PH. D. THESIS

DEEP LEARNING SOLUTIONS ON IMPERFECT DATASETS IN OPHTHALMOLOGY

by

YUKA KIHARA

Department of Robotic Science and Technology
Graduate School of Engineering
Chubu University

JANUARY 2023

Epigraph.

Dedication.

ACKNOWLEDGEMENTS

First and foremost, I would like acknowledge and give my warmest thanks to my two supervisors Hironobu Fujiyoshi and Aaron Lee. This dissertation would not have been possible without their support. Their immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. I would like to express gratitude to Cecilia Lee for her kind help and support that have made my study and life a wonderful time. I would like to thank all the members in Lee Lab and Machine Perception and Robotics Group that helped through comments and discussions during the writing of the various papers composing this thesis. I also want to thank my amazing co-authors and many other reviewers for their helpful comments, discussions, and reading countless drafts of (often last minute) papers and helping me understand clinical point of view. Finally, I would like to give special thanks to my parents, spouse, and my children as a whole for their continuous support and understanding when undertaking my research and writing my project.

ABSTRACT

Ophthalmology, with its heavy reliance on imaging, is an innovator in the field of artificial intelligence (AI) in medicine. Deep learning is what has made ophthalmology a pioneer in the field of implementing AI in medicine, because ophthalmologists are increasingly reliant on imaging tests to monitor patients. For the medical retina, imaging tests such as optical coherence tomography (OCT) are performed very frequently and have provided the clinical datasets to train, test, and then apply AI decision support systems. OCT, which can image retinal structures *in vivo*, has been widely applied in diagnostic ophthalmology due to its ease-of-use, lack of ionizing radiation, and high spatial resolution. There are approximately 30 million OCT procedures performed worldwide each year, with hundreds of consecutive B-scans comprising the majority of each procedure. With OCT, ophthalmologists can see each of the retina's distinctive layers. This allows ophthalmologists to map and measure the thickness of the layers and/or identify abnormal lesions around the layers. These clinical features aid with diagnosis, and also provide treatment guidance for glaucoma and many different diseases of the retina, including age-related macular degeneration (AMD) and diabetic eye disease. However, there are several challenges with handling such imaging datasets gathered from clinical practice settings; data-imbalance, lack of annotations and noise on the imaging or annotations. Furthermore, most of the existing AI approaches rely on *a priori* retinal layer segmentations being available or require additional imaging data from other devices such as optical coherence tomography angiography (OCTA). These retinal layer segmentations and OCTA instruments are not as widely available in clinical practices as routine structural OCT instruments. This thesis presents practical and interpretable machine learning systems and frameworks that do not require pre-processing nor expensive devices, but help with the mining of large clinical datasets from around the world, and address key problems in eye-disease treatment based on real world medical datasets. For Learning with imperfect datasets having limited annotations, I adopted semi-supervised techniques where pseudo-labels were applied to samples in the unlabeled set by using a model trained on the combination of the labeled samples and any previously pseudo-labeled samples, and iteratively repeating this process in a self-training cycle. For the datasets with various noise on the images, I present a multi-modal fusion network to maximally exploit the available information from the two imaging resources, including cases when either one was noisy or of poor quality. The model was trained with a sample reweighting scheme to mitigate bias in the datasets. All the presented models only rely on OCT which is the most commonly used imaging modality in vitreo-retina practice. Major study focuses include clinical outcomes research in AMD, glaucoma, and investigation of demographic and ocular factors relate to OCT.

TABLE OF CONTENTS

Chapter 1 Introduction **1**

1.1 Background 2

1.2 Research Purpose 3

1.2.1 Glaucoma study 5

1.2.2 UK Biobank Foveal/Macular Curvature analysis 6

1.2.3 Double Layer Sign detection for age-related macular degeneration (AMD) 7

1.3 The organization of the thesis 8

Chapter 2 Related Works **9**

2.1 Mitigating Data Imbalance Issues in Medical Image Analysis 10

2.1.1 Re-Weighting 10

2.1.2 Re-Sampling 11

2.1.3 Other types 12

2.2 Deep Multimodal Learning for Computer Vision 14

2.2.1 Data-level fusion (Early Fusion) 14

2.2.2 Feature-level fusion (Intermediate Fusion) 15

2.2.3 Decision-level fusion (Late Fusion) 15

2.3 Annotation-efficient deep learning for semantic segmentation 16

2.3.1 Semi-supervised learning with Entropy minimization 16

2.3.2 Semi-supervised learning with Consistency regularization 17

2.4 Summary 18

Chapter 3 Multimodal Deep Learning for Predicting Visual Fields from the Optic Disc and OCT Imaging **19**

3.1 Introduction 20

3.2 Multi-modal Policy model 21

3.2.1 Single modality DL model 22

3.2.2 Multi modality Policy DL model 22

3.3 Mitigating data imbalance 24

3.3.1 Sensitivity-weighted loss function 24

3.3.2 Risk-weighted loss function for policy network . . . 24
3.4 Experiment . . . 24
3.4.1 Data preparation . . . 24
3.4.2 Results . . . 26
3.5 Discussion . . . 32
3.5.1 Explainability . . . 33
3.5.2 Structure-Function Mapping . . . 33
3.5.3 Comparison with Previous Work . . . 34
3.5.4 Practical Implications and Future Directions . . . 34
3.6 Summary . . . 36

Chapter 4 Retinal layer Segmentation on large scale unlabeled data with Pseudo labeling 37
4.1 Introduction . . . 38
4.1.1 Foveal Curvature and Its Associations . . . 38
4.1.2 Macular Curvature and Dome-Shaped Configuration . . . 38
4.2 Curvature Analysis on UKBB . . . 39
4.2.1 Retinal Layer Segmentation with Pseudo Labels . . . 39
4.2.2 Center Point Retinal Thickness Analysis. . . . 42
4.2.3 Foveal Curvature Analysis . . . 42
4.2.4 Macular Curvature Analysis . . . 46
4.3 Results . . . 49
4.3.1 Results from Foveal Curvature Analysis . . . 49
4.3.2 Results from Macular Curvature Analysis . . . 55
4.4 Summary . . . 60

Chapter 5 Semantic Segmentation for Detection of Nonexudative Macular Neovascularization 62
5.1 Introduction . . . 63
5.2 Detection of Nonexudative Macular Neovascularization with ViT-based Segmentation model 63
5.2.1 ViT-based Segmentation model . . . 64
5.2.2 Classification criteria . . . 65
5.2.3 Results . . . 66
5.3 Detection of Nonexudative Macular Neovascularization with Semi-supervised Segmentation model . . . 70
5.3.1 Cross Pseudo Super Vision . . . 70
5.3.2 Preliminary validation of CPS . . . 71
5.4 Summary . . . 72

Chapter 6 Conclusions 73

References **75**

Publications and Awards **93**

LIST OF FIGURES

1.1 Optical Coherence Tomography. 1. White light source emits various wavelengths of light. 2. Beam splitter sends some light to 2B a reference arm and 2A the sample arm. 3. In the reference arm, light travels to a mirror where the distance to the mirror is known and constant. 4. In the sample arm, light travels to tissue of varying distance, such as different layers of the retina. The reflected frequencies and amplitudes of light are not constant, unlike the reference arm in 3. 5. The reflected light from the sample and reference arms return to a detector. At the detector, the phase interference is measured, and an a-scan is generated. 6. Multiple A-scans are stitched side-by-side together to create a b-scan, a cross-sectional representation of the retina. 4

