Theses and Dissertations

Spring 2019

Estimation of volumetric optical coherence tomography measurements from 2D color fundus photographs using machine learning

Samuel Steven Johnson
University of Iowa

Recommended Citation
Johnson, Samuel Steven. "Estimation of volumetric optical coherence tomography measurements from 2D color fundus photographs using machine learning." MS (Master of Science) thesis, University of Iowa, 2019.
https://doi.org/10.17077/etd.nwx1-l79t

ESTIMATION OF VOLUMETRIC OPTICAL COHERENCE TOMOGRAPHY
MEASUREMENTS FROM 2D COLOR FUNDUS PHOTOGRAPHS USING
MACHINE LEARNING

by

Samuel Steven Johnson

A thesis submitted in partial fulfillment of the
requirements for the Master of Science degree
in Electrical and Computer Engineering
in the Graduate College of
The University of Iowa

May 2019

Thesis Supervisor: Associate Professor Mona K. Garvin

www.manaraa.com

ACKNOWLEDGEMENTS

for all of the time that they have taken out of their normal schedules to review papers, provide data details and diagnosis, as well as letting me attend meetings with their staff. Without your help, our research would not have near the success or impact that it does. You have both helped tremendously in all clinical aspects of our work. Shout out to all of my Iowa City friends who have helped me maintain some semblance of a social life, when possible. I am still not confident that our trivia skills have improved, but the nights spent debating random questions have produced lifelong memories that I will always cherish.

Finally, and most importantly, I want to acknowledge the unwavering support I've received from my family. My parents, Elaine and Steve, my brother, Doug, and my grandmother Marilyn Johnson have always provided tremendous encouragements in all of my academic endeavours. You have always pushed me to do my best, consoled me when things didn't go well, and provided advice when needed. None of this would have possible without you.

ABSTRACT

The optic nerve head is the location in the rear of the eye where the nerves exit the eye towards the brain. Swelling of the optic nerve head (ONH) is most accurately quantitatively assessed via volumetric measures using 3D spectral-domain optical coherence tomography (SD-OCT). However, SD-OCT is not always available as its use is primarily limited to specialized eye clinics rather than in primary care or telemedical settings. Thus, there is still a need for severity assessment using more widely available 2D fundus photographs.

In this work, we propose machine-learning methods to locally estimate the volumetric measurements (akin to those produced by 3D SD-OCT images) of the optic disc swelling at each pixel location from only a 2D fundus photograph as the input. For training purposes, a thickness map of the swelling (reflecting the distance between the top and bottom surfaces of the ONH and surrounding retina) as measured from SD-OCT at each pixel location was used as the ground truth. First, a random-forest classifier was trained to output each thickness value from local fundus features pertaining to textural and color information. Eighty-eight image pairs of ONH-centered SD-OCT and registered fundus photographs from different subjects with optic disc swelling were used for training and evaluating the model in a leave-one-subject-out fashion.

Comparing the thickness map from the proposed method to the ground truth via SD-OCT, a root-mean-square (RMS) error of 1.66 mm^3 for the entire ONH region was achieved, and Spearman's correlation coefficient was R= 0.73. Regional volumes for the nasal, temporal, inferior, superior, and peripapillary regions had RMS errors of 0.64 mm^3, 0.61 mm^3, 0.74 mm^3, 0.71 mm^3, and 1.30 mm^3, respectively, suggesting that there is enough evidence in a singular color fundus photograph to estimate local swelling information.

Because of the recent success of deep-learning methods in imaging domains, a convolutional neural network was also trained using the same data as was used with

the random forest classifier. Because training data is used to help fine tune model parameters for deep learning, a subset of ten randomly selected patients was strictly withheld from the training process to be used for testing. Comparing the prediction results on the withheld data with the OCT ground truth, we achieved a root-mean-square (RMS) error of 2.07 mm^3 for the entire ONH region. Regional volumes for the nasal, temporal, inferior, superior, and peripapillary regions had RMS errors of 0.75 mm^3, 0.82 mm^3, 0.85 mm^3, 0.91 mm^3, and 1.62 mm^3, respectively. Although the errors are slightly higher than those from the random forest model, the test dataset was smaller as we could not use a leave-patient-out validation approach and this may not be representative of the whole dataset since results were not averaged as before. It is also known that deep learning models require larger training datasets to achieve similar results to traditional machine-learning methods. For these reasons, and the fact that the errors were close to those of traditional methods, we believe deep learning approaches for estimating local retinal thickness in cases of optic disc swelling still holds promise with larger datasets.

Both of the proposed approaches allow for clinicians to assess optic nerve edema in both a qualitative and quantitative manner using strictly fundus photography. The predictions allow for overall optic nerve head volume to be calculated as well as regional and local volumes which was not possible before.

PUBLIC ABSTRACT

The optic nerve head is the area in the rear portion of the eye where the optic nerve departs the eye towards the brain. There are a number of known conditions which can cause this region to swell, known as optic disc edema, which can potentially result in uncomfortable symptoms for the patient such as headaches, nausea, and even vision loss. For this reason, it is imperative to be able to diagnose and assess disease severity and progression.

Current methods of assessing optic nerve edema are limited and often subjective or expensive. One of the most prominent methods for assessment showing great promise is optical coherence tomography (OCT), a non-invasive 3D imaging modality which allows for a more quantitative evaluation. With the introduction of associate layer segmentation methods for OCT, and especially segmentation for swollen cases, we are able to calculate a total volume measurement for the swollen optic nerve head region. This allows clinicians to assess severity on a continuous scale with little variability. However, OCT technology is not available in many settings (e.g. telemedical) yet so there still exists a need for this quantitative assessment from other methods.

Fundus photographs are standard two-dimensional color photographs of the back of the eye and have traditionally been the medium used for clinician assessment of many ophthalmic conditions. They are relatively low cost, readily available, and require minimal discomfort from the patient. Using machine learning techniques, fundus photography has been proven to contain enough information to predict the swollen volume of the optic disc region.

In this thesis work, a method of predicting retinal thickness at each pixel location in a fundus photograph is proposed. This will allow for calculation of not only the total swollen volume across the optic nerve head, but also regional volumes (such as strictly the temporal or nasal regions). It can also provide a thickness map for visual qualitative analysis similar to those acquired via OCT.

Two algorithms were tested in this work: a random forest regression model as well

as a convolutional neural network. Eighty-eight patients with both OCT and fundus photographs taken at the same visit date were used for training and testing both approaches. For the random forest approach we train as many models as there are patients, leaving one out each time to be used for testing and then all model results are averaged. For the neural network, we strictly withhold ten randomly selected subjects for testing.

Comparing the prediction results from the random forest model on the withheld data with the OCT ground truth we achieved a root-mean-square (RMS) error of 1.66 mm^3 for the entire ONH region. Regional volumes for the nasal, temporal, inferior, superior, and peripapillary regions had RMS errors of 0.64 mm^3, 0.61 mm^3, 0.74 mm^3, 0.71 mm^3, and 1.30 mm^3, respectively. In the same analysis for the neural network, a RMS error of 2.07 mm^3 was obtained for the total ONH region. Regional volumes for the nasal, temporal, inferior, superior, and peripapillary regions had RMS errors of 0.75 mm^3, 0.82 mm^3, 0.85 mm^3, 0.91 mm^3, and 1.62 mm^3, respectively

Although the neural network has a slightly higher error rate, it is known that it requires much more training data for a neural network to reach the same accuracy as a traditional machine learning algorithm such as random forest. These results are very encouraging as a low cost alternative to OCT that still provides clinicians with both a quantitative and qualitative analysis of optic disc edema.

TABLE OF CONTENTS

ix

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Swelling of the optic disc, often referred to as optic disc edema, is an ophthalmic condition where the disc around which the optic nerve joins with the rear of the eye appears enlarged and bloated [6]. Optic disc edema can be indicative of a number of serious underlying conditions relating to retinal pathology. Many of these conditions can have a variety of negative physical effects on the patient such as abnormal headaches, nauseousness, and even permanent vision loss so it is crucial to acquire an accurate diagnosis as early as possible [7].

Despite being the only nerve visible on the exterior of the human body, diagnosing conditions of the optic disc is difficult for even expertly trained clinicians. One common technique for assessing optic nerve edema is called the Frisén scale [8, 9]. This is a subjective, discrete scale from zero to five (zero being normal and five being extremely severe swelling) used on optic disc edema diagnosed as papilledema, which is any swelling of the optic disc due to raised intracranial pressure (ICP). A clinician will either examine the back of the eye directly or, more commonly, will use a color fundus photograph (i.e. a photo of the optic disc and surrounding area) to assess the swelling. Because this scale relies upon a subjective and mostly qualitative assessment by the clinician, there can be a high amount of inter/intra-observer variability [10].

Alternative assessment methods have emerged in recent years that allow for a quantitative assessment. At the forefront of this is optical coherence tomography (OCT) which is a 3D imaging modality using light waves to capture high resolution images of the eye [11]. However, this technology is more resource intensive than traditional methods of assessment using fundus photographs. It is also not widely available in settings like family care facilities or telemedical applications. Clearly a quantifiable, lower cost method of assessment via fundus photographs would provide great value to both patients and clinicians alike.

In response to this need, Echegaray et al. proposed a method of semi-automatically predicting the Frisén grade given a color fundus photograph [12]. This approach extracted a total of 24 features from the fundus photograph that attempted to mimic clinical Frisén grading criteria, namely vessel, peripapillary, and ONH related features. These features were then used in training a random forest classifier that would predict the Frisén grade for that image. Although this was a step in the right direction for automatic quantification of the optic disc edema, we are still constrained by the inherent ordinal nature of the Frisén scale itself.

The introduction of OCT and associated retinal layer segmentation has allowed for more precise quantification of the degree of optic disc swelling [11,13]. By segmenting the different layers of the retina, we can calculate a volume measure for the swollen area, referred to as the total retinal volume (TRV) [14]. This TRV is defined as the sum of differences between the internal limiting membrane (ILM) and retinal pigment epithelium (RPE) at each voxel location within a $6 \times 6\text{mm}^2$ area, centered around the ONH. It has been shown that there is a high correlation between the TRV and the ordinal Frisén scale grades [14]. This new approach for assessing optic disc swelling provides huge benefits when compared with the Frisén grading system, namely that it is a continuous scale instead of discrete which removes subjectivity that was introduced by a manual diagnosis from a clinician on a discrete scale. However, assessing optic nerve edema using this approach requires that the patients have access to expensive machinery capable of capturing these OCT images. As this equipment is not available in many settings, such as emergency and primary care facilities, there still exists a need to mimic the volumetric assessment of optic disc swelling provided via OCT using more readily available fundus photography.

Previous works by Agne et al. [15] have sought to extend Echegary's work into the realm of this continuous scale of TRV. His approach uses more widely available 2D fundus photographs to predict the TRV. Twenty-seven features similar to those

Echegary used were extracted from the fundus image and then used to train a random forest regression model to predict the TRV for a fundus image. However, this prediction is limited in that it only provides a single quantitative value to describe the entire image. It cannot provide any qualitative anaylsis, nor can it describe regional swelling severity which has been proven useful for differentiation of causes of optic disc swelling [16].

In this work we propose machine learning methods to estimate local thickness values at each pixel location in a color fundus photograph. These approaches allow for both quantitative and qualitative analysis (thickness map) of the swelling, and not only allows you to calculate the TRV but also regional volumes on the optic disc. Regional volumes can help more accurately indicate disease progression and severity as it has been shown that in some cases of optic nerve edema (e.g. papilledema) swelling can start on the nasal side before progressing to the temporal side when the swelling becomes more severe [16].

For this thesis, two separate methods were developed for achieving the above stated goal: locally estimating retinal thickness for optic nerve edema assessment. First, a random forest regression model using manually selected features akin to those ophthalmologists look for was trained and analyzed. Second, a convolutional neural network was trained using automatically extracted high level features and again analyzed.

1.1 Main contributions

The main contributions of this thesis are two unique machine learning regression models to predict retinal surface thickness at each pixel location in a color fundus photograph. A random forest ensemble where features are manually selected comprises the first model, and a deep learning approach using high level features automatically obtained is the second. Both approaches allow for qualitative and quantitative evaluation of the severity of optic disc edema using only more widely available color fundus

photography.

1.2   Thesis overview

The rest of this thesis is organized as follows.

