Deep Learning for Optical Coherence Tomography Angiography: Quantifying Microvascular Changes in Diabetic Retinopathy

by

Julian Lo

BASc (Hons., Engineering Science), Simon Fraser University, 2018

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Applied Science

in the
School of Engineering Science
Faculty of Applied Sciences

© Julian Lo 2020
SIMON FRASER UNIVERSITY
Fall 2020

Declaration of Committee

Name: **Julian Lo**

Degree: **Master of Applied Science**

Title: **Deep Learning for Optical Coherence Tomography Angiography: Quantifying Microvascular Changes in Diabetic Retinopathy**

Committee: **Chair: Yifan Jian**
Adjunct Professor, Engineering Science

Marinko V. Sarunic
Supervisor
Professor, Engineering Science

Eduardo V. Navajas
Committee Member
Clinical Assistant Professor, Ophthalmology and Visual Sciences
University of British Columbia

Mirza Faisal Beg
Examiner
Professor, Engineering Science

Abstract

Optical Coherence Tomography Angiography (OCT-A) permits visualization of the changes to the retinal circulation due to diabetic retinopathy (DR), a microvascular complication of diabetes. Machine learning applications have directly benefited ophthalmology, leveraging large amounts of data to create frameworks to aid clinical decision-making. In this thesis, several techniques to quantify the retinal microvasculature are explored. First, high-quality, averaged, 6x6mm OCT-A *enface* images are used to produce manual segmentations for the corresponding lower-quality, single-frame images to produce more training data. Using transfer learning, the resulting convolutional neural network (CNN) segmented the superficial capillary plexus and deep vascular complex with performance exceeding inter-rater comparisons. Next, a federated learning framework was designed to allow for collaborative training by multiple participants on a de-centralized data corpus. When trained for microvasculature segmentation, the framework achieved comparable performance to a CNN trained on a fully-centralized dataset.

Keywords: optical coherence tomography, angiography, image processing, machine learning, deep neural networks, federated learning

Acknowledgements

The completion of this thesis could not have been accomplished without the support of my senior supervisor, Dr. Marinko V. Sarunic. He has provided me with his expertise and guidance throughout my graduate degree, and shared with me his passion for research. I would also like to sincerely thank Dr. Mirza Faisal Beg, whose expertise in medical image processing was essential for the completion of this thesis. I would also like to sincerely thank Dr. Eduardo V. Navajas, who provided key insights from a clinical perspective.

I am very grateful to be a part of the Biomedical Optics Research Group (BORG) at SFU, where I have had the opportunity to work with many dedicated and hardworking individuals. I am especially grateful to Dr. Morgan Heisler and Dr. Yifan Jian for their mentorship over the years, and to Dr. Da Ma for providing invaluable insights in our many discussions. I would also like to acknowledge Mr. Timothy Yu, who I mentored and worked with closely in recent months. To the rest of BORG, thank you for sharing your support and encouragement.

Lastly, I would like to thank my family and friends for their continuous support and encouragement in all my endeavours.

Table of Contents

List of Tables

List of Figures

List of Acronyms

CNN	Convolutional Neural Network
COVID-19	Coronavirus Disease 2019
DNN	Deep Neural Network
DSC	Dice Similarity Coefficient
DVC	Deep Vascular Complex
ETDRS	Early Treatment of Diabetic Retinopathy Study
FA	Fluoroscein Angiography
FAZ	Foveal Avascular Zone
FN	False Negative
FOV	Field-of-view
FP	False Positive
GCL	Ganglion Cell Layer
ICA	Inter-capillary Area
ILM	Inner Limiting Membrane
INL	Inner Nuclear Layer
IPL	Inner Plexiform Layer
MIP	Maximum Ischemic Point
OCT	Optical Coherence Tomography
OCT-A	Optical Coherence Tomography Angiography
OPL	Outer Plexiform Layer
RPC	Radial Peripapillary Capillary
SCP	Superficial Capillary Plexus
SNR	Signal-to-noise Ratio
TN	True Negative
TP	True Positive

Chapter 1.   Introduction

1.1. Motivation

The advent of machine learning and artificial intelligence has sparked significant global interest, resulting in tremendous leaps of progress across numerous areas. Machine learning uses representation learning to process raw data without explicit, programmed instructions. Allowing the machine to learn the features to perform a specific task captures minute nuances in real-world data that would be otherwise overlooked in manual feature engineering. Consequently, machine learning has been used in many applications found in daily life, including social media, navigation, and e-commerce.

Machine learning applications have directly benefited the healthcare industry, with tools developed for diagnosis and prognosis, drug development, and epidemiology. These tools leverage large amounts of data for methods which include medical image processing and pattern recognition, leading to large-scale frameworks that aid decision-making processes for clinicians. This thesis solely focuses on the usage of these frameworks in ophthalmology, which has seen a substantial increase in machine learning publications in recent years.

Vision is the most complex and developed of the five senses, and one that is frequently taken for granted. Diabetic retinopathy (DR) is one of the leading causes of vision loss, and the most common cause of vision loss among people with diabetes mellitus. Because the eye is an exceptionally delicate organ, there is significant interest in non-invasive imaging modalities to identify and track the progression of pathologies such as DR. An overview of DR is provided in Section 2.1.1. One of the most widely-used modalities for diagnostic imaging of the retina is the rapidly-evolving optical coherence tomography angiography (OCT-A), a variant of optical coherence tomography (OCT) which allows for volumetric imaging of the microvasculature of the retina, a light-sensitive tissue in the eye.

OCT-A images are information-rich, and prohibitively time-consuming to manually delineate the vessels for quantitative analysis and classification of DR. As such, the topic

of this thesis is focused on methods to automatically segment the microvasculature in OCT-A *enface* images, utilizing machine learning techniques such as deep learning, transfer learning, and federated learning.

1.2. Contributions

This master's thesis research project built upon the work in my bachelor's thesis, which involved the automated segmentation of vasculature in retinal funduscopic colour photos. The first major project of my post-graduate career was the development and maintenance of a deep neural network (DNN) for retinal microvasculature segmentation in commercial OCT-A images. Solutions for microvasculature segmentation already existed in commercial OCT systems, however the quality was inadequate for accurate analysis. In particular, the signal-to-noise ratio of images in the deep vascular complex (DVC) was too high for traditional image processing approaches.

Methods for semantic segmentation using DNNs have been widely explored in the literature. As a part of my research project, I investigated leveraging existing high-quality averaged OCT-A images to create additional training data. This was combined with a technique called transfer learning to significantly improve segmentation quality. The DNN was directly applied to patient data obtained from the Eye Care Center through retinal specialists Dr. Sonja Karst, Dr. Vinicius Vanzan, and Dr. Eduardo Navajas, allowing a unique opportunity to see how machine learning image processing techniques directly benefited clinicians and patients. My findings led to a first-authored publication [A1], with additional quantitative analyses, conducted by Dr. Karst, resulted in an additional co-authored publication [A2].

Through my work developing the DNN, it was apparent that the quantity of available data was unsuitable for a fully robust framework for microvasculature segmentation. To expand the capability of the DNN to be hardware-agnostic, additional datasets from different machines and institutions were required. Because patient data is securely kept behind privacy regulations, I investigated the usage of federated learning to allow for the usage of data from collaborators without direct interaction. I developed a federated learning framework, with a cloud-based storage solution in SFU Vault as a method of exchanging files between participants. The framework was initially tested on my microvasculature segmentation data, with eventual repurposing towards diabetic

retinopathy classification, a project handled by another member of BORG, Mr. Timothy Yu. This was done in collaboration with Dr. Yali Jia from the Oregon Health and Science University, as well as Dr. Aaron Lee from the University of Washington. I mentored Timothy as he worked with members of Dr. Jia and Dr. Lee's respective teams to deploy the framework on real patient data, obtaining results that were presented at the Gained in Translation 2020 conference.

The body of work done during my graduate career comprises one first-authored journal publication, and a co-author on three additional publications. During my graduate studies, I have given one podium presentation at an international conference. All peer-reviewed journal publications on which I am an author are listed in Appendix 1 [A1-A4]. These publications are representative of my contributions to the development and implementation of novel deep learning methods for ophthalmological image analysis, which is the focus of this thesis.

1.3. Thesis Organization

The remaining chapters of this thesis are organized as follows. Chapter 2 presents background information on topics discussed in this thesis which include: diabetic retinopathy and machine learning techniques. Chapter 3 outlines the use of a machine learning framework to segment and quantify the retinal microvasculature. This chapter demonstrates the use of a transfer learning method, combined with training data generated through a separate neural network to obtain high-quality segmentations of both the superficial capillary plexus and deep vascular complex. This chapter also outlines changes made to the framework since its submission to a peer-reviewed journal. Chapter 4 outlines the use of federated learning to develop a framework for hardware agnostic microvasculature segmentation, as well as possible expansion to diabetic retinopathy severity classification. Chapters 5 and 6 discuss future research directions in the field of ophthalmic image processing, as well as machine learning.

Chapter 2.   Background

2.1. Eye Anatomy

The eye is a complex sense organ that allows for the perception of light, and its subsequent conversion to electrical signals that the brain can interpret. Figure 2.1 shows a simple schematic of a human eye.
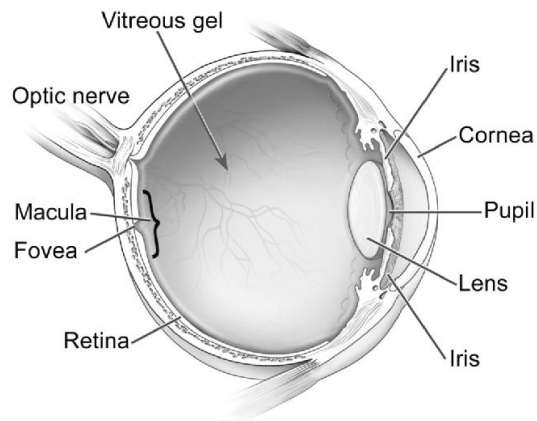
Figure 2.1.   Anatomy of the human eye

To enable vision, incoming light is focused by the cornea and the lens onto the retina. The retina is a layer of tissue that contains layers of photoreceptors that react to light, initiating a chain of chemical and electrical events that send neural impulses through the optic nerve to the visual cortex in the brain.

The retina is perfused by a network of capillaries, which can be differentiated into four capillary plexuses: the radial peripapillary, superficial, intermediate, and deep capillary plexuses. The radial peripapillary capillaries (RPCs) are the most superficial layer of capillaries lying in the inner part of nerve fiber layer (NFL), and feed the superficial nerve fibers surrounding the ONH. The superficial capillary plexus (SCP) is supplied by the central retinal artery and composed of various vessel types resides primarily in the ganglion cell layer (GCL) and anterior half of the inner plexiform layer (IPL). Below this is the ICP, which is supplied by vertical anastomoses from the SCP and contains the capillaries encapsulated by the posterior half of the IPL and the anterior half of inner nuclear layer (INL). Lastly, the DCP is also supplied via vertical anastomoses of

the SCP and is considered to be the vessels contained within the posterior half of the INL and the outer plexiform layer (OPL). Due to difficulties delineating the boundary between the intermediate and deep capillary plexuses after retinal volumetric imaging, these are frequently presented together as the deep vascular complex (DVC).

2.1.1. Diabetic Retinopathy

Diabetic retinopathy (DR) is a complication of diabetes mellitus, the most common cause of vision loss among people with diabetes, which affects 749,800 Canadians [1], DR damages the structure of the retinal microvasculature [2], leading to widespread areas of ischemia as it progresses. DR has numerous categorization schemes, however one of the most commonly used is the 5 stage International Clinical Diabetic Retinopathy categorization of no retinopathy, mild, moderate, severe non-proliferative retinopathy, and proliferative retinopathy [3]. Resultant decreases in perfusion density and enlargement of the foveal avascular zone (FAZ) have been correlated to a lower level of visual acuity [4], [5] in patients. Therefore, there is great clinical value in a framework that can accurately quantify these microvascular changes to improve DR treatment and patient prognoses.

2.1.2. Summary

Diabetic retinopathy has characteristic changes in retinal microvasculature which can be quantified through ophthalmic imaging. Methods of ophthalmic imaging of the retina are described in the next section.

2.2. Retinal Imaging

Obtaining clear images of the retina is essential in all ophthalmic clinics. With funduscopic photos being the gold standard in diabetic retinopathy diagnosis, optical coherence tomography provides clinicians with additional options for high-resolution imaging. In addition, angiographic imaging techniques such as optical coherence tomography angiography provide clinicians with additional details that cannot be otherwise obtained from structural imaging.

2.2.1. Optical Coherence Tomography

Optical coherence tomography (OCT) is a non-invasive, *in vivo* imaging technique that uses low-coherence near-infrared light to capture three-dimensional, volumetric tomographs. Cross-sections of these imaged volumes, termed B-scans, consist of depth scans, known as A-scans. OCT produces micrometer-resolution images up to 1 or 2 millimeters of penetration depth.

Swept source OCT has enjoyed recent popularity due to its increased acquisition speed over the previous standard, spectral domain OCT. These systems use a swept source laser with a wavelength centered at 1 µm that sweeps across a narrow band of wavelengths, with a point photodetector to detect the returning light waves. The photodetector detects the interference spectrum caused by the recombined backscattered light from the reference and sample arms. From this, the Fourier transform is calculated to obtain an A-scan, the reflectivity as a function of depth.

2.2.2. Optical Coherence Tomography Angiography

Optical Coherence Tomography Angiography (OCT-A) is an emerging variant of OCT with which the retinal microvasculature can be more easily visualized. OCT-A uses a speckle variance approach by computing the OCT speckle difference between multiple repeated B-scans at the same location. By applying the retinal layer boundaries found in the OCT structural volume to the OCT-A volume, the SCP and DVC can be resolved for analysis.

