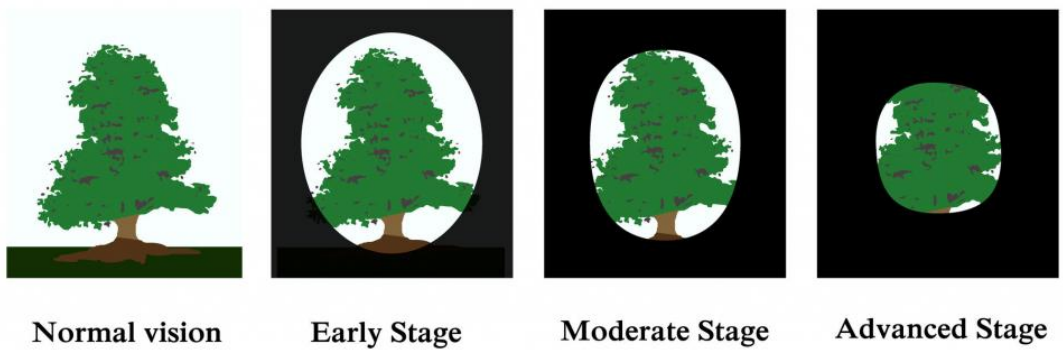


Figure 1.2: An example of different stages of vision loss from glaucoma

1.2 Retinal Imaging

Retinal imaging has evolved rapidly over past few years and has now become a mainstay in diagnosing many retinal diseases like glaucoma, diabetic retinopathy, age-related macular degeneration, etc. Two such popular retinal screening techniques are Fundus photography and Optical Coherence Tomography (OCT), that render support for the examination of inner details of the eye for detection and monitoring the progression of the disease. These two imaging modalities are known to provide complementary information.

1.2.1 Fundus photography

Fundus imaging is most commonly used technologies which yields a two-dimensional projection image of the interior surface of the eye, thus providing information of different retinal structures. Fundus is widely used by ophthalmologists to assist in the diagnosis and also for developing computer-aided diagnostic (CAD) based screening systems at a large scale due to its portability, cost-effectiveness and

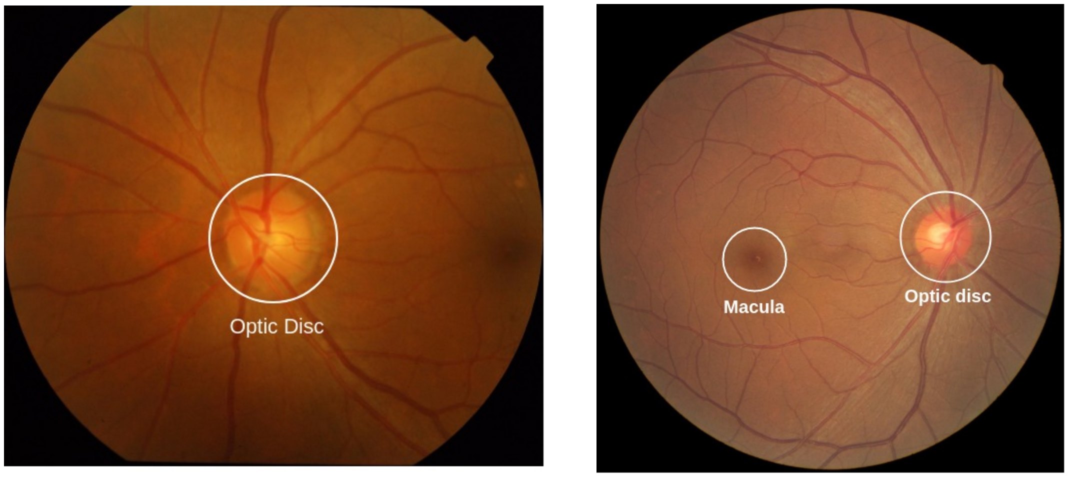
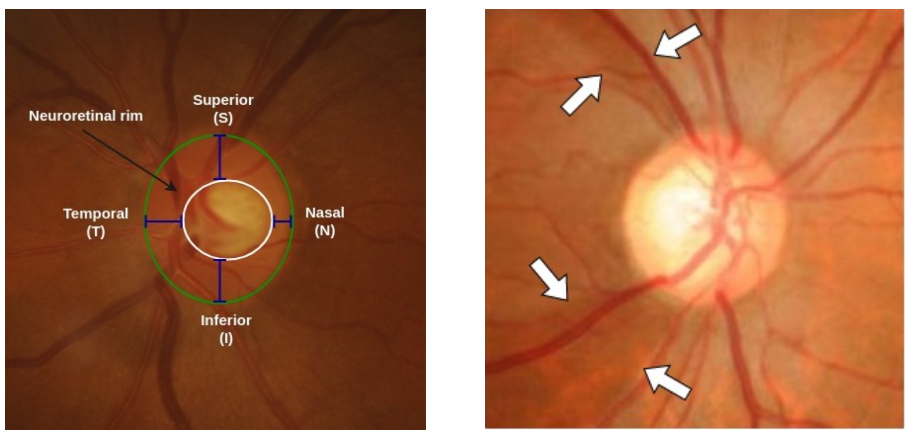


Figure 1.3: OD centric (left) and macula centric (right) fundus images of the retina



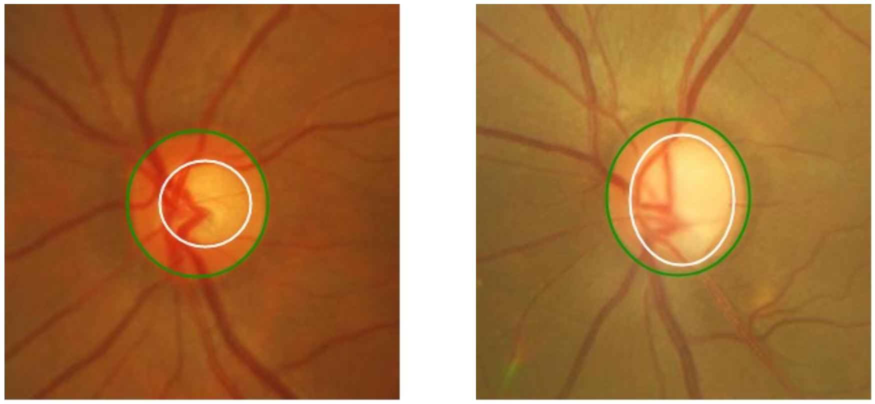
(a) Neuroretinal rim thickness

(b) Wedge shaped RNFL pattern

Figure 1.4: Indicators of glaucoma. (a) indicates the four sectors for rim thickness evaluation. RNFL pattern indicated by the arrows in (b).

easy acquisition. Moreover, the fundus images acquired from the fundus camera can facilitate the early detection of glaucoma in settings where more expensive equipment is not available.

The fundus image can be centered either at the bright ONH region or the dark circular region called macula as shown in Fig. 1.3. Therefore, in case of fundus image, the OD appears as a bright elliptical region, centered with a brighter optic cup. The tissue surrounding the cup separating it from the disc is called the neuroretinal rim. For determining the effect of neuroretinal rim thickness, the ONH is divided into four sectors: Inferior, Superior, Nasal and Temporal (ISNT) as shown in Fig. 1.4a. In a normal eye, the thickness is such that inferior region (I) $>$ superior region (S) $>$ nasal region (N) $>$ temporal region (T). The thinning of RNFL appears as dark wedge shaped bands in superior and inferior regions radiating out from the optic disc as shown in Fig. 1.4b. The enlargement of the cup in case of glaucoma compared to a normal subject is shown in Fig. 1.5.



(a) Normal (b) Glaucoma

Figure 1.5: Enlargement of the cup in glaucoma. Disc boundary marked in green and cup boundary in white.

1.2.2 Optical Coherence Tomography

OCT is a non-invasive imaging technique that renders high-resolution cross sectional view of structures in the retina in the form of optical slices. It works on a principle called low coherence interferometry. Similar to the way ultrasound measures the reflection of sound, OCT measures the reflection of laser light which produces detailed cross-sectional images of the retina. Thus, a three-dimensional reconstruction providing the depth information of the retina can be obtained, which in turn allows volumetric assessment of retinal structures, including ONH and macula.

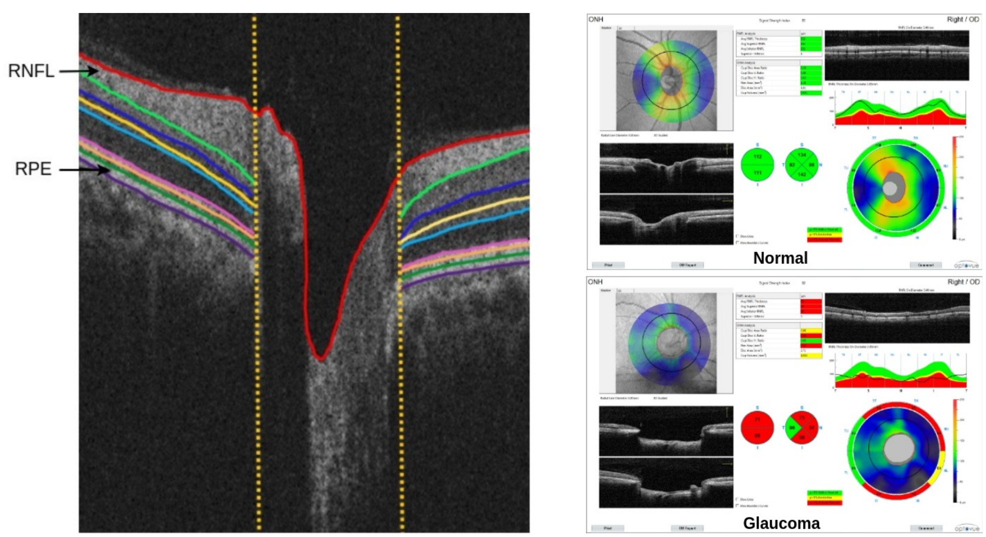
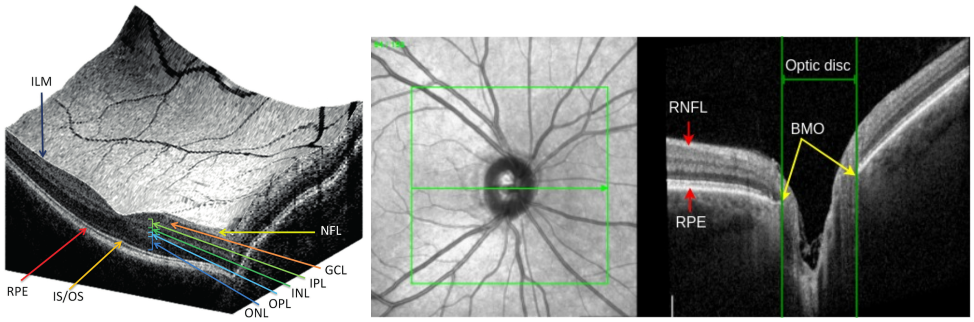


Figure 1.6: Different layers in ONH centered OCT scan (Left). Sample SD-OCT report of normal and glaucoma patient (Right).

Many technological developments have been made, ranging from time-domain (TD) to spectral-domain (SD). SD-OCT offers higher resolution and denser sampling, which increases its diagnostic



(a) OCT volume (b) OCT - Fundus correspondence

Figure 1.7: OCT scan on the right and the corresponding fundus image in the middle. The horizontal green line on the fundus image indicates the scan position.

utility, allowing detection of smaller, more subtle morphological changes and rapid collection of 3D datasets. OCT has been widely adopted by ophthalmologists and numerous studies demonstrated OCT to provide high diagnostic performance in discriminating between normal and glaucoma.

Since OCT provides cross sectional view, the different layers of the retina are visible as shown in Fig. 1.6. Different layers are affected in different retinal abnormalities. For example, in case of glaucoma the top most RNFL layer is affected, retinal pigment epithelium (RPE) and outer retinal layers are affected in case of Age-related macular degeneration (AMD), etc. It is also possible to identify the points of the Bruch's membrane opening (BMO), which is useful in determining other measurements relative to BMO. Thus acquisition of 3D images of the ONH region enables accurate and reproducible measurements of ONH parameters that include: disc and rim area measured using BMO, cup to disc ratio, cup depth and volume, thickness of different layers at every point etc as shown in Fig. 1.7. An example of the retinal report obtained from OCT for a healthy and glaucoma patient are displayed in Fig. 1.6.

1.3 Motivation of thesis

Glaucoma is caused by a number of highly complex factors, which are not feasible to capture by one modality alone. Nevertheless, most of the existing glaucoma screening systems use either one of fundus image or OCT volume, with the latter in a smaller proportion, and hence depending on the information provided by single modality. OCT allows quantitative assessment of the relevant layers, area and volume of cup using the depth information provided, whereas fundus image provides textural information in detail. Though both the modalities have their own advantages, they are still subjected to some limitations.

Fundus images are used in screening systems on a large scale due to its easy acquisition and cost effectiveness. Although OCT has been extensively used in ophthalmology, its efficacy in automated screening hasn't reached its full potential yet, because its acquisition is a highly expensive process.

Hence, the main objective of the thesis is to demonstrate the prominence of OCT in glaucoma diagnosis and consequently develop a strategy for a CAD model that unifies the strengths of both the modalities i.e., cost-effectiveness of fundus imaging and structural information from OCT.

1.4 Contributions

The major contributions of the thesis are associated with developing automated glaucoma assessment solutions, as mentioned below:

- Assessment of glaucoma directly from the OCT volumes using a deep neural network which successfully learns highly discriminative features.
- A mapping model which maps fundus image to the target OCT feature space that is obtained from Capsule network, hence developing a heterogeneous feature space.
- A novel CAD solution wherein information from both OCT and fundus modality are leveraged to learn a model, which is subsequently used to detect glaucoma given an image from only one modality (fundus).

1.5 Organization of thesis

The dissertation of the thesis is as follows: In Chapter 2, a deep learning approach for glaucoma classification is described, which directly operates on 3D OCT volumes. We show that highly discriminative features are learnt from OCT and thus establishing the prominence of this modality for automated glaucoma detection. Using the feature space obtained from the model developed in chapter 2, we present a feature fusion approach in Chapter 3 which utilises unpaired fundus and OCT data for mapping fundus images to the OCT feature space, hence bridging the gap between the two modalities. This is subsequently used for glaucoma detection provided input from fundus modality alone. In Chapter 4, conclusion and plausible ideas for further work are discussed.

Chapter 2

Glaucoma Assessment from OCT

2.1 Introduction

Glaucoma is a chronic optic neuropathy and one of the leading causes of blindness in the world. It results in progressive degeneration of the optic nerve fibers, leading to structural changes of the optic nerve head (ONH), which might result in permanent loss of vision. Since glaucoma is asymptomatic in the initial stages, assessment of the ONH is necessary for early detection. The disease is typically diagnosed in a clinic based on measurement of intraocular pressure, visual field, etc.

Screening using fundus photography has been explored as fundus images are easy to acquire and automated image analysis based diagnosis facilitates large-scale screening. Since, glaucoma leads to structural changes of the ONH, assessment of this region is necessary for early detection. ONH changes are typically detected from the ratio of vertical diameters of the optic disc and cup (CDR), and many methods have been proposed for automatic assessment from fundus images. Damage to the retinal nerve fiber layer (RNFL) is considered to be one of the root causes for the morphological changes in the ONH. The measurements in the ONH and peripapillary regions including global and sectoral RNFL thickness, lamina depth, and the neural rim width measured from the Bruch's membrane opening (BMO) [2] have been recognized as some of the predictors for early detection and monitoring. Both RNFL loss and BMO require a 3D view of the retina available via tomography.