- **Chapter 2** provides a clinical background on optic disc edema and outlines relevant prior technical work in the field of ophthalmic image analysis and thickness prediction. It also discusses techinical backgrounds of both random forest and neural networks.

- **Chapter 3** discusses preprocessing steps, implementation methods and details, and the resulting predictions made by the random forest regression model on the fundus images.

- **Chapter 4** presents a deep learning approach for predicting thickness across the fundus using the same data as the random forest regression model.

- **Chapter 5** offers closing remarks on the work presented here and describes potential avenues for future work.

CHAPTER 2
CLINICAL AND TECHNICAL BACKGROUND

2.1   Terminology

For convenience, acronyms that are used throughout this thesis are defined below:

- Optic nerve head (ONH): area, often referred to as a disc, where the optic nerve exits the back of the eye towards the brain

- Intracranial pressure (ICP): the pressure within the skull, and therefore within the brain tissue and cerebrospinal fluid

- Anterior ischemic optic neuropathy (AION): an ophthalmic condition involving interruption of blood flow to the anterior portion of the optic nerve, often presenting with sudden loss of vision

- Retinal pigment epithelium (RPE): outermost surface of the retina, consisting of multiple layers of pigmented hexagonal cells. Shown in green in Fig. 2.1.

- Internal limiting membrane (ILM): innermost surface of the retina which separates the retina from the clear gel-like substance which fills the interior of the eye, known as vitreous. Shown in red in Fig. 2.1.

- Cerebrospinal fluid (CSF): a clear fluid that is found within the brain.

- Convolutional Neural Network (CNN): a class of deep neural network, typically applied for applications of analyzing visual imagery

- Contrast limited adaptive histogram equalization (CLAHE): a traditional image processing technique used to improve contrast in images without over amplifying noise in near constant regions

- Root-mean-square error (RMS): a common measurement of the difference between values from an estimator and observed values

- Spectral-domain optical coherent tomography (SD-OCT, often referred to as just OCT): an imaging modality that produces high resolution three-dimensional images using low-coherence interferometry

- Total retinal volume (TRV): a common measurement of swollen optic discs, calculated by the difference between the ILM and RPE surfaces across the entire optic disc region

- Peripapillary regional volume (PRV): a common measurement of swollen optic discs, calculated by the difference between the ILM and RPE surfaces across only the peripapillary region (defined as the region inside a circle with radius 1.73 mm centered at the ONH)

Although the retina is made up of many different surfaces, for the purposes of this thesis we will only be concerned with the ILM (innermost surface) and RPE (outermost surface) as the thickness between them is what will be estimated. The RPE is made up of multiple layers, but for the purposes of this work we will refer strictly to the RPE as a single entity, though it is implied to mean the multi-layer structure. Fig. 2.1 provides an overview of the structure of all layers in the retina as well as their spatial relationship to the eye as a whole.

2.2 Optic disc swelling

The optic nerve head, also referred to as the optic disc, is the location in the rear of the eye where the nerves exit the eye towards the brain. Swelling of the optic disc can be caused by any one of a myriad of potentially serious underlying conditions. These include, but are not limited to anterior ischemic optic neuropathy (AION), papilledema, and pseudopapilledema. Despite presenting with slightly different symptoms, all of the conditions discussed here can present with some form of optic disc edema. Each of these conditions can also have lasting consequences if
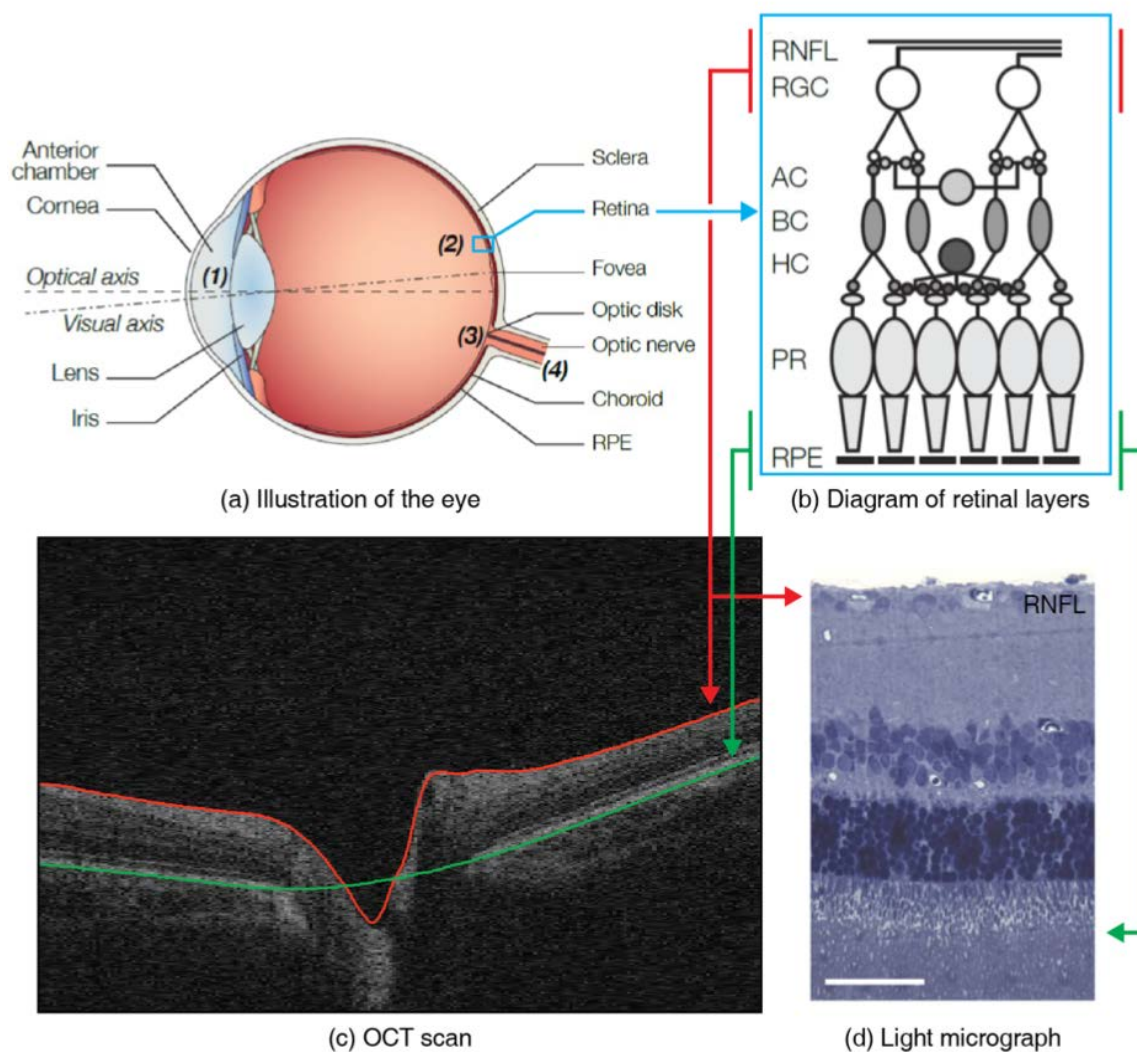
Figure 2.1: Layers of the retina. (a) The eye. (b) Diagram of the retinal layer cells. (c) OCT view of the optic nerve head. Highlighted surfaces are the ILM (red) and the RPE (green). (d) Light micrograph of the human retina. (Overall figure from John W. Miller's thesis [1], original panel images a, b, and d modified with permission from Daniel Palanker [2].)

not treated swiftly, so it is crucial to be able to diagnose and assess the severity and progression of the disease.

AION is general neuropathy of the optic nerve caused by reduced blood flow (i.e. ischemia) to the anterior region of the optic nerve [17]. In layman's terms this means that there is an impairment of blood circulation to the front of the optic nerve without true inflammation of the arteries which can cause disruption of information sent between the eye and the brain. AION can present in two different forms - arteric and non-arteric. Of these two, the non-arteric class makes up a large majority of all recorded cases - almost 95% [18]. Although studies have been performed, it is not yet clear whether any known treatment is effective for NAION [19].

Papilledema is the presentation of optic disc swelling with the specific cause being increased intracranial pressure (ICP) [6]. Because of the direct connection between the brain and the the eye via the optic nerve, if pressure within the skull increases it translate directly to the back of the eye as well [20]. Any disorder which can present with raised ICP (e.g. brain tumor, restricted cerebrospinal fluid, head injury, etc.) can also cause papilledema [21]. Due to this inherent nature a patient may have increased ICP without swelling of the optic disc, but cannot have papilledema without the presence of increased ICP. Because the underlying condition for papilledema must be raised ICP, the definitive means for diagnosis is a lumbar puncture (i.e. spinal tap) to measure CSF pressure. The threshold necessary to be classified as raised ICP is a lumbar CSF opening pressure of 250mm of water [22]. Although this method is the conclusive procedure necessary for diagnosis, it is expensive, time consuming, and can be very uncomfortable for the patient.

Pseudopapilledema presents in a very familiar manner as papilledema, however it is usually a benign condition [23]. Although many symptoms mimic those of traditional papilledema, it often presents with calcified deposits referred to as drusen which can help differentiate it [24]. Another noticeable difference between pseu-

dopapilledema and true papilledema is the shape of Bruch's membrane. In true papilledema raised ICP causes Bruch's membrane to angle towards the vitreous (i.e. the clear gel that fills the space between the retina and the lens of the eye), which is visually apparent in OCT imagery. Because pseudopapilledema is a benign condition while true papilledema can have serious underlying causes, initial diagnosis and assessment is paramount.

The data used for this thesis work is comprised primarily of papilledema patients, however there is at least one example of each of the listed conditions.

2.3 Fundus photography and associated Frisén grading scale

A monocular color fundus photograph is a standard two dimensional image of the posterior region of the retina, usually focused on either the ONH or the macular region of the selected eye. These images are traditionally used by various types of clinicians in assessing many conditions with visible signs on the optic nerve, macula, or any retinal vasculature. The advantages of this modality include quick acquisition time, lack of patient discomfort, and accessibility in many care settings. An example color fundus photograph is shown in Fig. 2.2 with the areas of interest marked.

The Frisén scale was first defined by Lars Frisén in 1982 as a consistent method for grading swelling of the optic nerve head in cases of papilledema [8]. This approach uses visually obvious features in a fundus photograph, such as optic disc border obscuration and vessel tortuosity, to assign an ordinal and discrete grade between zero and five to the given optic nerve head where zero is normal and five is extremely swollen. Three example fundus images are shown in Fig. 2.3 with swelling ranging from Frisén grade 0 (normal) where there is almost no visible swelling to Frisén grade 4 (very swollen) where there is a high degree of swelling and the optic disc is enlarged and the borders are obscured.