1.2 Normal retinal tissue layers image shot by OCT instrument. 5

1.3 The breakdown of OCT scans. These are the different types of scans: 1. A-scans are a pixel-thick line of interference values (1-D line). 2. A-scans stitched together yields a B-scan (2-D image composed of a series of adjacent A-scans). 3. Multiple, aligned B-scans represents a volumetric scan (3D-image composed of multiple B-scans). 4. In a volumetric scan, an en-face (C-scan) OCT image can be generated where different slabs of the retina can be examined. 5

1.4 Macula and Fovea [1] 7

3.1 An example of a disc (A) and OCT (B) images alongside the corresponding HVF (C). T = temporal; S = superior; N = nasal; I = inferior. 20

3.2 A) Process overview of VF prediction from the DL model using the IR SLO image (top) and the OCT image(bottom). B) Model architecture of SLO and OCT models. C) Model architecture of Policy model. The feature maps obtained via EfficientNet from the IR SLO and OCT images were concatenated. We then applied a 1x1 convolution to obtain a multi-modal feature representation while holding all the parameters in the base models "frozen" during the training of the policy model. The output of the policy network (54-dimensional vector) represents the policy preference for each VF location ranging from 0 to 1, generated by a fully connected layer. Values closer to 0 indicate a preference for the disc model, values closer to 1 indicate a preference for the OCT model. Since the target labels of the policy model were binary (0 = disc preferred, 1 = OCT preferred), we used a class weighted binary cross-entropy loss function. We also applied a pointwise "risk weight"" which reflects how much the error increases when the wrong choice is made. 23

3.3 Flowchart of the selection process from the full dataset of VF tests. FP = false positive; FT = full-threshold; IVT = intravitreal; SITA = Swedish interactive threshold algorithm; VF = visual field. 25

3.4 A, Predictions from the two sub-models and the policy network for each level of sensitivity. The diagonal line indicates perfect agreement. B, Relationship between the prediction Mean Absolute Error (MAE) from the two sub-models and the final average choice from the policy network. Higher MAE in the disc to HVF model is generally associated with a preference for the OCT to HVF model and vice-versa. The diagonal line indicates equivalent MAE. HVF = Humphrey visual field; MAE = mean absolute error; OCT = optical coherence tomography. 29

3.5 Prediction examples from the test set. The top of each panel reports the disc and OCT images used for predictions alongside the ground-truth HVF. The bottom of each panel shows, in order, the predictions from the two sub-models, the choice score of the policy network for each location, and the final prediction of the policy network. A, Both sub-models predict a similar location of the defect, but the OCT to HVF is more accurate in predicting the magnitude of glaucoma damage; the policy network correctly selects this prediction for the superior hemifield. B, the prediction from either individual sub-model is wrong, showing either little damage or diffused advanced loss; however, the policy network correctly selects predictions for each location to obtain a result very close to the ground truth, better characterizing the spared paracentral VF. HVF = Humphrey visual field; I = inferior; N = nasal; OCT = optical coherence tomography; S = superior; T = temporal. . . 30

3.6 Other predictions from the test set. A) Case of myopic disc with extensive peripapillary atrophy and full thickness retinal damage in the inferior sector. Interestingly, the superior loss due to the full thickness defect is captured by the OCT to HVF model and correctly selected by the policy network. B) Synergic interaction, where the policy network is able to select inputs from the two sub-models to achieve a result that would not be possible with either individual sub-model. C) OCT to HVF produces an exaggerated nasal inferior defect; the policy network correctly prefers the result of the Disc to HVF prediction D) Prediction failure: the disc prediction was closer to the ground-truth but the policy network relies more on the OCT prediction. Of notice, both models produced similar results but the Disc to HVF model predicts a shallower HVF defect compared to the ground truth. 31

3.7 Prediction performance (top) and policy choice (bottom) for different noise levels applied to disc and OCT images. As expected, adding noise to OCT images degraded the performance of the OCT prediction (red) while leaving the disc prediction unaltered (blue). The opposite was true when adding noise to the disc images. The policy network was able to maintain a stable performance throughout. This was achieved by changing the preference for inputs from the disc or the OCT prediction according to which provided more reliable results (i.e. contained less noise). In other words, the policy network was able to adapt its choice based on changes in the quality of the input image. The gaussian blur was added to the disc images (kernel size from 0 to 41 pixels from level 0 to 4) while brightness was changed on OCT images (intensities decreased by a value from 0 to 120 from level 0 to 4, over a maximum of 255 and clipped at 0). 32

3.8 A, Process for generating heatmaps for the disc SLO and OCT models. According to the location of the perturbation in the predicted VF, the magnitude of changes reflects where the network focuses on during the predictions. B, Heatmap for one patient-generated example based on our disc SLO. C, Heatmap for one patient-generated example based on the OCT model. D, Heatmap for one patient-generated example from the policy network obtained by occluding the disc SLO. E, Heatmap for one patient-generated example from the policy network obtained by occluding the OCT image. OCT = optical coherence tomography; SLO = scanning laser ophthalmoscopy; VF = visual field. 35

3.9 Structure-function mapping for the OCT to HVF model generated from the heatmaps for the test set for each testing point in the 24-2 VF pattern shown on polar coordinates. The map was built by plotting the distribution of the circumpapillary angles at which the highest peak of the heatmap was located. Other mapping schemes are reported for comparison. HVF = Humphrey visual field; I = inferior; N = nasal; OCT = optical coherence tomography; S = superior; T = temporal. 36

4.1 Diagram of machine learning metodology implemented to generate our automated OCT-derived FC quantification. 41

4.2 Process overview of Center Point Retinal Thickness Analysis. 42
4.3 Examples of detected curve of fovea. 43
4.4 Top: foveal curvature quantification from central b-scans for each curvature tertile (from right to left, flattest to steepest quantified curvature measurements). Bottom: histogram of foveal curvature distribution by FC tertile. 43
4.5 Examples of detected curve of macular. 47
4.6 The spectrum of macular curvature. The histogram (left) shows the distribution of macular curvature values; the colors highlight the four different quartiles. The optical coherence tomography B-scans (right) represent exemplary images for all four quartiles as used for validation. For the demonstrated exemplary set, the individual curvature value of the respective eyes is displayed. The bottom image shows an inward-directed macular curve associated with a negative curvature value, indicating a dome-shaped configuration. 49
4.7 Adjusted mean foveal curvature by deciles of covariates stratified by sex (annual income and Townsend index of deprivation shown in quintiles). Adjusted means (solid black dots), 95% confidence intervals (vertical solid lines), and regression line (dotted line) are from a multilevel model allowing for age, height, ethnicity, and UK Biobank center as fixed effects, and repeated foveal curvatur measurement for each person. *Ethnicity codes: W, white; B, black; A, Asian; M, mixed; C, Chinese; O, other. † Visual acuity shown in logMAR for visualization purposes. GBP, pound sterling; O/CSE, O levels, certificate of secondary education or equivalent. 53
4.8 Macular curvature and associated parameters. The panel reveals adjusted mean macular curvature by deciles of variables of interest. Adjusted means (solid black dots), 95% confidence intervals (vertical solid lines), and the regression line (dotted line) are from a multilevel model allowing for repeated macular curvature measurement for each person. 58
4.9 Inter-eye correlation. The scatter plots visualize the high inter-eye correlation for macular curvature, refractive error (as spherical equivalent; in diopters, dpt), visual acuity (in Logarithm of the Minimum Angle of Resolution, LogMAR), and intraocular pressure between right and left eyes of participants with both eyes included. There was a high inter-eye correlation for all parameters (macular curvature, r = 0.746, $P < 0.001$; refractive error, r = -0.451, $P < 0.001$; visual acuity, r = 0.368, $P < 0.001$; intraocular pressure, r = 0.595, $P < 0.001$). The diagonal symbolizes perfect inter-eye agreement. OD = Right eye; OS = Left eye. 60