Optical coherence tomography (OCT) is non-invasive and generates high-resolution 3D images of the retina. It allows for the direct examination of the 3D retinal structure and quantitative measurements of these regions. This is in contrast to fundus imaging which captures the structural changes in the top-most layer of the retina. It provide valuable information for understanding the changes occurring in the retina due to glaucoma, specifically, related to the retinal nerve fiber layer and the optic nerve head. Though OCT has been widely adopted in ophthalmology and clinical studies have shown it to be more sensitive than fundus photography for evaluating glaucoma progression [3], automated detection of glaucoma from fundus images, rather than OCT volumes, has received more attention possibly due to wider availability of fundus cameras. In this chapter, we explore a deep neural network approach for automatic detection of glaucoma directly from the OCT volumes.

2.2 Related work

Most of the existing work on OCT start with segmentation of retinal structures to extract features for machine learning techniques like Naive Bayes (NB), Support Vector Machine (SVM), Random Forests (RF), etc. The features that are extracted include peripapillary RNFL thickness, the cup-to-disc ratio obtained from detecting breakpoints in the Retinal Pigment Epithelium (RPE), wavelet-based features, macular GCIPL thickness, etc. Significant Discrete Wavelet Transform (DWT) coefficients are selected in [4] as features to be classified using SVM. The highest accuracy reported for this method is 90.75%. Parameters such as optic cup depth, average retinal thickness and CDR have also been proposed to be used for classification by a back-propagation neural network and SVM [5] with reported accuracies of 89.29% and 93.83% respectively. RNFL thickness maps extracted from OCT together with visual field test parameters were explored in [6] for classification using NB, RF, SVM, etc., with the highest AUC reported to be 0.946. A similar approach was used in [7] on a larger dataset of 500 images which included k-Nearest Neighbor among different classifiers. The highest AUC of 0.979 was reportedly achieved with RF. The good performance reported by these methods is contingent on the accurate segmentation and handcrafted features to represent the segmented structures. This is often difficult in low-quality scans or advanced glaucoma cases.

Deep learning is a vastly employed technique for many of the tasks in medical imaging. However, it has not been explored much for processing 3D OCT volumes for glaucoma assessment. The reasons for this could be because, it is a huge time and computationally complex process, low availability of OCT data, etc. More recently DeepMind introduced a framework [8], which is an ensemble of 5 segmentation and 5 classification networks, that analyses OCT scans. They demonstrated that operating on OCT scans directly resulted in a performance which was on par with the clinical experts. Convolutional Neural Networks (CNN) are state-of-the-art methods for many classification tasks as they have extensive learning capacity and can infer the nature of an input image without any prior knowledge. Recently, Sabour et al. introduced the Capsule network [9] to address some drawbacks of CNNs, namely, the requirement for large amounts of training data incorporating full-variations of the input and the usage of pooling operations which may discard relevant information that is valuable for classification. The main building blocks of this network are called ”capsules” which try to preserve the hierarchical association between features. Instead of applying a non-linear mapping to get a scalar output, a capsule outputs a vector by grouping multiple feature channels to represent a generalized set of related properties. While CNNs tend to detect the presence of features to make a decision, the Capsule networks' routing algorithm learns to find relations between the features. Many of the recent studies conducted on Capsule networks [10] [11] suggest that they can be trained with far less data for the same or better performance. An additional advantage is that they are more robust to an imbalanced class distribution, which is the case for many medical image datasets. In the case of glaucoma, along with the specific, localized features as mentioned before, it is also important to look at the subtle, diffuse changes. Therefore, contextual information, including correlations between features at different locations, may be required for good results. Hence, we explore the Capsule network architecture for automatic detection of glaucoma.

2.3 Methodology

2.3.1 Preprocessing

Preprocessing of the SD-OCT volumes included denoising, retinal curvature flattening and region of interest extraction. The overview of the preprocessing steps is shown in Fig. 2.1. The volumes are generally degraded by speckle noise whose characteristics vary depending on the structure of the tissue. Hence, for denoising, speckle reducing anisotropic diffusion [12] is used with iterations and timestep set to 30 and 0.1 respectively. It is followed by bilateral filtering to obtain a smooth image while preserving the edges.

The retinal layer region in the OCT has a natural curvature which varies both across volumes and also across the slices of each volume. This is corrected by flattening to obtain a normalized view of the retina using a technique introduced in [13]. Firstly, contrast enhancement is performed to get a better approximation of the boundaries. Next, the location of the brightest pixel in each column is obtained to get a rough estimate of the outer RPE layer boundary. The outlier pixels in the resultant vector are removed and a quadratic polynomial is fit, which is mapped onto a straight line by shifting the columns. Next, each slice is sequentially aligned to its previous by an exhaustive search of pixel translations that maximizes the normalized cross-correlation (NCC) between the slices.

Next, the region of interest (ROI) is extracted using a method introduced in [14]. The slices are projected along columns, and a Gaussian function is fitted on the resulting vector. The mean value of the resultant represents the y-coordinate of the center of the layer region. An ROI of 200 pixels in height centered around this coordinate and 304 pixels in width is extracted for further processing. To reduce the computational overhead during training, each volume is resized to 200x304x102.

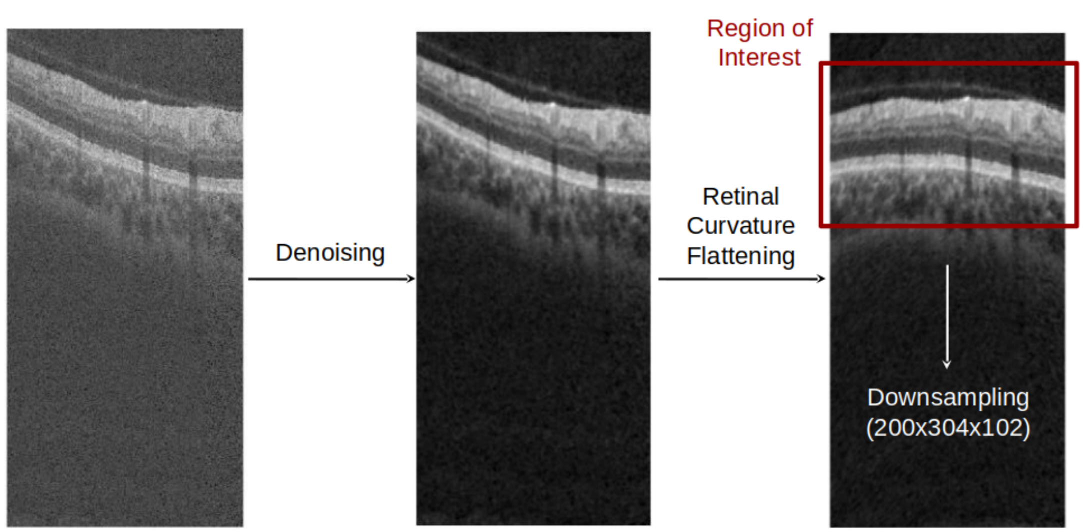


Figure 2.1: Overview of preprocessing steps

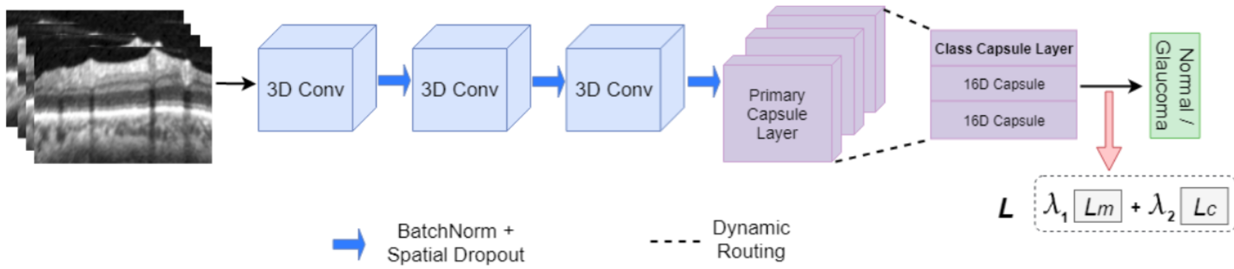


Figure 2.2: Overview of the proposed network for glaucoma assessment

2.3.2 Classification network

The original Capsule network was designed to process 2D images. Hence, the network is modified to make it suitable for processing 3D OCT data. The overview of the proposed network is shown in Fig. 2.3. It is a simple and shallow network consisting of three blocks of 3D Conv-BatchNorm-LeakyReLU-SpatialDropout layers and finally the dynamic routing layers. The three blocks are added before the primary capsule to shrink the volume in order to reduce the number of parameters in the model which decreases the training time.

The first block (Conv3D-1) performs 4 convolutional operations with a kernel size of 13x7x3 and strides of 3x3x1. The second block (Conv3d-2) performs 8 convolutional operations with a kernel size of 13x3x3 and strides of 3x2x2. The third block (Conv3d-3) performs 16 convolutional operations with a kernel size of 15x3x3 and strides of 3x2x2. The last block converts pixel intensities to the local features which are then used as inputs to the primary capsules. The primary capsule layer performs 256 (32x8) convolutional operations with a kernel size of 256x9x9 and a stride of 2 to produce 32 capsule maps each of size 6x6, each capsule being an 8D vector. Since the problem at hand is two-class classification, the class capsule layer is designed to have 2 capsules, one for each class. Each class is represented as a 16D capsule, where the length of each capsule determines the probability of the input belonging to a particular class. Each of these receives input from all capsules in the primary capsule layer. The dynamic routing algorithm is employed to achieve an agreement between the primary and class capsule layers. It is similar to a feedback algorithm which increases the contribution of those capsules that agree most with the parent output, thus strengthening their contribution. A non-linear squashing function is applied to the output vector of the class capsule layer. This ensures that the orientation of the vector is unchanged, while the length is between 0 and 1 in order to represent the probability that the entity represented by the capsule exists in the current input. The outputs from these squash functions tell us how to route data through various capsules that are trained to learn different concepts.

2.3.3 Loss functions

The network is trained with a weighted combination of two loss functions: margin loss and center loss.

$$L = \lambda_1 L_m + \lambda_2 L_c \ \ where \, \lambda_1 = 1 \, , \, \lambda_2 = 0.3$$

L_m is the margin loss from the original Capsule network paper [9], which is computed for each capsule. It is defined as

$$L_m = T_k \, max(0, \, m^+ - ||v_k||)^2 + \, (1 - T_k) \, max(0, \, ||v_k|| - m^-)^2$$

where $T_k = 1$ if and only if an input of class k is present and $m^+ = 0.9$ and $m^- = 0.1$. The first term is zero in case of a correct prediction with probability greater than m+, otherwise it is non-zero. Second term is zero in case of an incorrect prediction with probability less than m-, otherwise it is non-zero.

L_c is the center loss which learns the centers for each class simultaneously and penalizes the distances between the deep features learned and their corresponding class centers [15]. This is aimed at improving the discriminative power of the deep features. It is defined as

$$L_c = \frac{1}{2} \sum_{i=1}^{m} ||x_i - c_{y_i}||_2^2$$

where x_i denotes the i^{th} deep feature obtained as the output of class capsule layer and c_{y_i} denotes the center belonging to the y_i^{th} class, which is updated for each training batch and m is the batch size.

2.4 Datasets

The proposed method was evaluated on two privately sourced datasets of ONH-centered volumes:

- A set of 128 volumes (95 healthy and 33 glaucomatous) of size 640 x 304 x 304, were collected from Anand Eye Institute, India. These were acquired with an Optovue scanner (spectral domain RTVue - XR 100 OCT).
- A set of 125 volumes (69 healthy and 56 glaucomatous) of size 512 x 512 x 193 were obtained from Rotterdam Eye Institute. These were acquired by a Spectralis OCT system (Heidelberg Engineering, Germany).

2.5 Implementation details

2.5.1 Training and Testing sets

The two datasets were combined (253 volumes consisting of 164 healthy and 89 glaucomatous cases) and split randomly into three sets for training, validation and testing. The sets had roughly 2:1 ratio of healthy and glaucoma samples as shown in Table 3.3. All the volumes were reduced to 200x304x102.

	Total	Healthy	Glaucoma
Train	165	106	59
Validation	38	26	12
Test	50	32	18

Table 2.1: Distribution in train, validation and test sets

2.5.2 Evaluation metrics

Quantitative evaluation was done using the following criteria: Balanced Accuracy, Sensitivity, Specificity, and Area Under the ROC Curve (AUC).

$$Sensitivity = \frac{TP}{TP + FN} \qquad Specificity = \frac{TN}{TN + FP}$$

$$BalancedAccuracy = \frac{Sensitivity + Specificity}{2}$$

where TP, TN, FP and FN denote the number of true positives, true negatives, false positives and false negatives respectively. Along with these, the ROC graphs are also provided for visualizing the performance of different networks.

2.5.3 Hyper-parameter selection

The class imbalance in the dataset was handled by embedding class weights in the loss function. The number of iterations for routing is a hyperparameter for the Capsule network and was chosen to be 3 since larger values might lead to overfitting. The network is trained using 5-fold cross validation using Adam optimizer with a learning rate of 0.0001 which decays at the rate of 0.9 for faster training. The batch size was set to 3 because of memory constraints. The architecture has 243,736 network parameters, which is significantly smaller compared to many CNN-based architectures. The test run time per OCT volume is about 0.9 sec on average.

The proposed pipeline was implemented on an NVIDIA GTX 1080 GPU, with 11GB of GPU RAM.

2.6 Experiments and Results

The proposed system was assessed by conducting various experiments. These are described in the following sections.

2.6.1 Effect of preprocessing

To evaluate the effect of preprocessing on the performance, we compared the proposed network in two scenarios: with and without preprocessing. We considered the worst case scenario for the nopreprocessing case, where all the original volumes were resized to 200 x 304 x 102 and the proposed network was retrained on these volumes. The performance was assessed using the evaluation metrics and reported in Table 2.2. The tabulated results indicate that preprocessing is critical for good performance. This is to be expected as the (speckle) noise levels in OCT are high and the curvature of the tissues is variable depending on the image acquisition angle.

	BAcc.	**Sens.**	**Spec.**	**AUC**
Without preprocessing	0.6226	0.5714	0.6800	0.5036
With preprocessing	**0.9200**	**0.8900**	**0.9600**	**0.9730**

Table 2.2: Performance comparison of the network with and without preprocessing of OCT volumes

2.6.2 Assessment of different networks

First study was to assess the choice of network: Capsule network versus 3D CNN and 3D ResNet-50. The 3D CNN is composed of five 3D convolutional layers with leaky-relu activation, batch normalization and spatial dropout where each 3D Conv layer has 32 filter banks with filters of sizes 7-5-3-3-3 and with strides 2-1-1-1-1. Global Average Pooling layer and a dense layer are employed after the last 3D convolutional layer, and before the final softmax output layer for the prediction of class labels. Due to memory constraints, the OCT data is resized to 64x64x128 for training both these networks. The proposed Capsule network is also modified and trained on the resized input for a fair comparison of their performance. The test results shown in Table 2.3, indicate the Capsule network clearly outperforming both 3D CNN and 3D ResNet-50. The corresponding ROC curves are plotted and shown in Fig. 2.4 It also shows that the Capsule network operating on the whole OCT volume (200x304x102) instead of the resized has significant improvement in performance.