Despite the proven effectiveness of this grading system, it has some very obvious
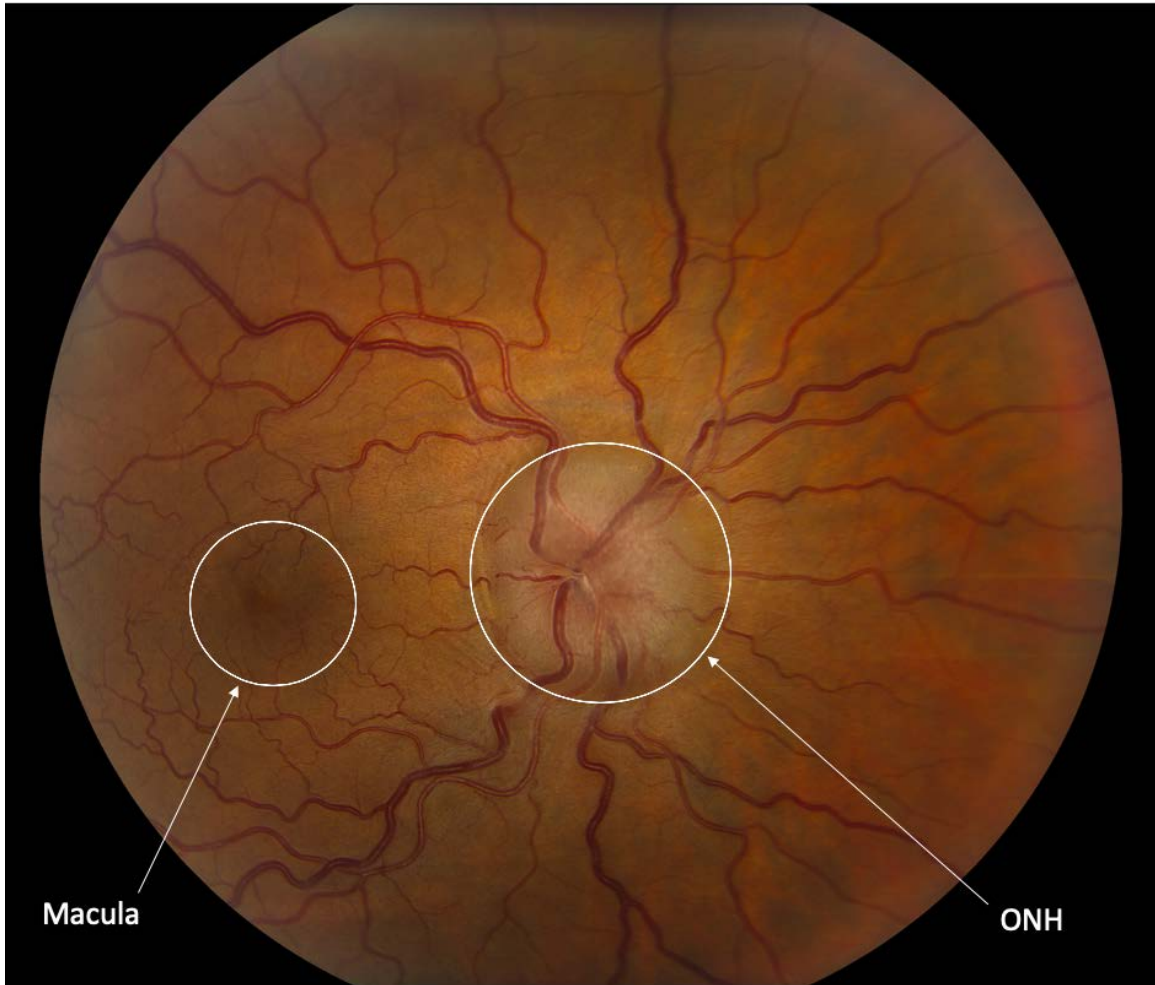
Figure 2.2: An example fundus photograph centered and focused around the optic nerve head with the ONH region and the macula region marked

Figure 2.3: Example fundus photographs with varying amounts of swelling. From left to right the Frisén grades are 0, 2, and 4. Figure modified, with permission, from Jason Agne's dissertation [3].

drawbacks [9, 10]. First, it is a discrete scale meaning that it cannot have a high degree of precision. It also requires expert training which takes a large amount of time and resources on the clinicians' part. This also can lead to a high inter/intra observer variability [10] since it is completely subjective to the current assessment by the clinician. Lastly, it is limited by the fundus photograph's quality because if the image is unintentionally blurred in any way it can skew the grade assigned to the patient. Because of all of these negative qualities, a more continuous and reliable method is desired.

2.4   Optical coherence tomography

Optical coherence tomography is an imaging modality introduced by Huang et al. in 1991 [11] that produces high resolution three-dimensional images using low-coherence interferometry, i.e. emitting a broadband light spectrum into the eye and then measuring its reflected response. Because only the emitted light interacts directly with the patient, this is considered a non-invasive methodology.

Initial versions of OCT operated in the time domain, known as TD-OCT, and were publicly available for first use in 2003. Refined versions of this technology were later released in 2007 that worked in the spectral domain (SD-OCT) which allowed for enhanced resolution as well as faster sampling times (implying less capture time, and therefore less patient involvement). Since it's release, SD-OCT has become the gold standard for non-invasive imaging of the back of the eye (i.e. fundus). Imaging methods used by OCT machines differ slightly for each manufacturer. All OCT images used in this thesis work are SD-OCT images collected on a Cirrus OCT device from Carl Zeiss Meditec, Inc. (Dublin, CA).

Although the above mentioned machines will do multiple different protocols, we are going to focus strictly on the volumetric scans. Each individual SD-OCT image cube is centered at the ONH region with $200 \times 200 \times 1024$ voxels covering a physical space of $6 \times 6 \times 2$ mm^3. Fig. 2.4 shows the relationship between the fundus region,
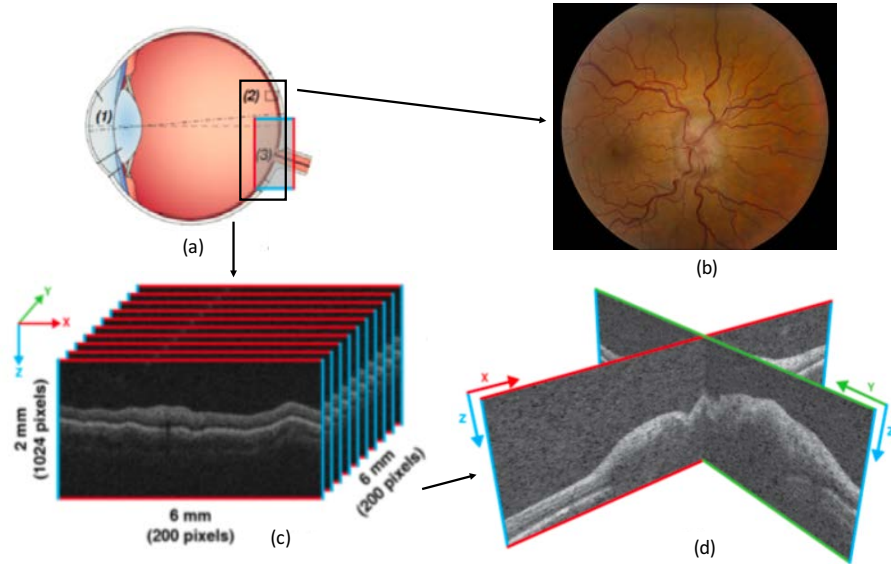
Figure 2.4: Physical relationships between the eye, ONH centered SD-OCT images and fundus photography. (a) Diagram of the eye with OCT scan region highlighted in blue and red, and the fundus photograph region marked in black (modified with permission from Palanker [2]). (b) An example fundus photograph displaying both the ONH and macular regions (c) Example SD-OCT B-scans with the x,y, and z dimensions labeled with their sizes (d) 3D visualization by displaying two cross-section B-scans perpendicularly ((c) and (d) modified with permission from [1])

a standard fundus photograph, and the orientation of the OCT scans.

2.5 Graph-based segmentation of retinal surfaces and associated volumetric measurements

Before we can predict retinal thickness across the fundus we must first define the upper and lower boundaries of the retina. Shortly after the OCT imaging modality emerged, so too did associated segmentation of the different surfaces visible within these images. Initial graph-search algorithms used to performed surface segmentation in 3D images were proposed by Li et al. [25] in 2006, shortly before SD-OCT became available. This work was later adapted by Garvin et al. [26] for normal (i.e. non-swollen) SD-OCT images. Wang et al. [14] went on to adapt these methods for use
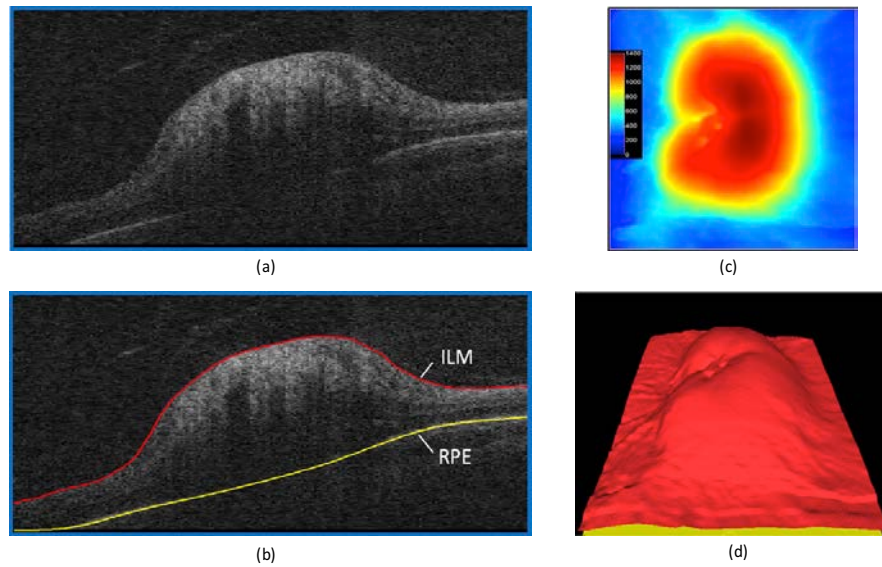
Figure 2.5: (a) An example SD-OCT B-scan. (b) An example of a OCT B-Scan shown with the ILM (in red) and RPE (in yellow) segmented (c) An example 2D thickness map generated from the surface segmentation shown in micrometers. (d) 3D surface rendering generated from the surface segmentation

on SD-OCT images with cases of optic disc swelling. Fig. 2.5 (b) demonstrates an example segmentation of the uppermost (ILM) and lowermost (RPE) surfaces in SD-OCT B-scan with optic disc swelling. The thickness we will be predicting as the main goal in this work is the thickness, in micrometers, between the ILM and the RPE.

Using these SD-OCT images and associated segmentation, it is possible to calculate a volume in the regions between surfaces. Figure 2.5 shows an example segmentation on an SD-OCT image, a two dimensional thickness map, and the associated three dimensional surface rendering. These generated thickness maps can be used to calculate the overall volume between surface in the ONH region, often referred to as total retinal volume (TRV). It is also possible to calculate specific regional volumes, such as strictly the temporal or nasal quadrant of the ONH region. In cases of optic disc swelling the TRV has shown high correlation with the Frisén grading scale [3],

and the regional volumes have shown promise in assessing disease progression in cases of papilledema [14].

As stated in Chapter 1, Echegaray proposed a semi-automated method of predicting the Frisén grades using fundus photography [12], and Agne extended this work by instead using a monocular fundus photograph to predict the TRV [3,15]. Tang et al. [27] has also proven that sterescopic fundus photographs can be used to estimate this volumetric measurement. However, all of these approaches are still limited in that the resulting prediction is a singular number and cannot be used in a qualitative fashion.

2.6   Random forest regression

Because the overarching goal of this thesis is to predict thickness at each pixel location in an image, a machine learning method is desired for this estimation. Due to the limited data available for this work, a random forest was chosen to avoid overfitting the data. Properly introduced by Breiman in 2001 [28], random forest is a general term which can apply to either classification or regression models. For the purposes of this thesis, we will only use regression applications, but all ideas are easily transferable to classification [29]. Built upon the ensemble learning method known as decision trees, random forests introduced noticeable improvements such as higher accuracy and a lower sensitivity to noise than traditional decision trees.

Traditional decision trees work by by repeatedly separating data into two groups. Each split in the data is known as a node in the tree. At each node in the tree, the model identifies features on which to split the data into partitions that continually get more homogeneous. A visualization of this process is shown in Fig. 2.6. The process of training the model determines which features to select at each level in the tree as well as what the threshold values to use for splitting the data. To determine both of these items, a measure known as the Gini impurity index, $I(t)$, is used to measure

impurity of a dataset or subset [30, 31]:

$$I(t) = \sum_{m=1}^{M} P(\omega_m|t)(1 - P(\omega_m|t)) \tag{2.1}$$

where $P(\omega_m|t)$ is the probability of finding a point, ω_m, from class m in a subset X_t and M is the total number of classes. X_t is the current set of points at node t, before the data are split into two purer subsets. The probability is approximated as

$$P(\omega_m|t) = \frac{N_t^m}{N_t}, m = 1, 2, \ldots, M$$

where N_t^m is the number of points from class m in a potential new subset X_{tY} and N_t is the complete number of points in X_t. The Gini impurity is maximized when $P(\omega_m|t)$ is constant. A tree is trained by selecting optimal features and thresholds that minimizes the Gini impurity index, which results in a tree that is optimally split. After looking at the theory behind decision trees, it is clear how this would apply to classification, but maybe less so how it would apply to a regression problem where our result needs to be on a continuous scale. Because the algorithm "bins" similar data together, it can predict similar values for data with similar feature values even on a continuous manner.