5.1 Framework of ViT segmentation model. 512×512 images were used as input. 16×16 image patches were projected to a sequence of embeddings and then encoded with a transformer and reshaped into a segmentation map. 65

5.2 Process overview. A. DLSs and drusen were extracted from B scans by a segmentation model. The detected DLSs and drusen were depicted in blue and yellow, respectively. B. After processing predictions for all B-scans, we generated en-face maps. C. eyes with MNV were then identified based on the en-face projection map. 66
5.3 Threshold vs. Classification errors. On each 128×128 en-face projection map on the validation set, we observed how the number of classification errors changed with regard to the threshold. Between threshold values were 40 to 60, the errors got minimum, and 65 was the middle point. 66
5.4 Prediction examples from the validation set. Grounded Truth(left) and predicted masks(right) overlaid on input B scan. 1 to 4th rows: The detected DLSs and drusen were depicted in blue and yellow, respectively. Our segmentation model mostly was able to extract DLSs and drusen at different locations/volumes properly. 5-th row: Failure case examples. Our segmentation model tends to mislabel drusen to DLSs on eyes with geographic atrophy. . . 68
5.5 Improvements over the supervised baseline. 72

LIST OF TABLES

3.1 Descriptive Statistics of the Selected Sample. Continuous variables are reported as median [interquartile range]. F = female; FP = false positive; IOP = intraocular pressure; M = male; MD = mean deviation; OCT = optical coherence tomography; PSD = pattern standard deviation. 26

3.2 Prediction performance using other fusion algorithms. Mean: average the two base model outputs, Blending: calculate multiplication by predicted probability to blend model outputs proportionally, Policy (selected approach): use either prediction from disc model or OCT model with threshold=0.5. For each cell, MAE with 95% Confidence Intervals are reported. 27

3.3 Prediction performance on other EfficientNet types. For each cell, MAE with 95% Confidence Intervals are reported. 27

3.4 Prediction performance using DenseNet. Two different architectures were used to train the disc model and OCT model. The parameter size of DenseNet 201 is comparable with EfficientNet B4, and the parameter size of DenseNet 121 is comparable with EfficientNet B2. EfficientNet achieved a better MAE. For each cell, MAE and 95% Confidence Intervals are reported. 28

3.5 Prediction performance with different allowed time gaps between imaging and visual field test. For each cell, MAE with 95% Confidence Intervals are reported. 28

4.1 Regression table for females with fovea curvature$\times 100$ as dependent variable. 45

4.2 Regression table for males with fovea curvature x 100 as dependent variable. 46

4.3 Patient Level Characteristics Stratified by Sex 52

4.4 Eye Level Characteristics Stratified by Sex 52

4.5 Standardized difference in fovea curvature per specified differences in covariates ([95% CI]; p-value) for females. 54

4.6 Standardized difference in fovea curvature per specified differences in covariates ([95% CI]; p-value) for males. 55

4.7 Standard deviation differences in macular curvature by change in independent variables Model 1 stands for a multilevel model that adjusts for demographical parameters (age, sex, and ethnicity) as fixed effects and a random effect for person to allow for within person eye measurements. Model 2 adjusts as model 1 plus functional (visual acuity and fluid intelligence) as well as ocular measures (spherical equivalent, corneal curvature and intraocular pressure). Model 3 adjusts as model 2 plus infancy factors (maternal smoking and birthweight). The marginal R^2 (i.e., variance explained by the models) and the conditional R^2 were 0.004 and 0.747 for model 1, 0.161 and 0.794 for model 2, and 0.167 and 0.797 for model 3, respectively. CI = confidence interval, LogMAR = Logarithm of the Minimum Angle of Resolution; mmHg = millimeters of mercury 57
4.8 Dome-Shaped Macular Configuration and Refractive Error 59

5.1 Presence of Subclinical Macular Neovascularization in Test Eyes with Nonexudative Age-Related Macular Degeneration. 68
5.2 Sensitivity, Specificity, and Predictive Values of Double-Layer sign for Identifying Subclinical Macular Neovascularization in Nonexudative Age-Related Macular Degeneration . . . 69
5.3 The number of the labeled and unlabeled data for each training size on SS-OCT dataset. . 71
5.4 The number of the labeled and unlabeled data for each training size on SD-OCT dataset. . 71

CHAPTER 1

INTRODUCTION

1.1 Background

Every day, large amounts of health-related data are generated by medical institutions, health and care services, and from other sources such as academic studies and research. These datasets consist of medical treatment information, scans, or more general information such as local or national population data, the number of babies being born or how many people are admitted to hospital on any day. The medical databases are being utilized to gain profound medical knowledge. By accessing large amounts of collected data, there have been impressive research results in a range of medical fields such as genetics, radiology, neuroscience, diagnosis, patient outcome prediction or drug discovery. Particularly deep learning systems, composed of millions of trainable parameters, require large amounts of data to learn meaningful representations robustly. Aside from quantity, the quality of the available patient-level data is particularly essential for medical research. Highly diverse and well-curated training data empowers researchers to produce generalizable insights and reduces the risk of biased predictions when applied in practice.

Among all of this medical data, medical imaging contributes significantly to progress in scientific discoveries and medicine. The purpose of medical imaging is to provide reliable information for accurate diagnosis and subsequent clinical decisions for effective patient care. Optical coherence tomography (OCT) has become ubiquitous in ophthalmic diagnostic imaging over the last three decades [2, 3]. However, clinical OCT image-quality is highly variable and often degraded by inherent speckle noise [4, 5], bulk-motion artifacts [6, 7, 8], and ocular opacities/pathologies [9, 10]. Poor image-quality can limit visualization and introduce errors in quantitative analysis of anatomic and pathologic features-of-interest.

In addition, medical datasets often suffer from the imbalance problem in which the majority of data instances (examples) belong to one class and far fewer instances belong to others. While in many applications, the minority instances actually represent the concept of interest, a classifier induced from an imbalanced data set is more likely to be biased towards the majority class and show very poor classification accuracy on the minority class. Despite extensive research efforts, data imbalance remains one of the most challenging problems in data mining and machine learning.

Another issue is annotation cost. Semantic segmentation partitions raw image data into structured and meaningful regions and thus enables further image analysis and quantification, which are critical for various applications, including anatomy research, disease diagnosis, treatment planning, and prognosis monitoring. However annotating medical images is a time-consuming, labor-intensive, and expensive process. Depending on the complexity of the regions of interest to segment and the local anatomical structures, minutes to hours may be required to annotate a single image. Furthermore, label noise is inevitable in real world applications of deep learning models. Such noise can result from systematic errors of the annotator, as well as from inter-annotator variation. More than three domain experts are typically needed to generate trustworthy annotations as any biases in the data can be transferred to the outcomes of the learned models. Consequently, the lack of large and high quality labeled datasets has been identified as the primary limitation of the application of supervised deep learning for medical imaging tasks. Crowdsourcing has become

popular in this context as it is based on outsourcing cognitive tasks to many anonymous, untrained individuals so-called "workers", from an online community. It has proven to be a valuable tool for cost-effective large-scale image annotation, in particular when the data cannot be processed by computers and is too large to be annotated by individuals. While expert labeling is expensive and time-consuming, collecting labels from crowdsourcing results in noise which may degenerate the accuracy of deep learning models. In many applications, labels are acquired from nonexperts (e.g., Amazon Mechanical Turk [11]) and sometimes automatically generated from the source information (e.g., downloading from social media with tags [12], extracting labels for x-ray images from associated radiology reports [13], [14]). These processes could introduce potential error or label noise into the model training.

In the field of medical image, imaging techniques such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and SinglePhoton Emission Computed Tomography (SPECT) have provided clinicians with information of the human body's structural characteristics, soft tissue, and so on. Different imaging methods keep different characteristics, and different sensors obtain different imaging information of the same part. The limited information provided by single modal medical images often cannot meet the need of clinical diagnosis which requires a large amount of information, making medical image fusion research become a hot field.