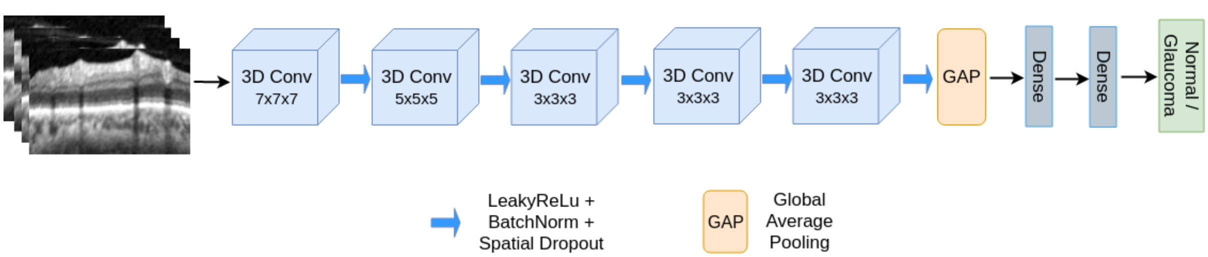


Figure 2.3: 3D CNN used for comparison

	BAcc.	Sens.	Spec.	AUC
3D CNN	0.64	0.61	0.65	0.619
3D ResNet-50	0.78	0.61	**0.96**	0.899
Capsule Network (Resized input)	0.84	0.722	**0.96**	0.945
Capsule Network (Proposed)	**0.92**	**0.89**	**0.96**	**0.973**

Table 2.3: Performance comparison of different networks

For all the networks, 5-fold cross validation is performed and the obtained training and validation loss curves for each fold are plotted as shown in the figures, Fig. 2.5 and Fig. 2.6 below. The corresponding box plots for training loss, training accuracy, validation loss, validation accuracy are shown in Figure 2.7.

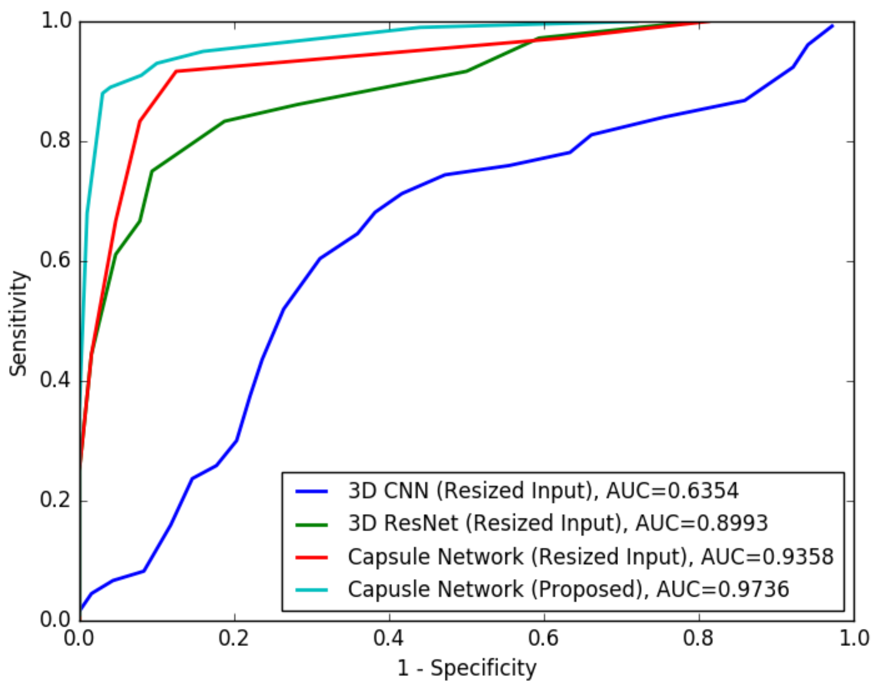
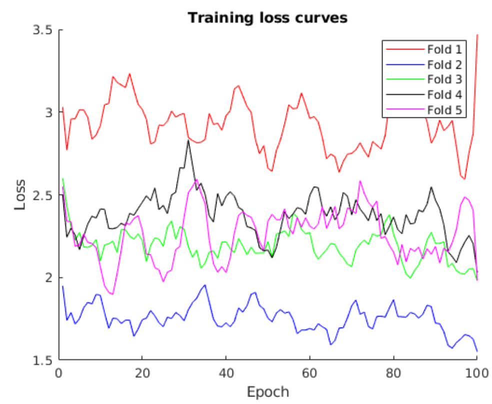
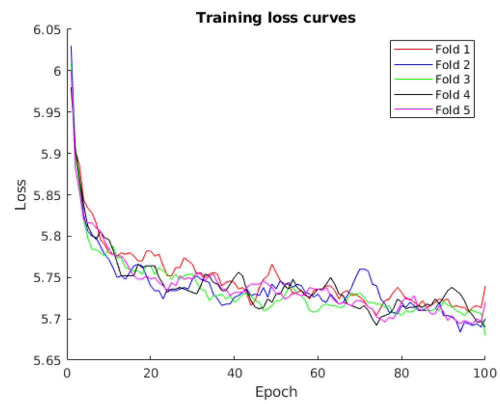


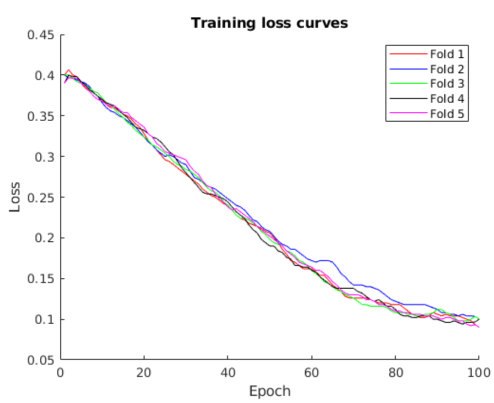
Figure 2.4: ROC curves comparing the performance of different networks



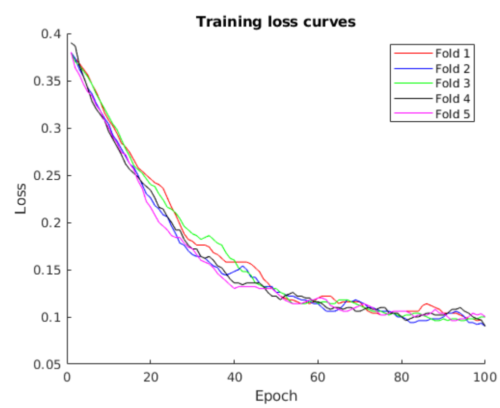
(a) Train loss curves for 3D CNN



(b) Train loss curves for 3D ResNet-50

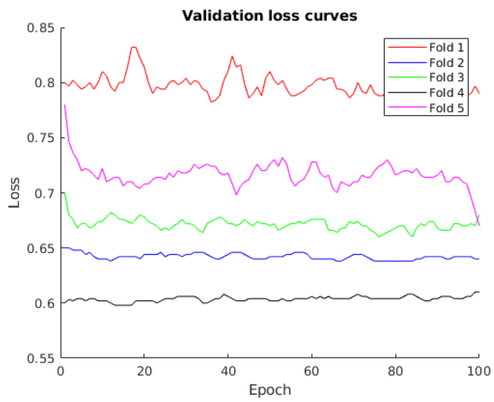


(c) Train loss curves for resized Capsule network

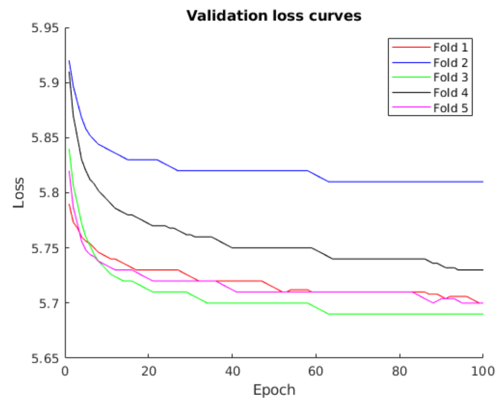


(d) Train loss curves for proposed Capsule network

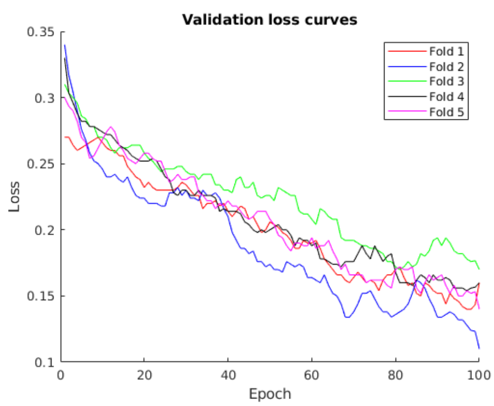
Figure 2.5: Training loss curves for different networks



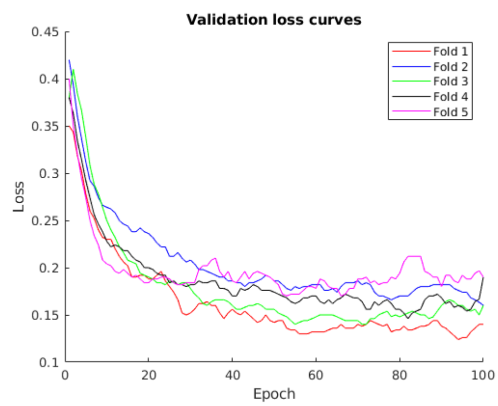
(a) Validation loss curves for 3D CNN



(b) Validation loss curves for 3D ResNet-50

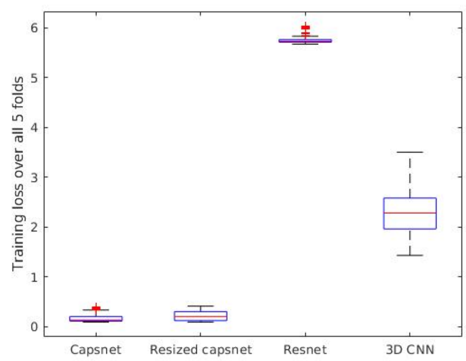


(c) Validation loss curves for resized Capsule network

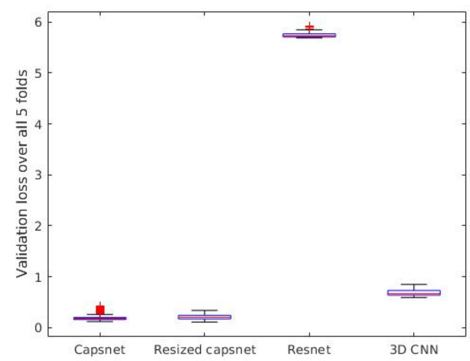


(d) Validation loss curves for proposed Capsule network

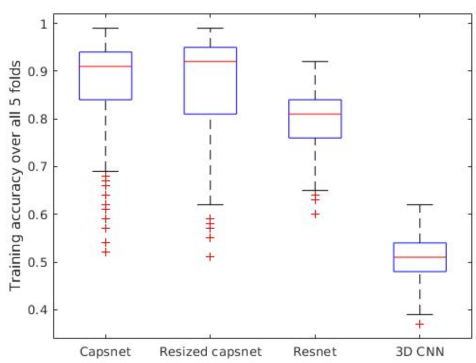
Figure 2.6: Validation loss curves for different networks



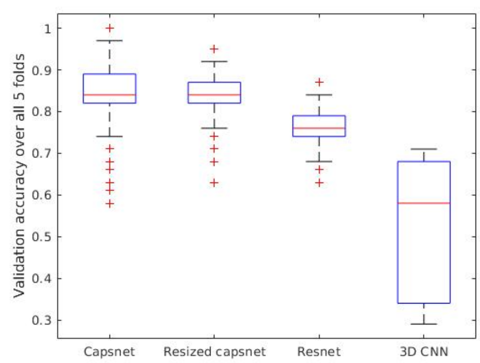
(a) Box plot for training loss over 5 folds



(b) Box plot for validation loss over 5 folds



(c) Box plot for training accuracy over 5 folds



(d) Box plot for validation accuracy over 5 folds

Figure 2.7: Box plots

2.6.3 Assessment of different regions

The second study aimed at understanding the contributions of different regions (intra-disc, peripapillary) to the final decision. Each volume was manually divided into two parts: one with only the slices of the superior and inferior peripapillary regions and another with slices from within the optic disc. The former can provide information on any early loss of RNFL while the latter has information on optic disc/cup morphology and BMO. Fig. 2.9 display few slices corresponding to the disc region (blue) and the peripapillary region (green) respectively that are sampled at regular intervals. The corresponding maximum value projections forming the enface are shown in Fig. 2.8. The size of the volumes in the two cases was standardized using padding with blank slices. The proposed model was trained with these two sets separately and the assessment results are reported in Table 2.4 and the corresponding ROC curves in Fig. 2.10. The results indicate that the disc and the peripapillary regions are nearly equal in importance with the latter having a slight edge (in terms of AUC) and best classification performance is achieved only with both these information.

	Sens.	**Spec.**	**Acc.**	**AUC**
All slices	0.89	0.96	0.94	0.973
Peripapillary region slices	0.77	0.96	0.90	0.949
Disc region slices	0.83	0.90	0.88	0.920

Table 2.4: Performance comparison with restrictions in input data

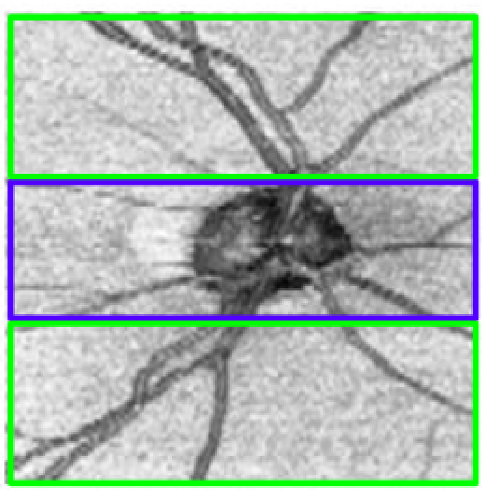


Figure 2.8: OCT enface corresponding to different regions mentioned in Table 2.4. The enface areas marked my green correspond to the Peripapillary region slices. The enface area marked by blue corresponds to the Disc region slices.

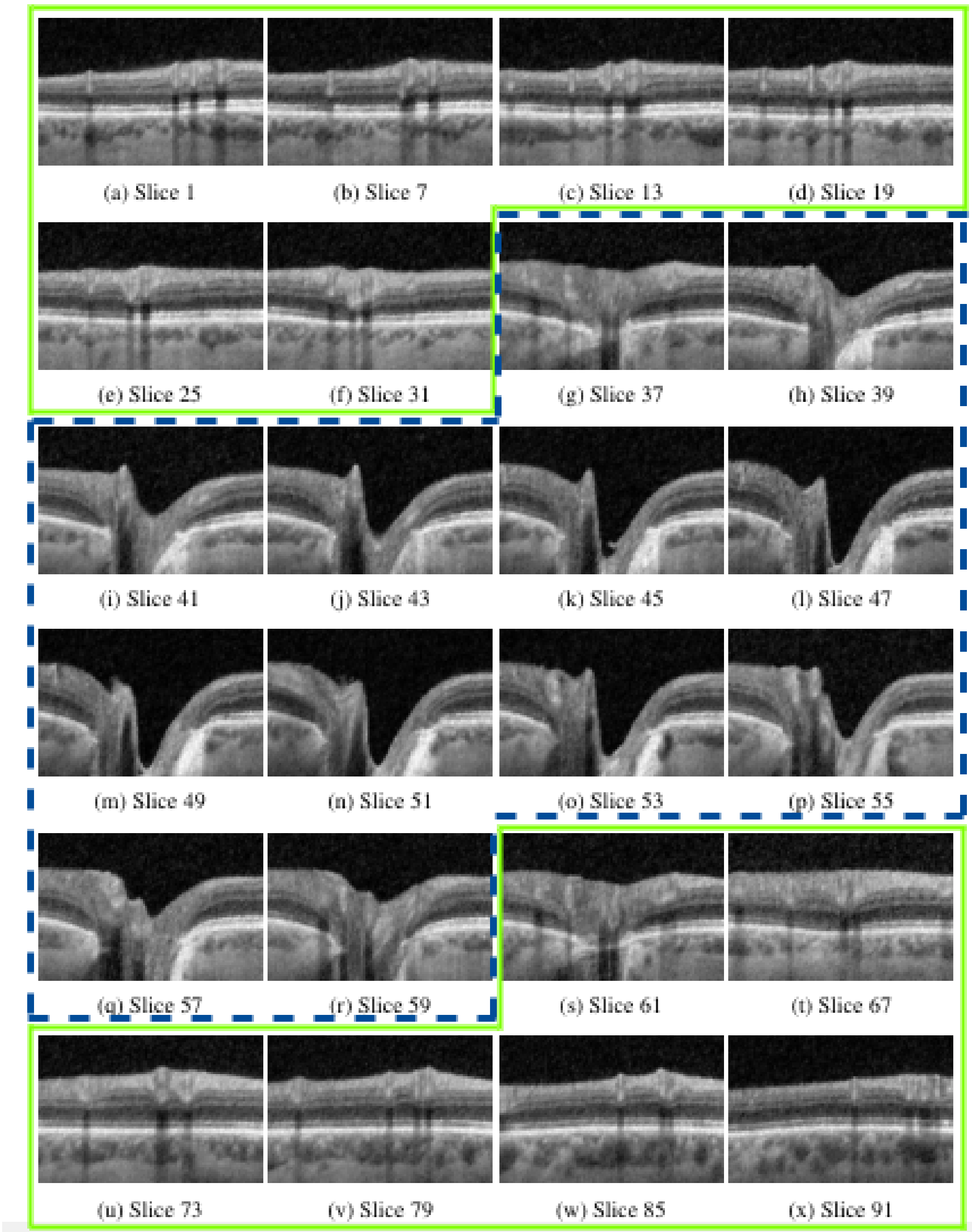


Figure 2.9: (a) - (f) and (s) - (x): Slices corresponding to the BMO, optic disc and cup regions. (g) - (r): Slices corresponding to the peripapillary region