Decision trees maintain a few advantages over other historically successful machine learning models. The theory, as shown above, is simple and easily understood. Inherently, it is also easy to implement and modify for both classification and regression [29]. However it also has a few drawbacks, namely that it can be sensitive to small changes in training data and compared with methods such as support vector machines it can have lower accuracy. Breiman's random forest technique improves on both of these problems.

Random forests improve upon traditional decision trees by using a bootstrapped ensemble of decision trees as well as a random subset of features during the training
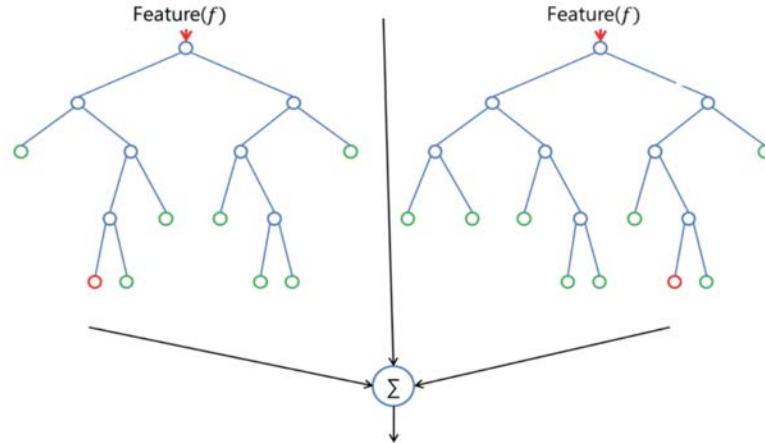
Figure 2.6: An example visualization of the structure of a basic random forest with only two trees [4]

process to produce high accuracy and a higher resistance to noise in the data. The overall structure remains the same as in decision trees (shown in Fig. 2.6), but only a randomly selected subset of features is considered at each node. Random forest also often implement bagging, or bootstrap aggregating, to reduce variance in the final model [28]. This is the practice of training multiple trees within a given data and then averaging the results of the trees.

Random forest has proven success in many different applications [32], as well as in ophthalmic image analysis. It has been used to generate cost images for layer segmentation [33], predict the total retinal volume of swollen optic nerve heads from fundus photographs [15], estimate papilledema severity using retinal regional volumes and layer shapes [34], and more recently to differentiate between causes of optic disc swelling using retinal layer shape features [1].
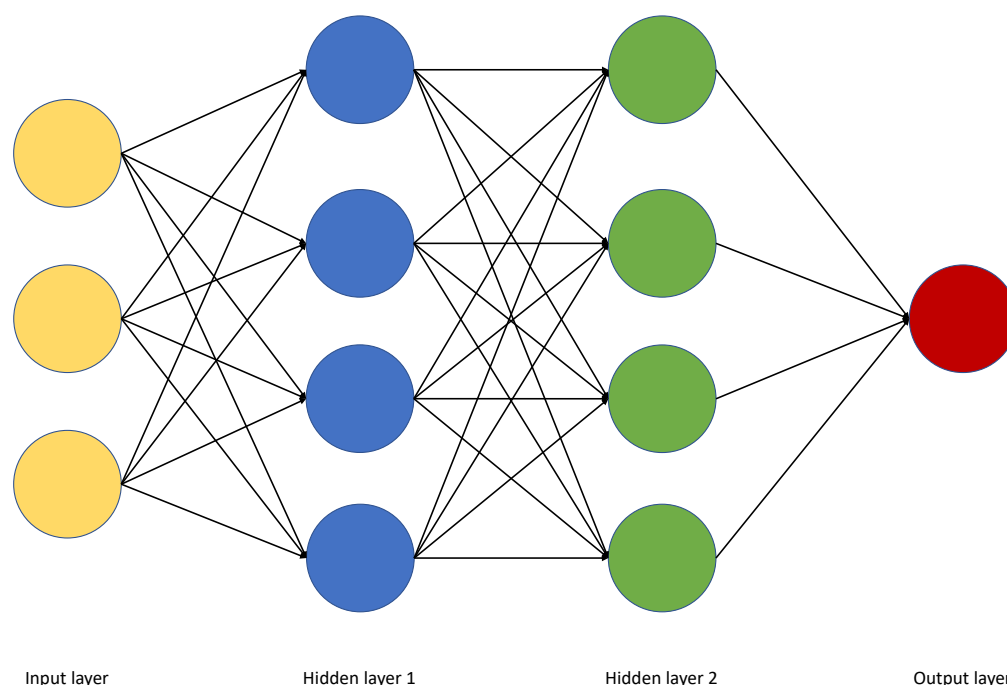
Figure 2.7: A connection diagram of neurons in a basic neural network with two hidden layers

2.7 Convolutional neural networks and U-net

2.7.1 Basic theory and convolutional neural networks

Deep learning is a recently popularized type of machine learning that has shown great promise in tasks such as natural language processing and medical image analysis [5, 35, 36]. At a very high level, deep learning involves artificial neural networks that contain multiple hidden layers [37]. Artificial neural networks were inspired by the neurological process of neurons transmitting information throughout the brain. Each artificial neuron receives a signal (or multiple signals) from the previous layer, performs some computation (often some type of filtering), and then passes the modified signal on to the next layer. Fig. 2.7 shows a very simple illustration of this concept.

A specific deep learning framework, known as convolutional neural networks (CNN),

have recently shown encouraging results when applied to image-based tasks for both regression and classification [38]. CNN's are deep networks that use stacks of learnable filters which can automatically learn features from within the training data. First introduced by Lecun et al. [38] in 1998 and later modified by Krizhevsky et al. in 2012 [36], these networks specifically attempt to mimic the structure of the primary visual cortex.

To provide a thorough analysis of the problem of predicting retinal thickness from fundus images, we would be remiss if we did not attempt to use a CNN approach. It should be noted however that that neural networks tend to use larger amounts of training data to reach similar levels of accuracy as classical machine learning algorithms [36].

2.7.2 Fully convolutional networks and U-net architecture

Early versions of CNNs used a batch-based approach that forced the network to run on each pixel, resulting in longer run times. A remedy approach was suggested by Sermanet et al. [35] in 2013 called OverFeat. This involved replacing the fully connected layers with convolution layers. However, this approach was still limited by its output resolution which was significantly lower than the input resolution. Fully convolutional neural networks [39] were developed based on this need to improve output resolution. To increase the output resolution, deconvolution layers (which are a transposed process of standard convolution) were added in the latter portion of the network. These layers essentially perform up-sampling before the network output allowing for increased precision in the output.

An extension on fully convolutional networks, known colloquially as U-net, was specifically proposed for medical imaging tasks at the MICCAI conference in 2015 by Ronneberger [5] with very promising results. A U-net is defined by its asymmetric structure and by the up-sampling layers in the second half of the network that replace
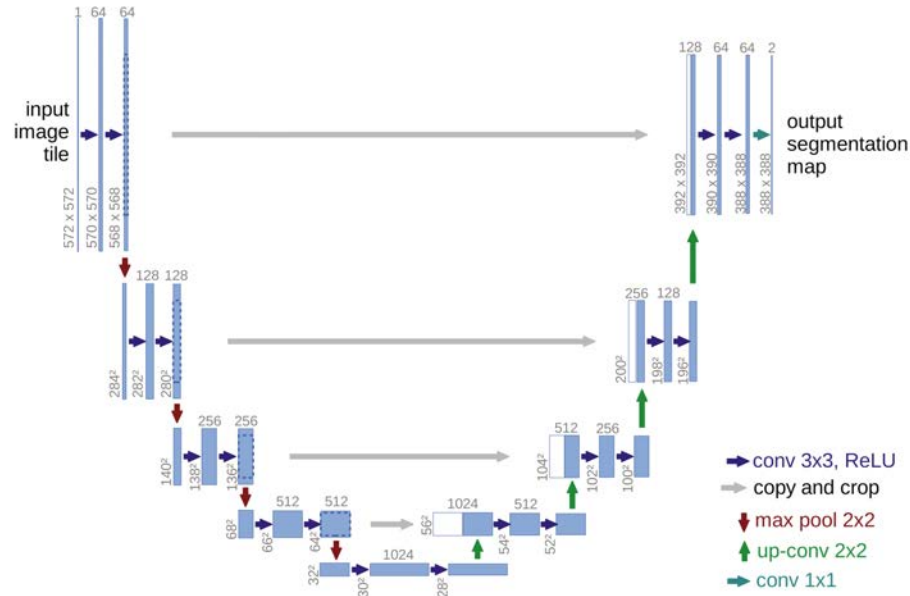
Figure 2.8: The original architecture of Ronneberger's proposed U-net used for segmentation [5]

the pooling layers in a standard CNN and help form the U-shape which gives the network its name. Ronneberger's initial proposed network architecture is shown in Fig. 2.8. The up-sampling layers enable the output to be a similar resolution to the input image, which was an issue with previous architectures as standard convolutional layers reduce resolution. The network only uses the valid part of each convolution without any fully connected layers [39]. For predictions in the border regions of the output image, the context that is missing is extrapolated directly from the original image. This framework is ideal for many medical imaging tasks because of the precision allowed by the gained resolution, and has already shown merit in tasks involving pixel-wise regression [40].

CHAPTER 3
PREDICTING FUNDUS THICKNESS FROM 2D FUNDUS PHOTOGRAPHY USING RANDOM FOREST REGRESSION

Note: A version of this work has previously been published and presented at the MICCAI conference's Ophthalmic Medical Image Analysis workshop in 2018 [41].

3.1   Motivation

For many years, color fundus photographs have been a common imaging modality for ophthalmologists to examine the back of the eye in cases of optic disc swelling [Fig. 3.1(a)] [42]. Traditionally, assessment of swelling via fundus photographs has been a very challenging task due to the specialized expertise required and the inability to compute volumetric measures of the swelling [9]. With the introduction of spectral-domain optical coherence tomography (SD-OCT) [Fig. 3.1(b)] and development of associated retinal-layer-segmentation algorithms [14, 26] [Fig. 3.1(c)], SD-OCT-based volumetric measures have recently shown great promise in overcoming many of the limitations of a traditional fundus-based assessment.

However, SD-OCT is not always available as its use is primarily limited to specialized eye clinics rather than in primary care or telemedical settings. Thus, the need for better approaches for the assessment of optic disc swelling via fundus photographs alone still exists. In this area, Echegaray et al. [12] proposed a decision-tree system that automatically grades/stages the optic disc swelling using image features directly extracted from monocular fundus photographs. However, this approach only provides ordinal severity grades rather than continuous volumetric measures. Tang et al. [27] demonstrated that the stereoscopic color fundus photographs have the potential to reconstruct the depth information and allow the volumetric estimation for the optic disc swelling, but the requirement of carefully acquired stereo images rather than monocular images limits its applicability. More recently, Agne et al. [15] proposed a regression approach, which can directly estimate the total retinal volume (TRV)
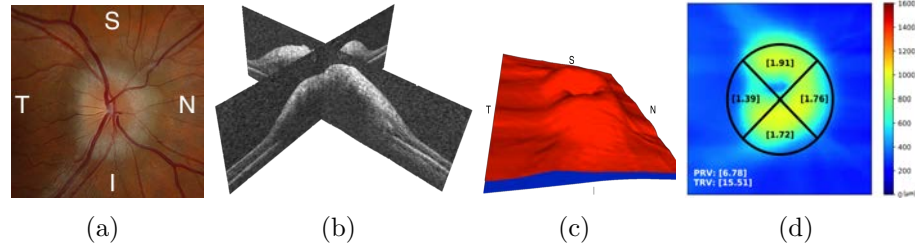
Figure 3.1: (a) An example registered and cropped fundus photograph at the optic-nerve-head (ONH) region with the nasal (N), temporal (T), superior (S), and inferior (I) sides marked. (b) A 3D rendering of the corresponding OCT image depicting a swollen optic disc. (c) A surface rendering of the internal limiting membrane (ILM) shown in red and the retinal pigment epithelium complex (RPE) shown in blue. (d) The ONH thickness map with a grid showing regional volumes (in mm^3) displayed within the grid, peripapillary volume (PRV), and total retinal volume (TRV) both shown in the bottom left-hand corner (also in mm^3).

at the optic-nerve-head (ONH) region by only inputting a single fundus photograph; however this approach doesn't predict local thickness values as may be needed to compute regional volumes.