2.2.3. Summary

High-resolution images of the retina are facilitated through the imaging modalities explored in this section. Retinal perfusion can be visualized through angiographic imaging. Machine learning techniques that can be used for quantification and analysis of these images are explained in the next section.

2.3. Machine Learning Techniques

Machine learning is an application of artificial intelligence that involves algorithms that improve through experience, as opposed to traditional rule-based programming. This thesis will focus on a subset of machine learning, called supervised learning, where a mapping function is approximated based on a set of desired inputs and outputs (e.g. images and labels), known as the training data. The objective is to be able to accurately infer outputs from a new set of inputs. These mapping functions can be represented through numerous algorithms, which include support vector machines, random forest classifiers, and neural networks.

2.3.1. Neural Networks

A neural network (model) is a machine learning algorithm which learns representations of the data to extract useful features to perform tasks such as regression, classification, and simple processing. Neural networks are composed of layers of neurons or nodes, and are modelled after biological neural networks, which contain neurons that receive inputs through dendrites and send outputs through axons. Typically, neural networks contain an input layer, and output layer, with intermediate layers in between. Each connection between each node has a learnable weight, and each nodal output is a linear combination of each of the outputs of the preceding nodes and their corresponding weights, subsequently multiplied by an activation function. Basic neural networks that contain only fully-connected layers, where each node in a given layer is connected to all nodes in the previous layer, are known as multilayer perceptrons. In general, neural networks with more than one intermediate layer are called deep neural networks (DNNs). Different DNN architectures can be separated into additional categorizations, including recurrent neural networks, which are best-suited for temporal data, and convolutional neural networks, which are best-suited for images.
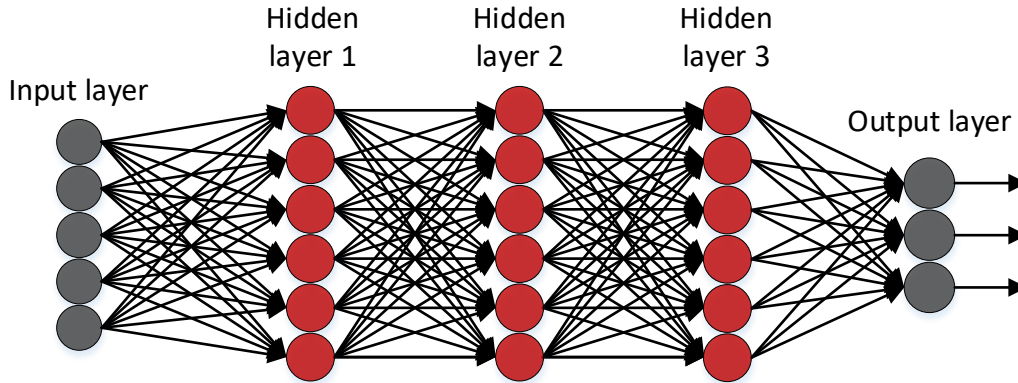
Figure 2.2.   **Multilayer perceptron with three hidden layers**

Convolutional Neural Networks

Convolutional neural networks (CNNs) are a class of DNNs, which are generally applied to image processing. CNNs are able to assign weights and biases to individual features in the image as part of the training process and can take into account spatial dependencies between these features. CNNs are most notably used for image classification and have been shown to outperform traditional image processing methods due to their ability to capture minute nuances that would be otherwise missed in algorithm-based programming. CNNs typically make use of convolutional layers and pooling layers, in addition to the fully-connected layers typically seen in multilayer perceptrons.

Network Architectures

The first network architecture used in this thesis is U-Net [6], a fully-convolutional network (meaning it lacks any fully-connected layers) developed for semantic segmentation. Its structure consists of a contracting and expanding path. The contracting path follows a typical CNN structure, where convolutional layers extract features, and pooling layers progressively down-sample feature maps to extract context. The symmetric expanding path uses convolutional layers and up-convolutional layers that propagate contextual information to higher-resolution layers. In addition, span connections are added for each level, where high resolution weights are combined with the upsampled, generalized weights in the corresponding level of the expanding path. This allows the network to retain the learned localization information and better segment smaller, more detailed structures in the image.

Residual blocks have seen success in ResNet [7] and provide a solution to the vanishing gradient problem that has emerged as a result of using excessive numbers of layers in CNNs. As the gradient backpropagates to earlier layers during training, successive multiplications reduce the gradients to zero. As shown in Figure 2.3, an added skip connection bypasses weight layers, which allow gradients to flow back to earlier layers in the network. This reduces the complexity of the network (and its corresponding mapping function), smoothing the "loss landscape" [8] and allowing local minima to be more easily found. As a result, training and test accuracy degradation are mitigated even when more layers are added.
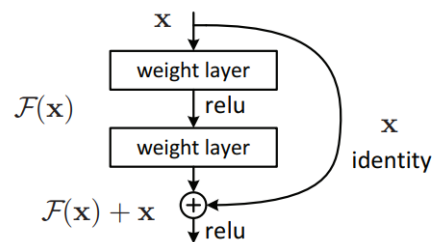
Figure 2.3.   Residual block, as shown in [7].

Residual blocks can be subsequently added to the U-Net architecture [9], which is shown in Figure 2.4. The identity function is implemented through a 1x1 convolutional layer, which ensures the representation shapes are compatible when adding.
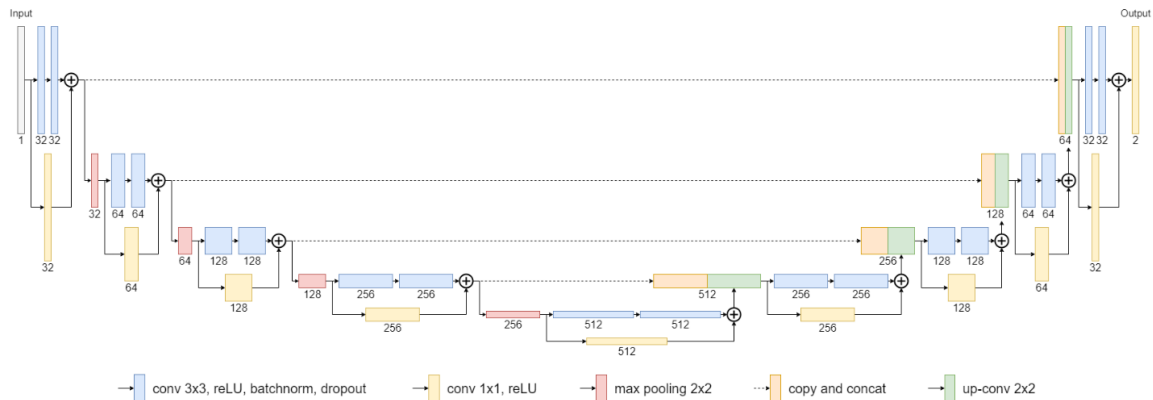
Figure 2.4.   Residual U-Net architecture

9

2.3.2. Deep Learning

Deep learning is a subset of machine learning which uses DNNs. Due to the larger number of weights and parameters involved with DNNs, these allow for deep abstractions and high-level representations within its nodes. DNN's benefit from larger datasets, where an increased number of training examples improves its ability to generalize, meaning its ability to apply learned representations from its training data to new data.

2.3.3. Federated Learning

As deep learning applications increase in complexity, the amount of data required to train a robust, accurate network increases significantly. However, for medical applications, data is securely guarded behind privacy regulations and additional legal hurdles, which include the European Union's General Data Protection Regulation and the United States Health Insurance Portability and Accountability Act. Federated learning is a method of collaborative deep learning between groups that is designed to allow multiple parties to jointly train a model without directly sharing sensitive data. As shown in Figure 2.5, each participating group (client) trains a model on a secure local server only using their own data. Only the weights are sent to a central aggregate server, where all client model updates are globally aggregated. The resulting aggregate model is subsequently distributed to all clients to repeat the process.

2.3.4. Summary

Machine learning techniques provide essential tools for quantitative image analysis. In the next chapter, I will discuss how these techniques help aid the segmentation and quantification of the retinal microvasculature.
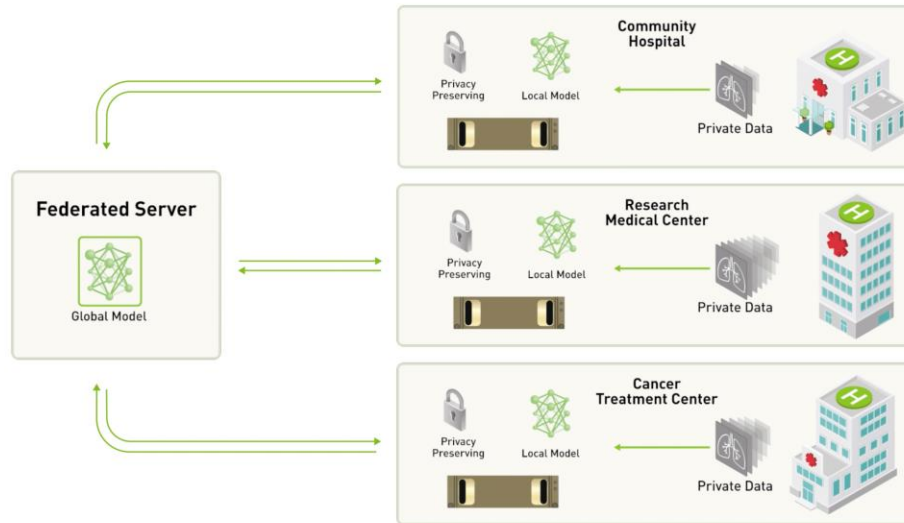
Figure 2.5. **General federated learning schematic. Image from [10]**

Chapter 3.   Microvasculature Segmentation and Inter-capillary Area Quantification of the Deep Vascular Complex using Transfer Learning[1]

3.1. Abstract

3.1.1. Purpose:

Optical Coherence Tomography Angiography (OCT-A) permits visualization of the changes to the retinal circulation due to diabetic retinopathy (DR), a microvascular complication of diabetes. We demonstrate accurate segmentation of the vascular morphology for the superficial capillary plexus (SCP) and deep vascular complex (DVC) using a convolutional neural network (CNN) for quantitative analysis.

3.1.2. Methods:

The main CNN training dataset consisted of retinal OCT-A with a 6x6mm field of view (FOV), acquired using a Zeiss PlexElite. Multiple-volume acquisition and averaging enhanced the vasculature contrast used for constructing the ground-truth for neural network training. We used transfer learning from a CNN trained on smaller FOVs of the SCP acquired using different OCT instruments. Quantitative analysis of perfusion was performed on the resulting automated vasculature segmentations in representative patients with DR.

3.1.3. Results:

The automated segmentations of the OCT-A images maintained the distinct morphologies of the SCP and DVC. The network segmented the SCP with an accuracy and Dice index of 0.8599 and 0.8618, respectively, and 0.7986 and 0.8139 for the DVC.

---

[1] This work has been published in *Translational Vision Science & Technology.* [J. Lo, M. Heisler, V. Vanzan, S. Karst, I. Z. Matovinović, S. Lončarić, E. V. Navajas, M. F. Beg, and M. V. Šarunić, "Microvasculature Segmentation and Intercapillary Area Quantification of the Deep Vascular Complex Using Transfer Learning," Transl. Vis. Sci. Technol., vol. 9, no. 2, p. 38, Jul. 2020.]

The inter-rater comparisons for the SCP had an accuracy and Dice index of 0.8300 and 0.6700, respectively, and 0.6874 and 0.7416 for the DVC.

3.1.4. Conclusions:

Transfer learning reduces the amount of manually-annotated images required, while producing high quality automatic segmentations of the SCP and DVC that exceed inter-rater comparisons. The resulting inter-capillary area quantification provides a tool for in-depth clinical analysis of retinal perfusion.

3.1.5. Translational Relevance:

Accurate retinal microvasculature segmentation with the CNN results in improved perfusion analysis in diabetic retinopathy.

3.2. Introduction

Diabetic retinopathy (DR) is a complication of diabetes mellitus, the most common cause of vision loss among people with diabetes, which affects 749,800 Canadians [1]. DR damages the structure of the capillaries in the retina [2], leading to widespread areas of retinal ischemia as it progresses. Optical coherence tomography angiography (OCT-A) is a rapidly emerging imaging technology that allows for the retinal microvasculature to be seen volumetrically in micrometer-scale detail [11], [12]. OCT-A has shown to produce images that closely relate to histology[13]–[16], and presents a noninvasive and dye-free alternative with a lower risk of complications [17] when compared to the current gold standard, fluoroscein angiography (FA).

Analysis and quantification of the retinal microvasculature benefits from multi-scale imaging with fields-of-view (FOV's) ranging from ~2x2mm to ~6x6mm. At a smaller FOV, the capillaries that comprise the structure of the microvasculature can be individually resolved, whereas with a larger FOV, macroscopic features, including regions of capillary non-perfusion, can be identified. With recent research hypothesizing that early manifestations of DR form in the retinal periphery [18], improving the vessel segmentations and tools for quantification in wider fields of view for both the superficial capillary plexus (SCP) and deep vascular complex (DVC) [19] are important assets to

13

clinicians. The DVC resulted from combining the intermediate and deep capillary plexuses due to difficulty in resolving each plexus individually, and has shown a higher correlation to retinal ischemia in DR [20], [21].