In summary, learning with imperfect datasets consisting of noisy images/labels, highly imbalanced labels, having limited or lacking target domain annotations, or multimodal data fusion are four of the most frequently encountered challenges in clinical applications.

1.2 Research Purpose

In this thesis, we aimed to present algorithms designed to gain profound medical knowledge from real world ophthalmic imaging datasets where these data challenges are present. Ophthalmic imaging is a highly specialized field of ophthalmology which helps doctors to diagnose and manage a wide variety of eye conditions. This is a rapidly developing area, with new instruments and techniques allowing physicians to better understand eye diseases and their treatments. One of the most remarkable developments in the field of ophthalmic imaging is optical coherence tomography (OCT). OCT is an imaging technique which generates cross-sectional images of tissue with high resolution. The basic OCT system and an example of retinal layers captured by OCT instrument are illustrated in Figure 1.1 and Figure 1.2. In OCT many one-dimensional scans (A-scans) are performed at several depths to create a two-dimensional image (B-scan). Those B-scans can be translated into a volumetric image (C-scan) of a retina (Figure 1.3). Since OCT is completely noninvasive, it provides *in vivo* images without impacting the tissue that is imaged. OCT has been applied in clinical research and practice in the ophthalmic field since the 90's, and has become the standard tool for imaging macular disease, diabetic retinopathy and glaucoma, to name a few examples from a wide range of retinal applications. Time-domain (TD) OCT was the first model of *in vivo* evaluation of the human retina and optic nerve until 1995, when Fourier-domain (FD) OCT was introduced for ocular

imaging. The advantages of the Fourier-domain OCT are improved speed and resolution. Because each A-scan is acquired all at once in Fourier-domain OCT, the acquisition rate is much higher at 16,000 to 40,000 A-scans per second. This acquisition rate allows for much faster scanning times, reducing motion artifacts, and enables denser and other novel scan patterns. Fourier-domain OCT is also able to improve axial resolution from 8-10 µm to 3-6 µm, which improves the ability to visualize intraretinal structures. Fourier-domain OCT can be further divided into spectral-domain OCT (SD-OCT) and swept-source OCT (SS-OCT). SD-OCT devices use a broadband near-infrared superluminescent diode as the light source with a spectrometer as the detector, while SS-OCT instruments apply a tunable swept laser as the light source with a single photodiode detector. OCT is now the standard-of-care in the management of macular diseases and glaucoma. In our studies, we have developed deep learning models based on three real-world OCT image datasets.

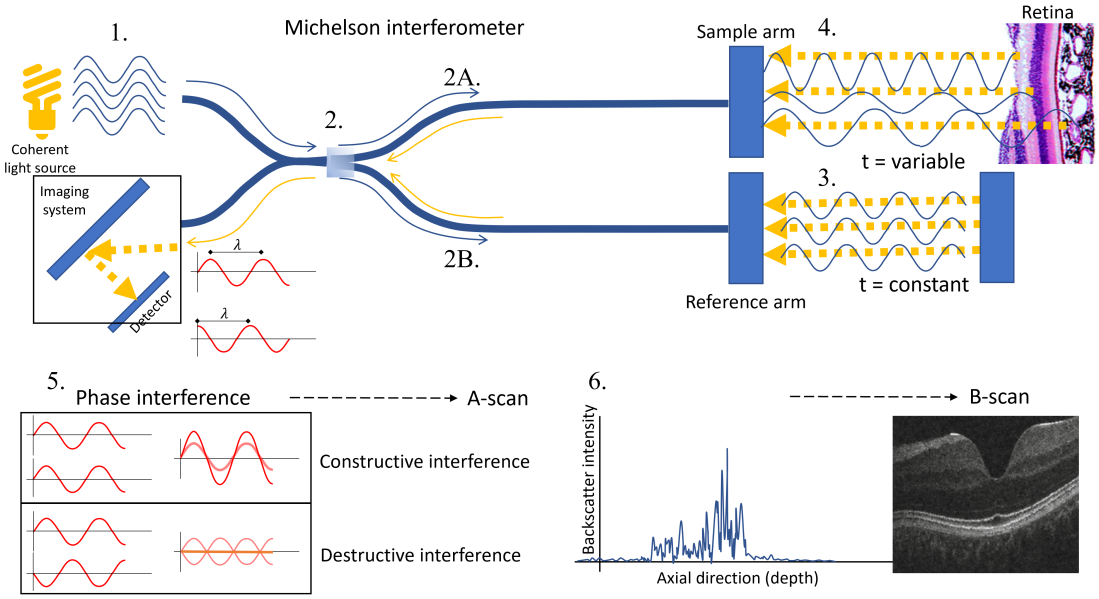

Figure 1.1: Optical Coherence Tomography. 1. White light source emits various wavelengths of light. 2. Beam splitter sends some light to 2B a reference arm and 2A the sample arm. 3. In the reference arm, light travels to a mirror where the distance to the mirror is known and constant. 4. In the sample arm, light travels to tissue of varying distance, such as different layers of the retina. The reflected frequencies and amplitudes of light are not constant, unlike the reference arm in 3. 5. The reflected light from the sample and reference arms return to a detector. At the detector, the phase interference is measured, and an a-scan is generated. 6. Multiple A-scans are stitched side-by-side together to create a b-scan, a cross-sectional representation of the retina.

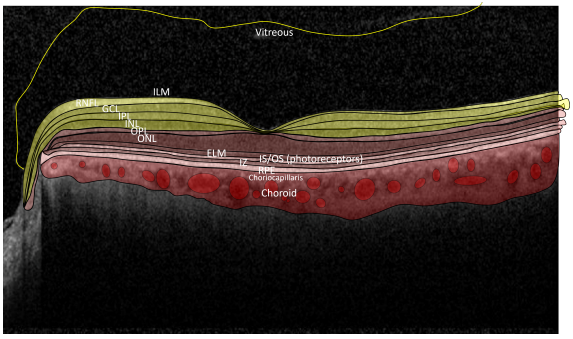

Figure 1.2: Normal retinal tissue layers image shot by OCT instrument.

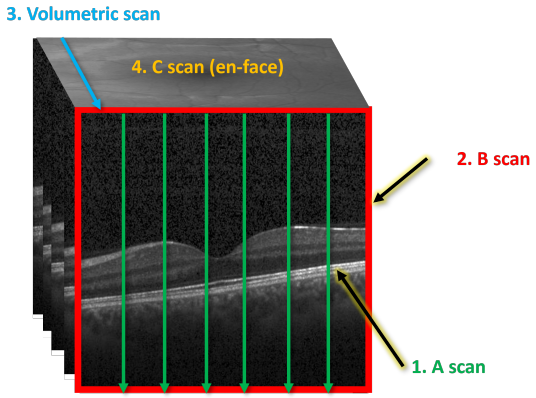

Figure 1.3: The breakdown of OCT scans. These are the different types of scans: 1. A-scans are a pixel-thick line of interference values (1-D line). 2. A-scans stitched together yields a B-scan (2-D image composed of a series of adjacent A-scans). 3. Multiple, aligned B-scans represents a volumetric scan (3D-image composed of multiple B-scans). 4. In a volumetric scan, an en-face (C-scan) OCT image can be generated where different slabs of the retina can be examined.

1.2.1 Glaucoma study

Glaucoma is a disease of the optic nerve head (ONH) characterized by loss of retinal ganglion cells and their axons. Glaucoma patients undergo a chronic progressive optic neuropathy with corresponding and characteristic patterns of visual field (VF) loss. VF tests are the benchmark for detecting and monitoring glaucoma. In the majority of patients, VF changes are initially localized and as the disease progresses, these focal areas become wider, deeper, and more numerous. However, test–retest variability of VF data seriously degrades the capacity to recognize true VF progression.