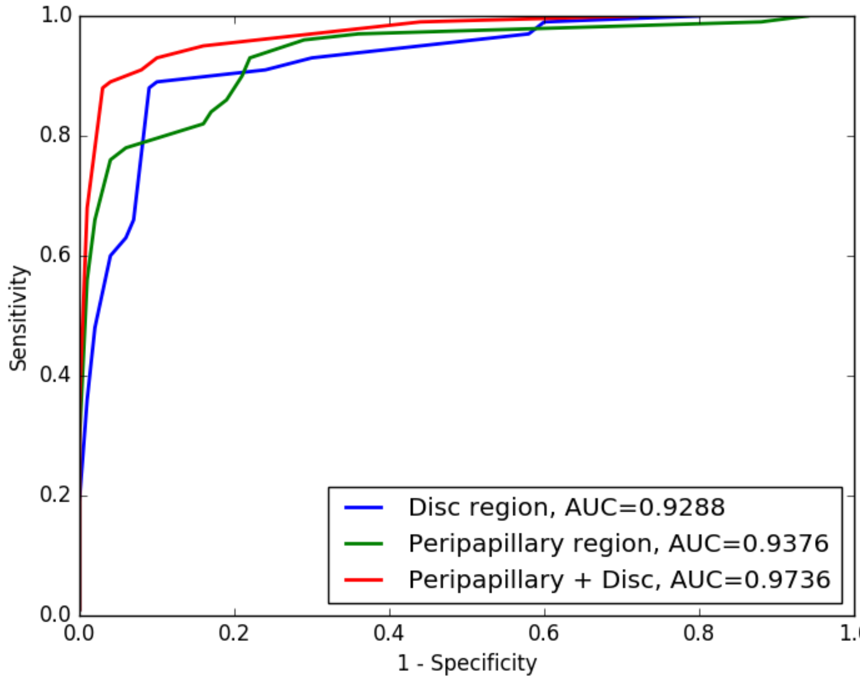


Figure 2.10: ROC curves comparing the performance of Capsule network with restrictions in input data

2.6.4 t-SNE Visualization

In order to understand what is being learned, it helps to look at the visualization of the outputs of different layers. But, as we go deeper into the network, the interpretability is lost due to complexity of the network. Hence, to visualize the outputs of Primary capsule and Class capsule layers, a non-linear dimensionality reduction technique called t-SNE [16] is applied, because it has been proven to provide better results compared to other dimensionality methods like PCA. Using t-SNE, the output vectors of the primary and class capsule layers are converted to 2D and plotted as shown in Fig. 2.11 and Fig. 2.12. From the plots, it can be observed that, although few samples are overlapping, the Primary capsule layer differentiates the features of healthy and glaucoma reasonably well which is further improved by the Class capsule layer, by tightly clustering the samples belonging to the same class.

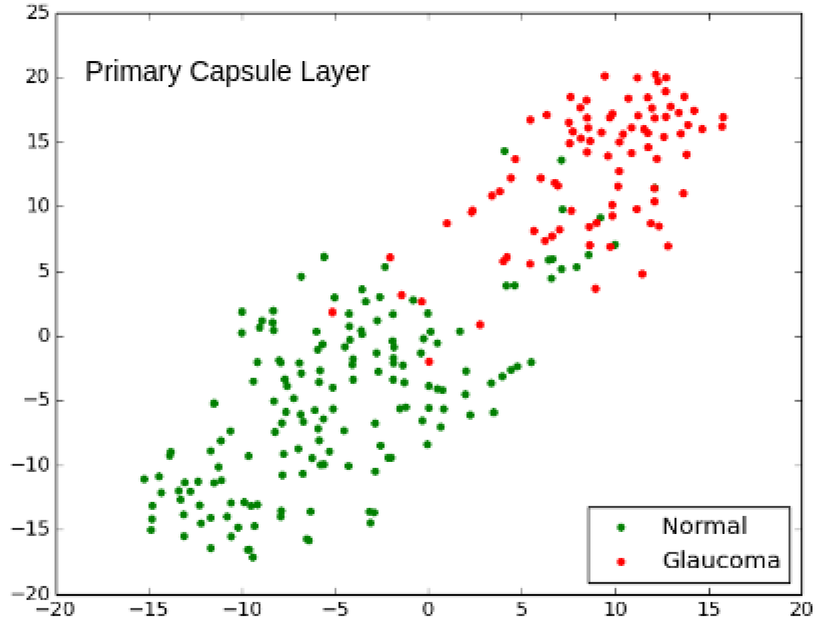


Figure 2.11: t-SNE visualization of features learned by the Primary Capsule layer

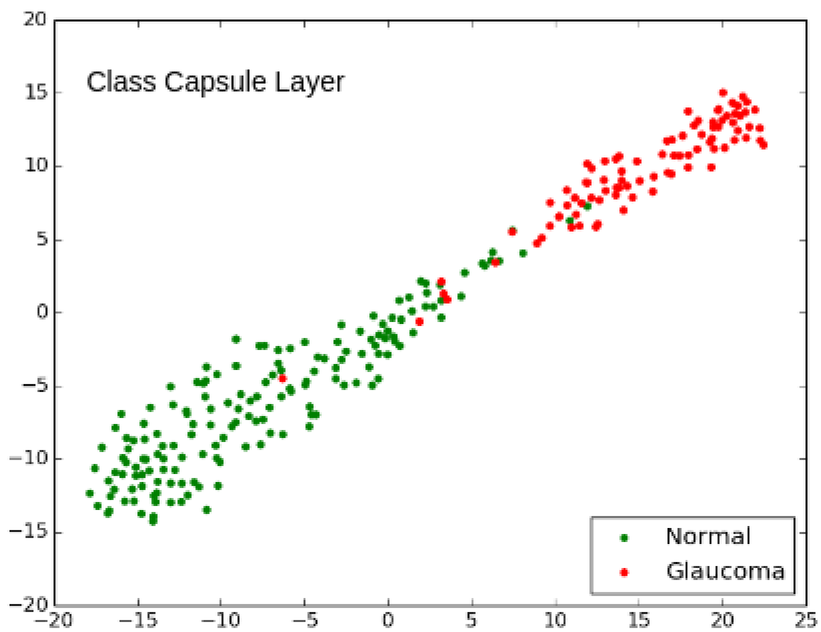


Figure 2.12: t-SNE visualization of features learned by the Class Capsule layer

2.7 Conclusions

We proposed a solution for the assessment of glaucoma directly from the OCT volumes. This is motivated by the fact that clinical diagnosis relies on multiple factors, which is difficult to derive via features from segmented regions. The known bio-markers such as RNFL thinning, rim thinning within the ONH, BMO appear at different locations. The Capsule network was seen to be able to integrate such spatially distributed information better than a traditional convolutional network and also achieve good results in very few training epochs. Hence, it can be concluded that the proposed Capsule network successfully learns highly discriminative features from OCT. Currently, attempts by different groups are focused towards miniaturization of OCT. When this matures, the obtained results can pave the way for a reliable OCT-based screening solution on a large scale.

Chapter 3

Glaucoma Assessment from Fundus Images with Fundus to OCT Feature Space Mapping

3.1 Introduction

Damage to the optic nerve leads to glaucoma, a major eye disease which gradually leads to an irreversible loss in vision. Hence, early detection of glaucoma is required to prevent degradation of the optic nerves and arrest the progression of the disease. Fundus and Optical Coherence Tomographic (OCT) are the two retinal imaging modalities typically used separately for automated assessment of the disease. OCT provides high-resolution 3D images with detailed ocular structures in both the anterior and posterior segments of the eye whereas fundus captures the structures of the interior surface of the eye. This complementary nature of both the modalities pave the way for formulating solutions that combine information from such orthogonal sources which may have the potential for a more accurate diagnosis.

Existing automated systems rely largely on fundus imaging for assessment of glaucoma due to their fast acquisition and cost-effectiveness compared to OCT imaging. Many studies showed that OCT images provide vital and unambiguous information about nerve fibre loss and optic cup morphology which are essential for disease assessment. However, the high cost of OCT is a deterrent for deployment in screening at large scale. This implies that it might not be feasible to have a combination of modalities available in a glaucoma screening scenario. We therefore look at the problem of glaucoma diagnosis anew keeping this constraint in mind and pose it as a missing data problem. In other words, the CAD solution aims to learn a model that leverages information from Fundus and OCT images during a training phase and perform an accurate assessment during a subsequent online detection phase in the absence of OCT data. This extends the possibilities of deploying the CAD systems where the OCT devices are too expensive for widespread usage.

This chapter gives a detailed description of the proposed novel glaucoma assessment solution. The strategy we adopt is to learn a model that maps fundus to a OCT-feature space such that given a fundus image, the model is able to generate additional features from this space. This is however, done without relying on paired fundus-OCT data, another perk of our method since acquiring paired data is challenging. Finally, to provide a more accurate diagnosis, a deep neural network model is employed which generates the aforementioned fundus-to-OCT mapped features along with the fundus-specific features which are then fused and used for classifying the input as normal or glaucomatous. We also demonstrate that our approach has a significant improvement in classification accuracy over that obtained from a corresponding unimodal (fundus) approach.

3.2 Related work

Majority of the computer aided diagnostic (CAD) solutions proposed for detecting glaucoma are based on fundus camera images given their cost effectiveness and wider availability. Most of these solutions rely on two distinct approaches: analysis of the OD region, with or without segmentation of the OD and OC. In the first approach, hand-crafted features are extracted and then classified by various machine learning strategies. Features explored include higher-order spectral features, fractal features, wavelet-based features, texture features, statistical features like entropy, uniformity, smoothness, etc, either from the entire image or a specific region of interest [17, 18, 19, 20, 21, 22]. Classification is done with methods like Support Vector Machine (SVM), k-Nearest Neighbours (k-NN), Naive Bayes (NB), etc. Advances in neural networks have enabled learning the features using deep convolutional neural network (CNN) architectures [23, 24]. Variants of approaches can be found in recent literature. These include a network [25] consisting of four convolution layers and two fully connected layers, along with response normalization and overlapping pooling layers that were adapted to reduce overfitting. Two private datasets have been used in their experiments with one used in training. The reported AUC are 0.83 and 0.88. A novel encoder-decoder based design [26], where the encoder is used to obtain a latent-space distribution which is then used by a classifier for glaucoma detection. Given that several architectures (VGG-16, InceptionNet and ResNet50) have already been demonstrated to be successful at classification in the computer vision domain, transfer learning with these networks has also been explored [27]. Image synthesis using a Semi-Supervised Deep Convolutional Generative Adversarial Network (DCGAN) is another novel attempt to generate labelled retinal images cropped around the OD. Such a network has been trained on a large dataset (over 80000 images), to design a semi-supervised classifier for glaucoma assessment in [28] and the reported Sens./Spec./AUC are 0.8290/0.7986/0.9017. Another approach based on transfer learning/fine tuning is reported in [29]. Popular architectures such as VGG16, ResNet, GoogleNet have been experimented with and without transfer learning on a combined dataset consisting of images from Drishti-GS1, RIM-ONE and a private set. The combined dataset of around 1700 images was split into two sets for training and testing. The highest performance was found to be obtained using a fine tuned VGG16 whose BAcc./Sens./Spec./AUC are 0.88/0.87/0.89/0/94.

The second approach is based on parametric analysis using CDR and rim profile which are also used in the clinical evaluation process. This requires the segmentation of OD and OC. Segmentation solutions that have been proposed range from super-pixel classification [30] [31], weighted reconstruction of OD and OC [32] and depth based joint estimation [33]. The U-net is a neural network architecture proposed for deep learning-based segmentation. Adapting this has also been attempted for glaucoma assessment, either with very little [34] or a more extensively modified form [35]. The framework proposed by [35], M-Net is a pixel-wise classification based approach which first classifies each pixel as OD or OC or non-target, and then uses ellipse-fitting to approximate the boundaries and then CDR was extracted for glaucoma detection. While the OD segmentation methods report very high accuracy, cup segmentation is still an open problem and errors in this stage has a bearing on the final diagnosis. A few approaches have combined both segmentation and image-based features for glaucoma classification by fusing them in various ways. For instance, [36] proposed a framework that fuses both segmentation-based and image-based features within a co-training based semi-supervised setting to classify the input. Another framework, DE-Net introduced by [37] integrates four independent neural network streams using U-Net and ResNet-50 whose predictions are combined to obtain the final screening result.

In contrast to the automated diagnosis, clinical evaluation of glaucoma relies on a wider range of parameters for early detection of glaucoma. While the RNFL thickness was one such parameter, recent studies have shown the location of the Bruch's membrane opening (BMO) [2] and lamina cribrosa displacement [38] also to have some role in early detection. These require cross sectional information of retina which is available only in a tomographic image. Since fundus image provides the structural information of only the top-most layer of the retina, there is a possibility that early symptoms like changes in RNFL thickness might not be captured. OCT provides a detailed cross-sectional view of the retinal tissues. It has advantages over fundus imaging with regards to the assessment of glaucoma. Many clinical studies have shown OCT to be more sensitive compared to fundus photography for assessing glaucoma. [3]. In addition to RNFL thickness changes, BMO [2], neuroretinal rim width [39], lamina cribrosa depth [38], and OC depth, a key indicator of glaucoma, can also be measured in OCT. Most of the existing work on OCT-based assessment is based on extracting features such as discrete wavelet coefficients [4], from the segmented retinal structures and classifying them using machine learning techniques. More recently, DeepMind introduced a framework [8], which is an ensemble of 5 segmentation and 5 classification networks, that analyses OCT scans for both the tasks, and reported a performance on par with the clinical experts. They have demonstrated that their system offers potential usage in a clinical setting.

3.3 Methodology

When a structure is imaged with 2 different modalities the information captured across the two may be different (even complementary) but statistically or otherwise related. This is true even if one is a functional imaging modality while the other is a structural imaging modality. This is at the heart of the use of mutual information for cross modal registration. In the problem at hand, both OCT and fundus imaging are structural imaging modalities with one representing the retinal surface while the other its cross section. Despite this difference, it should be possible to relate the two in a feature space. Further, it is desirable that this relation is defined (or learnt) at a semantic level (normal versus glaucomatous class) as the abstraction will then be agnostic to variations within the classes and across individual eyes (retina). Thus, the OCT and fundus images need not be paired, a key advantage as sourcing paired data is more difficult. Hence, in our design, we propose to use a network to derive features from labelled OCT volumes and a second network to perform an adversarial-learning based mapping between features derived from labelled fundus images to the OCT features. It notable that this is different from domain adaptation technique wherein both the images are mapped to a common feature space. Finally, we design a classification network that fuses features and performs the glaucoma assessment.