OCT-A images are information-rich, and time-consuming for a clinician to trace the vessels for detailed analysis. For the cases of highest clinical interest, small changes in the capillaries need to be detected. Consequently, accurate automated methods of microvasculature segmentation are an essential step towards quantification. However, the efficacy of traditional image processing algorithms can vary based on artifacts present in the image, most notably from noise. Segmentation of microvasculature in funduscopic photos, as well as FA have been examined [22], however fewer specific algorithms developed for OCT-A have been developed. Simple thresholding of the OCT-A image intensity has been applied [23], [24], but these approaches pose numerous drawbacks in its invariance to microvasculature features and performance when applied to lower-quality images with a low signal-to-noise ratio (SNR). The disadvantages of thresholding can be seen in Figure 3.1, where an image of the DVC in a patient with mild diabetic retinopathy was processed with Otsu's method [25]. This representative example of thresholding demonstrates that vessels do not maintain continuity, and a significant portion of the speckle is erroneously delineated as a vessel.
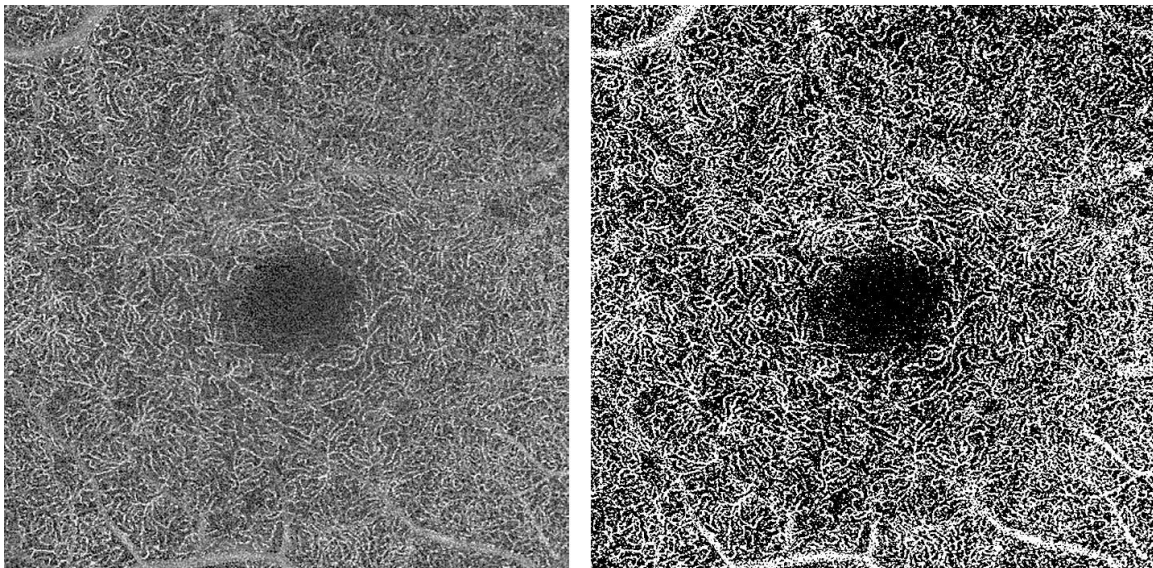
Figure 3.1.   **Left: original single-frame image of the deep vascular complex. Right: image thresholded using Otsu's Method.**

Methods using vesselness filters have been developed [26], [27], but are either similarly limited by SNR, or require manual correction. A tophat filter and optimally oriented flux method for segmenting the vessels [28] has been implemented and demonstrated for brain imaging in mice. In addition, some commercial OCT systems also provide segmentation of the vessels but generally face the same issues with images with low SNR.

Machine learning is a rapidly growing field, showing promising results for numerous ophthalmological applications. A few recent reports that are related to this topic include: retinal layer [29]–[31] and capillary plexus [32] segmentation, cone photoreceptor identification [33], [34], macular fluid segmentation [35], geographic atrophy segmentation [36], OCT image categorization [37], diagnosis and referral for retinal disease patients [38], [39], and synthesis of funduscopic images [40]. Additionally, recent reports have published online tools to improve the accessibility of machine learning-based retinal layer segmentation through intuitive user interfaces that can be used directly by clinicians [41]. Machine learning algorithms have also been applied towards OCT-A segmentation, with a recent approach (MEDnet) [42] applying a convolutional neural network (CNN) to identify and segment avascular areas in wide-field images of the SCP. We have also previously published a method of using a CNN to segment 1x1mm images of the SCP [43]. Machine learning algorithms using CNN's are well suited to address the issues of vessel segmentation through a series of trainable filters. These filters allow the segmentation to be sensitive to vessel boundaries, and hence also have the potential to preserve the vessel widths. However, even with the strengths of machine learning, the quality of the OCT-A images will have a significant impact on the results of the vessel segmentation and quantification.

In our previous work, we proposed a method to register and average multiple sequentially acquired OCT-A images in order to significantly improve image quality and vessel discernibility [44]. Related works in the Literature have also investigated averaging of OCT-A images to improve vessel contrast [45] and automated biomarker identification algorithms [46]. However, this requires prolonged imaging sessions, which is not always possible, particularly in a high-volume clinical environment. Therefore, there is greater clinical utility in the development of an algorithm that can accurately segment, and subsequently quantify, the vasculature and corresponding inter-capillary

areas (ICA's) from one single-frame OCT-A image. Quantification of individual ICA's has been previously explored [23], [26], but used 3x3mm images of the SCP. Similarly, quantification of the vasculature in the DVC has been explored, but did not include ICA's [27].

With many approaches presenting accurate analysis of the ICA's in the SCP, the contribution of this report is to describe an original and novel method to accurately and automatically segment and quantify these regions in the DVC. This used an approach of transfer learning, referred to as fine-tuning, for the segmentation of retinal microvasculature in single-frame, wide-field 6x6mm OCT-A images for the purposes of ICA quantification. The developed framework allows for the adaptation of an initial segmentation network to a new dataset with significantly fewer manually graded training examples. We combined the approach of OCT-A averaging to generate high contrast images of the vascular networks with supervised learning to provide the CNN with accurate ground truth data in order to guide the vessel segmentation even in the case of a single (unaveraged) OCT-A image. The computer-generated segmentations were qualitatively examined by retinal specialists and compared to manual segmentations from a trained rater. The outputs of the automated vessel analysis provide near-immediately available, quantitative information on the microvasculature and ICA's from a single OCT-A volume, and hence can potentially accelerate treatment plans and improve DR prognosis.

3.3. Methods

3.3.1. Subject Criteria and Data Preparation

Subject recruitment and imaging took place at the Eye Care Centre of Vancouver General Hospital, and North Shore Eye Associates. The project protocol was approved by the Research Ethics Boards at the University of British Columbia and Vancouver General Hospital, and the experiment was performed in accordance with the tenets of the Declaration of Helsinki. Written informed consent was obtained by all subjects.

Subjects in the control group (n = 8) displayed no evidence of retinal or ocular pathology upon examination by an experienced retina specialist. Subjects classified as diabetic (n = 28) were diagnosed with DR based on the international DR severity scale

[3]. All subjects were screened for clear ocular media, ability to fixate, and ability to provide informed consent before imaging. In addition, patients with diabetic macular edema were not included in the study.

Table 3.1. **Demographics of the control dataset used in this study.**

Gender	n	Mean Age (Standard Deviation)
Male	4	24.5 (3)
Female	4	53.5 (23.1)

3.3.2. Optical Coherence Tomography Instrumentation

The data used for this study was acquired with the ZEISS PlexElite (Carl Zeiss Meditec, Dublin, CA) with software version 1.7.31492. The nominal 6x6mm scanning protocol was used, sampling at a 500x500 resolution at a rate of 100,000 A-scans per second at a visual angle of 20.94 degrees. Each B-scan was repeated twice at the same position and the optical micro-angiography (OMAG) implemented on the commercial imaging system was used to generate the angiographic information. The A-scan depth was 3mm with an axial resolution of 6.3 µm and a transverse resolution of 20 µm, as described in the product specifications.

The inner limiting membrane (ILM), and posterior boundary of the outer plexiform layer (OPL) were used as the segmentation boundaries for the commercial device, with the inner plexiform layer (IPL)/inner nuclear layer (INL) complex used as the SCP/DVC boundary. The SCP and DVC were subsequently extracted, with projection artifacts removed, via a built-in software feature in the Zeiss PlexElite and exported at a 1024x1024 resolution. Scans were only included in the study if the system specified signal strength was 8 (out of 10) or higher.

3.3.3. Network Architecture

The network for vessel segmentation used a variation of the U-Net [6] architecture, which was adapted for two classes: vessel and background. The basic U-Net architecture is shown in Figure 3.2 and consists of convolutional and pooling layers. The convolutional layers consist of a series of trainable filters, which are correlated across the image and subsequently passed through a rectifier linear unit activation with

units capped at 6 (ReLU-6) [47]. Each convolutional layer was followed by a batch normalization layer, as well as a dropout layer with a coefficient of 0.5 [47]. Pooling layers were inserted to increase the receptive field of the subsequent filters in the convolutional layers, helping with generalization to prevent overfitting.
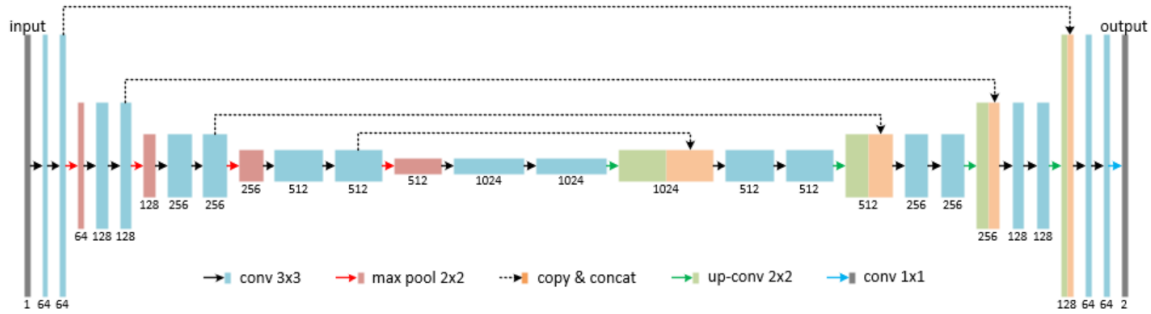
Figure 3.2.   U-net architecture.

The U-Net architecture provides high-resolution feature extraction through its structure, which consists of a contracting and expanding path. For each level in the contracting path, high resolution weights are combined with the upsampled, generalized weights (span connections) in the corresponding level of the expanding path. This allows the network to retain the learned localization information and better segment smaller, more detailed structures in the image. In addition, methods using residual blocks for each convolutional block were experimented, however this resulted in similar or lower performance (data not shown).

3.3.4. Training

Two OCT-A datasets were used for training the network. To construct the initial weights, data were acquired from a previous study [48]. Briefly, this data consisted of 29 images with a 2x2mm FOV acquired with a prototype swept source OCT instrument [49] and 47 images with a 3x3mm FOV acquired with a commercial spectral domain OCT instrument. Each OCT-A image was manually segmented using a Microsoft Surface Pro tablet and GNU image manipulation program (GIMP) by one trained rater and verified and accepted by two additional trained raters.

To construct the initial weights, the network hyperparameters were optimized using 3-fold cross-validation. This resulted in a network trained over 120 epochs using

the Adam optimizer, with an initial learning rate of 10^{-4}, and a custom epsilon value of 10^{-5}. Evaluation was performed qualitatively on a set of acquired 3x3mm FOV OCT-A images across all devices based on images most recently acquired at the clinic. Segmentation of a single 3x3mm or 2x2mm image using the network took approximately two seconds on a Nvidia RTX 2060 GPU, with a possible decrease to 0.3 seconds per image when segmented in larger batches of ~10 images.

The network with the initial weights was subsequently fine-tuned on the 6x6mm FOV images acquired with the PlexElite. This was done in two stages: First, a smaller dataset of 10 single-frame OCT-A images of each of the SCP and DVC for which there existed a corresponding high-quality averaged image was identified. As described in our previously-published study [44], images were registered and averaged based on a template image that was free of microsaccadic motion. This allowed us to use the averaged OCT-A images to construct ground-truth labels for each single-frame template image. These labeled ground truth vessel segmentations were subsequently paired with the single-frame template OCT-A images to train the deep neural network to perform segmentations approaching the quality of averaged images, while only using single-frame images.
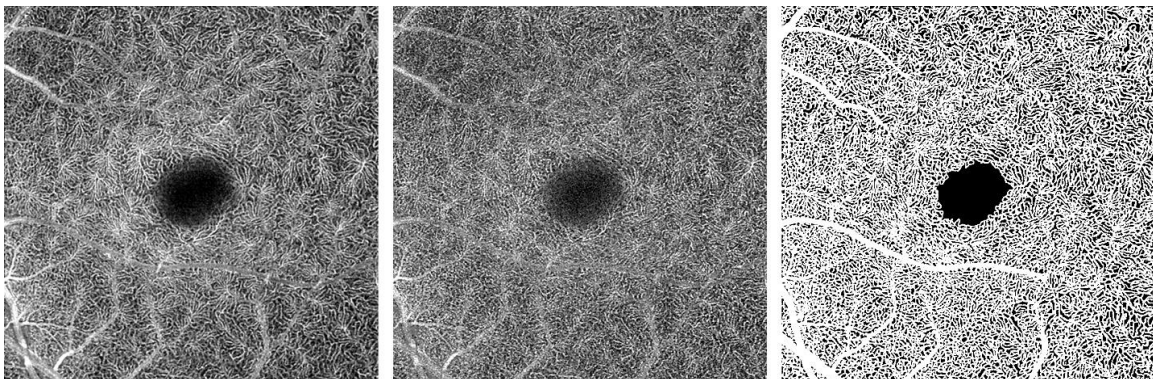
Figure 3.3.　　Left: 10-frame averaged 6x6mm image of the deep vascular complex (DVC). Centre: Single-frame template image of the same region. Right: Manually-segmented averaged image, to be paired with the template image for training.

For the SCP, the automated segmentations of the averaged images generated by the network with initial weights were adequate, as determined by a separate group of trained raters; hence these results were fed back into the network as additional training examples. However, the automated segmentations on the DVC required additional

19

manual correction due to the lower SNR of the images in this layer and the different morphological features of the vasculature relative to the SCP. Using the initial weights as a guideline for the manual raters would introduce biases that could negatively impact further stages of training; hence, the vasculature of the DVC images was instead first segmented through Otsu's method [25]. Another masked and trained rater manually corrected the resulting segmentation using a Microsoft Surface Pro tablet and GIMP. All segmentations were reviewed and accepted by two of three other trained raters. A second trained rater segmented three images to obtain inter-rater metrics.