On the other hand, glaucomatous damage in turn can lead to progressive thinning of the retinal nerve fiber layer (RNFL) accompanied by structural changes within the optic nerve head (ONH). The ability of OCT to capture retinal layers allows for thickness measurement, which improves glaucoma diagnosis. Hence,

SD-OCT is a valuable clinical tool for glaucoma diagnosis and detection of progression. Circumpapillary OCT (Cp-OCT) scans are a quick and common OCT scanning pattern that capture a circular OCT section around the ONH. We thus considered that it would be beneficial if each point on VF could be accurately predicted from Cp-OCT imaging (Figure 3.1).

A large database of anonymized clinical data was curated from three regionally different National Health Service glaucoma clinics in the United Kingdom. Clinical data, imaging data, and VF data were extracted for a total of 24,248 patients. After applying inclusion and exclusion criteria, the total number of eligible study patients was 6,437 (11,025 eyes). Several challenges exist within the dataset. First, the dataset is highly imbalanced; the majority of the patients have mild symptoms, and the number of cases that have severe symptoms is limited. Second, quality assessment of the images was not performed in order to build an algorithm which works on images that would be reflective of those acquired in standard clinical settings.Therefore the dataset includes a certain amount of noisy images. To address these challenges, we took a multi-modal approach to make our model robust on the noisy images, and handled data imbalance by introducing weighting in the loss function. These will be discussed in detail in Chapter 3.

1.2.2 UK Biobank Foveal/Macular Curvature analysis

This project investigated foveal and macular curvature based on OCT imaging using data from the UK Biobank (UKBB), one of the world's largest single resources for comprehensive study of health and disease. The UKBB includes data from over 500,000 recruited adults and OCT data from more than 60,000 participants in the United Kingdom. For foveal curvature (FC), we aimed to understand the interindividual variation in the general population, and what factors may be associated with these differences. The fovea is a highly specialized retinal region at the center of the macula responsible for driving high visual acuity and color vision [15]. Although our understanding of the mechanisms and functional implications of cytoarchitectural and morphological foveal alterations is driven by studies in patients with absent or poorly formed foveal depressions with presence of inner retinal layers in selective settings, the interindividual variation of FC in the general population has not been explored well yet. The studies analyzing the OCT-derived foveal slope have been limited to using small to moderate sample sizes (typically with less than 400 subjects). Our objective is to address a gap in our knowledge by exploring the associations of sociodemographic, ocular, and early life factors and OCT-derived FC features in healthy individuals. For macular curvature (MC), we aimed to comprehensively and systematically investigate macular curvature in particular, the prevalence of the dome-shaped configuration and to evaluate potential associations that can be tested in later studies. Further insights into macular configurations and development might be of particular scientific interest for myopia research, the study of ocular development, and to investigate the many ocular diseases associated with this particular region, including AMD, the most common cause of legal blindness in the developed world [16].

For the calculation of FC and MC, the boundaries of the retinal pigment epithelium (RPE) and inter-

nal limiting membrane (ILM), and the RPE and RPE choroid (RPE/C) must be detected (Figure 1.2). In clinical research, OCT image segmentation (delineation of boundaries to allow measurements) is often performed manually by trained image graders. While highly accurate, such an approach is time-consuming and therefore not feasible for large studies such as the UKBB. As OCT imaging is increasingly incorporated into large, population-based epidemiological studies, approaches to allow for rapid, automated, quantitative analysis of OCT image sets will become increasingly necessary. In this study, I present a fast and fully-automated framework for generting OCT-derived FC and MC quantification that does not require any human annotations. These will be discussed in detail in Chapter 4.

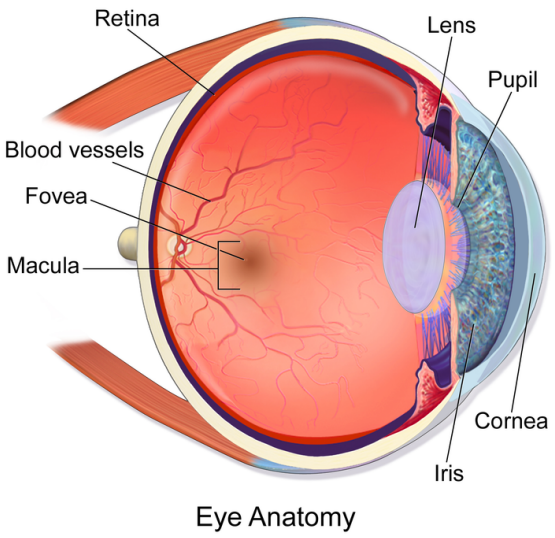

Figure 1.4: Macula and Fovea [1]

1.2.3 Double Layer Sign detection for age-related macular degeneration (AMD)

AMD is the fourth most prevalent ocular disease resulting in vision loss in the macula [17]. The macula is located in the optical center of the human eye and is an important part of the retina. It is required for reading, driving, watching TV, and performing many other daily activities [18]. Of all cases of blindness worldwide, 8.7% are caused by AMD and the number of patients with AMD was estimated at around 196 million in 2020, which is predicted to rise to 288 million by 2040 [19]. AMD is broadly classified into non-exudative or dry AMD and exudative or wet AMD. The difference between dry and wet AMD is that dry AMD does not have any blood or serum leakage. Around 85% to 90% of AMD cases are dry [20]. Patients suffering from dry AMD have a significant anomaly known as drusen in the RPE layer. The formation of drusen leads to a thinning and drying out of the macula, which results in the loss of macular function. Although patients with dry AMD may still have a good central vision, they may have significant functional

limitations, including limited night vision, vision fluctuations, and reading difficulties due to a limited area of central vision. Moreover, a certain percentage of dry AMD cases may develop into wet AMD as time goes by [21]. In wet AMD, patients may see dark spots in their central vision due to blood or fluid leakage under the macula. The main pathogenesis of wet AMD is choroidal neovascularization (CNV), which occurs under the retina and macula. This neovascularization may lead to macular swelling and a reversible loss of vision, or bleeding, which can be highly toxic to the overlying photoreceptors, sometimes even causing irreversible vision loss [22, 23]. In wet AMD, vision loss may be rapid and progressive. Once CNV has developed in one eye, the other eye is in a high-risk state and requires periodic eye examination [24]. Type 1 macular neovascularization (MNV) is the most frequent MNV occurring as a complication of AMD. This neovascularization arises from the choroid and grows under the RPE and resides between Bruch's membrane (BM) and the RPE [25]. In AMD, these nonexudative neovascular lesions have an increased risk of progressing to exudation and vision loss [26, 27, 28]. Since early detection and treatment of exudative AMD has been shown to result in better visual acuity outcomes, it is important to identify and closely follow nonexudative MNV (neMNV) before exudation develops so that treatment can be initiated once symptomatic exudation arises.

Type 1 neMNV is associated with the presence of a double layer sign (DLS) on OCT imaging. The DLS consists of 2 highly reflective layers that correspond to a separation between the RPE and another highly reflective layer beneath the RPE, which is presumed to be the BM. The presence of a DLS on structural OCT scans has been shown to be a useful strategy for identifying non-exudative MNV [29]. The specificity and sensitivity for identifying non-exudative MNV can be as high as 88% [29]. Thus we aimed to develop a deep learning algorithm to detect DLS based on cross-sectional structural OCT B-scans. These will be discussed in detail in Chapter 5.