Thus, to overcome the limitations above, we propose a machine-learning method that estimates the local volumetric information by only requiring a single monocular color fundus photograph [Fig. 3.1(a)] as the input. The proposed method outputs a thickness map with 200×200 pixels covering 6×6 mm^2 at the ONH region. Based on the resulting thickness map, the volumes of the peripapillary region, the nasal, temporal, inferior, and superior quadrants, as well as the TRV can be computed [shown in Fig. 3.1(d)]. Results are quantitatively assessed using the root-mean-square errors between the model's outputs and the OCT ground truths, as well as Spearman's rank correlation coefficients. Visualizations of the predicted thickness maps are also provided for qualitative assessment.

3.2 Methods

3.2.1 Overview

For the purpose of a fair comparison between the two image modalities, the input fundus photographs were registered to the SD-OCT image domain and centered/cropped at the ONH [Fig. 3.1(a)]. Next, the blood vessels were inpainted and the resulting fundus image had features pertaining to textural and color information extracted. Based on these selected features, random forest classifiers, which will be discussed more thoroughly in Section 3.3, were trained to estimate the depth information of the retina at the pixel level and be able to output an ONH thickness map that makes regional volumetric measurements computable [Fig. 3.1(d)].

3.2.2 Preprocessing

We first registered the input fundus photograph with the ONH-centered SD-OCT *en-face* image. In particular, the SD-OCT images were segmented using 2D/3D graph-theoretical algorithms [14, 26], and the *en-face* image was created by averaging the pixel intensities along each A-scan within the retinal pigment epithelium (RPE) complex. After that, we applied blood vessel inpainting on the ONH-centered/registered fundus photograph [Fig. 3.2(a)] to suppress the negative effects from blood vessels on the predicted thickness map. During the processes of vessel inpainting, a blood vessel probability map was computed using a deep learning based approach using U-Net [5] [Fig. 3.2(b)]. Then, this vessel probability map was thresholded ($p = 0.5$) into a blood vessel mask. Next, a binary morphology dilation (spherical filter size: $r = 1$ pixel) was used to ensure that the vessels were completely encompassed by the mask. By overlapping the cropped fundus photograph with the dilated blood vessel mask [Fig. 3.2(c)], a blood vessel inpainted image using second order interpolation was created [Fig. 3.2(d)].

<div align="center">(a)   (b)   (c)   (d)</div>
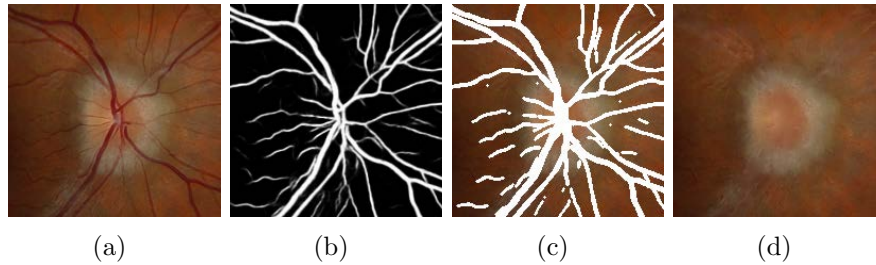
Figure 3.2: (a) An example cropped ONH-centered fundus image. (b) The corresponding blood vessel probability map obtained from a deep-learning neural network. (c) Overlapped dilated vessel mask on the cropped fundus photograph. (d) The blood vessel inpainted image.

3.2.3   Feature extraction

In the processes of assessing optic disc swelling via fundus photographs, neuro-ophthalmologists grade the swelling severity by inspecting key observable features on the image. Similarly, in this work, several feature sets (categories include: image intensity, color representations, gradient, and texture information) were extracted from the inpainted image to help the proposed classifier to estimate the thickness information at pixel-level; a total of 58 features were used as listed in Table 3.1.

To quantify the textural information, Gabor filters [43] are commonly used to analyze image objects with specific combinations of frequencies, directions, and regions of interest. Here, Gabor magnitude responses were computed at 0°, 45°, 90°, and 135° with wavelengths of two and four pixels at each orientation [Fig. 3.3].

Textural features were also obtained via use of gray-level co-occurrence matrices (GLCM) which involves statistically considering the spatial relationship of pixels [44]. The GLCMs were computed for each pixel in the inpainted image at an offset of one pixel at the right using three different neighborhood sizes: 10×10, 15×15, and 20×20. For each GLCM, statistical properties, including variance, contrast, entropy [Fig. 3.4(a)], homogeneity [Fig. 3.4(b)], and inertia, were used to create different feature images.

Table 3.1: Complete list of features for regression analysis.

Feature List			
Number	Description	Number	Description
1-5	GLCM Properties† (10×10)	51	Green Channel
6-10	GLCM Properties† (15×15)	52	Blue Channel
11-15	GLCM Properties† (20×20)	53	Hue (HSV)
16-27	Entropy	54	Saturation (HSV)
28	Gradient Magnitude	55	Value (HSV)
29	Gradient Direction	56	Histogram Equalized Image
30-49	Gabor Responses	57	Saturated Image
50	Red Channel	58	Grayscale Intensity

† GLCM properties include: variance, contrast, entropy, homogeneity, and inertia

(a) (b) (c) (d)
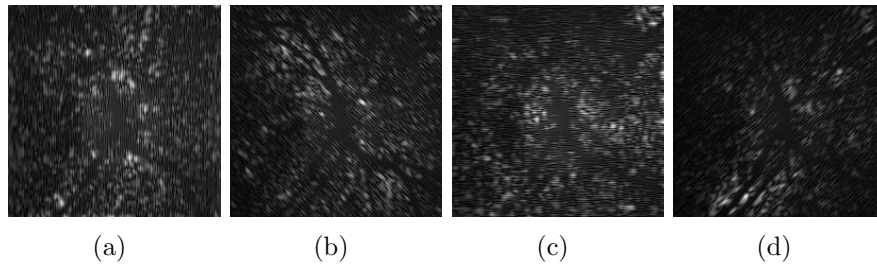
Figure 3.3: Example Gabor responses with a wavelength of two pixels with directions (a) 0°, (b) 45°, (c) 90°, (d) 135°.

(a) (b) (c) (d)
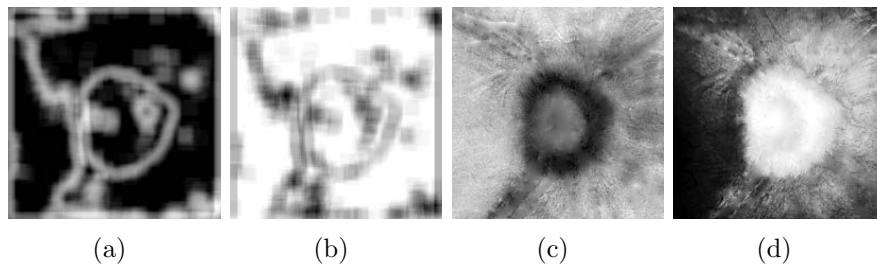
Figure 3.4: Example feature images. (a) GLCM entropy (15×15). (b) GLCM homogeneity (15×15). (c) Saturation image. (d) Histogram equalized fundus image.

In addition, entropy can also be used to evaluate the image information in a particular region of interest. In this work, both small and large sliding window sets were applied on the inpainted fundus photographs to compute image entropy in regions[1]. For the smaller windows, these computed entropy values were interpreted as quantitative indicators of the presence of homogeneous textures; for the larger windows, the computed entropy values indicated the presence of heterogeneous textures. The image gradient direction and magnitude were also included as features as well to enhance the borders among the regions with texture and/or color changes. Along with the red, green, and blue intensity values, the inpainted fundus image was also converted to the hue-saturation-value (HSV) color space, and the separate channel values were used as features as well. Differences in color contrast can be accentuated in the hue and saturation [Fig. 3.4(c)] channels, which are highly indicative of texture since an opaque texture is more associated with swelling. In addition, the histogram-equalized image [Fig. 3.4(d)] as well as an intensity mapped image with the top and bottom one percent of pixels saturated were used as features. Both images work to emphasize contrast in color between regions of differing colors or differing intensities.

3.3   Experimental methods

A total of 88 subjects with optic disc swelling having both volumetric SD-OCT as well as color fundus images were used for experimental analysis. The true thickness information at the ONH (i.e., the ground truth) for each subject was calculated based on the segmented internal limiting membrane (ILM) and the lower bounding surface of the RPE complex in the SD-OCT image. With a total of 58 input features from the input registered fundus image, a random forest bagged ensemble regression model was trained using 500 learning cycles and feature importance was calculated as part of the training process. To reduce computational complexity, the model was trained

---

[1]The small sliding window sizes include: 5×5, 7×7, 11×11, 13×13, 15×15, 17×17, 21×21, 25×25; the large sliding window sizes include: 37×37, 49×49, 73×73, 101×101

on two and a half percent of the pixels in each image chosen randomly. Evaluation was performed in a leave-one-subject-out approach so that the model for each subject was obtained by training the classifier on the images from the remaining 87 subjects. After predictions were made for individual pixel locations, volumes were calculated for the peripapillary, nasal, temporal, inferior, and superior regions. The peripapillary region was defined as the region inside a central circle with radius 1.73 mm. The nasal, temporal, inferior, and superior regions were defined as the four interior quadrants of the peripapillary circle using the 135° and 45° lines as boundaries. The overall TRV was calculated as well. Errors and correlations were then calculated for each individual region.

3.4   Results

When comparing the total retinal volume (TRV) calculated from the retinal thickness predictions generated from the described model and the ground truth from OCT images, a root-mean-square-error of 1.66 mm³ was achieved. Spearman's correlation coefficient was $R = 0.73$. When comparing regional volumes, the nasal, temporal, inferior, superior, and peripapillary regions had root-mean-square-errors of 0.64 mm³, 0.61 mm³, 0.74 mm³, 0.71 mm³, and 1.30 mm³, respectively. The correlations (R) were 0.71 (nasal), 0.72 (temporal), 0.61 (inferior), 0.65 (superior), and 0.75 (peripapillary). Examples of comparisons between the total retinal thickness maps from the SD-OCTs (i.e, ground truths) and from the monocular fundus photographs are shown in Fig. 3.5 and Fig. 3.6.

Average feature importance across all models was calculated by permuting the features and looking for change in the model error. Top features were found to be entropy in large neighborhoods as well as features that accentuate color change, such as hue, saturation [Fig. 3.4(c)], or the histogram equalized image [Fig. 3.4(d)]. All features that had distinctly different values for the optic disc compared to the peripheral area were helpful in distinguishing swollen regions from non-swollen regions.

3.5   Discussion

In this preliminary study, we have shown that the proposed method demonstrates the monocular fundus photographs as a potentially lower cost but more available alternative to the SD-OCT in the primary care or telemedical settings in cases of assessment for optic disc swelling. Using features pertaining to textural and color information obtained directly from the fundus photographs, akin to properties neuro-ophthalmologists use, local thickness measurements can be estimated using a regression model trained on a variety of subjects with varying degrees of swelling. The ability to assess depth information at the swollen regions can help clinicians identify onset of diseases in earlier stages; for example, papilledema (a particular type of optic disc swelling due to increased intracranial pressure) often presents with swelling in the nasal quadrant relatively early [16]. In addition, the predicted thickness maps provide a future opportunity to create 3D retinal shape models directly from the 2D monocular fundus photographs. The retinal shape information is hypothetically helpful for distinguishing the different causes of the optic disc swelling.

The limitations of this work may include: 1) the lack of subjects with severe optic disc swelling causes the trained classifiers to slightly underestimate the retinal thickness at the extremely swollen regions, and 2) the thickness predictions at the regions with imperfect vessel inpainting are less accurate due to the inconsistent texture information between the swollen retinal tissue and the vessel residual. Potential future work can involve collecting more SD-OCT and fundus image pairs with various severity of optic disc swelling or creating 3D retinal shape models directly from the 2D fundus photographs to help identify causes of optic disc swelling.

Figure 3.5: Input cropped and registered fundus photographs with their accompanying thickness maps generated by SD-OCT (used as the ground truth) as well as the predicted thickness maps generated by the model with regional volumes shown in the grid, and peripapillary plus total volumes shown in the lower left hand corner (in mm³) shown in order of increasing swelling. Each column represents a different subject.