3.3.1 Overview of the proposed method

The proposed framework consists of three modules: (i) preprocessing of OCT volumes and fundus images as described in section 3.3.2; (ii) establishing the OCT-feature space to which the fundus images are mapped using a generative adversarial technique as described in section 3.3.3; (iii) an end-to-end model for deriving the decision on input fundus images. This is described in section 3.3.4.

3.3.2 Preprocessing

Fundus images often suffer from non-uniform illumination. This was corrected by luminosity and contrast normalization [40] as follows. For each pixel ($I(x, y)$), the mean ($\mu(x, y)$) and standard deviation ($\sigma(x, y)$) within a fixed-size window is computed. These statistics are used to identify the background pixels using the Mahalanobis distance $D(x, y)$ and finally the corrected image $I_c(x, y)$. D and I_c are computed as

$$D(x,y) = \left| \frac{I(x,y) - \mu(x,y)}{\sigma(x,y) - \sigma_\epsilon} \right| \qquad I_c(x,y) = \left| \frac{I(x,y) - \mu_B(x,y)}{\sigma_B(x,y) - \sigma_\epsilon} \right|$$

where $\mu_B(x, y)$ and $\sigma_B(x, y)$ are the mean and the standard deviation of the background pixels; a pixel I is defined to belong to the background if $D(x, y) < 1.3$.

Since, we are interested only in the optic nerve head and the peripapillary regions, a region of interest enclosing these was extracted from the given fundus image as follows. Firstly, the center of the optic disc (OD) is determined using a Hough transform based OD localization method [41]. Next, a region of 304 pixels in height and width is cropped around this center.

3.3.3 GAN based mapping of fundus to OCT feature space

3.3.3.1 Establishing the OCT-feature space

Our previously developed Capsule network, as shown in Fig. 3.1 **A** was used for setting up the OCT-feature space. For each of the OCT image, a 32D feature vector is obtained after the class capsule layer, which is used in the next stage.

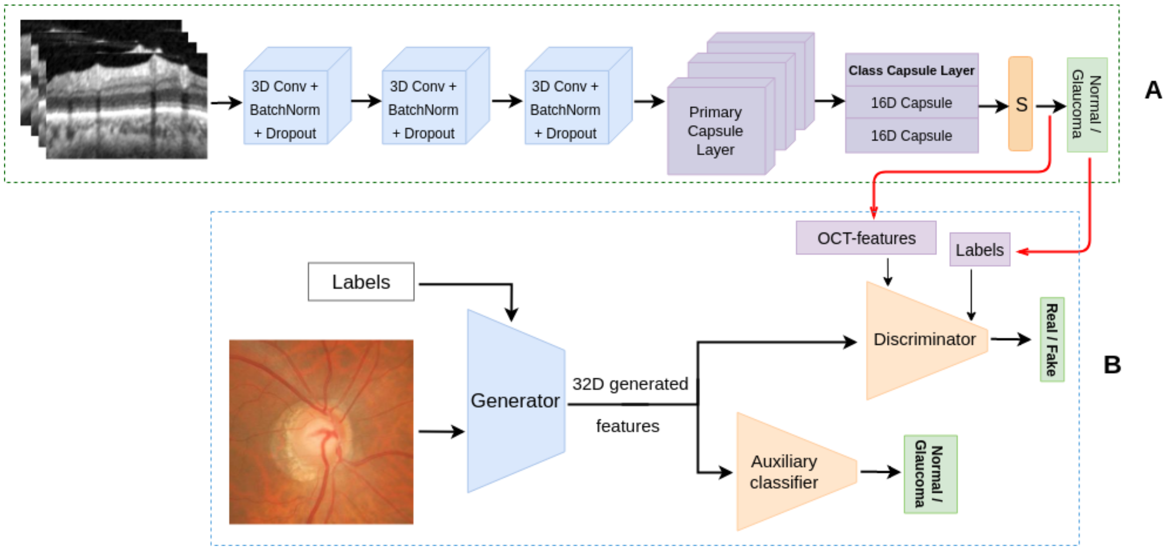


Figure 3.1: Two-stage system for mapping fundus to the OCT-feature space. **(A)** represents Capsule network generating the OCT-features. **(B)** The Mapping network: shows the GAN model mapping fundus to the OCT-feature space.

3.3.3.2 Mapping the fundus to OCT-feature space using GAN

Although, highly discriminative features are learnt from OCT, operating on such high dimensional data has high computational and time complexity. To avoid this, we propose a strategy to transfer the generalization ability of the Capsule network into a smaller model which operates on fundus. This is done using the capsule features generated, since they represent the valuable information of OCT in a lower dimensional space. For this purpose, we make use of an adversarial network to learn a mapping from fundus to the OCT-feature space.

Generative Adversarial Networks (GAN), introduced by [42] are generative models that learn a mapping from random noise vector to output. Conditional GAN (CGAN) introduced by [43] extends the GAN model allowing the generation of samples with certain attributes. It learns mapping from a random noise vector z and observed input x, to a required output y. Here, the generator and discriminator are conditioned on any information related to the target samples, which is provided as additional input layer in both the models. Hence, the generator is guided to produce fake samples with a specific condition rather than a generic sample from an unknown noise distribution.

The conventional CGAN was modified to suit the problem at hand. The generator synthesizes features from the fundus image whereas the discriminator tries to distinguish between these generated features and the features obtained in the earlier stage from OCT volumes. The generator is conditioned on the input fundus image with the corresponding labels to produce the required feature vector. For the discriminator, feature vector and their corresponding labels are presented as inputs. The goal is to learn a generator that maps the fundus to the OCT-feature space. An auxiliary classifier network having the same architecture as the discriminator is included which acts as a regularizer. It takes the generated

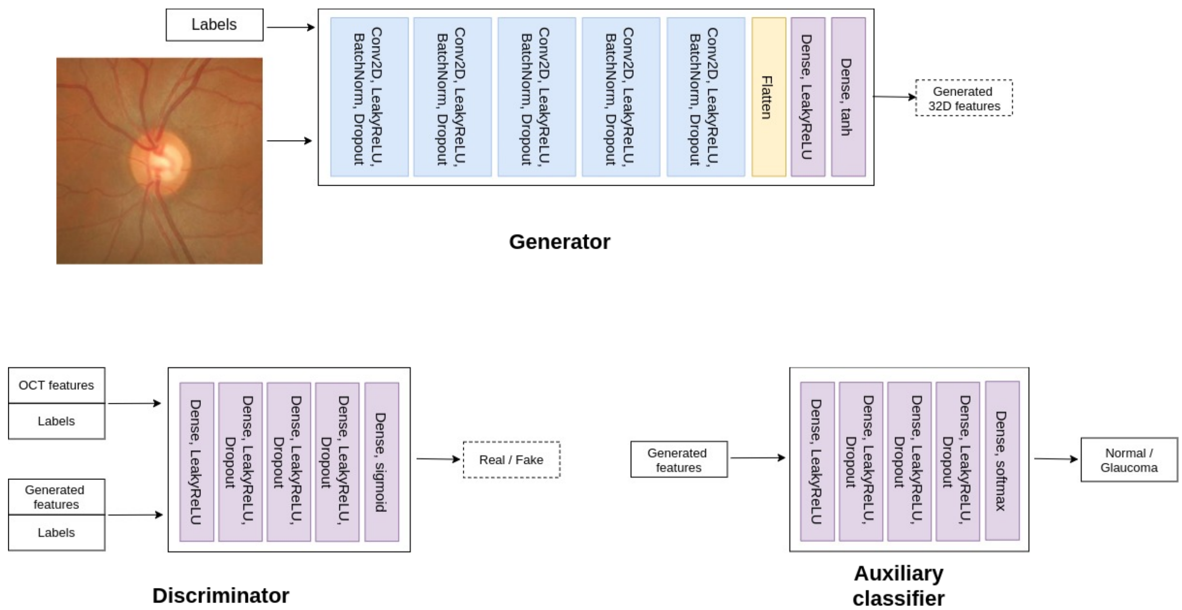


Figure 3.2: Architectures of Generator, Discriminator and Auxiliary classifier

features as input and outputs the probability of the class labels. This additional classifier network is incorporated to ensure that the generated features are class specific.

The proposed network is shown in Fig. 3.1 **B**. At the end of this stage, we have a mapping from the fundus to the OCT-feature space.

Generator (G) Architecture: The generator (G) takes input as fundus image of size 304x304 and its corresponding label and outputs a 32D feature vector. The base architecture (B, which is again used in the later stages) consists of four Conv-BatchNorm-LeakyRelu-Dropout blocks and MaxPool layer after the second and fourth blocks. After flattening the output of the second MaxPool layer, a Dense layer is employed which is connected to a final Dense layer with *tanh* activation function that outputs the required feature vector. The network is regularized with a dropout of 0.4 on all the Conv and Dense layers. After the training, the output feature vector obtained from G for each of the fundus input are used as soft labels, which are used during the training of the proposed classification network in the next stage. The architecture of the generator is shown in Figure 3.2.

Discriminator (D) and Auxiliary classifier (C): A network consisting of four Dense layers with LeakyRelu activation functions and a dropout of 0.4 for the last two layers is employed. This base architecture is shared between the discriminator (D) and auxiliary classifier (C) to make the training process computationally efficient. The inputs to D are the 32D features and their corresponding labels, whereas C takes the inputs as only the 32D feature vectors. On top of the shared network, D contains a final Dense layer with a *sigmoid* activation to determine whether the input is real or synthesized with binary cross entropy as loss function. In case of C, two dense layers are used, where the last dense layer uses *softmax* activation to output the label by minimizing the cross entropy loss. C acts like a regularizer retaining the class information of the generated features. The architectures are shown in Figure 3.2.

3.3.4 Proposed model for classification

The overview of the model proposed for classification is illustrated in Fig. 3.3. Since Feature fusion is the basis of the final classification result in this model, we refer to it as the **Fefu** model. A detailed view of the Fefu model shown in Fig. 3.4. It is a combination of three sub-networks: (i) F_{f-o}, which learns the fundus-to-OCT mapped features, (ii) F_f, which learns features specific to fundus images and (iii) a Fully Connected Network (FCN) which combines the features from F_{f-o} and F_f to derive the output label. Fine-tuned VGG-19 network is employed for both F_{f-o} and F_f. The last dense layers in the original VGG-19 are replaced with Global Average Pooling (GAP) and three dense layers, where *softmax* activation is used in the last dense layer, as shown in Fig. 3.4). Both F_{f-o} and F_f networks have an identical architecture to maintain symmetry and also for making the training process simpler. The feature vectors obtained from both the networks are merged and fed to the FCN, which consists a set of four dense layers with *ReLU* activation. A probability distribution of the class label is derived by the FCN.

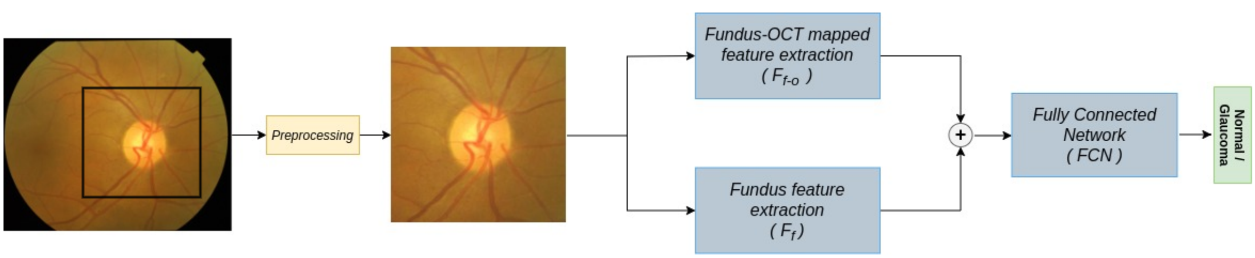


Figure 3.3: Overview of the proposed classification network. F_{f-o} represents the network generating fundus-to-OCT mapped features. F_f represents the network generating fundus-specific features.

3.3.4.1 Description of F_{f-o}

The soft labels generated for fundus images (in the training set) from the previous stage are used to train this network. The main aim of this network is to generate features from the OCT-feature space along with retaining the class information from the fundus input. For this purpose, two loss functions are employed: (i) Mean squared error (L_{MSE}), which minimizes the error between the generated feature vector and the soft label; (ii) Margin loss (L_M), which minimizes the error between the predicted and true labels. This is included to ensure the classification is robust.

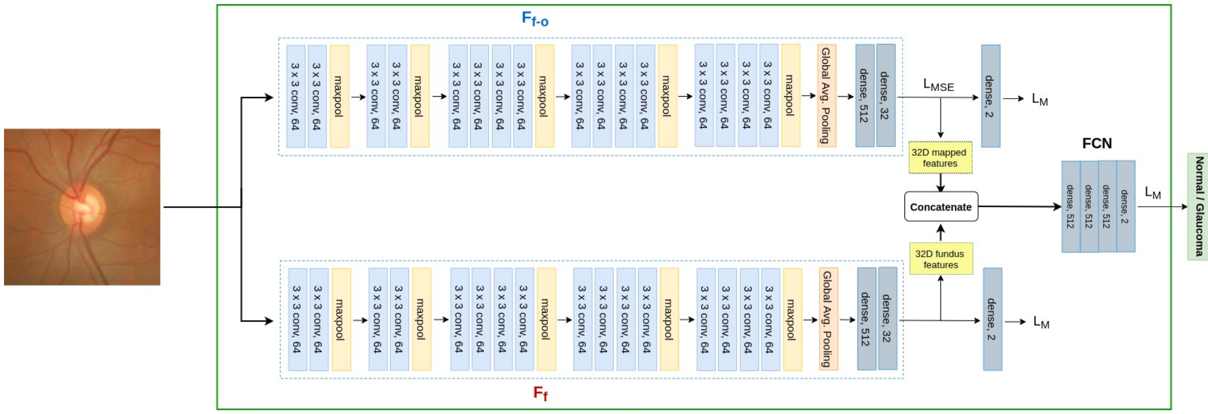


Figure 3.4: Architecture of the proposed Fefu model. L_{MSE} is the mean squared error loss and L_M is the margin loss.

3.3.4.2 Description of F_f

This network generates features specific to fundus images. A margin loss (L_M) is employed to minimize the error between the predicted and true labels.

Both F_{f-o} and F_f networks are initialized with the ImageNet weights. For the final prediction at the output of FCN, margin loss is used. All the four loss functions labelled in Fig. 3.4 are optimized in each epoch. In this way, the model incorporates both fundus and OCT image representations that are obtained from fundus images alone to improve the classification ability.

3.4 Datasets

A total of 2168 images were obtained by combining one private and three publicly available datasets, namely, Drishti-GS1 [44], RIM-ONE v3 [45] and REFUGE. The private dataset consisted of 1490 *OD-centric* images along with eye/image level diagnosis. Amongst the public datasets, Drishti-GS1 and RIM-ONE v3 have 101 and 111 *OD-centric* images respectively, whereas REFUGE has 400 *macula centric* images. The REFUGE dataset was released as part of a challenge [46]. The distribution of normal and glaucomatous cases in these datasets are given in Table 3.1. It can be noted from this table that the image sources, acquisition protocol (view, resolution) vary widely. Hence, these were combined to create a training and a testing dataset. The contribution of the four datasets to the training and testing sets is shown in Table 3.2. It can be observed from Table 3.1 that major proportion of the images are OD-centric as it captures the structures (OD and OC) and the peripapillary regions without any distortion.