A dataset of annotated structural B scans was fed into the algorithm to train the model. One hundred eyes with nonexudative AMD from 94 patients were used in this study. Patients were enrolled in a prospective OCT imaging study at the Bascom Palmer Eye Institute (BPEI). B-scans were annotated by graders from BPEI. The institutional review board of the University of Miami Miller School of Medicine approved this study. Informed consent was obtained from all patients. The study was performed in accordance with the tenets of the Declaration of Helsinki and complied with Health Insurance Portability and Accountability Act of 1996.

1.3 The organization of the thesis

The thesis comprises six chapters. Chapter 1 (this chapter) provides a general introduction to the medical datasets along with the motivation for this research. The main body, starting from Chapter 3 through Chapter 5, is devoted to the discussion of three groups of mutually interrelated topics. We shall finally provide a summary of our findings and some concluding remarks on Chapter 6.

CHAPTER 2

RELATED WORKS

2.1 Mitigating Data Imbalance Issues in Medical Image Analysis

While AI has seemingly limitless potential to gather new insights from medical images, implementing the technology in clinical settings comes with some major barriers. Many medical datasets suffer from the imbalance problem [30] which hampers the detection of outliers (e.g. rare health care events). Most data analysis methods assume an equal occurrence of classes and a standard classification learning algorithm might be biased towards the majority class and ignore the importance of the minority class (class of interest), which generally leads to the wrong diagnosis. In this way, analyzing medical datasets with imbalanced data is a significant challenge for the machine learning and deep learning community. Generally, data re-balancing during model training can be implemented in two ways: re-weighting or re-sampling. Re-weighting focuses on tuning the loss for different classes to redefine the importance of each class or sample. Re-sampling focuses on reconstructing a balanced dataset by either oversampling the minority classes or under-sampling the majority classes.

2.1.1 Re-Weighting

Re-weighting (cost-sensitive learning) focuses on tuning cost or loss to redefine the importance of each class or sample. Cost-sensitive re-weighting methods assign different weights to samples to adjust their importance. The naive practice of dealing with an imbalanced dataset is to directly use label frequencies of training samples for loss re-weighting, instead of heuristically using the number of classes [31] introduced the concept of "effective number" of a class, which takes the information overlapping into consideration for re-weighting. While these methods can successfully assign more weights to the minority samples, they assign the same weights to all samples belonging to the same class, regardless of each sample's importance. Such a loss can be further improved by tuning the "influence" of label frequencies on loss weights, based on a sample's influence on forming a decision boundary [32] or the distribution alignment between model prediction and a balanced reference class distribution [33]. Instead of using label frequencies of training data, LADE [34] proposed to use the label frequencies of test data (if available) to post-adjust model outputs, so that the trained model can be calibrated for arbitrary test class distribution.

Besides using a pre-defined weighting function, the class weights can also be learned from data. Shu *et al*. [35] proposed a meta-learning process to learn a weighting function guided by a balanced validation set. The weighting function which is approximated by a one-layer MLP is updated for fitting the long-tailed distribution, so that a well-performed model on the uniform test set can be learned. The recent paper by [36], based on empirical observations, raises a major concern about the impact of importance weighting for the deep learning models which are generally over-parameterized. They observe from experiments that there is little impact of importance weights on the converged deep neural network, if the data can be separated by the model using gradient descent. An intuitive alternative is to shift the separator closer to a

dominant class [37]. [38] proposed to integrate per-class margin into the cross-entropy loss (LDAM loss). The margin is inversely proportional to the prior probability of a class and thus can enforce larger margins between a tail class and other classes.

Another issue is that the classifier trained by the widely applied CrossEntropy Loss is highly biased on long tailed datasets. The major reason is that gradients brought by positive samples are overwhelmed by gradients from negative samples on tail classes. [39] devises Equalization Loss, which argues that the poor performance of tail classes originates from the over-suppression of samples from head classes. Since tail classes only contain few samples, they receive much more negative gradients than positive ones during training, thus they are consistently in a state of being suppressed in most of the training time. In order to prevent tail classifiers from being over-suppressed, Equalization Loss proposes to randomly drop some scores of tail classes in the Softmax function that can effectively help balance the positive and negative gradients flowing through the score outputs. Recently, [40] proposes Seesaw loss to adaptively re-balance the gradients of positive and negative samples with two re-weighting factors, i.e., a mitigation factor and a compensation factor. The mitigation factor reduces punishments to tail categories w.r.t. the ratio of cumulative training instances between different classes, and the compensation factor up-weights the penalty of misclassified instances to avoid false positives of tail categories.

2.1.2 Re-Sampling

The basic idea of resample-based methods is to either oversample the minority categories or to under-sample the frequent categories in the training process. Under-sampling discards part of the data in head classes and over-sampling repetitively samples data from the tail classes. Oversampling can lead to over-fitting to the tail classes [41] while under-sampling can potentially lose information about the head classes but may yield good results if each sample of a head class is close to other samples of the same class. Class-aware sampling [42] proposes to choose samples of each class with equal probabilities, which is widely used in vision tasks [43]. It first uniformly samples a class from the whole classes, and then samples an example from the selected class randomly. This process runs iteratively in each training epoch. In addition, [44] proposes to increase the sampling rate for classes with low performance after each training epoch and balance the feature learning for under-privileged classes. While in class-balanced sampling, each class has an equal probability of being selected, in instance-balanced sampling, each sample has an equal probability of being sampled. In addition, square-root sampling [45] is a variant of instance-balanced sampling, where the sampling probability for each class is related to the square root of sample size in the corresponding class. Progressively-balanced sampling [46] interpolates progressively between instance- and class-balanced sampling. Dynamic Curriculum Learning (DCL) [47] develops a new curriculum strategy to dynamically sample data for class rebalancing. To be specific, the more instances from one class that are sampled as training goes on, the lower the probabilities of this class being sampled later. Following this idea, DCL first conducts random sampling to learn general representations, and then samples more

tail-class instances based on the curriculum strategy to handle long-tailed class imbalance. Balanced meta-softmax [48] developed a meta learning based sampling method to estimate the optimal sampling rates of different classes for long-tailed learning. Specifically, the proposed meta-learning method, a bi-level optimization strategy, learns the best sample distribution parameter by optimizing the model classification performance on a balanced meta validation set. Feature Augmentation and Sampling Adaptation (FASA) [49] proposed to use the model classification loss on a balanced meta validation set (as a metric) to adjust feature sampling rate for different classes, so that the under-represented tail classes can be sampled more. Bilateral-Branch Network (BBN) [44] develops two network branches (i.e., a conventional learning branch and a re-balancing branch) to handle class imbalance based on a new bilateral sampling strategy. To be specific, BBN applies uniform sampling for the conventional branch to simulate the original long-tailed training distributions; meanwhile, BBN applies a reversed sampler for the rebalancing branch to sample more tail-class samples for improving tail-class performance. The final prediction is the weighted sum of the two network branches. Besides sampling for bilateral branches, Balanced Group Softmax (BAGS) [50] puts classes with similar numbers of training instances into the same group and computes group-wise softmax cross entropy loss respectively. BAGS achieves relative balance within each group, thus avoiding a severely biased classifier due to imbalance. Instead of division into several balanced groups, Ally Complementary Experts (ACE) [51] divide samples into several skills diverse subsets, where one subset contains all classes, one contains middle and tail classes, and another contains only tail classes. Based on these subsets, ACE trains different experts to have specific and complementary skills for ensemble learning.

Another recent line of research is re-sampling in the feature space rather than in the input space, such as Deep Oversampling (DOS) [52], Feature-space Augmentation (FSA) [53], and Meta Semantic Augmentation (MetaSAug) [54]. These methods aim to augment minority classes in the feature space by sampling from the in-class neighbors in the linear subspace, using learned features from pretrained networks [53], or using an implicit semantic data augmentation (ISDA) algorithm [55].