Due to memory limitations when training, each 6x6 image was separated into 4 quadrants, which were saved as 4 separate images. The same method of augmentation and cross-validation was used. This resulted in the network being trained through the Adam optimizer, using an initial learning rate of 10^{-2}, and a custom epsilon value of 10^{-2}. The forward inference segmentation of a single 6x6mm image using the network took approximately four seconds, with a possible decrease to 0.5 seconds per image when segmented in larger batches.

To further reinforce the manually-segmented dataset, extensive data augmentation was performed. Each OCT-A image (along with its corresponding manual segmentation) in the training set was rotated 90 degrees three times with no processing. Next, to account for noise, each image was rotated 90 degrees an additional 5 times with various contrast adjustments, which included contrast-limited adaptive histogram equalization, as well as the built-in imadjust function in MATLAB. To account for motion, each rotated image was also separated into randomly-sized strips, which were re-ordered randomly to simulate motion artifacts in the image. The probability maps resulting from the automated segmentations were binarized at a value of 0.5, the default class cut-off in the probability map. After binarization, isolated clusters of less than 30 pixels were deemed false positives and removed.

Additional training data were required to improve the performance of the network on the DVC using a single frame OCT-A image. The intermediate network (after training on the 10 manually-segmented images) was able to segment additional 6x6mm averaged OCT-A images of the DVC. Subsequently, the next stage of fine-tuning the CNN involved using this intermediate network to generate additional training data, using manual inspection for quality, but not requiring laborious manual segmentation at the

scale of capillaries. We applied the intermediate network trained on the manual segmentations to all 50 averaged OCT-A images in our 6x6mm dataset. After manual inspection, we identified 39 images for each of the SCP and DVC with adequate automated segmentations that constituted the new training set. A summary of the training sets used is provided in Table 3.2.

Similarly to the first stage, the weights were initialized with the original network trained with the first dataset of 2x2mm and 3x3mm images. The same methods of image augmentation and training were applied, and cross-validation resulted in an initial learning rate of 10^{-2}, and a custom epsilon value of 10^{-2} using the Adam optimizer.

Table 3.2. **Overview of the three datasets used to train the fine-tuned network**

	2x2/3x3mm dataset	**First 6x6mm dataset**	**Second 6x6mm dataset**
Training images	SCP: 76 averaged and single-frame	SCP: 10 single-frame DVC: 10 single-frame	SCP: 39 single-frame DVC: 39 single-frame
Ground-truth segmentations	SCP: manual	SCP: automated DVC: manual	SCP: automated DVC: automated

3.3.5. Performance Evaluation

To evaluate the automated segmentation performance, a number of metrics were calculated. The number of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) were calculated using pixel-wise comparison between a reference manual segmentation and the thresholded binary output of the neural network. To calculate inter-rater metrics, these metrics were calculated by comparing one manual segmentation to another by a different rater. In the context of this study, pixels corresponding to vessels and the background comprised the positive and negative classes, respectively. Using the TP, FP, FN, and TN numbers we can calculate the accuracy of the segmentation, as shown in Equation 1:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{1}$$

Additionally, we can compute the Dice similarity coefficient (DSC), which quantifies the similarity between two segmented images, through measuring the degree

of spatial overlap. The DSC value ranges from 0, indicating no spatial overlap, and 1, indicating complete overlap and is calculated by Equation 2:

$$DSC = \frac{2TP}{2TP + FP + FN}.$$

(2)

Three methods were tested: using only the initial weights, using a network solely trained on the 6x6mm dataset, and the fine-tuned network, trained on both datasets.

3.3.6. Post-Processing of the Automated Vessel Segmentation

The neural network generated segmentation of the vessels in the OCT-A images, but further processing was required for quantitative analysis. Next, the ICA's were identified, determined by the non-vessel pixels. The largest ICA within a small region in the center of the image was defined as the foveal avascular zone (FAZ). All erroneously segmented pixels within the FAZ were set to a non-vessel classification, with the centroid then used as the center of the Early Treatment of Diabetic Retinopathy Study (ETDRS) grid.

Two metrics were of interest when quantifying ICA's: the area of the region and maximum ischemic point (MIP), defined as the point of maximum distance to nearest vessel within the ICA. The metrics were measured for each ETDRS region and compared to a database of healthy eyes, for which the SCP and DVC were extracted. As outlined in Section 3.3.1, 8 healthy controls were recruited, resulting in a possible 16 eyes. Of these, we were able to obtain high-quality averaged images for 12 to construct the reference database and were also included in the training dataset outlined in Table 3.2. Each measured ICA was color-coded and overlaid on the original image based on number of standard deviations from the mean. Perifoveal vessel density (for each ETDRS region) was also calculated as the proportion of measured area occupied by pixels which were classified by the algorithm as a vessel. In addition, the projection artifacts remaining in segmentations of the DVC were excluded for the calculation of vascular metrics including vessel density and vessel index. This was done through an automated MATLAB post-processing step using image erosion and dilation.

22

3.4. Results

3.4.1. Network Performance Evaluation

Quantitative Segmentation Comparison

Table 3.3 and Table 3.4 show comparative quantitative results when segmenting the SCP and DVC respectively. The network trained on the initial dataset of 2x2mm and 3x3mm images, the network solely trained on the 6x6mm dataset, and the network trained with our proposed transfer-learning method were enumerated as Network A, B, and C, respectively. The accuracy and Dice index for Network C showed a high similarity between segmentations of the single-frame template images and the averaged images. The inter-rater metrics were only conducted on the manually-segmented datasets and are intended to be a representative number illustrating the difficulty of this problem and the variation in the metrics between human raters. Table 3.5 shows the same networks but evaluated on the original 2x2mm and 3x3mm dataset.

Table 3.3. **Comparative quantitative results of the segmentation of the SCP between three networks: Network A consisted of only the initial weights, Network B was trained on only the images from the 6x6mm dataset, and Network C was the fine-tuned network using our proposed transfer learning method.**

	Network A	Network B	Network C	Inter-rater
Accuracy	0.8141	0.8534	0.8599	0.8300
Dice similarity index	0.8060	0.8586	0.8618	0.6700

Table 3.4. **Comparative quantitative results of the segmentation of the DVC between three networks: Network A consisted of only the initial weights, Network B was trained on only the images from the 6x6mm dataset, and Network C was the fine-tuned network using our proposed transfer learning method.**

	Network A	Network B	Network C	Inter-rater
Accuracy	0.6934	0.7822	0.7986	0.6874
Dice similarity index	0.6469	0.8065	0.8139	0.7416

23

Table 3.5. Comparative quantitative results of the segmentation of the 2x2mm and 3x3mm dataset between three networks: Network A consisted of only the initial weights, Network B was trained on only the images from the 6x6mm dataset, and Network C was the fine-tuned network using our proposed transfer learning method.

	Network A	Network B	Network C	Inter-rater
Accuracy	0.8677	0.8329	0.8350	0.8300
Dice similarity index	0.8395	0.8059	0.8066	0.6700

Qualitative Segmentation Comparison

The fine-tuned network was qualitatively evaluated on data unseen by the CNN during training on control and DR patients. Figure 3.4 focuses on a peripheral area of the SCP located close to the optic nerve head, where the elongated vascular structure of radial peripapillary capillaries (RPC's) are visible. As shown in Figure 3.4-C2, Network C (the fine-tuned network) preserves the features characteristic of the RPC's, and segments larger vessels more accurately than Network A (the initial weights). The differences between Network B (trained solely with the 6x6mm dataset) and Network C are less pronounced, due to the higher SNR present in images of the SCP.

Figure 3.4.   **A1: 2x2mm window of an averaged 6x6mm image taken of the superficial capillary plexus (SCP). A2: 2x2mm window of the corresponding region in the equivalent single-frame template image. B1: averaged image segmented using the initial weights (Network A). B2: single-frame image segmented using Network A. C1: averaged image segmented using the fine-tuned network (Network C). C2: single-frame image segmented using Network C. D: comparison between the automated segmentations between the averaged and single-frame images produced by Network C, represented by magenta and green respectively, with white representing the union. E: comparison between single-frame segmentations between Network B and C, represented by magenta and green respectively, with white representing the union.**

Figure 3.5 shows an additional enlarged comparison for an image of the SCP. It can be seen here that Network C is able to accurately segment the areas of ischemia observed in the averaged images, and as also shown in Figure 3.4, segments larger vessels more accurately.

Figure 3.5.   **A1: 2x2mm window of an averaged 6x6mm image taken of the superficial capillary plexus (SCP). A2: 2x2mm window of the corresponding region in the equivalent single-frame template image. B1: averaged image segmented using the initial weights (Network A). B2: single-frame image segmented using Network A. C1: averaged image segmented using the fine-tuned network (Network C). C2: single-frame image segmented using Network C. D: comparison between the automated segmentations between the averaged and single-frame images produced by Network C, represented by magenta and green respectively, with white representing the union. E: comparison between single-frame segmentations between Network B and C, represented by magenta and green respectively, with white representing the union.**

Figure 3.6Figure 3.6 shows an enlarged comparison of the segmentations results obtained by Network A and Network C when segmenting the elongated, lobular capillary structure of the DVC in a lower-quality image. As shown in Figure 3.6-B2, certain clusters of vessels were erroneously treated as noise by Network A, resulting in regions of false negatives. This is characteristic of single-frame OCT-A images; the blurred-out regions were replaced by a discernible vessel structure when using the averaged OCT-A images. The results presented in Figure 3.6-C2 are representative of the outputs from Network C, which eliminated a portion of these false negatives.

Residual projection artifacts from the SCP were automatically identified and are highlighted in cyan in Figure 3.6-C1 and Figure 3.6-C2. As outlined in Section 3.3.6, these regions obscure the capillaries underneath and are subsequently excluded as a post-processing step when calculating the vessel density and vessel index metrics.

When segmenting with Network A, these projection artifacts were erroneously segmented as additional capillaries. This can also be seen in Figure 3.6-E where the green spots along the projection artifacts indicate that the segmentation produced by Network B is not as continuous as Network C. With accurate and continuous projection artifact delineation, these can be more accurately removed in post-processing.
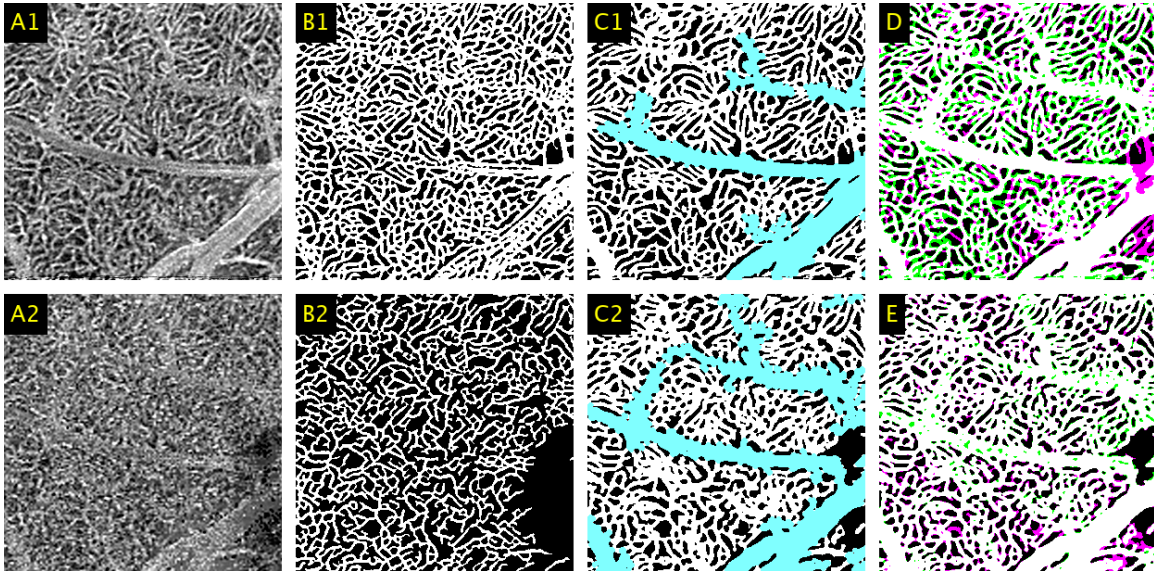
Figure 3.6. **A1: 2x2mm window of an averaged 6x6mm image taken of the deep vascular complex (DVC). A2: 2x2mm window of the corresponding region in the equivalent single-frame template image. B1: averaged image segmented using the initial weights (Network A). B2: single-frame image segmented using Network A. C1: averaged image segmented using the fine-tuned network (Network C) with projection artifacts to be excluded highlighted in cyan. C2: single-frame image segmented using Network C with projection artifacts to be excluded highlighted in cyan. D: comparison between the automated segmentations between the averaged and single-frame images produced by Network C, represented by magenta and green respectively, with white representing the union. E: comparison between single-frame segmentations between Network B and C, represented by magenta and green respectively, with white representing the union.**

Figure 3.7 shows an additional enlarged comparison of the results when segmenting the DVC with different versions of the deep neural network. The images segmented by Network C, shown in Figure 3.7-C1 and Figure 3.7-C2, more closely follow the elongated, lobular ICA morphology of the DVC and the results are less prone to over-segmenting noise. This presents a substantial improvement over the images