Evaluation of the network on unseen data was carried out on two more datasets (last two rows of Table 3.1): A public dataset, namely, sjchoi86-HRF obtained from [47] which consists of *macula-centric* images and a locally sourced (Private-2) dataset consisting of *OD-centric* images acquired with

Dataset	Resolution	View	Total	Normal	Glaucoma
Private-1	1996x1494	OD	1490	706	784
Drishti-GS1	2896x1944	OD	101	31	70
RIM-ONE v3	1072x1424	OD	111	78	33
REFUGE	2124x2056	Macula	400	360	40
sjchoi86-HRF	2592x1728	Macula	401	300	101
Private-2	4288x2848, 2196x1958	OD	383	214	169

Table 3.1: Details of the fundus image datasets

two different fundus cameras. In order to handle the wide variations in image resolution, all the images were preprocessed to a common format using the same process as described in 3.3.2.

The combined dataset consisting of 2168 images was randomly split into training and testing sets in our experiments. Following the conventional method, diverse datasets were combined to ensure the proposed model is unbiased. The distribution of train, validation and test sets for the capsule network, CGAN and Fefu is provided in Table 3.3.

	Normal (Train / Test)	Glaucoma (Train / Test)
Private-1	430 / 299	715 / 112
Drishti-GS1	20 / 11	40 / 30
RIM-ONE v3	50 / 28	20 / 13
REFUGE	300 / 60	25 / 15
Total	800 / 398	800 / 170

Table 3.2: Decomposition of the combined set used for the Fefu model's training/testing in terms of the individual datasets.

3.5 Implementation details

The mapping architecture (CGAN) consists of 1,604,639 parameters and was trained for 700 epochs using Adam optimizer with a batch size of 100, learning rate of 0.001 and decaying at the rate of 0.8. F_f and F_{f-o} have 20,040,866 parameters each and the FCN network consists of 559,874 parameters. The train set was passed through the CGAN to obtain their respective soft-labels (32D features) for the

F_{f-o} module of the Fefu model. The Fefu model was trained for 100 epochs using SGD optimizer with learning rate of 0.001, momentum of 0.9, decay rate of 1e-8 and a batch size of 32 with 5-fold cross validation. Quantitative evaluation for all the stages was done using the following metrics: Balanced Accuracy (BAcc.), Sensitivity (Sens.), Specificity (Spec.), and AUC.

$$Sensitivity = \frac{TP}{TP + FN} \qquad Specificity = \frac{TN}{TN + FP}$$

$$BAccuracy = \frac{Sens. + Spec.}{2}$$

where TP, TN, FP and FN denote the number of true positives, true negatives, false positives and false negatives, respectively. Along with these, ROC plots were also derived. The optimum threshold was calculated by using the Youden index ([48]), where the sum of the Specificity-1 and Sensitivity is maximum. All frameworks were implemented on Keras with Tensorflow backend on a NVIDIA GTX 1080 GPU, with 11GB of GPU RAM.

	Train (N / G)	**Test (N / G)**
Mapping network	1278 (605/673)	212 (101/111)
Fefu model	1440 (800/800)	568 (398/170)

Table 3.3: Distribution of datasets for training/testing/validation of the different modules. **N** and **G** denote the number of normal and glaucomatous samples.

3.6 Experiments and Results

The proposed system was assessed by conducting various experiments. These are described in the following sections.

The correlation between the OCT-features and mapped features was also assessed with a set of 112 paired OCT and fundus images were used to derive the OCT features (from the capsule network) and the mapped fundus features ($f - o$). To visually assess the correlation, t-SNE was used to convert the 32D feature vectors to 2D which are plotted in Fig. 3.5. It can be observed that the the healthy and glaucoma classes can be differentiated reasonably well in the mapped feature space.

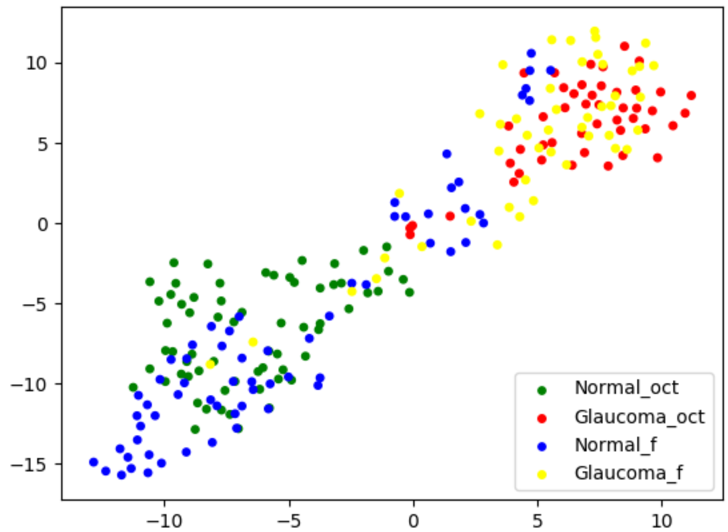


Figure 3.5: t-SNE visualization of features. The red and green represent the OCT-features. The blue and yellow represent features obtained from the mapped network (F_{f-o})

3.6.1 Performance on the combined dataset

A 5-fold cross validation was performed to assess the Fefu model in Fig. 3.3 using the combined test set consisting of 568 images (328 normal and 170 glaucomatous). The average value for each of the metrics is reported in Table 3.4 and the corresponding ROC curves are shown in Fig. 3.6a. The individual contribution of the two networks F_{f-o} and F_f was also assessed to gauge the level of contributions of the Fundus (f) only and Fundus to OCT ($f-o$) mapped features. The two networks were trained for 50 epochs individually on the combined training set for this study. In both these cases, 5-fold cross validation was performed. The tabulated results indicate the improvement in AUC (with f as baseline), is 4% with mapped features ($f-o$) and 6% with combined features. Hence, it can be concluded that mapped features are superior to unmapped features for classification of an image as being normal or glaucomatous and combining features is beneficial.

	BAcc.	**Sens.**	**Spec.**	**AUC**
Only F_f	0.8252	0.8573	0.7930	0.8882
Only F_{f-o}	0.8661	0.8809	0.8514	0.9240
Fefu model (Proposed)	**0.8909**	**0.9044**	**0.8773**	**0.9429**

Table 3.4: Performance of the proposed Fefu model on the combined test set

3.6.2 Performance on individual datasets

The results reported in Table 3.5 are based on detailed analysis of the performance of the Fefu model and its modules on individual datasets that constituted the combined test dataset. The Sens. and Spec. figures are reported (as before) at the optimum threshold. Overall, it can be noted that the Fefu model has consistently high performance ($0.85 \leq BAcc. \leq 0.96$; $0.91 \leq AUC \leq 0.99$) on all datasets which is much higher than that of F_f ($0.76 \leq BAcc. \leq 0.85$; $0.82 \leq AUC \leq 0.90$). It is also worth noting that there is a 5-7% improvement in the performance of F_{f-o} relative to F_f for all the datasets, implying a superiority of the Fundus-OCT feature space over the fundus-feature space for glaucoma assessment.

	Private-1 (N = 299, G = 112)				Drishti-GS1 (N = 11, G = 30)			
	B Acc.	**Sens.**	**Spec.**	**AUC**	**B Acc.**	**Sens.**	**Spec.**	**AUC**
Only F_f	0.7578	0.7649	0.7507	0.8204	0.8258	0.8333	0.8182	0.9030
Only F_{f-o}	0.8021	0.8175	0.7867	0.8633	0.8997	**0.9267**	0.8698	0.9473
Fefu model	**0.8511**	**0.8860**	**0.8162**	**0.9067**	**0.9167**	0.8333	**1.0000**	**0.9667**
	RIM-ONE v3 (N = 28, G = 13)				REFUGE (N = 60, G = 15)			
	B Acc.	**Sens.**	**Spec.**	**AUC**	**B Acc.**	**Sens.**	**Spec.**	**AUC**
Only F_f	0.8489	0.7692	0.9286	0.9011	0.8333	0.8000	0.8667	0.8878
Only F_{f-o}	0.9466	0.9077	**0.9857**	0.9664	0.9250	**0.8800**	0.9700	**0.9346**
Fefu model	**0.9643**	**1.0000**	0.9286	**0.9945**	**0.9333**	0.8667	**1.0000**	0.9100

Table 3.5: Performance of the proposed Fefu model on the individual test sets. **N** and **G** denote the number of normal and glaucomatous samples, respectively.

3.6.3 Comparison with state-of-the-art method

The proposed system was cross validated and compared with DE-Net [37], a state-of-the-art method. DE-Net derives the classification decision by combining the predictions of four independent neural network streams using ResNet-50. The pre-trained weights for the DE-Net available online [49] were used in this experiment.

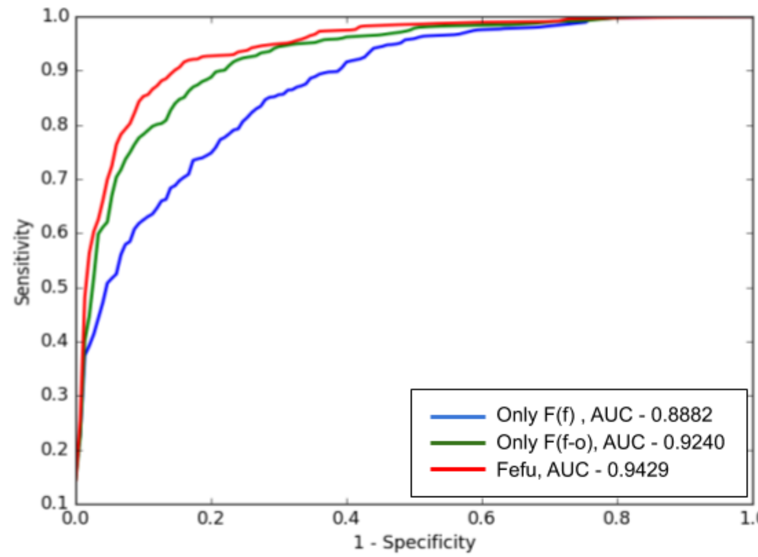
We begin with cross validation results of the DE-Net. The $BAcc/AUC$ was found to be 0.67/0.72 for Private-1, 0.83/0.85 for Drishti-GS1, 0.72/0.78 for RIM-ONE v3 and 0.76/0.83 for REFUGE datasets. These cannot be directly compared with the figures in Table 3.5 as the Fefu model was trained on a combined set generated by splitting these datasets. Nevertheless, the figures exhibit a downward trend. For a fairer comparison, two independent (i.e. neither were part of the combined set used for training the proposed networks.) datasets were chosen for cross validation: a public (sjchoi86-HRF) dataset and a locally sourced one named as Private-2. The obtained results presented in Table 3.6 show that the proposed model outperforms the DE-Net on both the unseen datasets. Specifically, in terms of AUC, the difference is 18.6% on the Private-2. Relative to this, the proposed model's performance on sjchoi86-HRF is lower perhaps due to the fact that nearly 80% of the training images for our model were OD-centric, whereas the sjchoi86-HRF has only macula-centric images. However, our model's AUC is still 8% higher than that of DE-Net.

	Private-2 (N = 214, G = 169)				sjchoi86-HRF (N = 300, G = 101)			
	B Acc.	**Sens.**	**Spec.**	**AUC**	**B Acc.**	**Sens.**	**Spec.**	**AUC**
Only F_f	0.7675	0.7219	0.8131	0.8356	0.6840	0.6700	0.6980	0.7290
Only F_{f-o}	0.8323	0.8047	0.8598	0.9048	0.7358	0.7400	0.7315	0.7841
Fefu model	**0.8707**	**0.8817**	**0.8598**	**0.9162**	**0.7558**	**0.7700**	**0.7416**	**0.8089**
DE-Net [37]	0.7182	0.6607	0.7757	0.7724	0.6997	0.7327	0.6667	0.7490
MNet [35]	0.3655	0.3495	0.3846	-	0.6992	0.7018	0.6915	-

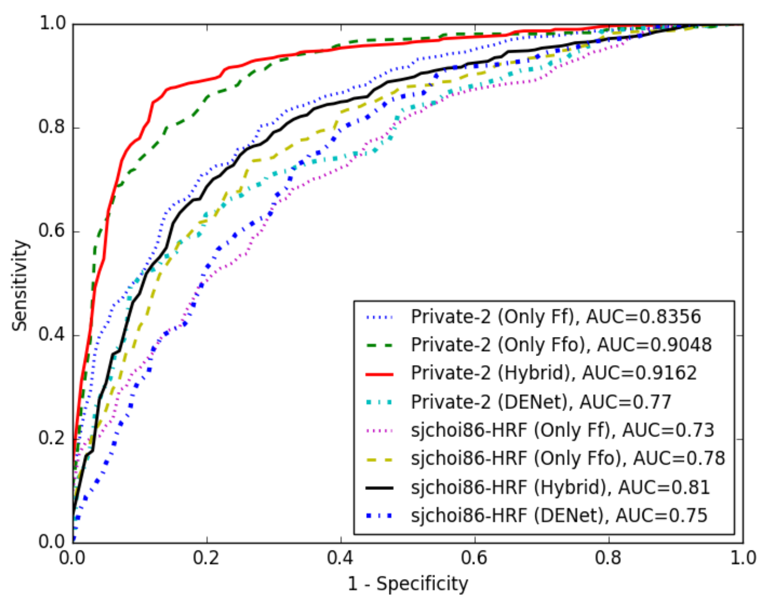
Table 3.6: Cross-validation performance comparison of different networks on the individual test sets.

It is also noteworthy that the AUC values obtained for F_{f-o} is 17.1% and 4.6% higher than DENet on the Private-2 and sjchoi86-HRF. This underscores the superiority of the OCT-Fundus feature space over feature space obtained using only fundus images.

Our model was also compared against CDR -based classification as it is popular for glaucoma. The pre-trained MNet [35] model (available online) was used to derive the CDR. Next, the International Classification of Diseases (ICD-9) rule on CDR was used to classify the images into 2 classes: normal ($0 \leq CDR < 0.5$) and glaucomatous ($0.5 \leq CDR \leq 1$). The results reported in Table 3.6 indicate our model to be superior.



(a) ROC curves along with AUC comparing the performance of different networks on the combined test set corresponding to Table 3.4



(b) ROC curves for cross-validation performance of different networks, corresponding to the first 2 rows of Table 3.6

Figure 3.6: ROC plots of different experiments performed

Besides the DE-Net, few other deep neural network architectures have been reported for glaucoma assessment from Fundus images. Table 3.7 lists the performance figures for the most recent methods for completeness of information on the prevailing scenario. A direct comparison with these methods is not feasible since the datasets, models/parameter settings are unavailable.

Method	**Dataset type**	**BAcc**	**Sens.**	**Spec.**	**AUC**
[25]	Private	–	–	–	0.88
[37]	$Private_a$ (train set)	–	–	–	–
	$Private_b$	0.8429	0.8478	0.8380	0.9183
	$Private_c$	0.7495	0.7876	0.7115	0.8173
[28]	Various public and private	0.8138	0.8290	0.7986	0.9017
[29]	2 public and one private	0.8801	0.8701	0.8901	0.9420
Proposed	One private + 3 public	0.8909	**0.9044**	0.8773	**0.9429**

Table 3.7: Performance figures reported in literature by recent methods.

Class Activation Mapping (CAM) is a visualization technique that allows us to understand what networks look at when they classify an input image, thus making it possible to visually understand why the network chooses one class over another. More recently, improving upon CAM, Gradient-weighted Class Activation Mapping (Grad-CAM) [50] is introduced. In particular, Grad-CAM uses class specific gradient information flowing into the final convolutional layer of the network to produce a coarse localization map of the important regions in a given input image. Grad-CAMs were computed for F_f and F_{f-o} networks, resized and overlayed on the input fundus, to identify the regions that the networks sees to be important for the classification decision. Fig. 3.7 - 3.14 depict the Grad-CAMs for few glaucoma samples from Private-2 and sjchoi86-HRF sets. As illustrated in the localization maps of the correctly predicted glaucoma samples, it can be observed from Fig. 3.7 and 3.9 that F_f seems to identify the cup and superior and inferior RNFL regions, whereas in case of F_{f-o} it is the disc region (Fig. 3.11 and 3.13). These regions correspond well to the common biomarkers for glaucoma diagnosis. The localization of the wrongly predicted samples are displayed in Fig. 3.8, 3.10, 3.12, 3.14, where insignificant regions are highlighted.