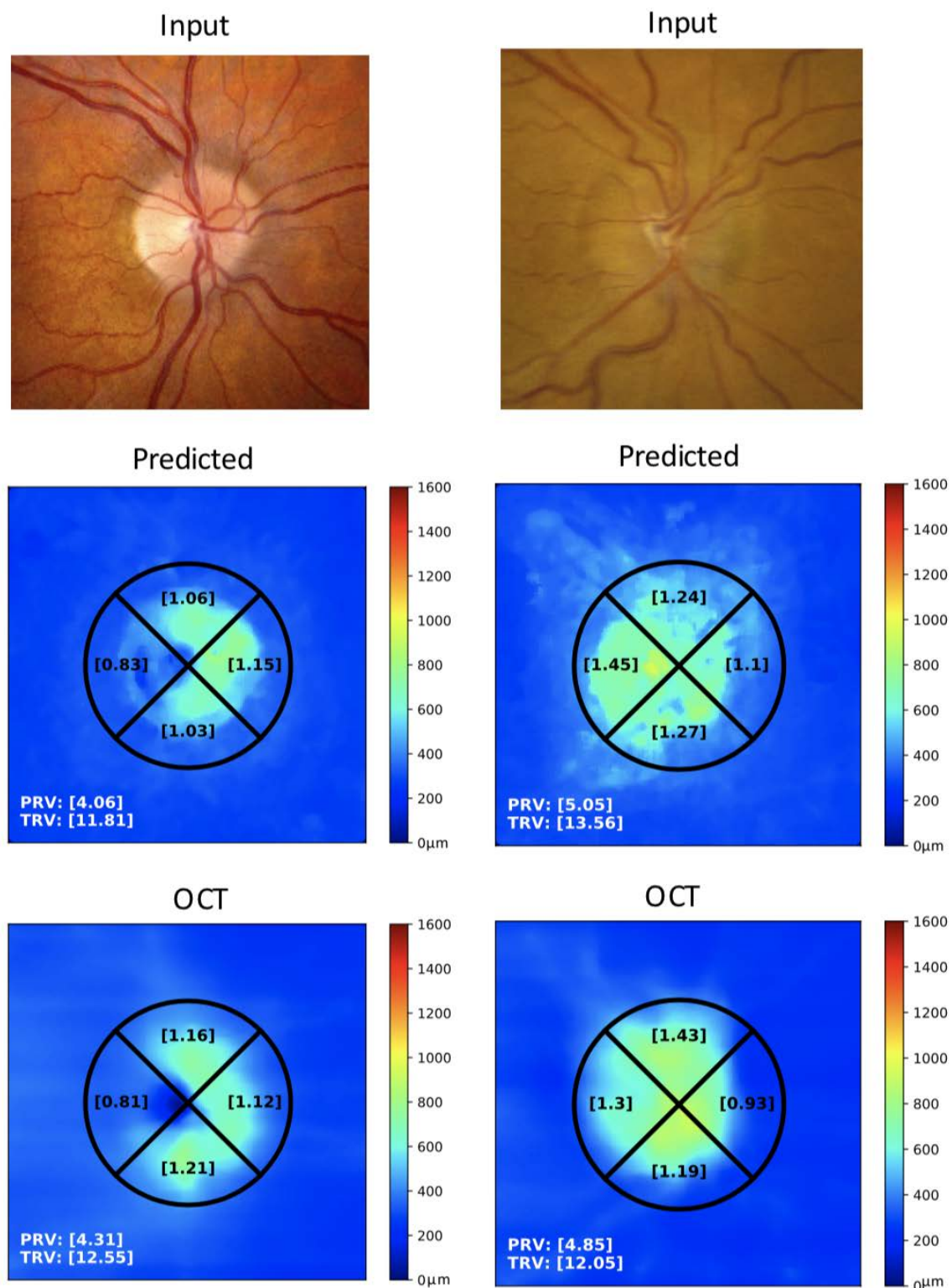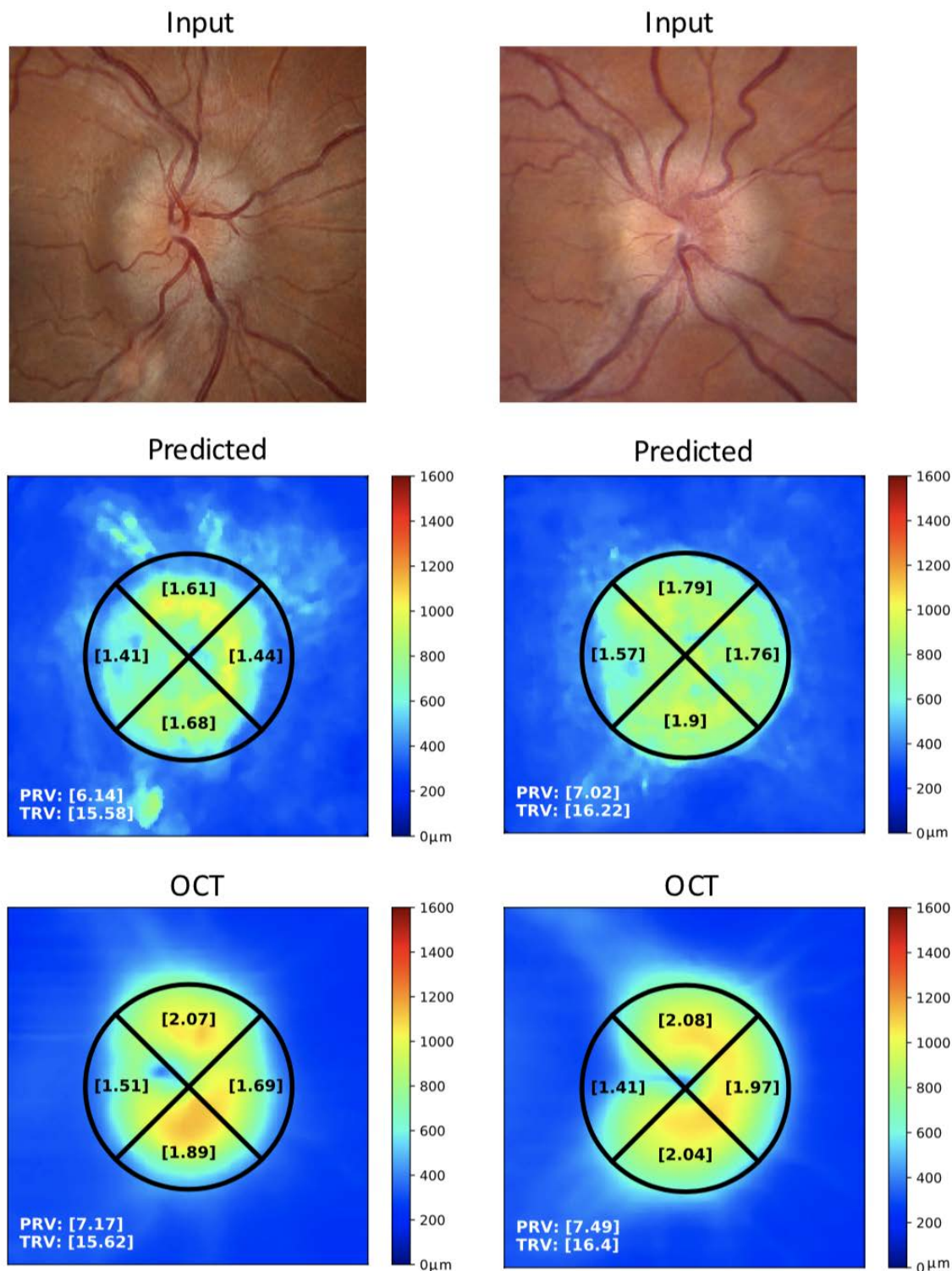Figure 3.6: Input cropped and registered fundus photographs with their accompanying thickness maps generated by SD-OCT (used as the ground truth) as well as the predicted thickness maps generated by the model with regional volumes shown in the grid, and peripapillary plus total volumes shown in the lower left hand corner (in mm^3) shown in order of increasing swelling. Each column represents a different subject.

CHAPTER 4
A DEEP LEARNING APPROACH FOR RETINAL THICKNESS ESTIMATION USING FUNDUS PHOTOGRAPHY

Note: A version of this work has been accepted for presentation at the ARVO conference in 2019 [45].

4.1   Motivation

Deep learning, or more technically artificial neural networks with multiple hidden layers, have shown great promise in tasks involving visual imagery. More specifically, convolutional neural networks (CNNs) have been proven successful for a variety of tasks in the medical imaging domain [46]. Image classification, segmentation, registration, and detection are just a few of the many applications which have had positive results after applying CNNs [35, 46].

As particularly relevant to this thesis, CNNs have shown promise with the use of fundus photography applications [47]. It has been shown that CNNs can, with a relatively high degree of accuracy, predict diagnoses such as diabetic retinopathy, macular edema, and even predict the location of retinal vessel structures [47]. It is worth noting however, that to get results of this magnitude a very large amount of data is needed to train these complex mathematical algorithms.

4.2   Implementation

4.2.1   Preprocessing

As a form of basic data augmentation, for each image there was a 50% chance that we flip the image across the Y axis (a right to left flip). Because all of the images used in this work are right eyes, this gives approximately half of our dataset the orientation of left eyes. This allows our model to be trained on more robust data, as it will have seen both right and left eye orientations.

Because fundus photographs are usually focused in a single area (typically the macula or ONH), it is possible to get a partially unfocused image in some areas

causing small, but important, image features to be blurred or dark. To help remedy this we perform contrast limited adaptive histogram equalization (CLAHE) on the cropped and registered fundus images. An example original and equalized image are shown in Fig. 4.1. CLAHE has been proven to help enhance key image features in medical images [48].

It is also known that deep learning requires more training data than traditional machine learning algorithms [36], so we will take some additional preprocessing steps so that we have a maximal number of training instances. More specifically we break our cropped and registered fundus images that are 200×200 and break them up into 25 separate sub-images, each with size 40×40. An example is shown in Fig. 4.2.

Once the original images have been preprocessed using the steps described above, we finally normalize the pixel values from zero to one. For purposes of training, we also transform the ground truth OCT thickness maps from physical dimensions (micrometers) to a normalized scale of, again, zero to one by mapping the maximum physical thickness in the dataset to one and them minimum to zero (although this was already the minimum). Once the model has made its predictions, we transform the predictions back to physical space.

4.2.2   Network architecture

The network chosen for this application is a well known network architecture originally proposed by Ronneburger at MICCAI in 2015 [5] and shown in Fig. 2.8. The U-Net consisted of a total of 15 neural layers, 13 convolution layers, and two max-pooling layers. The neural network was designed to obtain image features in different resolutions by passing the input image through a contracting path (i.e. the first half of the network; repeatedly using the combination of the convolutional layers, exponential linear units (ELU), and a max-pooling layer) followed by an "up-sampling" path (i.e. the second half of the network; repeatedly using the combination of the convolutional layers, ELU, and up-convolutional layers). In the up-sampling path,
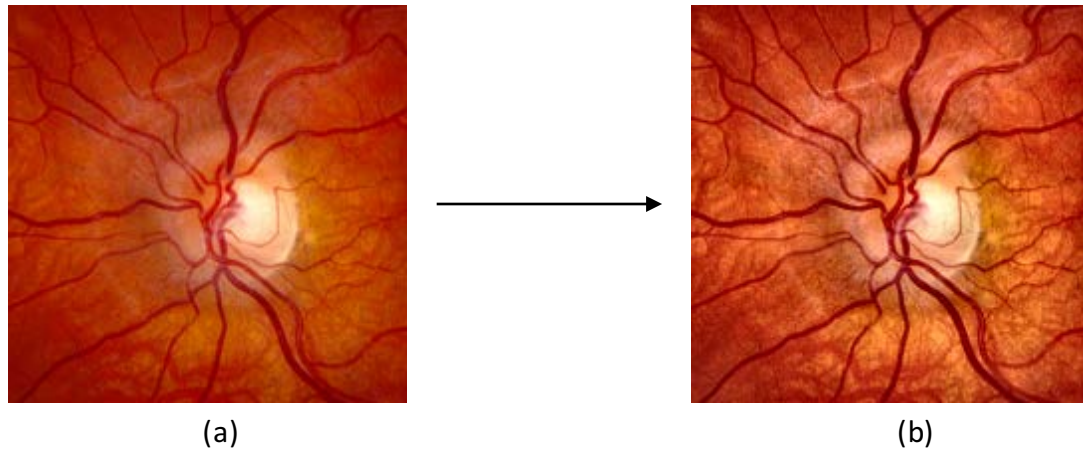
Figure 4.1: (a) Original cropped and registered fundus photograph. (b) Contrast Limited Adaptive Histogram Equalized fundus photograph

the "up-convolutional" layers are applied to double the dimensions of the feature map, and then the outputting feature maps are concatenated with the ones in the contracting path to reconsider past features and help fill in missing context in the prediction with information from the original image. It was also necessary to change the activation function in the final convolutional layer to a linear activation function since this is a regression problem instead of the original classification task.