27

segmented by Network A, as shown in Figure 3.7-B1 and Figure 3.7-B2, the results of which incorrectly apply the branching structure characteristic of the SCP to the DVC.
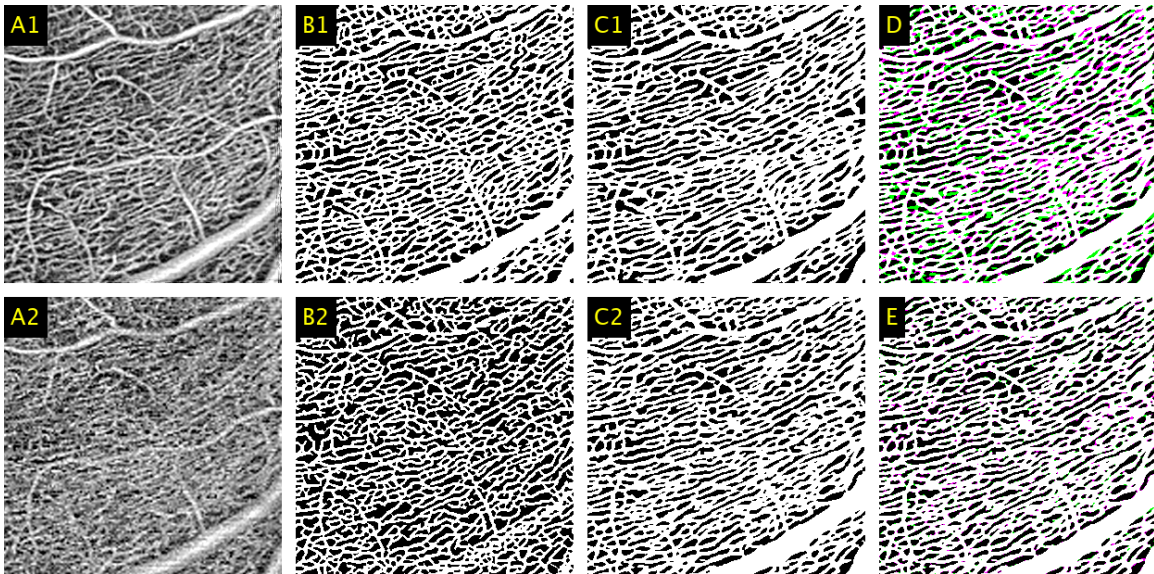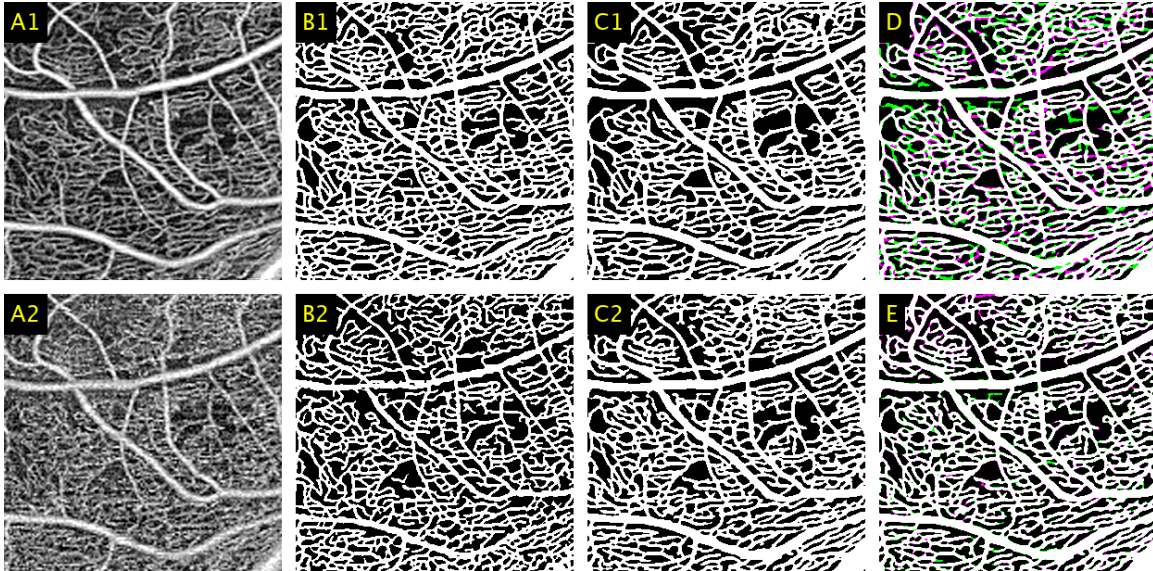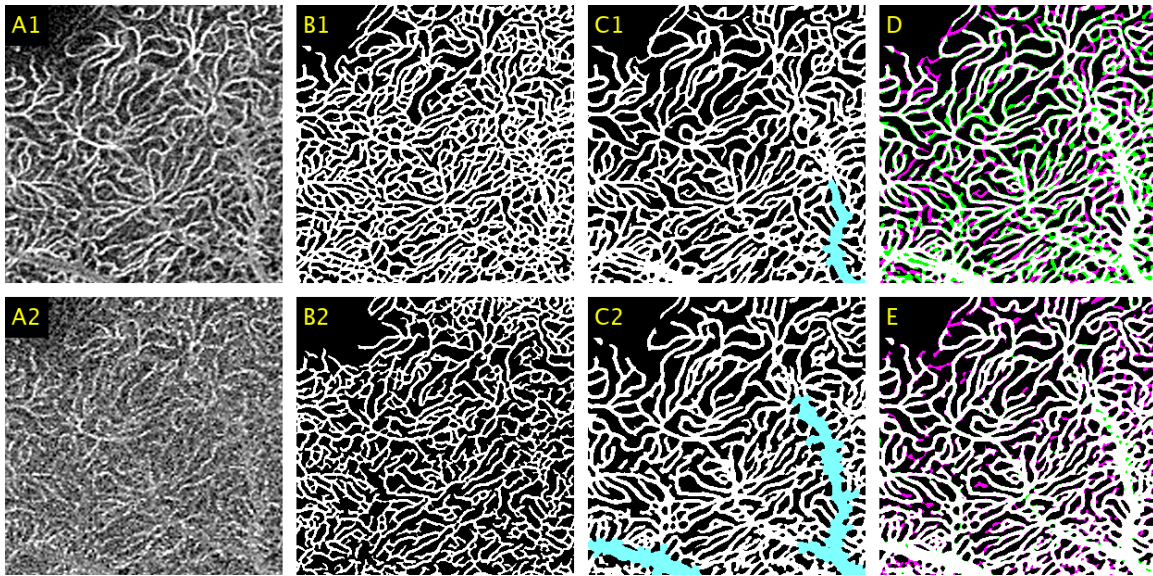

Figure 3.7.   **A1: 2x2mm window of an averaged 6x6mm image taken of the deep vascular complex (DVC). A2: 2x2mm window of the corresponding region in the equivalent single-frame template image. B1: averaged image segmented using the initial weights (Network A). B2: single-frame image segmented using Network A. C1: averaged image segmented using the fine-tuned network (Network C) with projection artifacts to be excluded highlighted in cyan. C2: single-frame image segmented using Network C with projection artifacts to be excluded highlighted in cyan. D: comparison between the automated segmentations between the averaged and single-frame images produced by Network C, represented by magenta and green respectively, with white representing the union. E: comparison between single-frame segmentations between Network B and C, represented by magenta and green respectively, with white representing the union.**

3.4.2. Inter-capillary Area Evaluation

Figure 3.8 shows representative images, segmentations, and standard deviation maps for diabetic subjects without DR, with mild/moderate non-proliferative NPDR, and with severe non-proliferative DR as graded by a retina specialist.
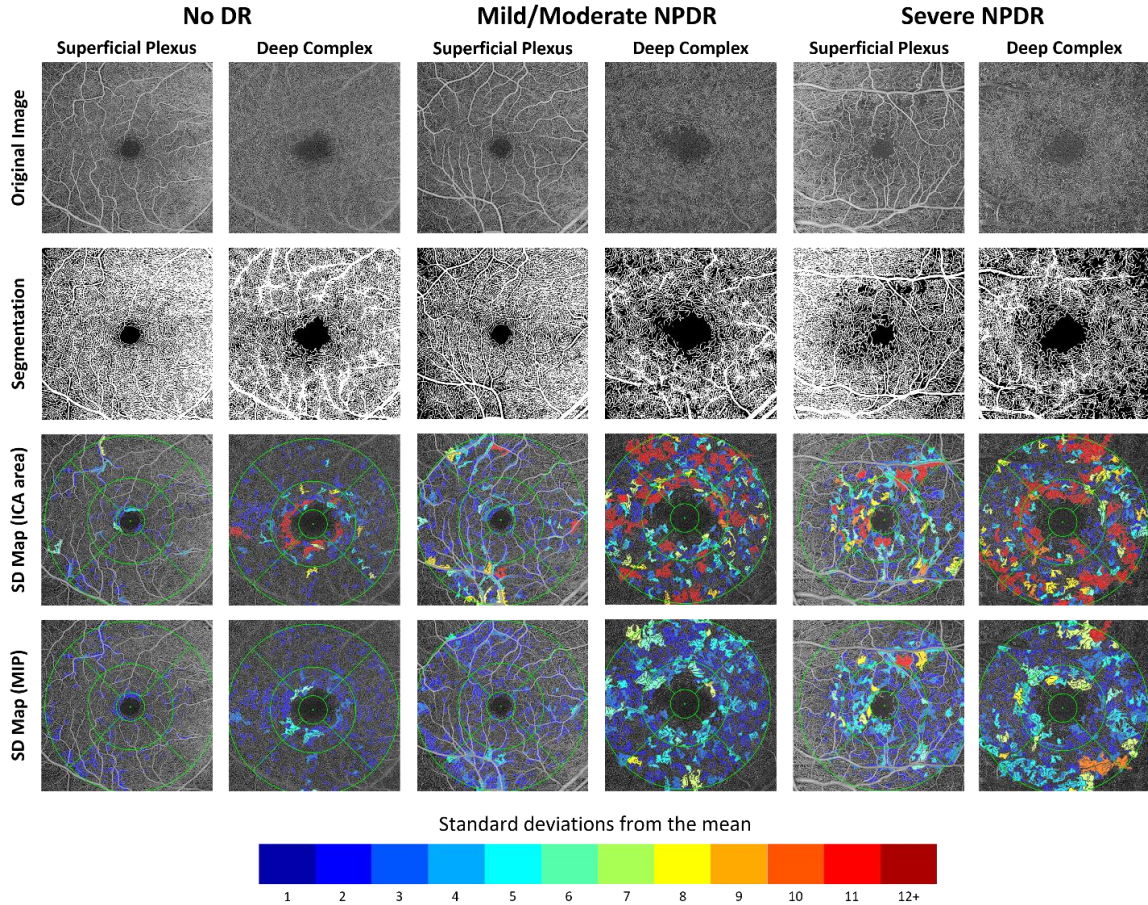
Figure 3.8. **Labeled standard deviation maps for subjects with no diabetic retinopathy, mild/moderate non-proliferative diabetic retinopathy, and severe non-proliferative diabetic retinopathy. Original images of the deep vascular complex have been brightened for clarity. All inter-capillary areas are labeled based on the number of standard deviations its maximum ischemic point and area exceeded the reference mean.**

3.5. Discussion

Early detection of DR is paramount to ensuring effective treatment and improved patient quality-of-life. Detecting changes in both the SCP and DVC have been identified as potential early biomarkers of DR. As a result, accurate segmentation and quantification of increasingly wide-field images of both the SCP and DVC will allow for further insight into the emergence and progression of DR.

We designed a transfer learning-based framework for automated segmentation of the microvasculature in the SCP and DVC, as well as quantification of the ICA's in

6x6mm single-frame OCT-A images. The framework consists of two convolutional neural networks: an initial network, trained on 2x2mm and 3x3mm images; and a second network, which utilized the pre-existing weights and fine-tuning on a smaller dataset of 6x6mm images of both the SCP and DVC. This approach allowed for accurate feature detection despite a limited training set, with results that exceed the intra-rater accuracy. In particular, fine-tuning from an existing set of data provided more robust projection artifact delineation in the DVC, allowing for removal in post-processing when computing vascular metrics. The resulting ICA quantifications allow for a closer investigation into suspected areas of low perfusion but does not expressly define what constitutes such areas.

A prevailing limitation of many machine learning problems is training dataset acquisition. For our study, manually segmenting an individual 6x6mm image of the DVC took each rater an average of 4 hours to complete, which can pose a significant challenge for problems requiring larger datasets. Solely training a new network on our limited, manually-segmented, 6x6mm dataset would overfit to the training set and including this new dataset with the original would result in a heavy data imbalance. The introduction of additional automated segmentations of averaged images greatly increased the size and quality of our training set, from 10 images of each of the SCP and DVC to 39. This allowed for a larger variation in training samples, consequently improving network performance.

The impact of the training examples is most evident in the segmentations of the DVC, where the initial weights produced segmentations that significantly differed from the images produced by the fine-tuned network. As seen in Figure 3.6 and Figure 3.7, vessels segmented by the initial weights closely resemble the denser morphological characteristics of the SCP. In particular, the ICA's in the DVC follow a lobular pattern, which is reflected more accurately in the segmentations generated by the fine-tuned network.

Another limitation is the quality of the data. Currently, images with an SSI of 8 or lower, as well as images with excessive microsaccadic eye motion were omitted from the study. If there are excessive microsaccadic artifacts, microvascular features begin to blur, and can be subsequently classified as noise by the network. This emphasizes the importance of using the averaged 6x6mm images as the ground-truth data obtained from

manual segmentations because it will be the most anatomically accurate. Our previously-published method of averaging and registering single-frame images based on a template [44] allowed for segmentations of averaged images to be paired with single-frame training data, greatly improving the quality of our training samples. Segmentation quality appeared to be independent of location in the image, as automated segmentation accuracy was consistent across the 6x6mm FOV in the absence of additional artifacts.