2.1.3 Other types

[56] reveals that stronger augmentation for minority classes is beneficial to mitigate overfitting and improve tail-class performance without sacrificing head-class performance. To generate diverse minority data, recent works [57, 58] have proposed adversarial augmentations by adding small noise to the input images. To this end, Majorto-minor Translation (M2m) [57] transfers knowledge from majority classes using a pre-trained network, and Balancing Long-Tailed datasets (BLT) [Kozerawski] uses a gradient-ascent image generator based on the confusion matrix. Since the data augmentation method is closely related to the oversampling methods, some recent long-tailed recognition methods have used the mixup method. [44] uses the mixup as a baseline method, and MiSLAS [59] uses mixup in its Stage1(=representation learning) training. While these methods apply mixup without any adjustments, Remix [60] assigns a label in favor of the minority classes when mixing two samples, and [61] proposes a balance-oriented data augmentation

named Uniform Mixup (UniMix) to promote mixup in long-tailed scenarios, which adopts advanced mixing factor and sampler in favor of the minority. [62] proposes a new oversampling method called Context rich Minority Oversampling (CMO) to transfer the rich contexts of majority samples to minority samples by applying CutMix [63] data augmentation; to diversify the limited context of the minority samples, an original image from a minority class is cropped in various sizes and pasted onto the various images from majority classes. Recently a new line of work approaches long-tailed image classification by decoupling the representation learning and classifier learning into two stages[Kang, Zhou, Zhang]. Decoupling [Kang] was the pioneering work that introduced the two-stage training scheme, which first trains a good representation network with natural sampling and then finetunes the classifier with class-balanced sampling. Parallel to this, the work [44] obtained similar conclusions empirically. In addition, a bilateral-branch network is proposed in [44], where one branch uses random sampling to learn head data and the other branch uses revered sampling to emphasize tailed data. The shared motivation of such work is that image feature learning and classifier learning may favor different data sampling strategies and thus the focus is to identify suitable sampling strategies for these two tasks. Specifically, they find that under cross-entropy loss, random data sampling can benefit feature learning more while class-balanced sampling is a better option for classifier learning. There are also attempts to modify the classifier to improve the performance on tail classes, e.g., using different classifiers for different groups of classes [50], or use two classifiers trained with different data samplers [64]. However, multi-stage training methods may rely on heuristic design. More recently, multi-expert frameworks have received increasing concern, e.g., BBN [44], RIDE [65], TADE [66] and ACE [51]. BBN [44], which assigns two branches with normal and reversed sampling, respectively, incorporates a cumulative learning strategy to adjust the bilateral training. BBN merges the two-stage methods into one, but still suffers from the same drawback of slight degradation of the head's accuracy. RIDE [65] are multi-expert architectures that learn diverse classifiers in parallel, combining knowledge distillation and distribution-aware expert selection. ACE [51] proposes to improve on this strategy by training different experts optimized for specific class groups. TADE [66] explores the multi-expert scheme and proposed to learn ensembling weights in an unsupervised manner at test time via contrastive learning to handle test distribution-agnostic long-tailed recognition, where the test class distribution can be either uniform or long-tailed. Ensembling strategies tend to achieve superior performance, which can be partly attributed to increased capacity -each expert learns a different set of features in the last two blocks of ResNet models, consistently increasing model capacity as new experts are introduced but can also leading to scalability concerns.

Recent studies also explore contrastive learning for long-tailed problems. KCL [67] proposed a k-positive contrastive loss to learn a balanced feature space, which helps to alleviate class imbalance and improve model generalization. Following that, Hybrid [68] introduced a prototypical contrastive learning strategy to enhance long-tailed learning. Parametric contrastive learning (PaCo) [69] further innovated supervised contrastive learning by adding a set of parametric learnable class centers, which play the same role as a classifier if regarding the class centers as the classifier weights. DRO-LT [70] extended the prototypical

contrastive learning with distribution robust optimization [71], which made the learned model more robust to data distribution shift.

2.2 Deep Multimodal Learning for Computer Vision

Unstructured real-world data can inherently take many forms, also known as modalities, often including visual and textual content. In this section, we mainly focused on visual modalities, such as images as a set of discrete signals from a variety of image sensors observing a common phenomena. There are many practical tasks that benefit from the use of multimodal data. In ophthalmology, multimodality imaging is widely considered to involve the incorporation of two or more imaging modalities concurrently or in a short period of time, for the purpose of diagnosis, prognostication, management and monitoring of disease. There has been rapid development in the field, from colour fundus photography, fluorescein angiography (FA), near-infrared (NIR) reflectance, fundus autofluorescence (FAF) and indocyanine green angiography (ICG) to high-resolution and swept-source optical coherence tomography (OCT), OCT-angiography (OCT-A) and adaptive optics.

The underlying motivation to use multimodal data is that complementary information could be extracted from each of the modalities considered for a given learning task, yielding a richer representation that could be used to produce much improved performance compared to using only a single modality. According to the fusion level in deep multimodal learning, we can distinguish three families of approaches: early fusion, where the raw modalities are combined ahead of feature extraction; intermediate fusion where the features respective to each modality are concatenated before classification; and late fusion, where the modality-wise classification results are combined [72].

2.2.1 Data-level fusion (Early Fusion)

Data level fusion is a traditional way of fusing multiple sources of data before being used as input to a machine-learning algorithm, either by removing correlations between modalities or representing the fused data in a lower-dimensional common subspace [73]. Techniques that accomplish one or both of these objectives include principal component analysis (PCA), independent components analysis, and canonical correlation analysis. The fused data are then presented to a machine-learning algorithm.

Early fusion is applicable for raw data or pre-processed data obtained from sensors. Data features should be extracted from the data before fusion, otherwise the process will be challenging, especially when the data sources have different sampling rates between the modalities. For instance, the sampling rate between different sensors could vary, or synchronized data from multiple data sources might not be available if one source produces discrete data while another source provides a continuous data stream. Data can be joined in many ways, however early fusion typically occurs through concatenation or pooling [74, 75]. The

assumption behind early data fusion is the conditional independence between the states of various sources of information. This assumption is not always true as multiple modalities can have highly correlated features. Another paper [76] also states that different modalities can contain information that are correlated to each other at a higher level. Thus, the outputs of each modality can be assumed to be processed independently of one another. Poria *et al.* [77] implemented early stage data fusion which involved concatenation of the features in a multimodal stream; this can be assumed to be the simplest form of early stage data fusion.

2.2.2 Feature-level fusion (Intermediate Fusion)

Feature-level multimodal fusion includes all the approaches which combine the available input data before performing the objective task [78, 79, 80]. In this case, the number of features extracted from different modalities must be combined in a unique vector (output) which will be considered as a unique input by the objective task. Feature-level fusion in a deep learning multimodal context is a fusion of different modality representations into a single hidden layer so that the model learns a joint representation of each of the modalities. An important characteristic of deep learning is its ability to learn hierarchical representations(features) from raw data. This characteristic can be exploited in multimodal learning to enable fine-grained control over how learned features are fused. Features can be learned from different kinds of layers including: 2D convolution, 3D convolution and fully connected. Different modalities can be fused simultaneously into a single shared representation layer or this can be performed gradually using one or multiple modalities at a time.

The main advantages of the feature-level multimodal fusion techniques are the need for a unique learning phase for the combined feature vector and the possibility of taking advantage of the correlation between multiple features from different modalities. Although this method is the most flexible method, allowing for data fusion at different stages of model training, it may lead to model overfitting or the network may fail to learn the relationship between each modality.