4.2.3 Experimental methods

The same data set as used for the experimental analysis in chapter 3 was used for this analysis consisting of 88 subjects with optic disc swelling having both volumetric SD-OCT as well as color fundus images were used for experimental analysis. However, because deep learning algorithms use the training data to help tune the network, it is not a fair analysis of those subjects are also used for evaluating the network. So for the purposes of testing, ten randomly selected subjects were strictly withheld to be used as a test set. The remaining 78 images were split into a training set (70) and a validation set (8). The network was then trained using the training and validation
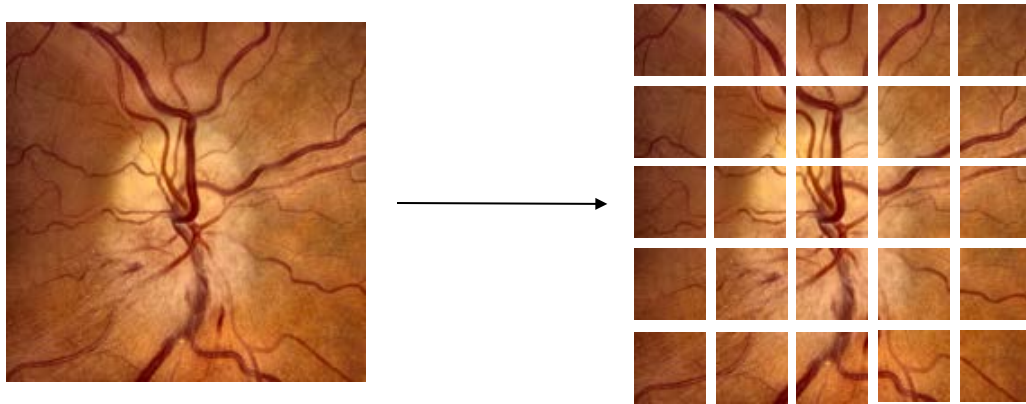
Figure 4.2: CLAHE cropped fundus image shown on the left, and the same image split into 25 equal sub-images to be used for training

sets using a learning rate of 10^{-3} for 500 epochs with a batch size of 8 images. The ground truth thickness information around the ONH region for each subject was again calculated using the segmented internal limiting membrane (ILM) and the lower bounding surface of the RPE complex in the SD-OCT image. After predictions were made for individual pixel locations in the ten withheld images, volumes were again calculated for the peripapillary, nasal, temporal, inferior, and superior regions. As stated in section 3.3, the peripapillary region was defined as the region inside a central circle with radius 1.73 mm. The nasal, temporal, inferior, and superior regions are defined as the four interior quadrants of the peripapilary circle using the 135° and 45° lines as boundaries. A map image is shown in Fig. 4.4 The overall TRV and PRV were calculated as well. Errors and correlations were then calculated for each individual region.

4.3 Results

Model predictions on the test images were compared with OCT-derived volumetric measures: total retinal volume as well as regional volumes were calculated. For total

retinal volume, a root-mean-square-error (RMSE) of 2.07 mm^3 was achieved. When comparing regional volumes, the nasal, temporal, inferior, superior, and peripapillary regions had RMSE's of 0.75 mm^3, 0.82 mm^3, 0.85 mm^3, 0.91 mm^3, and 1.62 mm^3 respectively. Examples of three different patients with input fundus photographs, the model predictions, as well as the OCT ground truth thickness maps are shown in Fig. 4.3. As you can see from the predictions, both quantitative measures (in the form of total retinal volume and regional volumes) and qualitative measures (overall color thickness map) are available.

4.4   Discussion

Although the quantitative results from our deep learning are slightly more error-prone than those shown by its random forest counterpart, it is not completely accurate to say that the model performed worse. Because we evaluated the random forest model using a leave-one-patient-out approach and the deep learning method with a strictly withheld randomly selected subset the results cannot be directly compared. Because there was a test set withheld from the deep learning model, this meant that there was actually less data to be used for training the model itself - only 78 images (70 training, 8 validation) to be exact compared with the full 88 used for the random forest.
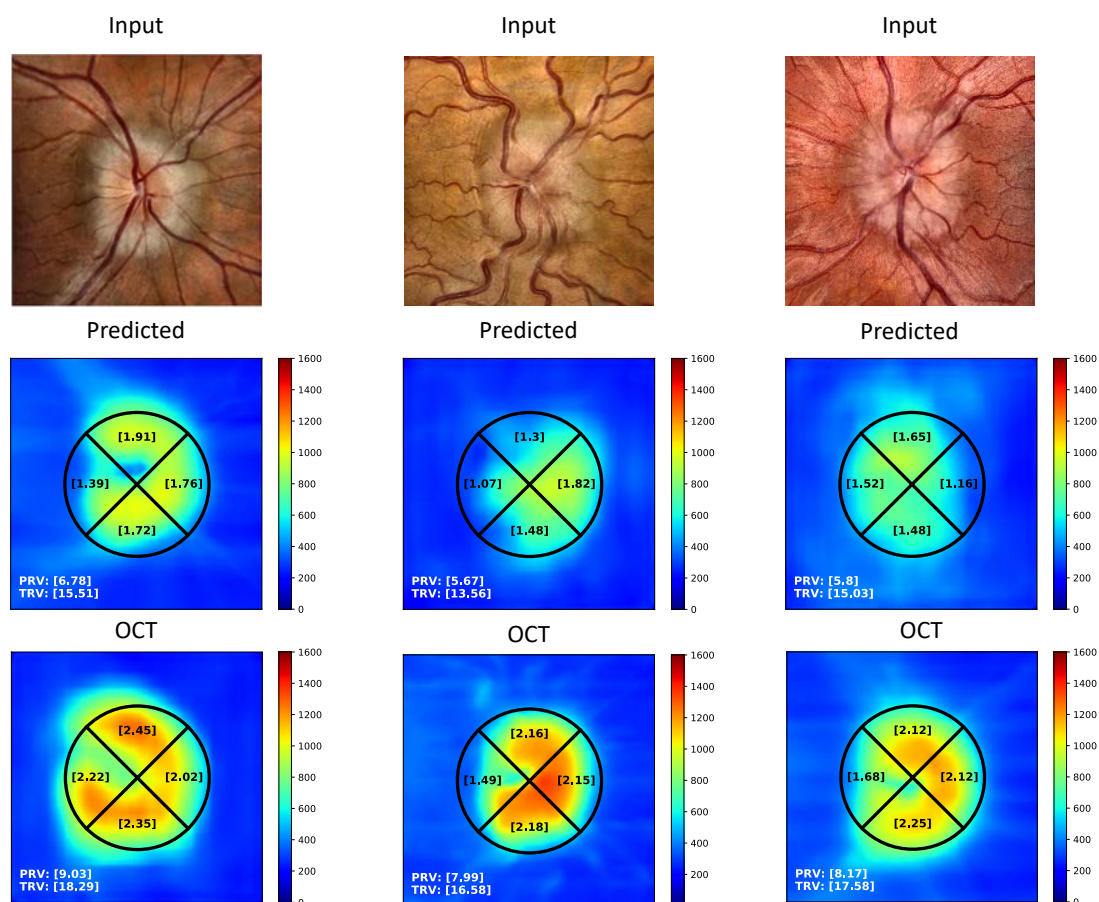
Figure 4.3: Top: Input 200x200 ONH region fundus photographs. Middle: The predicted thickness map for the above fundus image. Local predictions are shown in micrometers. Bottom: OCT reference standard thickness map. All volumes shown are in mm^3. PRV (peripapillary regional volume) refers to the cumulative volume of all 4 quadrants in the peripapillary circle.

Map Image
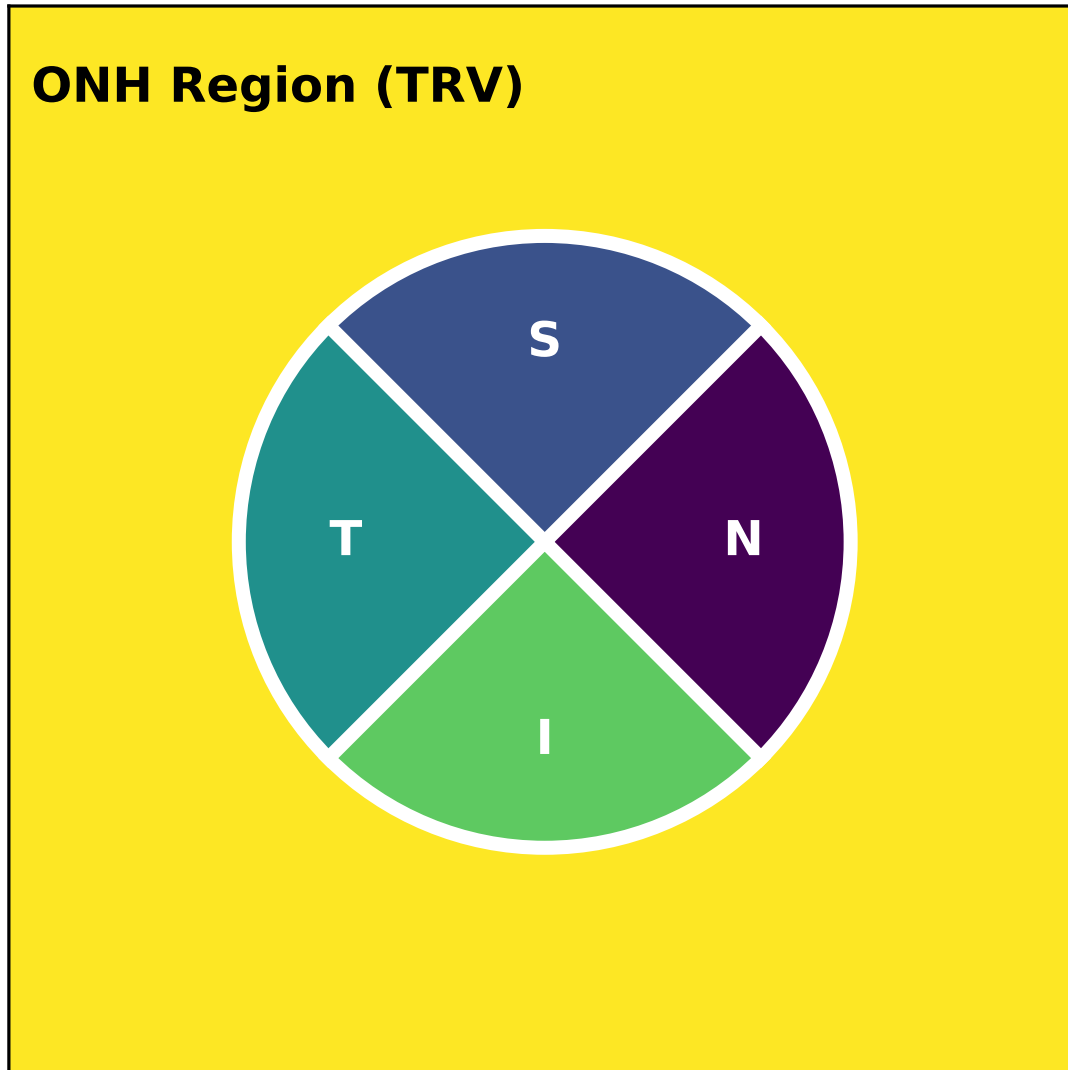
Figure 4.4: Map image displaying the reference regions of the ONH area. S refers to the superior quadrant, I the inferior quadrant, T the temporal quadrant, N the nasal quadrant, peripapillary refers to the cumulative area of all 4 quadrants, and TRV refers to the entire image area.