To summarize, we designed a machine learning framework to accurately segment and quantify the retinal microvasculature in the SVP and the DVC. It produces immediately-available segmentations, which provide clinicians with a tool for in-depth analysis of ICA's and the level of retinal perfusion. Through this framework, patient care for DR can be adapted individually, improving compliance and patient prognosis. In addition, visualization and quantification of retinal vasculature at a high level of accuracy provide more information about disease activity and therefore may add to individualized patient care.

3.5.1. Contributions

The details of the methods were conceived and designed by myself, with help from Dr. Morgan Heisler and Dr. Marinko Sarunic. The data was acquired at the Eye Care Centre by Dr. Sonja Karst. All machine learning scripts and accompanying MATLAB scripts were written by me, as well as the eventual manuscript, which was subsequently published.

After the publication of these methods, I changed the architecture to a residual U-Net approach, as shown in Figure 2.4. I also implemented a method of converting the Keras model files to a MATLAB-compatible version, allowing other members of BORG to use the segmentation tool without my assistance.

Chapter 4.   Federated learning

4.1. Introduction

The usage of deep learning has been extended to numerous problems in the medical industry, such as described in Chapter 3. However, as deep learning applications increase in complexity, the amount of data required to train a robust, accurate model increases significantly. However, as institutions and corporations increase the amount of data collected per user or participant, the consequences of a breach could prove disastrous. As a result, for medical applications, data is securely guarded behind privacy regulations and additional legal hurdles, which include the European Union's General Data Protection Regulation and the United States Health Insurance Portability and Accountability Act. This presents a nearly insurmountable hurdle for collaborative data sharing between institutions. Additionally, there exists a possibility that deep neural networks (DNNs) trained solely on medical data available in its own "data island" are significantly overfitted, especially if all the data originates from one source. This is the case for image processing algorithms, where images originating from one source can have distinct features that can facilitate overfitting as training progresses.

Federated learning is a distributed machine learning approach which enables DNN training on a large corpus of de-centralized data originating from different sources without directly accessing sensitive data. Cross-device federated learning was originally developed by Google for usage with large amounts of data present in phones and tablets for the development of intelligent applications [50]. It was proposed that a central server could distribute copies of a machine learning model to a set of clients, where each client could locally perform one or multiple steps of gradient descent on local data, and subsequently return its results to the central server to be averaged among the rest of the client base. Frameworks for developing federated learning algorithms such as Tensorflow Federated, NVIDIA Clara, as well as PySyft [51] exist, however, for medical applications with a small number of participating (collaborating) institutions, this approach can be simplified. This is termed cross-silo federated learning [52]. Compared to cross-device federated learning, the small number of collaborative participants in a cross-silo setting simplifies execution by allowing for the training to be synchronous. The

cross-silo setting also assumes that the participants are trusted, and do not present an adversarial risk towards federated training, which can include white-box and black-box inference attacks [53], or exploiting the gradients to reconstruct the training data [54].

Development of a cross-silo approach to federated learning for medical image analysis has been explored by various groups, most notably for the coronavirus disease 2019 (COVID-19) pandemic. A collaborative federated learning platform for CT scan-based COVID-19 diagnosis using a 3D dense convolutional neural network (CNN) has developed [55]. In addition, a federated approach to both L1 regularization and multilayer perceptron models were applied to electronic health records to predict COVID-19 mortality, showing improvement over models trained locally [56]. A federated learning framework was also developed for functional magnetic resonance imaging analysis using domain adaptation [57]. In the field of OCT for ophthalmic imaging, to the best of our knowledge, there is no previous report examining the use of federated learning. A related model-to-data approach was applied to intraretinal fluid segmentation in OCT volumes [58] with significant success. his report represents the next step in the progression of model-to-data to federated learning.

In this study, we investigate the federated learning approach to apply the microvasculature segmentation in Chapter 3 to multiple datasets across multiple institutions in a simulated environment. The performance of the federated model will be compared to models trained solely on local data, models trained on a fully centralized dataset, as well as alternative methods of collaborative deep learning.

4.2. Methods

4.2.1. Main components

The main components of the framework involve the central server and the individual clients from which the private data is sourced. The central server serves as the 'hub' of the framework, through dictating when training occurs, and defining hyperparameters for each client. The central server also collects an aggregate of each client's updates, and performs averaging to compute a new global model for distribution. Model updates from each client, as well as each iteration of the aggregated global model are saved locally, and is not accessed by other participants. In the cross-silo federated

learning setting, clients are different organizations, which include hospitals, research institutions, or geo-distributed datacenters, and it is assumed that clients are reliable and trusted participants with no adversarial motives. Data used for training is stored directly on servers that the client owns, and is not accessed by outside parties. Each individual client executes a persistent script that continually checks for instructions from the central server, and trains for an epoch on local data when directed.

4.2.2. Secure File Transfer

The training cycle in federated learning consists of regular communication between the central server and client, necessitating a secure method of programmatic file transfer. Direct SSH access to either the client or central server was deemed to pose a high security risk, especially if the traditional SSH keypair approach was used. In particular, providing direct access to the computer files was considered a risk in the case of an adversarial client (or a compromised client) potentially gaining access to models trained by other clients. Furthermore, the central client would have access to each client's training data. To solve this problem, a cloud-based drop-off folder was implemented, with the configuration as shown in Figure 4.1.

Central server   Drop-off folder (SFU Vault)   Client server
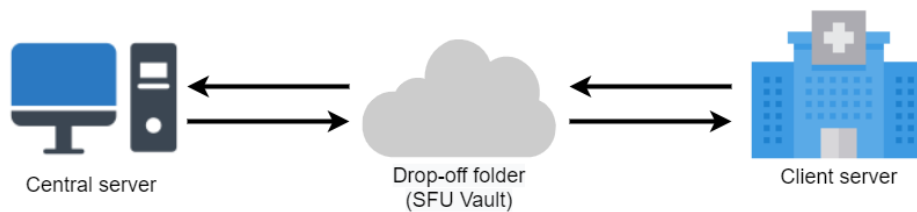
Figure 4.1.   **Overview of the data flow configuration between the central server and the client server. Each client is assigned its own drop-off folder, where files are uploaded to be downloaded by its recipient.**

The drop-off folder serves as a middle point for both the central server and the client, providing a medium on which to exchange files without direct SSH access. Drop-off folders are hosted on SFU Vault, which is an implementation of Owncloud, and can be subsequently utilized through its API. Each client is assigned its own folder, with a secure link and password that is provided and accessed programmatically.

4.2.3. Aggregation

The aggregation algorithm overview is as shown in Figure 4.2. For each training cycle, the central server would distribute the aggregate model, as well as instructions in a configuration file (INI was used, but formats such as JSON or XML would also be suitable) for each client. This configuration file included information such as the current epoch, as well as the learning rate at which that epoch would be trained. Each client would receive the model and configuration file, perform data augmentation, train for a full epoch on its training set, and validate on its validation set. The resulting model and CSV file containing the loss and accuracy for both training and validation would be returned to the central server. Once the central server received the model and CSV file from each client, it would further validate each client model on a separate dataset with high-quality, averaged OCT-A images in order to isolate models that could disrupt the overall training process. This can occur from a model simply diverging, or through adversarial attacks such as white-box and black-box attacks. Consequently, any client model that scored below a specified tolerance value would be omitted from the average. The paths to saved models and training data are defined in a configuration file located locally on the client server and are not accessed by the aggregator.
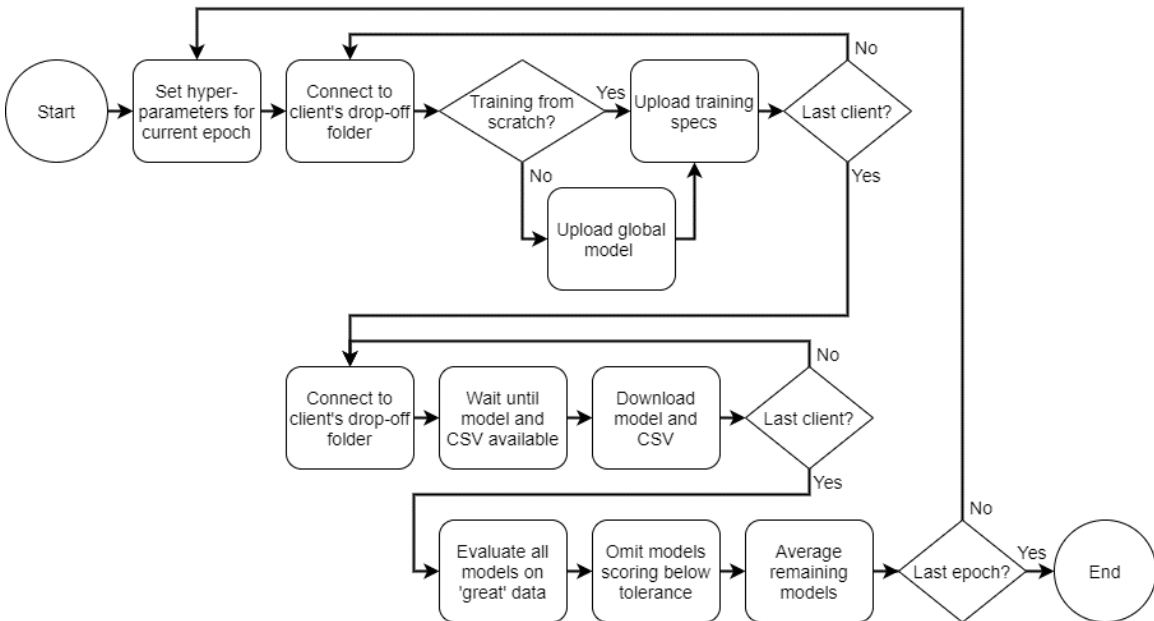
Figure 4.2. **Overview of the aggregation algorithm that oversees the federated learning training process.**

4.2.4. Training

To examine the efficacy and clinical utility of a federated learning approach, the resulting model was compared to alternative collaborative approaches, which include pooling all the data into a centralized location. Because of this, the model was trained on four base datasets that were locally available and sourced and are described in detail in Section 3.3.4 and are summarized in Table 4.1.

Table 4.1. **Overview of the four base datasets used for the federated learning simulation. Images in the PlexElite 6x6mm dataset were split after partitioning into training, validation, and test set.**

Image Source	Field-of-view	Included capillary plexuses	Number of available images	Dataset partitions (quadrants)
SFU prototype swept-source OCT-A	2x2mm	SCP	30	18 training 6 validation 6 testing
RTVue XR Avanti (Optovue, Inc.)	3x3mm	SCP	26	16 training 5 validation 5 testing
Angioplex (Carl Zeiss Meditec)	3x3mm	SCP	24	14 training 5 validation 5 testing
PlexElite 9000 (Carl Zeiss Meditec)	6x6mm	SCP and DVC	73 (292 quadrants)	42 (168) training 15 (60) validation 16 (64) testing

Each base dataset was split into training, validation, and testing partitions, after which each 6x6mm image was divided into quadrants to ensure image size consistency. Four simulated clients were assigned one each of the four base datasets, with each client occupying a compute node on Cedar (Compute Canada supercomputer cluster). A local computer acted as the central server. The architecture used on each client was the residual U-Net outlined in Section 2.3.1. Training was performed over a maximum of 1000 epochs, after which the model with the lowest validation loss was selected for evaluation. The clients all trained with stochastic gradient descent with a cyclic learning rate schedule that used warm restarts. The learning rate would decrease from 0.1 to 0.001 with a decay factor of 0.1, and repeat. Each client employed the same data augmentation steps, which included 5-10% random pixel dropout, linear and gamma

contrast adjustment, rotations from -15 to 15 degrees, and horizontal and vertical flips using the ImgAug Python library.

The performance of the federated learning model was investigated by using the data to train several scenarios. The model-to-data approach was investigated by training a model on each base dataset individually, but without transfer learning as in [58]. A combined dataset with all available images was used for training in the ideal case, where all data is available. The effects of the size of the data sets (and the resulting diversity in training examples) was explored through constructing two additional combined datasets with an equal number of images: one with 14 images randomly sampled from each source (the maximum possible, as the Angioplex 3x3mm training set contained 14 images), and one with only 4 images randomly sampled from each image source, to approximate the size of a smaller base dataset. Lastly, the federated model was compared to a model trained on all four base datasets sequentially in the order shown in Table 4.1, to simulate a naïve collaborative deep learning approach.