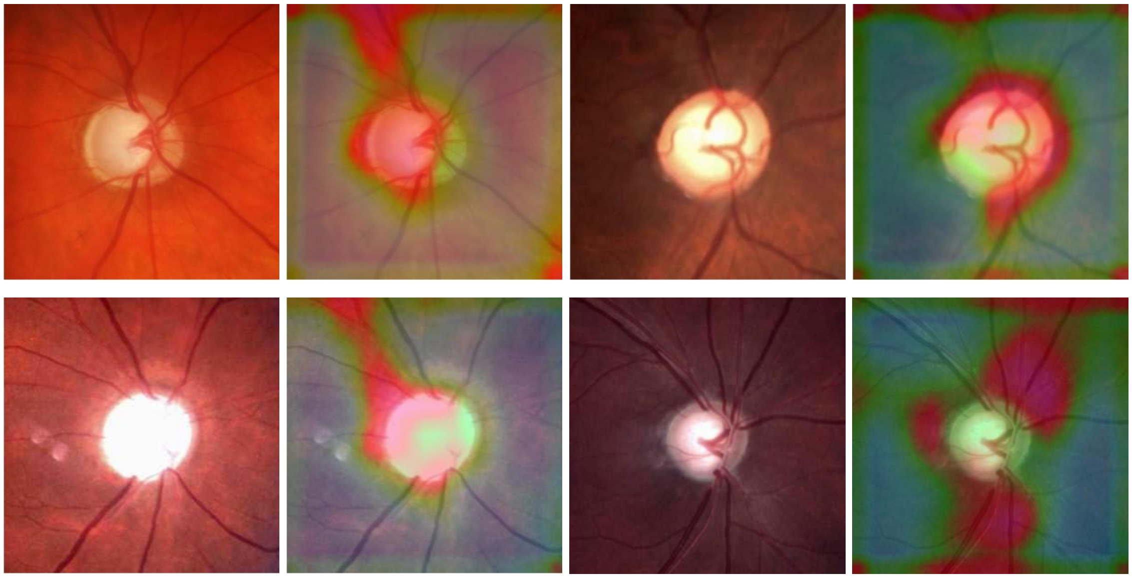


Figure 3.7: Grad-CAM visualization of glaucoma samples from Private-2 set that are **correctly** predicted by F_f

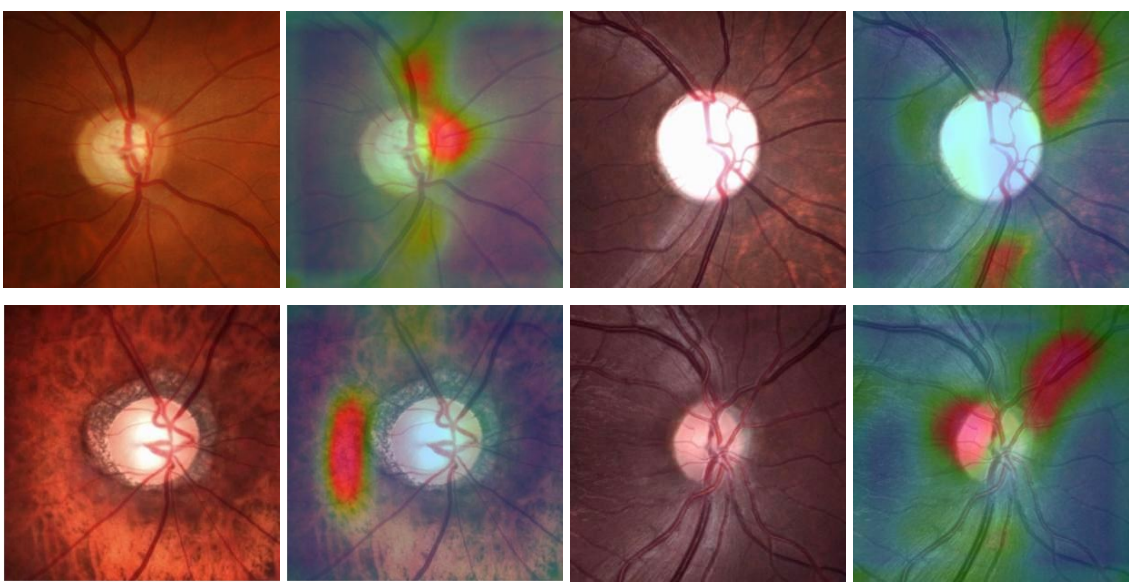


Figure 3.8: Grad-CAM visualization of glaucoma samples from Private-2 set that are **wrongly** predicted by F_f

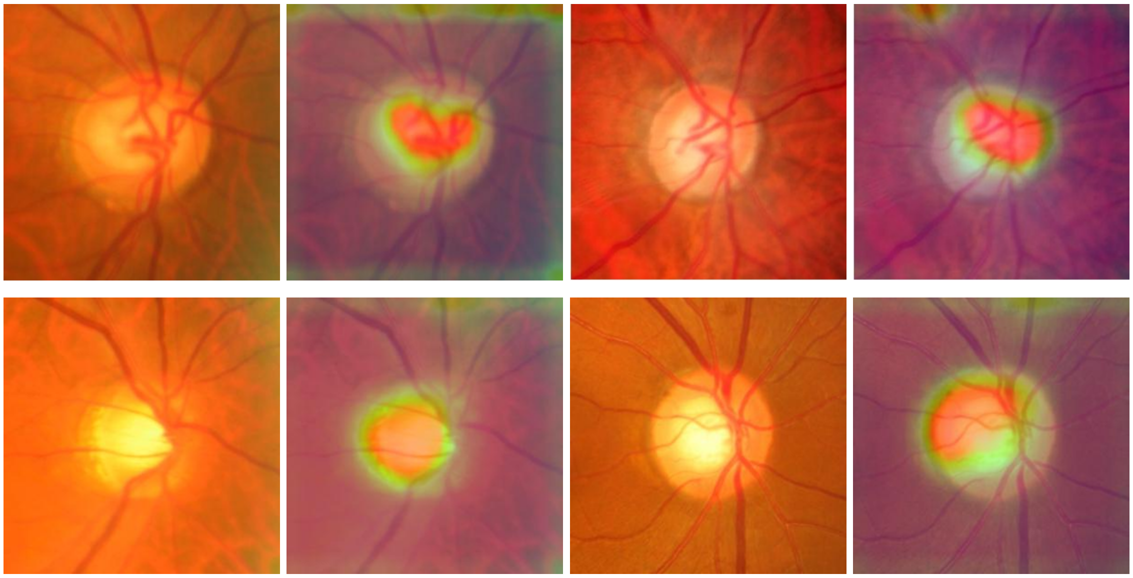


Figure 3.9: Grad-CAM visualization of glaucoma samples from sjchoi86-HRF set that are **correctly** predicted by F_f

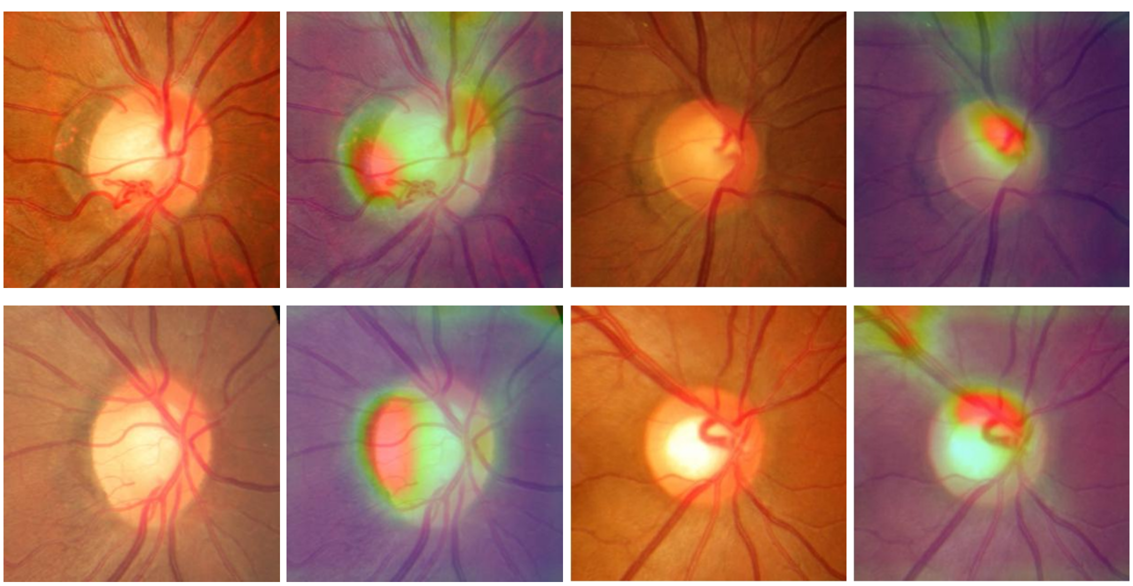


Figure 3.10: Grad-CAM visualization of glaucoma samples from sjchoi86-HRF set that are **wrongly** predicted by F_f

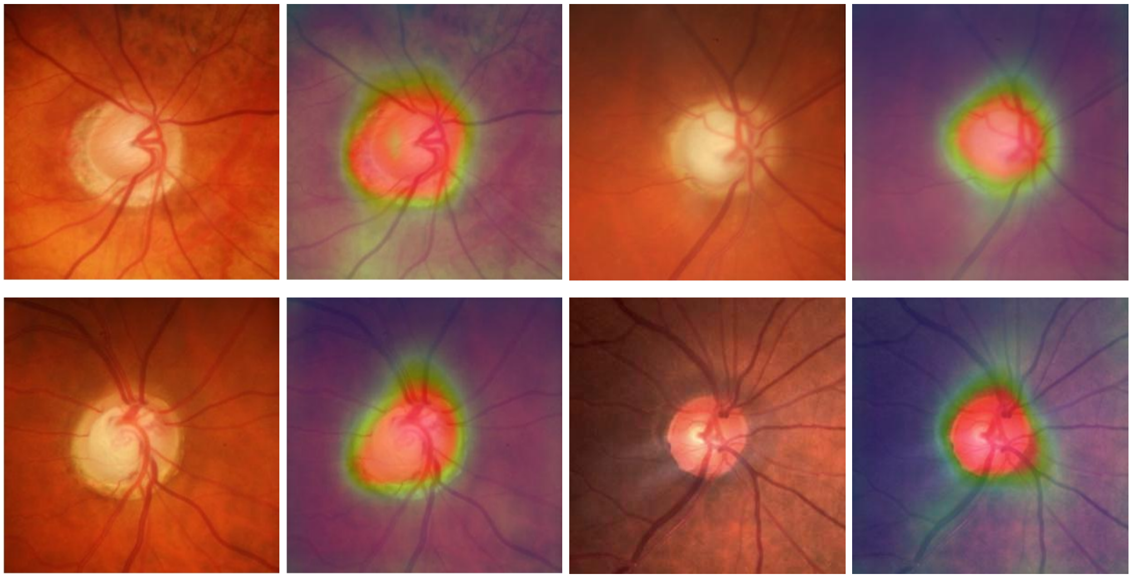


Figure 3.11: Grad-CAM visualization of glaucoma samples from Private-2 set that are **correctly** predicted by F_{f-o}

Figure 3.12: Grad-CAM visualization of glaucoma samples from Private-2 set that are **wrongly** predicted by F_{f-o}

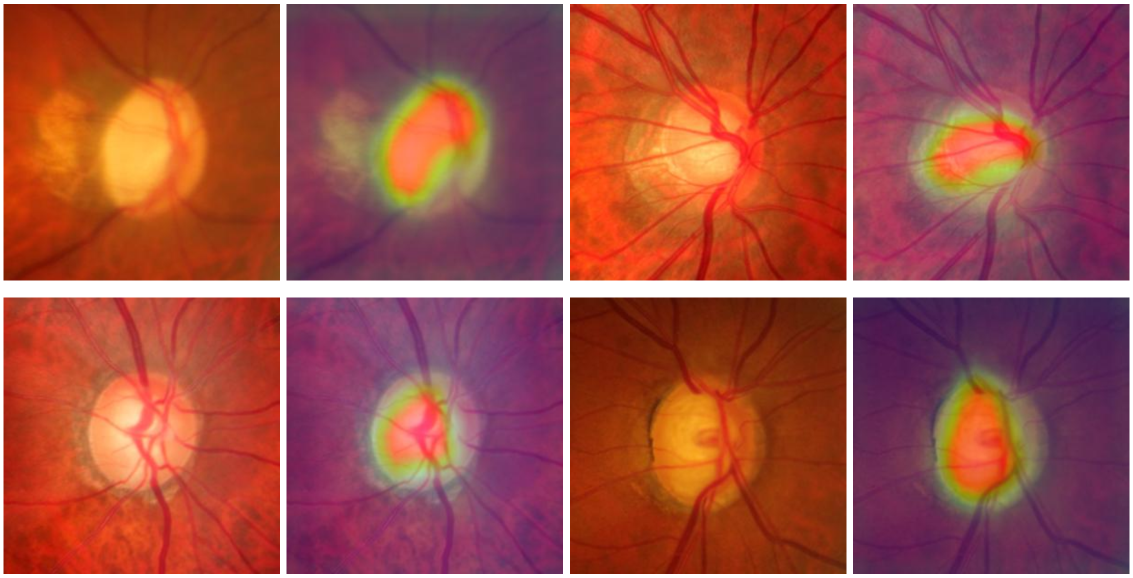


Figure 3.13: Grad-CAM visualization of glaucoma samples from sjchoi86-HRF set that are **correctly** predicted by F_{f-o}

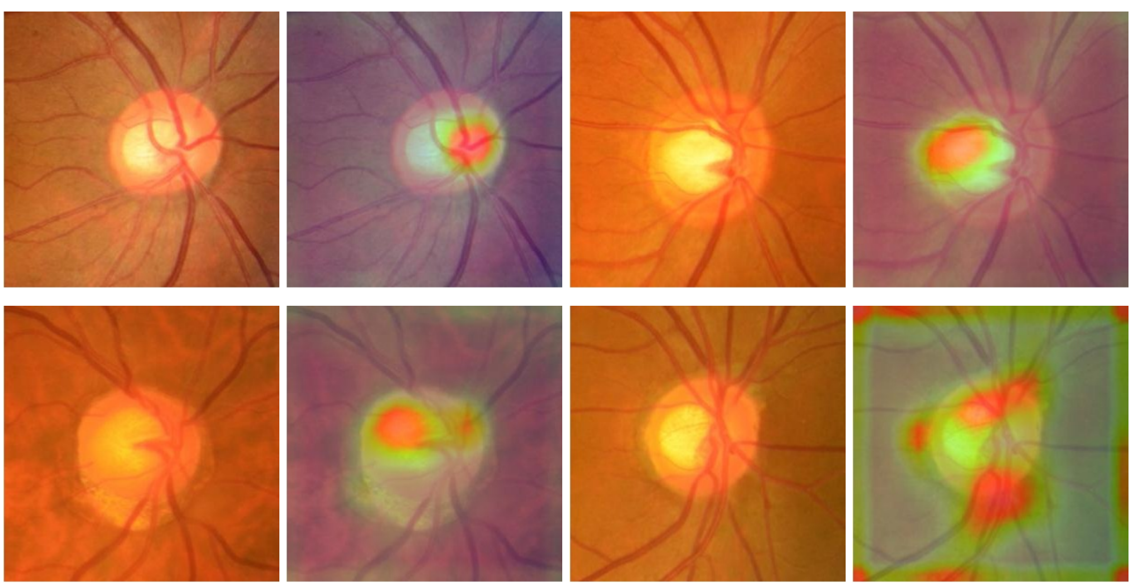


Figure 3.14: Grad-CAM visualization of glaucoma samples from sjchoi86-HRF set that are **wrongly** predicted by F_{f-o}

3.7 Conclusions

We presented a solution for glaucoma assessment which utilises unpaired data for mapping fundus to OCT feature space, hence bridging the gap between the two modalities. It is noteworthy from the results that the fundus-OCT feature space has more discriminative ability compared to the fundus-feature space, which was further boosted on the combination of both. Another important point to note is that a strong and consistent (across datasets) performance can be obtained with a lighter network with fewer parameters by using such mapping approach and that doesn't rely on segmentation. This is in contrast to existing methods (that rely only on fundus images) which employ multiple networks and large number of parameters. Smaller networks are attractive in screening scenarios as the online computing can then be done on the imaging device (such as a smartphone-based fundus imaging system) itself. The proposed idea was implemented largely with conventional networks (capsule, CGAN, VGG, etc.) without any need for segmentation of structures. Hence, in the future, other custom designs for the networks can be explored to implement the proposed ideas as well.