CHAPTER 5
DISCUSSION AND FUTURE WORK

5.1   Discussion

This thesis work focuses on instituting machine learning methods which allow for accurate evaluation of optic disc edema in a manner similar to those used in OCT imagery, but instead using more readily available color fundus photography.

Although the algorithms used in this thesis work (random forest regression ensemble, CNN, U-Net) are not proposed for the first time, the application on which they are applied in this work is completely new. Never before have machine learning methods been applied for retinal thickness estimation on a pixel-by-pixel basis in the context of optic disc edema.

The proposed approaches have clear merit in the area of low cost assessment, as well as applications such as telemedicine where the clinician may only have access to fundus photography. Although these approaches only produce predictions and would not be used as a gold standard assessment, they could provide a highly desired instrument in a clinician's arsenal of diagnostic tools to help evaluate (both qualitatively and quantitatively) and assess the severity of optic disc edema in patients using relatively low cost and widely available fundus photographs.

The most obvious limitation of this work is the limited amount of data available for training and evaluation of both proposed methods. As more data is becomes available and more is learned about textural and depth information contained in two dimensional photographs, the accuracy of predictions of retinal layer thickness will improve and will enable clinicians to rapidly and reliably assess severity of optic disc edema using strictly fundus photography. Also due to the clear risk for overfitting on a small dataset, there was a very finite amount of hyper-parameter tuning that could be done for each model (e.g. learning rate, number of trees, etc.). Additionally, because of our limited evaluation methods (leave-one-patient-out cross validation for

random forest and a randomly selected strictly withheld subset for deep learning), it is difficult to fairly evaluate the models side-by-side.

Speaking just about the random forest regression model, another limitation is the selection and knowledge of relevant features for the model to train on. Model results are completely based on manual feature selection which requires expert knowledge about three dimensional contextual information contained in a two dimensional image.

This limitation is overcome by using a method that automatically selects features from the data, such as a deep learning approach, but as stated before these approaches typically require more training data to learn those features. They also potentially introduce a new limitation in the form of necessary hardware. Typically a high-powered and expensive GPU is needed for these types of algorithms since they perform many floating point calculations.

5.2   Conclusions and future work

The main contribution of this thesis is the innovative application of machine learning methods for assessing the severity, in both a qualitative and quantitative manner, of optic disc edema. In Chapter 3 we presented a traditional machine learning approach in the form of a random forest ensemble regression model to estimate thickness between retinal layers which was published as a conference paper in the MICCAI conference's Ophthalmic Medical Image Analysis workshop [41] in 2018 and was also given as an oral presentation. For comparison, and due to its recent success in both image classification and medical image analysis, a deep learning method was implemented on the same data used for the random forest, which is described in Chapter 4 and was accepted as a conference abstract at the ARVO conference in 2019 [45]. Both methods show considerable promise in predicting local and global volumetric measurements akin to those procured from OCT imagery.

Future avenues for extending this work could include analysis of additional textural and depth-related features from two dimensional images. Regional volume predic-

tions could also potentially be used as differentiating features between causes of optic disc swelling. It is also possible that with a different neural network architecture or more data augmentation strategies, less training data would be needed to achieve similar results as those shown with traditional machine learning methods. Irregardless of which method is used, more data will allow future analysis to become even more robust and accurate for qualitatively and quantitatively assessing optic disc edema via fundus photography.

REFERENCES

[1] J. W. Miller, "Differentiation between causes of optic disc swelling using retinal layer shape features," Master's thesis, University of Iowa, 2018.

[2] Palanker, "Diagram of the eye and placement of the retinal implants," 2016, wikimedia Commons.

[3] J. Agne, "Automated image-based estimation of severity and cause of optic disc edema," Ph.D. dissertation, University of Iowa, 2017.

[4] N. Donges and N. Donges, "The random forest algorithm," Feb 2018. [Online]. Available: https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 1–8. [Online]. Available: http://arxiv.org/abs/1505.04597

[6] G. P. V. Stavern, "Optic disc edema." *Seminars in neurology*, vol. 27 3, pp. 233–43, 2007.

[7] M. Wall and D. George, "Idiopathic intracranial hypertension: a prospective study of 50 patients," *Brain*, vol. 114, no. 1, pp. 155–180, 1991.

[8] L. Frisén, "Swelling of the optic nerve head: a staging scheme," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 45, no. 1, pp. 13–18, 1982.

[9] C. J. Scott, R. H. Kardon, A. G. Lee, L. Frisén, and M. Wall, "Diagnosis and grading of papilledema in patients with raised intracranial pressure using optical coherence tomography vs clinical expert assessment using a clinical staging scale," *Archives of Ophthalmology*, vol. 128, no. 6, pp. 705–711, 2010.

[10] A. J. Sinclair, M. A. Burdon, P. G. Nightingale, T. D. Matthews, A. Jacks, M. Lawden, A. Sivaguru, B. J. Gaskin, S. Rauz, C. E. Clarke, and A. K. Ball, "Rating papilloedema: an evaluation of the frisén classification in idiopathic intracranial hypertension," *Journal of Neurology*, vol. 259, no. 7, pp. 1406–1412, Jul 2012. [Online]. Available: https://doi.org/10.1007/s00415-011-6365-6

[11] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito *et al.*, "Optical coherence tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[12] S. Echegaray, G. Zamora, H. Yu, W. Luo, P. Soliz, and R. Kardon, "Automated analysis of optic nerve images for detection and staging of papilledema," *Investigative Ophthalmology and Visual Science*, vol. 52, no. 10, pp. 7470–7478, 2011.

[13] M. K. Garvin, "Automated 3-D segmentation and analysis of retinal optical coherence tomography images," Ph.D. dissertation, 2008.

[14] J. K. Wang, R. H. Kardon, M. J. Kupersmith, and M. K. Garvin, "Automated quantification of volumetric optic disc swelling in papilledema using spectral-domain optical coherence tomography," *Investigative Ophthalmology and Visual Science*, vol. 53, no. 7, pp. 4069–4075, 2012.

[15] J. Agne, J.-K. Wang, R. H. Kardon, and M. K. Garvin, "Determining degree of optic nerve edema from color fundus photography," in *Proc. SPIE 9414, Medical Imaging 2015: Computer-Aided Diagnosis*, 2015, pp. 94 140F_1–94 140F_9.

[16] J.-K. Wang, M. S. Miri, R. H. Kardon, and M. K. Garvin, "Automated 3-D region-based volumetric estimation of optic disc swelling in papilledema using spectral-domain optical coherence tomography," in *Proc. SPIE 8672, Medical Imaging 2013: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 2013, pp. 867 214_1–867 214_8.

[17] M. Fontal, J. Kerrison, R. Garcia, and V. Oria, "Ischemic Optic Neuropathy," *Seminars in Neurology*, vol. 27, no. 3, pp. 221–232, jul 2007.

[18] M. G. Hattenhauer, J. A. Leavitt, D. O. Hodge, R. Grill, and D. T. Gray, "Incidence of Nonarteritic Anterior Ischemic Optic Neuropathy," *American Journal of Ophthalmology*, vol. 123, no. 1, pp. 103–107, 1997.

[19] E. J. Atkins, B. B. Bruce, N. J. Newman, and V. Biousse, "Treatment of Nonarteritic Anterior Ischemic Optic Neuropathy," *Survey of Ophthalmology*, vol. 55, no. 1, pp. 47–63, 2009.

[20] H. E. Killer, G. P. Jaggi, and N. R. Miller, "Papilledema revisited: is its pathophysiology really understood?" *Clinical & Experimental Ophthalmology*, vol. 37, no. 5, pp. 444–447, 2009.

[21] M. Rigi, S. J. Almarzouqi, M. L. Morgan, and A. G. Lee, "Papilledema : epidemiology , etiology , and clinical management," *Journal of Eye and Brain*, vol. 2015, no. 7, pp. 47–57, 2015.

[22] D. I. Friedman and D. M. Jacobson, "Diagnostic criteria for idiopathic intracranial hypertension," *Neurology*, vol. 59, no. 10, pp. 1492–1495, 2002.

[23] J. Chiang, E. Wong, A. Whatham, M. Hennessy, M. Kalloniatis, and B. Zangerl, "The usefulness of multimodal imaging for differentiating pseudopapilloedema and true swelling of the optic nerve head: a review and case series," *Clinical and Experimental Optometry*, vol. 98, no. 1, pp. 12–24, 2015. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/cxo.12177

[24] B. L. Lam, C. G. M. Jr., and J. Pasol, "Drusen of the optic disc," *Current Neurology and Neuroscience Reports*, vol. 8, pp. 404–408, 2008.

[25] K. Li, W. Xiaodong, D. Chen, and M. Sonka, "Optimal Surface Segmentation in Volumetric Images-A Graph-Theoretic Approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 119–134, Jan 2006.

[26] M. K. Garvin, M. D. Abràmoff, X. Wu, S. R. Russell, T. L. Burns, and M. Sonka, "Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images," *IEEE Transactions on Medical Imaging*, vol. 28, no. 9, pp. 1436–1447, 2009.

[27] L. Tang, R. H. Kardon, J. K. Wang, M. K. Garvin, K. Lee, and M. D. Abràmoff, "Quantitative evaluation of papilledema from stereoscopic color fundus photographs," *Investigative Ophthalmology and Visual Science*, vol. 53, no. 8, pp. 4490–4497, 2012.

[28] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct 2001. [Online]. Available: https://doi.org/10.1023/A:1010933404324

[29] ——, *Classification and regression trees*, ser. Wadsworth statistics/probability series. Wadsworth International Group, 1984.

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[31] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*, ser. .NET Developers Series. Elsevier Science, 2015.

[32] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning, 12th printing*, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2017.

[33] M. S. Miri, M. D. Abràmoff, K. Lee, M. Niemeijer, J.-K. Wang, Y. H. Kwon, and M. K. Garvin, "Multimodal Segmentation of Optic Disc and Cup from SD-OCT and Color Fundus Photographs Using a Machine-Learning Graph-Based Approach," *IEEE Transactions on Medical Imaging*, vol. 34, no. 9, pp. 1854–1866, 2015.

[34] J.-K. Wang, "Estimation of papilledema severity using spectral-domain optical coherence tomography," Ph.D. dissertation, University of Iowa, 2016.

[35] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *arXiv preprint arXiv:1312.6229*, 2013.

[36] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[37] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[38] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner *et al.*, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[39] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[40] W. Yao, Z. Zeng, C. Lian, and H. Tang, "Pixel-wise regression using u-net and its application on pansharpening," *Neurocomputing*, vol. 312, pp. 364–371, 2018.

[41] S. S. Johnson, J. Wang, M. S. Islam, M. J. Thurtell, R. H. Kardon, and M. K. Garvin, "Local estimation of the degree of optic disc swelling from color fundus photography," in *Computational Pathology and Ophthalmic Medical Image Analysis - First International Workshop, COMPAY 2018, and 5th International Workshop, OMIA 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16-20, 2018, Proceedings*, 2018, pp. 277–284. [Online]. Available: https://doi.org/10.1007/978-3-030-00949-6_33

[42] S. S. Hayreh, "Optic disc edema in raised intracranial pressure - v. pathogenesis," *Archives of Ophthalmology*, vol. 95, no. 9, pp. 1553–1565, 1977. [Online]. Available: http://archopht.jamanetwork.com/article.aspx?articleid=632282

[43] R. Roslan and N. Jamil, "Texture feature extraction using 2-D Gabor filters," in *2012 International Symposium on Computer Applications and Industrial Electronics (ISCAIE)*, Dec 2012, pp. 173–178.

[44] R. M. Haralick, I. Dinstein, and K. Shanmugam, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.

[45] S. S. Johnson, J. Wang, M. S. Islam, M. J. Thurtell, R. H. Kardon, and M. K. Garvin, "Deep-Learning-Based Estimation of Regional Volumetric Information from 2D Fundus Photography in Cases of Optic Disc Swelling," *Investigative Ophthalmology & Visual Science*, 2019.

[46] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018.

[47] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega, and D. R. Webster, "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 12 2016. [Online]. Available: https://doi.org/10.1001/jama.2016.17216

[48] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Muller, M. P. Braeuning, and S. M. Pizer, "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *Journal of Digital Imaging*, vol. 11, no. 4, p. 193, Nov 1998. [Online]. Available: https://doi.org/10.1007/BF03178082