4.3. Results

4.3.1. Training and Validation

Figure 4.3 shows the training and validation loss curves over the entire training process of 1000 epochs. For the training loss curve, it can be seen that the clients converge as expected, with minimal oscillations along the length of the curve, which can be attributed to the cyclic learning rate schedule. These oscillations decrease in amplitude as training progresses. Convergence is also seen in the validation loss curve, with the minimum validation loss occurring at the 654th epoch. The validation losses for clients 0 to 2 (with the 2x2mm and 3x3mm base datasets) decrease as expected, with a plateau observed after approximately 300 epochs, while the loss for client 3 (with the 6x6mm base dataset) plateaus at approximately 0.47 and begins to decrease to 0.4 after the 350th epoch. This suggests that the gradient may have slowly traversed from a plateau or saddle point to a nearby local minimum in the loss landscape.
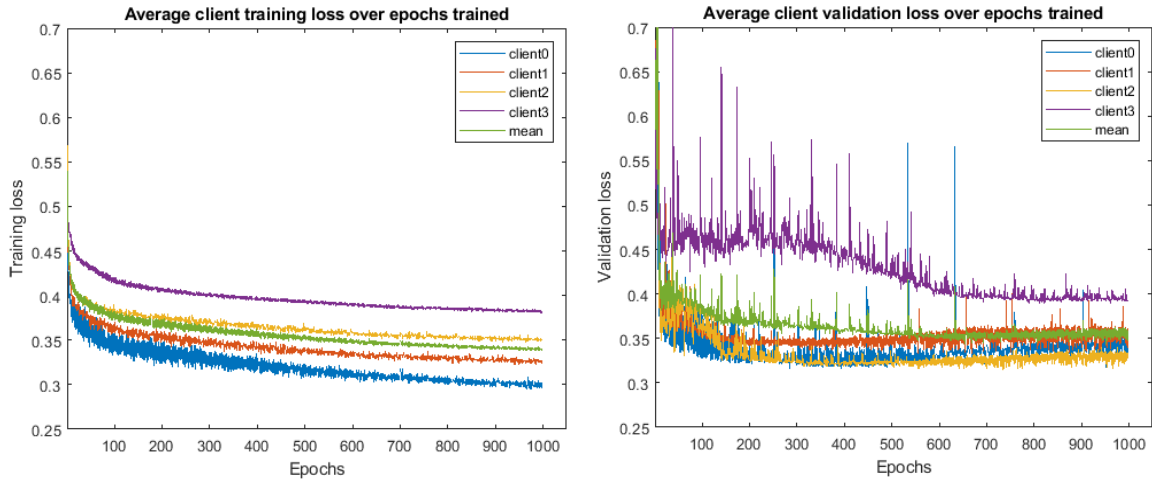
Figure 4.3.   **Training and validation loss curves for the federated model. The order of clients in the figure legends follows the same order of base datasets in Table 4.1, where client0 contains the prototype 2x2mm dataset, client1 contains the Optovue 3x3mm dataset, client2 contains the Angioplex 3x3mm dataset, and client3 contains the PlexElite 6x6mm dataset.**

4.3.2. Quantitative Analysis

Table 4.2 shows the accuracy when all 9 training methods are evaluated on each base dataset's test set. In this case, the baseline performance (shown in red) is defined as the performance of a model naively trained, and subsequently tested, on one base dataset. It can be seen that the model trained with federated learning attained accuracy scores comparable to the baseline performance, which was the best-performing model for the 3x3mm and 6x6mm base datasets. It also achieved similar performance to the combined datasets, suggesting that pooling all datasets into a centralized location may only provide a marginal benefit over federated learning. Additionally, the federated model outperformed the sequentially-trained model for the 2x2mm and 3x3mm base datasets, as the model biased towards the most recently-seen 6x6mm base dataset. A similar trend is seen when calculating the Dice index, as detailed in Table 4.3. It can be seen that the federated learning model exceeds the baseline performance for the 2x2mm base dataset, and achieves comparable scores for the two 3x3mm datasets. The federated learning model also achieves comparable scores to all combined datasets, similar to the results shown in Table 4.2.

38

Table 4.2. **Accuracy for each training method, evaluated on each base dataset's test set. Numbers in brackets in the first column represent number of images. Bolded values represent highest number in each column, and values in red represent the baseline performance (i.e. a model trained and tested on images from one base dataset).**

Model training method (number of images)	SFU prototype (2x2mm)	Optovue (3x3mm)	Angioplex (3x3mm)	PlexElite (6x6mm)
Federated learning	0.8568	0.8148	0.8502	0.7836
Sequential	0.8350	0.7889	0.8212	0.8098
Only SFU prototype (18)	0.8575	0.7208	0.8077	0.5741
Only Optovue (16)	0.8305	**0.8293**	0.8010	0.6211
Only Angioplex (14)	0.8180	0.8165	**0.8582**	0.7132
Only PlexElite (168)	0.8140	0.8001	0.8285	**0.8168**
Combined (18/16/14/168)	0.8549	0.8037	0.8425	0.8056
Combined equally (14/14/14/14)	**0.8611**	0.8285	0.8573	0.7951
Combined equally (4/4/4/4)	0.8554	0.8292	0.8532	0.7850

Table 4.3. **Dice similarity index for each training method, evaluated on each base dataset's test set. Numbers in brackets in the first column represent number of images. Bolded values represent highest number in each column, and values in red represent the baseline performance (i.e. a model trained and tested on images from one base dataset).**

Model training method (number of images)	SFU prototype (2x2mm)	Optovue (3x3mm)	Angioplex (3x3mm)	PlexElite (6x6mm)
Federated learning	**0.7727**	0.8138	0.8239	0.7608
Sequential	0.7515	0.7817	0.7815	0.7983
Only SFU prototype (18)	0.7561	0.6629	0.7460	0.2652
Only Optovue (16)	0.6714	0.8362	0.7343	0.4388
Only Angioplex (14)	0.7353	0.8179	0.8363	0.6368
Only PlexElite (168)	0.7382	0.8006	0.8008	0.8157
Combined (18/16/14/168)	0.7547	0.7973	0.8063	0.7965
Combined equally (14/14/14/14)	0.7719	0.8353	0.8347	0.7839
Combined equally (4/4/4/4)	0.7723	**0.8380**	0.8314	0.7822

4.4. Discussion

As deep learning applications grow in complexity, the need for labelled ground-truth data increases exponentially. In many cases, a single institution does not have enough resources to procure the data needed to train a robust model. However, for

medical images, this data is securely guarded behind various regulations and legal hurdles, resulting in a significant barrier to collaborative data sharing between institutions. There also exists a possibility that models trained mainly within one "data island" are significantly overfitted, which limits its eventual application. In this chapter, we designed a robust federated learning framework for hardware-agnostic microvasculature segmentation in OCT-A images, using a cross-silo approach. The framework involved individual clients training a model on its local training data and sending the weights to a central server. The central server would aggregate the weights from all clients and re-distribute the new global model. Secure file transfer was handled through a cloud-based drop-off folder hosted on SFU Vault, eliminating the need for direct SSH access between the central server and each participant.

To evaluate the performance of this framework, we simulated its performance on four base datasets from four separate imaging devices, consisting of fields-of-view ranging from 2x2mm to 6x6mm. The resulting federated model achieved performance comparable to the baseline – a model trained and tested on one single base dataset, showing that a model can converge despite training in a de-centralized manner. The federated model also achieved similar performance to models trained on centralized datasets, where multiple configurations were tested: combining all training images naively, and randomly sampling an equal number of images from each base dataset. The federated model was also compared to the sequential training method outlined in Section 4.2.3, providing better performance on the first three base datasets in the sequence due to its bias towards the most recently seen fourth base dataset. It was also observed that the difference in Dice index scores (in Table 4.3) for both models trained on randomly-sampled combined datasets varied by a maximum of 0.4% for each base dataset. This suggests that increasing the diversity of examples by adding images from additional imaging sources (through combining the datasets or federated learning) benefits the overall model more so than simply expanding a pre-existing dataset.

Because a model trained with federated learning achieved comparable performance to a model trained on any centralized dataset, it can be inferred that federated learning provides the most benefit to clients with the least amount of data (assuming that the data are similar). Clients with larger datasets are shown more unseen cases, allowing for better identification of edge cases. The resulting model is expected to

generalize more so than a model trained on a dataset originating from a single data source, resulting in better performance on images from a fifth image source. However, one area that merits further exploration is how each model would perform on data acquired by the same machines in Table 4.1, but at different institutions with different imaging protocols. If computing resources are available, this can be used in conjunction with a "filter bank" implementation, where individual models are trained specifically to segment images from only one image source, and openly shared among collaborators. This appears to be the most appropriate solution for a set of clients with a large variance in their respective datasets, however, we expect this approach to achieve lower performance on new and unseen data when compared to federated learning. An alternative approach is to use the federated learning model as a starting point for transfer learning, allowing each institution to tailor the model to their data.

Similar to traditional deep learning, a limitation of federated learning is the quantity and diversity of the training data. As shown in Table 4.1, the 6x6mm base dataset contained significantly more images than the other three base datasets, resulting in one client iterating over more steps per epoch during training. However, when aggregating the client models, each was weighted equally during averaging. This was done to ensure the federated model does not bias towards a single data source, despite the data imbalance. A potential solution to this is to perform additional augmentation on the smaller datasets, however, because each client model is aggregated into a global model, the benefits may be minimal.

With any attempts of collaborative deep learning, there exist several risks if a presumed collaborator has an adversarial motive. In federated learning, adversaries are motivated to disrupt the training process through white-box and black-box attacks [53], which we mitigate through a framework-wide validation step before averaging. In addition, "dummy" gradients can be introduced from a client to leak private training data [54]. Presently, this only applies to image classification problems with low batch sizes and low-resolution images (up to 64x64), which is unsuitable for the high-resolution images seen in medical data, and the substantially higher number of parameters seen in image segmentation architectures.

4.4.1. Contributions

The details of the methods were conceived and designed by myself, with advice from Dr. Marinko Sarunic, Dr. Da Ma, and Mr. Timothy Yu. I was present during the initial exploratory phase of the project, experimenting with different federated learning frameworks, before opting to code everything myself. Additional meetings with Dr. Aaron Lee, from the University of Washington (UW), provided insight on security concerns and constraints. I wrote all the code, including the aggregator and client scripts, and supervised Timothy as he adapted the framework for DR classification with our collaborators from Dr. Lee's lab, as well as Dr. Yali Jia's lab from the Oregon Health and Science University. All tests for vasculature segmentation for the purposes of evaluating the federated learning performance were organized and run by myself.

Chapter 5.   Conclusions

5.1. Summary

The main purpose of this thesis is to investigate methods of quantifying the microvasculature of the retina in OCT-A *enface* images using deep learning. The low SNR found in single-frame OCT-A images (most notably in the DVC) provides a substantial challenge to traditional image processing techniques, paving the way for deep learning algorithms which improve through experience. However, as deep learning applications increase in complexity, the amount of manually-labeled data may prove inadequate. The works described in Chapters 3 and 4 outline methods to overcome limitations through inadequate amounts of labeled data, through leveraging existing datasets to federated learning.

First, the basic concepts of the eye anatomy, ophthalmic imaging techniques, and machine learning were described as it pertains to this thesis. A comprehensive overview of each concept is beyond the scope of this thesis, however multiple review articles are available in the Literature for machine learning in ophthalmology [59] as well as federated learning [52].

Chapter 3 examined the usage of transfer learning to leverage pre-existing high-quality, averaged, 6x6mm OCT-A *enface* images to obtain detailed manual segmentations. When paired with the equivalent high-SNR single-frame images, and pre-existing weights obtained from a model trained on 2x2mm and 3x3mm images, we were able to train a convolutional neural network to segment both the SCP and DVC. The resulting segmentations were analyzed to produce detailed perfusion maps based on the maximum ischemic point and area observed in each inter-capillary area. Through this study, we were able to show that the resulting convolutional neural network was able to segment both the SCP and DVC, with performance exceeding inter-rater comparisons.

Because obtaining labeled data for complex tasks such as semantic segmentation is often arduous and resource-intensive, constructing a robust model relies heavily on collaborations between institutions. However, medical data sharing is severely limited by privacy regulations and additional legal hurdles. Chapter 4 expanded

43

on the previous method for microvasculature segmentation through the design and implementation of a federated learning framework to allow for multi-client training on a de-centralized data corpus. The framework consisted of a central aggregator server, as well as a series of clients to perform training on secure local data. A drop-off folder, hosted on SFU Vault, facilitated secure file transfer without direct SSH access between the central server and each participant. We showed that the federated learning framework achieves performance comparable to models trained on a fully-centralized dataset, while providing a model that is more generalized than a system of "filter banks".

The framework can also be adapted to other deep learning applications and is currently being implemented for DR severity classification. Available DR datasets can vary greatly across institutions based on location demographics and patient availability; thus, classification of multiple severity levels of DR can prove challenging due to the high potential of encountering unseen features. Pooling all available data through a decentralized dataset is essential to capture these edge cases, increasing the robustness of the resulting model, as well as its accuracy.

In summary, the major contributions of this work are (1) the development of a tool using transfer learning to obtain accurate automated segmentations of the retinal microvasculature, and (2) the implementation of a robust federated learning framework for microvasculature segmentation, with possible adaptation to other applications.

5.2. Future Works

While several deep learning techniques were investigated for ophthalmological image analysis, there exist several additional topics that have yet to be explored.

Expanding on the research presented in Chapter 3, there exist several potential directions for research using quantitative microvasculature data. As more data is available, longitudinal studies could be developed using these automated segmentations, which could provide insight into disease progression and prognosis. As imaging technologies improve, reducing the SNR for higher FOVs such as 9x9mm to 20x20mm, this segmentation tool can be expanded to encompass additional regions of the retina. This would allow for the identification of additional risk factors, leading to more

precise risk stratification – namely identifying patients likely to progress to proliferative disease, and increasing the frequency of follow-up visits for treatment.

As cross-silo federated learning is still in its infancy in terms of development, many additional experiments can be conducted to improve its clinical utility. It is possible that using different loss functions for each client could impact the results (in particular, the model's ability to be fully hardware-agnostic), as well as a modified averaging scheme during aggregation. In addition, further measures could be taken if a client model scores below the specified validation threshold once it is collected by the central server to salvage the valuable data learned during that specific epoch while remaining resistant to adversarial white-box and black-box attacks. Encryption of weights using secure multi-party computation would increase the overall security of the framework, however the usage of a secure drop-off folder partially mitigates this issue. Looking into the future, encryption could be relevant if multiple clients share a communal space when transferring models, or when security concerns exist during model exchange. Furthermore, as the work presented in Chapter 4 utilizes synchronous learning, further experiments could be conducted on asynchronous training, where clients "subscribe" to the framework and receive updates from the central server.

The examples outlined in this section remain small steps from the work presented in this thesis but present an optimistic outlook for the future of this field. As research and development in machine learning techniques continues to accelerate, the resulting discoveries can revolutionize ophthalmology beyond image processing and analysis, as well as direct diagnosis. The ultimate goal for any artificial intelligence application is to be able to produce a report, individualized for each patient, that details disease progression, expected prognosis, and a list of treatments the patient is most likely to respond to.

References

[1]   "Blindness in Canada | CNIB." [Online]. Available: https://cnib.ca/en/sight-loss-info/blindness/blindness-canada?region=bc.