Chapter 4

Conclusions and Future Work

This thesis explores methods for automatic detection of glaucoma using two widely used retinal imaging techniques, OCT and Fundus. First, we demonstrate the efficacy of the OCT volumes in glaucoma detection using Capsule network. The OCT feature space obtained from this network is further used such that fundus images are mapped to this space, thus bridging the gap between the two modalities. Next, a novel CAD solution is proposed integrating information from OCT and fundus modalities to provide a credible diagnosis.

We demonstrated that glaucoma detection from OCT volumes directly is attainable with a performance comparable or better than the existing approaches that rely on hand-crafted features extracted after segmentation of the required structures from OCT. We performed experiments extensively and showed that our model achieves high performance when compared to other deep learning architectures even with less data. In one of the experiments, we also demonstrated that both optic nerve head (ONH), providing disc and cup information and peripapillary region, providing RNFL thickness information are required for accurate detection of glaucoma, which is on par with the clinical experts. Thus, our deep learning model learned highly discriminative features from OCT and showed promising detection results. Such a feature-agnostic method has a potential to widen the range of applications. Although OCT showed superior performance when compared to many of the existing fundus-based screening methods, it may not be used in a practical setting due to its highly expensive acquisition process.

Although OCT showed superior performance when compared to many of the existing fundus-based screening methods, it may not be used in a practical setting due to its highly expensive acquisition process whereas fundus images are cost-effective and thus, widely available. We try to address this problem of OCT in such a manner that the information from OCT is used to assist the fundus-based screening system. Since, both modalities provide complementary information, with one representing the retinal surface while the other its cross section, integrating both in a feature space is possible which could potentially improve the detection accuracy. Thus, we proposed a network to perform an adversarial-learning based mapping between features derived from labelled fundus images and the OCT features. Finally, we designed and validated a deep learning classification network that fuses features and performs the glaucoma assessment when only one modality (fundus) was provided. This is a major advantage because our model learns information from multiple sources and provides the screening result using only fundus, since OCT cannot be easily acquired. Thus we are leveraging information from OCT in such a way that it is not required for providing the diagnosis in a practical setting thus enabling the screening process to be carried out on a large scale. We also provided the activation maps that display the important regions for the model, which correspond closely to the common biomarkers for glaucoma diagnosis.

Future Work

Although, the thesis provides promising results which are competitive with the existing techniques and the mapping of fundus to OCT feature space does offer an attractive option and provided some encouraging results for improving glaucoma assessment, this is still a preliminary work and there is scope for improvement in many areas.

- Firstly, we propose a glaucoma screening model using the OCT volumes directly. The algorithm performance can be improved by obtaining and experimenting with more data. Experiments with OCT data from different scanners can also be included and the performance can be validated. Further study of the model can be done by visualizing it to see which areas of OCT are highlighted which could provide scope for improving the model accordingly.
- A mapping model was proposed to map fundus images to the OCT feature space. For this purpose, CGAN with auxiliary network was trained. Various other adversarial-learning methods can be explored for improving the mapping mechanism. Instead of using OCT features generated from the Capsule network at the discriminator of the proposed CGAN, it would be interesting to explore and develop a model which can directly integrate 2D (fundus) and 3D (OCT) information.
- The proposed Fefu model for glaucoma detection is trained a diverse fundus dataset, majorly consisting of OD-centric and few macula-centric images. More macula-centric fundus images can be acquired and included in the training stage so that the model is more robust to the view-point. The Fefu model comprises of two VGG-19 model, hence, other deep learning architectures can be explored and their effect on the screening performance can be studied.
- Since, smaller networks are attractive in screening scenarios as the online computing can then be done on the imaging device itself, other smaller network architectures can be explored.
- We proposed a two-stage network: the first stage maps the fundus images to the OCT feature space and the second stage classifies the fundus images into healthy and glaucomatous. This can be further improved to build a single end-to-end model

Publications

List of Publications

- **Glaucoma Assessment from OCT using a Capsule Network**
  Divya Jyothi Gaddipati, Alakh Desai, Jayanthi Sivaswamy, Koenraad A. Vermeer
  41st IEEE Engineering in Medicine and Biology Conference (EMBC), 2019.

- **Glaucoma Assessment from Fundus Images using Fundus to OCT Feature Space Mapping**
  Divya Jyothi Gaddipati and Jayanthi Sivaswamy
  IEEE Journal of Biomedical and Health Informatics (JBHI), 2019. In review

Other Related Publications

- **Construction of a Retinal Atlas for Macular OCT Volumes**
  Arunava Chakravarty, Divya Jyothi Gaddipati, Jayanthi Sivaswamy
  International Conference Image Analysis and Recognition (ICIAR), 2018.

Bibliography

[1] A. A. Salam, T. Khalil, M. U. Akram, A. Jameel, and I. Basit. Automated detection of glaucoma using structural and non structural features. *SpringerPlus*, 5(1), 2016.

[2] A. S. Reis, N. O'Leary, H. Yang, G. P. Sharpe, M. T. Nicolela, C. F. Burgoyne, and B. C. Chauhan. Influence of clinically invisible, but optical coherence tomography detected, optic disc margin anatomy on neuroretinal rim evaluation. *Investigative Ophthalmology and Visual Science*, 53(4): 1852–60, 2012.

[3] J. Kotowski, G. Wollstein, L. S. Folio, H. Ishikawa, and J. S. Schuman. Clinical use of oct in assessing glaucoma progression. *Ophthalmic surgery, lasers and imaging*, 42:S6–S14, 2011.

[4] A. Rajan and G. P. Ramesh. Automated early detection of glaucoma in wavelet domain using optical coherence tomography images. *Biomedical Pharmacol Journal*, 12:2821–8, 2015.

[5] T. R. Ganesh, S. Devi, and R. Venkatesh. Optic nerve head segmentation using fundus images and optical coherence tomography images for glaucoma detection. *Biomedical papers of the Medical Faculty of the University Palacky*, 159:607–15, 2015.

[6] F. R. Silva, V. G. Vidotti, F. Cremasco, M. Dias, E. S. Gomi, and V. P. Costa. Sensitivity and specificity of machine learning classifiers for glaucoma diagnosis using spectral domain oct and standard automated perimetry. *Arquivos brasileiros de oftalmologia*, 76:170–4, 2013.

[7] J. S. Kim, H. Ishikawa, K. R. Sung, J. Xu, G. Wollstein, R. A. Bilonick, M. L. Gabriele, J. S. Kagemann, L. Duker, J. G. Fujimoto, and J. S. Schuman. Retinal nerve fibre layer thickness measurement reproducibility improved with spectral domain optical coherence tomography. *British Journal of Ophthalmology*, 93:1057–63, 2009.

[8] J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, and et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, 24:1342–1350, 2018.

[9] S. Sabour, N. Frosst, and G. E. Hinton. Dynamic routing between capsules. *Advances in Neural Information Processing Systems (NIPS)*, pages 3856–66, 2017.

[10] A. Jimenez-Sanchez, S. Albarqouni, and D. Mateus. Capsule networks against medical imaging data challenges. In *CVII-STENT/LABELS@MICCAI*, 2018.

[11] P. Afshar, A. Mohammadi, and K. N. Plataniotis. Brain tumor type classification via capsule networks. *25th IEEE International Conference on Image Processing (ICIP)*, pages 3129–3133, 2018.

[12] Y. Yu and S. T. Acton. Speckle reducing anisotropic diffusion. *IEEE Transactions on Image Processing*, 11:1260–70, 2002.

[13] A. Chakravarty, D. J. Gaddipati, and J. Sivaswamy. Construction of a retinal atlas for macular oct volumes. *15th International Conference on Image Analysis and Recognition*, pages 650–8, 2018.

[14] K. Gopinath and J. Sivaswamy. Segmentation of retinal cysts from optical coherence tomography volumes via selective enhancement. *IEEE Journal of Biomedical and Health Informatics (JBHI)*, 23:273–282, 2019.

[15] Y. Wen, K. Zhang, Z. Li, and Y. Qiao. A discriminative feature learning approach for deep face recognition. pages 499–515, 2016.

[16] L. Maaten and G. E. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, pages 2579–2605, 2008.

[17] S. Maheshwari, R. B. Pachori, and U. R. Acharya. Automated diagnosis of glaucoma using empirical wavelet transform and correntropy features extracted from fundus images. *IEEE Journal of Biomedical and Health Informatics (JBHI)*, 21:803–13, 2017.

[18] S. Maheshwari, R. B. Pachori, V. Kanhangad, S. V. Bhandary, and U. R. Acharya. Iterative variational mode decomposition based automated detection of glaucoma using fundus images. *Computers in Biology and Medicine*, 88:142–9, 2017.

[19] S. Dua, U. R. Acharya, P. Chowriappa, and S. V. Sree. Wavelet-based energy features for glaucomatous image classification. *IEEE Transactions on Information Technology in Biomedicine*, 16: 80–87, 2012.

[20] J. Odstrcilik, R. Kolar, R. P. Tornow, J. Jan, A. Budai, M. Mayer, M. Vodakova, and et al. Thickness related textural properties of retinal nerve fiber layer in color fundus images. *Computerized Medical Imaging and Graphics*, 38:508–516, 2014.

[21] M. A. Ali, T. Hyrtut, T. Faucon, and F. Cheriet. Glaucoma detection based on local binary patterns in fundus photographs. *Proceedings of SPIE Medical Imaging*, 2014.

[22] A. Septiarini, D. M. Khairina, A. H. Kridalaksana, and H. Hamdani. Automatic glaucoma detection method applying a statistical approach to fundus images. *Healthcare informatics research*, 24:53–60, 2018.

[23] X. Chen, Y. Xu, S. Yan, D. W. K. Wong, T. Y. Wong, and J. Liu. Automatic feature learning for glaucoma detection based on deep learning. *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 669–677, 2015.

[24] J. I. Orlando, E. Prokofyeva, M. del Fresno, and M. Blaschko. Convolutional neural network transfer for automated glaucoma identification. *Proc. SPIE, 12th International Symposium on Medical Information Processing and Analysis*, 2016.

[25] X. Chen, Y. Xu, D. W. K. Wong, T. Y. Wong, and J. Liu. Glaucoma detection based on deep convolutional neural network. *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 715–718, 2015.

[26] A. Pal, M. R. Moorthy, and A. Shahina. G-eyenet: A convolutional autoencoding classifier framework for the detection of glaucoma from retinal fundus images. *25th IEEE International Conference on Image Processing (ICIP)*, pages 2775–2779, 2018.

[27] M. Cristopher, A. Belghith, C. Bowd, J. A. Proudfoot, M. H. Goldbaum, R. N. Weinreb, C. A. Girkin, J. M. Liebmann, and L. M. Zangwill. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. *Scientific Reports*, 8, 2018.

[28] A. Diaz-Pinto, A. Colomer, V. Naranjo, S. Morales, Y. Xu, and A. F. Frangi. Retinal image synthesis and semi-supervised learning for glaucoma assessment. *IEEE Transactions on Medical Imaging (TMI)*, 2019.

[29] J. Gomez-Valverde, A. Anton, G. Fatti, B. Liefers, A. Herranz, A. Santos, C. Sanchez, and M. Ledesma-Carbayo. Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning. *Biomedical Optics Express*, pages 892–913, 2019.

[30] J. Cheng, J. Liu, D. Tao, F. Yin, D. W. K. Wong, Y. Xu, and T. Y. Wong. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. *IEEE Transactions on Medical Imaging (TMI)*, 32:1019–1032, 2013.

[31] N. M. Tan, Y. Xu, B. Goh, and J. Liu. Robust multi-scale superpixel classification for optic cup localization, 2015.

[32] Y. Xu, L. Duan, S. Lin, X. Chen, D. W. K. Wong, T. Y. Wong, and J. Liu. Optic cup segmentation for glaucoma detection using low-rank superpixel representation. *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 17:788–795, 2014.

[33] A. Chakravarty and J. Sivaswamy. Joint optic disc and cup boundary extraction from monocular fundus images. 147:51–61, 2017.

[34] A. Sevastopolsky. Optic disc and cup segmentation methods for glaucoma detection with modification of u-net convolutional neural network. 27:618–624, 2017.

[35] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Transactions on Medical Imaging (TMI)*, 37:1597–1605, 2018.

[36] A. Chakravarty and J. Sivaswamy. Glaucoma classification with a fusion of segmentation and image-based features. *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 689–692, 2016.

[37] H. Fu, J. Cheng, Y. Xu, C. Zhang, D. W. K. Wong, J. Liu, and X. Cao. Disc-aware ensemble network for glaucoma screening from fundus image. *IEEE Transactions on Medical Imaging (TMI)*, 37:2493–2501, 2018.

[38] D. W. Kim, J. W. Jeoung, Y. W. Kim, M. J. A. Girard, J. M. Mari, Y. K. Kim, K. H. Park, and D. M. Kim. Prelamina and lamina cribrosa in glaucoma patients with unilateral visual field loss. *Investigative Ophthalmology and Visual Science*, 57:1662–1670, 2016.

[39] E. Tsikata, R. Lee, E. Shieh, H. Simavli, C. J. Que, R. Guo, Z. Khoueir, J. de Boer, and T. C. Chen. Comprehensive three-dimensional analysis of the neuroretinal rim in glaucoma using high-density spectral-domain optical coherence tomography volume scans. *Investigative Ophthalmology and Visual Science*, 57:5498–508, 2016.

[40] M. Foracchia, E. Grisan, and A. Ruggeri. Luminosity and contrast normalization in retinal images. *Medical Image Analysis*, 9:179–190, 2005.

[41] G. D. Joshi, R. Gautam, J. Sivaswamy, and S. Krishnadas. Robust optic disk segmentation from color retinal images. *Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP)*, pages 330–336, 2010.

[42] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. *International Conference on Neural Information Processing Systems (NIPS)*, pages 2672–2680, 2014.

[43] M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[44] J. Sivaswamy, S. R. Krishnadas, G. D. Joshi, M. Jain, and A. U. S. Tabish. Drishti-gs: Retinal image dataset for optic nerve head(onh) segmentation. *IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pages 53–56, 2014.

[45] F. Fumero, S. Alayon, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez. Rim-one: An open retinal image database for optic nerve evaluation. *24th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 1–6, 2011.

[46] REFUGE. Retinal fundus glaucoma challenge. https://refuge.grand-challenge.org/Home/, 2018.

[47] Sungjoon. sjchoi86-hrf. https://github.com/yiweichen04/retina_dataset, 2016. [Accessed: March-2019].

[48] R. Fluss, D. Faraggi, and B. Reiser. Estimation of the youden index and its associated cutoff point. *Biometrical Journal*, 47:458–472, 2005.

[49] H. Fu and et al. Code for disc-aware ensemble network for glaucoma screening from fundus image. https://github.com/HzFu/DENet_GlaucomaScreen, 2018. [Accessed 2018].

[50] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. *IEEE International Conference on Computer Vision (ICCV)*, pages 618–626, 2017.