[2]   M. M. Nentwich and M. W. Ulbig, "Diabetic retinopathy - ocular complications of diabetes mellitus.," *World J. Diabetes*, vol. 6, no. 3, pp. 489–99, Apr. 2015.

[3]   C. P. Wilkinson *et al.*, "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales.," *Ophthalmology*, vol. 110, no. 9, pp. 1677–82, Sep. 2003.

[4]   O. Arend, "The Relationship of Macular Microcirculation to Visual Acuity in Diabetic Patients," *Arch. Ophthalmol.*, vol. 113, no. 5, p. 610, May 1995.

[5]   C. Balaratnasingam *et al.*, "Visual Acuity Is Correlated with the Area of the Foveal Avascular Zone in Diabetic Retinopathy and Retinal Vein Occlusion," *Ophthalmology*, vol. 123, no. 11, pp. 2352–2367, Nov. 2016.

[6]   O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," pp. 1–8, 2015.

[7]   K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2015.

[8]   H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein, "Visualizing the Loss Landscape of Neural Nets," Dec. 2017.

[9]   Z. Zhang, Q. Liu, and Y. Wang, "Road Extraction by Deep Residual U-Net," Nov. 2017.

[10]   N. Rieke, "What Is Federated Learning?," 2019. [Online]. Available: https://blogs.nvidia.com/blog/2019/10/13/what-is-federated-learning/.

[11]   A. H. Kashani *et al.*, "Optical coherence tomography angiography: A comprehensive review of current methods and clinical applications.," *Prog. Retin. Eye Res.*, vol. 60, pp. 66–100, 2017.

[12]   Y. Jia *et al.*, "Quantitative optical coherence tomography angiography of vascular abnormalities in the living human eye.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 18, pp. E2395-402, May 2015.

[13]   D. An *et al.*, "Quantitative comparisons between optical coherence tomography angiography and matched histology in the human eye.," *Exp. Eye Res.*, vol. 170, pp. 13–19, 2018.

[14]   C. Balaratnasingam *et al.*, "Comparisons Between Histology and Optical Coherence Tomography Angiography of the Periarterial Capillary-Free Zone.," *Am. J. Ophthalmol.*, vol. 189, pp. 55–64, May 2018.

[15]   P. E. Z. Tan *et al.*, "Quantitative Comparison of Retinal Capillary Images Derived By Speckle Variance Optical Coherence Tomography With Histology.," *Invest. Ophthalmol. Vis. Sci.*, vol. 56, no. 6, pp. 3989–96, Jun. 2015.

[16]   D.-Y. Yu *et al.*, "Retinal capillary perfusion: Spatial and temporal heterogeneity.," *Prog. Retin. Eye Res.*, Feb. 2019.

[17]   L. A. Yannuzzi *et al.*, "Fluorescein angiography complication survey.," *Ophthalmology*, vol. 93, no. 5, pp. 611–7, May 1986.

[18]   P. S. Silva *et al.*, "Peripheral Lesions Identified on Ultrawide Field Imaging Predict Increased Risk of Diabetic Retinopathy Progression over 4 Years.," *Ophthalmology*, vol. 122, no. 5, pp. 949–56, May 2015.

[19]   J. P. Campbell *et al.*, "Detailed Vascular Anatomy of the Human Retina by Projection-Resolved Optical Coherence Tomography Angiography," *Sci. Rep.*, vol. 7, no. 1, p. 42201, Mar. 2017.

[20]   A. Carnevali *et al.*, "Optical coherence tomography angiography analysis of retinal vascular plexuses and choriocapillaris in patients with type 1 diabetes without diabetic retinopathy.," *Acta Diabetol.*, vol. 54, no. 7, pp. 695–702, Jul. 2017.

[21]   N. Hasegawa, M. Nozaki, N. Takase, M. Yoshida, and Y. Ogura, "New Insights Into Microaneurysms in the Deep Capillary Plexus Detected by Optical Coherence Tomography Angiography in Diabetic Macular Edema.," *Invest. Ophthalmol. Vis. Sci.*, vol. 57, no. 9, pp. OCT348-55, 2016.

[22]   C. L Srinidhi, P. Aparna, and J. Rajan, "Recent Advancements in Retinal Vessel Segmentation," *J. Med. Syst.*, vol. 41, no. 4, p. 70, Apr. 2017.

[23]   B. D. Krawitz *et al.*, "Parafoveal Nonperfusion Analysis in Diabetic Retinopathy Using Optical Coherence Tomography Angiography.," *Transl. Vis. Sci. Technol.*, vol. 7, no. 4, p. 4, Jul. 2018.

[24]   T. S. Hwang *et al.*, "Automated Quantification of Capillary Nonperfusion Using Optical Coherence Tomography Angiography in Diabetic Retinopathy.," *JAMA Ophthalmol.*, vol. 134, no. 4, pp. 367–73, Apr. 2016.

[25]   N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Syst. Man. Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[26]   J. Schottenhamml *et al.*, "AN AUTOMATIC, INTERCAPILLARY AREA-BASED ALGORITHM FOR QUANTIFYING DIABETES-RELATED CAPILLARY DROPOUT USING OPTICAL COHERENCE TOMOGRAPHY ANGIOGRAPHY.," *Retina*, vol. 36 Suppl 1, pp. S93–S101, Dec. 2016.

[27]   T. P. Zhu *et al.*, "COMPARISON OF PROJECTION-RESOLVED OPTICAL COHERENCE TOMOGRAPHY ANGIOGRAPHY-BASED METRICS FOR THE EARLY DETECTION OF RETINAL MICROVASCULAR IMPAIRMENTS IN DIABETES MELLITUS," *Retina*, p. 1, Oct. 2019.

[28] A. Li, J. You, C. Du, and Y. Pan, "Automated segmentation and quantification of OCT angiography for tracking angiogenesis progression.," *Biomed. Opt. Express*, vol. 8, no. 12, pp. 5604–5616, Dec. 2017.

[29] A. G. Roy *et al.*, "ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," *Biomed. Opt. Express*, vol. 8, no. 8, p. 3627, 2017.

[30] P. M. Maloca *et al.*, "Validation of automated artificial intelligence segmentation of optical coherence tomography images.," *PLoS One*, vol. 14, no. 8, p. e0220063, 2019.

[31] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search.," *Biomed. Opt. Express*, vol. 8, no. 5, pp. 2732–2744, May 2017.

[32] Q. Zhu *et al.*, "A New Approach for the Segmentation of Three Distinct Retinal Capillary Plexuses Using Optical Coherence Tomography Angiography," *Transl. Vis. Sci. Technol.*, vol. 8, no. 3, p. 57, Jun. 2019.

[33] M. Heisler *et al.*, "Automated identification of cone photoreceptors in adaptive optics optical coherence tomography images using transfer learning.," *Biomed. Opt. Express*, vol. 9, no. 11, pp. 5353–5367, Nov. 2018.

[34] D. Cunefare, A. L. Huckenpahler, E. J. Patterson, A. Dubra, J. Carroll, and S. Farsiu, "RAC-CNN: multimodal deep learning based automatic detection and classification of rod and cone photoreceptors in adaptive optics scanning light ophthalmoscope images.," *Biomed. Opt. Express*, vol. 10, no. 8, pp. 3815–3832, Aug. 2019.

[35] T. Schlegl *et al.*, "Fully Automated Detection and Quantification of Macular Fluid in OCT Using Deep Learning.," *Ophthalmology*, vol. 125, no. 4, pp. 549–558, Apr. 2018.

[36] Z. Ji, Q. Chen, S. Niu, T. Leng, and D. L. Rubin, "Beyond Retinal Layers: A Deep Voting Model for Automated Geographic Atrophy Segmentation in SD-OCT Images," *Transl. Vis. Sci. Technol.*, vol. 7, no. 1, p. 1, Jan. 2018.

[37] W. Lu, Y. Tong, Y. Yu, Y. Xing, C. Chen, and Y. Shen, "Deep Learning-Based Automated Classification of Multi-Categorical Abnormalities From Optical Coherence Tomography Images," *Transl. Vis. Sci. Technol.*, vol. 7, no. 6, p. 41, Dec. 2018.

[38] J. De Fauw *et al.*, "Clinically applicable deep learning for diagnosis and referral in retinal disease.," *Nat. Med.*, vol. 24, no. 9, pp. 1342–1350, Sep. 2018.

[39] D. S. Kermany *et al.*, "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning.," *Cell*, vol. 172, no. 5, pp. 1122-1131.e9, 2018.

[40] H. Zhao, H. Li, S. Maurer-Stroh, and L. Cheng, "Synthesizing retinal and neuronal

images with generative adversarial nets.," *Med. Image Anal.*, vol. 49, pp. 14–26, Oct. 2018.

[41]   G. Ometto *et al.*, "ReLayer: a Free, Online Tool for Extracting Retinal Thickness From Cross-Platform OCT Images," *Transl. Vis. Sci. Technol.*, vol. 8, no. 3, p. 25, May 2019.

[42]   Y. Guo, A. Camino, J. Wang, D. Huang, T. S. Hwang, and Y. Jia, "MEDnet, a neural network for automated detection of avascular area in OCT angiography.," *Biomed. Opt. Express*, vol. 9, no. 11, pp. 5147–5158, Nov. 2018.

[43]   P. Prentašic *et al.*, "Segmentation of the foveal microvasculature using deep learning networks," *J. Biomed. Opt.*, vol. 21, no. 7, p. 075008, Jul. 2016.

[44]   M. Heisler *et al.*, "Strip-based registration of serially acquired optical coherence tomography angiography.," *J. Biomed. Opt.*, vol. 22, no. 3, p. 36007, 2017.

[45]   A. Uji *et al.*, "Multiple enface image averaging for enhanced optical coherence tomography angiography imaging.," *Acta Ophthalmol.*, vol. 96, no. 7, pp. e820–e827, Nov. 2018.

[46]   T. G. Schmidt, R. E. Linderman, M. R. Strampe, T. Y. P. Chui, R. B. Rosen, and J. Carroll, "The Utility of Frame Averaging for Automated Algorithms in Analyzing Retinal Vascular Biomarkers in AngioVue OCTA," *Transl. Vis. Sci. Technol.*, vol. 8, no. 1, p. 10, Jan. 2019.

[47]   A. Krizhevsky, "Convolutional Deep Belief Networks on CIFAR-10," 2010.

[48]   M. Heisler *et al.*, "Deep learning vessel segmentation and quantification of the foveal avascular zone using commercial and prototype OCT-A platforms," Sep. 2019.

[49]   J. Xu *et al.*, "Retinal angiography with real-time speckle variance optical coherence tomography.," *Br. J. Ophthalmol.*, vol. 99, no. 10, pp. 1315–9, Oct. 2015.

[50]   H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," Feb. 2016.

[51]   T. Ryffel *et al.*, "A generic framework for privacy preserving deep learning," Nov. 2018.

[52]   P. Kairouz *et al.*, "Advances and Open Problems in Federated Learning," Dec. 2019.

[53]   M. Nasr, R. Shokri, and A. Houmansadr, "Comprehensive Privacy Analysis of Deep Learning: Passive and Active White-box Inference Attacks against Centralized and Federated Learning," Dec. 2018.

[54]   L. Zhu, Z. Liu, and S. Han, "Deep Leakage from Gradients," Jun. 2019.

[55] Y. Xu *et al.*, "A collaborative online AI engine for CT-based COVID-19 diagnosis.," *medRxiv Prepr. Serv. Heal. Sci.*, May 2020.

[56] A. Vaid *et al.*, "Federated Learning of Electronic Health Records Improves Mortality Prediction in Patients Hospitalized with COVID-19.," *medRxiv Prepr. Serv. Heal. Sci.*, Aug. 2020.

[57] X. Li, Y. Gu, N. Dvornek, L. H. Staib, P. Ventola, and J. S. Duncan, "Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results," *Med. Image Anal.*, vol. 65, p. 101765, Oct. 2020.

[58] N. Mehta *et al.*, "Model-to-Data Approach for Deep Learning in Optical Coherence Tomography Intraretinal Fluid Segmentation," *JAMA Ophthalmol.*, Aug. 2020.

[59] D. S. W. Ting *et al.*, "Artificial intelligence and deep learning in ophthalmology," *Br. J. Ophthalmol.*, vol. 103, no. 2, pp. 167–175, 2019.

Appendix A. Peer Reviewed Journal Publications

[A1] **J. Lo**, M. Heisler, V. Vanzan, S. Karst, I. Z. Matovinović, S. Lončarić, E. V. Navajas, M. F. Beg, and M. V. Šarunić, "Microvasculature Segmentation and Intercapillary Area Quantification of the Deep Vascular Complex Using Transfer Learning," *Transl. Vis. Sci. Technol.*, vol. 9, no. 2, p. 38, Jul. 2020.

[A2] S. G. Karst, M. Heisler, **J. Lo**, N. Schuck, A. Safari, M. V. Sarunic, D. Maberley, and E. V. Navajas, "Evaluating Signs of Microangiopathy Secondary to Diabetes in Different Areas of the Retina with Swept Source OCTA," *Investig. Opthalmology Vis. Sci.*, vol. 61, no. 5, p. 8, May 2020.

[A3] M. Heisler, M. Bhalla, **J. Lo**, Z. Mammo, S. Lee, M. Ju, M. F. Beg, and M. V. Sarunic, "Semi-supervised deep learning based 3D analysis of the peripapillary region," Biomed. Opt. Express, vol. 11, no. 7, p. 3843, Jul. 2020.

[A4] M. Heisler, S. Karst, **J. Lo**, Z. Mammo, T. Yu, S. Warner, D. Maberley, M. F. Beg, and E. V. Navajas, M. V. Sarunic, "Ensemble Deep Learning for Diabetic Retinopathy Detection Using Optical Coherence Tomography Angiography," *Transl. Vis. Sci. Technol.*, vol. 9, no. 2, p. 20, Apr. 2020.