2.2.3 Decision-level fusion (Late Fusion)

Decision-level fusion refers to the aggregation of decisions from multiple regressors/classifiers, each trained on separate modalities, and the goal is to find a decision rule that selects one one them, which is inspired by the popularity of ensemble classifiers [81]. In general, these decision-level fusion strategies are much simpler to implement than other fusion strategies, particularly when the different modalities vary significantly in terms of data dimensionality, unit of measurement, and sampling rates. Various rules exist to determine how to finally combine each of the independently trained models. Some of the most well-known fusion techniques include linear weighted fusion [82, 83], Support Vector Machines (SVM) [84, 85], and Bayesian inference [86, 87]. A drawback to late fusion is that direct interaction effects between the variables are lacking because the variables of the model for one modality are never updated based on data from other

modalities.

2.3 Annotation-efficient deep learning for semantic segmentation

In medical imaging, data is commonly sparse, and labeling it is costly. Additionally, many problems are semantic segmentation problems, a task where each pixel in the image needs to be classified. Annotating image data for a segmentation task is more time consuming, and in some domains like medical imaging, has to be done by experts.

Recently, semi-supervised learning has gained attention in the medical image computing community. A lot of semi-supervised methods have been proposed for medical image analysis [88, 89, 90, 91, 92]. All of these methods combine both labeled and unlabeled data to train powerful and robust convolutional neural networks (CNNs) models.

The core challenge in semi-supervised settings lies in how to effectively utilize the unlabeled images. Entropy minimization and consistency regularization are two common strategies for SSL. The entropy minimization strategy argues that the unlabeled data can be used to ensure classes are well-separated, which can be achieved by encouraging the model to output low-entropy predictions. Popularized by the self-training pipeline, this strategy leverages unlabeled data in an explicit bootstrapping manner, where unlabeled data is assigned with pseudo labels to be jointly trained with manually labeled data. On the other hand, the intuition behind consistency based approaches is that the model output should remain unchanged when the input is perturbed. The consistency regularization enforces the current optimized model to yield stable and consistent predictions under various perturbations on the same unlabeled data.

2.3.1 Semi-supervised learning with Entropy minimization

Self-training via pseudo labeling is an explicit and classical method originating from around a decade ago. In the semi-supervised setting, particularly, it has been revisited in several tasks, including image classification [93], and object detection [94]. Recently it has been applied for semi-supervised segmentation [95, 96, 97, 98]. Pseudo segmentation maps on unlabeled data are obtained from a segmentation model previously trained on labeled data and then used to retrain the segmentation model. The process can be iterated several times. Various schemes are used for the pseudo segmentation maps. A naive way to generate pseudo segmentation maps is to directly use prediction from the model previously trained on labeled data, and apply iterative training for further improvement [89]. However, such pseudo labels are unreliable. Observing that it is difficult for a single model to counter its own prediction errors, [95, 96] use multiple networks leveraging inter-model disagreement between different models to locate pseudo label errors by training with diversity loss [95] or a dynamically re-weighted loss function, called Dynamic

Mutual Training [96]. [97] applied strong data augmentation to make better use of the unlabeled data. These strong augmentations are likely to shift the distribution of natural images and lead to a domain gap between training images and testing images. In order to avoid the distribution shift caused by strong data augmentation, they propose distribution-specific batch normalization (DSBN). [98] proposes a self-training framework that performs selective re-training via prioritizing reliable unlabeled samples to safely exploit the whole unlabeled set in an easy-to-hard curriculum learning manner; stability of produced pseudo masks during the supervised training phase serves as a measurement for reliability. There is another group of works applying Generative Adversarial Networks (GANs) [99] based methods to tackle this task, using the discriminator for distinguishing the predictions and the ground-truth segmentation to select high-confident segmentation predictions on unlabeled images as pseudo segmentation. [100] used a GAN framework to add large fake visual data to enforce real samples to be close in the feature space, which, in turn, improves multiclass pixel classification. [90] proposed an adversarial network that consists of a segmentation network and an evaluation network which assess segmentation quality by distinguishing between segmentation results of unannotated images and annotated ones. [101] and [102] turn to adversarial learning, and a discriminator or a multi-label mean teacher (MLMT) branch is added to select reliable predictions as pseudo labels. [103] extended the GAN Framework and added a secondary model to correct the predictions from the segmentation model.

2.3.2 Semi-supervised learning with Consistency regularization

Consistency regularization is widely studied in semi supervised semantic segmentation. It enforces the consistency of the predictions with various perturbations, e.g., input perturbation by augmenting input images [104, 105], feature perturbation [Ouali], and network perturbation [106]. The effectiveness of consistency regularization is often attributed to the cluster assumption, which states that decision surfaces should lie in low density regions of the data distribution. This typically holds in classification tasks, but there are often no low-density regions along class boundaries in semantic segmentation. [104] argue that the cluster assumption to which effectiveness of consistency regularization has been partially attributed does not hold in semantic segmentation. They augment the input images randomly and impose the consistency constraint between the predictions of augmented images, so that the decision function lies in the low-density region by applying Cutout [107] and CutMix [63]. [108] also studied pseudo supervision in a way similar to [104] with the CutMix augmentation but using consistency between dual differently initialized models. PseudoSeg [109] adapts the weak-to-strong consistency to segmentation scenarios and further applies a calibration module to refine the pseudo masks. The approach combines the idea of both consistency regularization and pseudo-labeling in segmentation; refining pseudo-labels with self-attention grad-CAM and a calibrated prediction fusion, and consistency training by enforcing pseudo labels to be robust to strongly-augmented data.

One potential weakness of consistency learning is that it assumes accurate predictions for unlabelled

images; incorrect predictions from one model will deteriorate the training for the other model, and vice versa. [110] extended the Mean Teacher model to improve the segmentation accuracy of unlabelled training images with a new auxiliary teacher and the replacement of Mean Teacher's MSE loss (as a consistency loss) with a stricter confidence-weighted CE loss (ConfCE) that allows strong convergence and overall better training accuracy. [111, 112, 108] manage to tackle this task with simpler mechanisms, such as enforcing similar predictions under multiple perturbed embeddings [111], under two different contextual crops [112], and between dual differently initialized models [108].

2.4 Summary

In this chapter, we summarized three common challenges that deep learning faces in medical image datasets and discussed previous scientific studies, in a systematic manner, which have attempted to address them. In the following chapters 3, 4, and 5, we present the deep learning frameworks we have developed for each case study where we faced these same challenges.

CHAPTER 3

MULTIMODAL DEEP LEARNING FOR PREDICTING VISUAL FIELDS FROM THE OPTIC DISC AND OCT IMAGING

.

3.1 Introduction

Glaucoma is a disease of the optic nerve head (ONH) characterized by loss of retinal ganglion cells and their axons. This in turn can lead to progressive damage to the visual field (VF) with specific patterns determined by the arcuate trajectories followed by the axons as they enter the ONH. In clinical practice, glaucomatous damage can be detected and monitored through functional and structural evaluations. VF tests are known for requiring strong cooperation from the patient and, as currently implemented, are affected by considerable test-retest variability, especially in people with advanced VF damage. Structural evaluations of the macular region and the ONH are often done with imaging devices, most commonly fundus cameras and Spectral Domain Optical Coherence Tomography (SD-OCT). The latter has the advantage of providing cross-sections of the retina and of the ONH that allow precise evaluation of their layers. Structural damage from glaucoma often manifests itself as thinning of the retinal nerve fiber layer (RNFL, where the ganglion cell axons reside). Circumpapillary OCT (Cp-OCT) scans are a quick and common OCT scanning pattern that capture a circular OCT section around the ONH. In some devices, this is coupled with a two dimensional Scanning Laser Ophthalmoscopy (SLO) image of the ONH, often used to track eye movements during the acquisition. An example of Cp-OCT, SLO images, and VF test of a patient with glaucoma is showen in Figure 3.1.

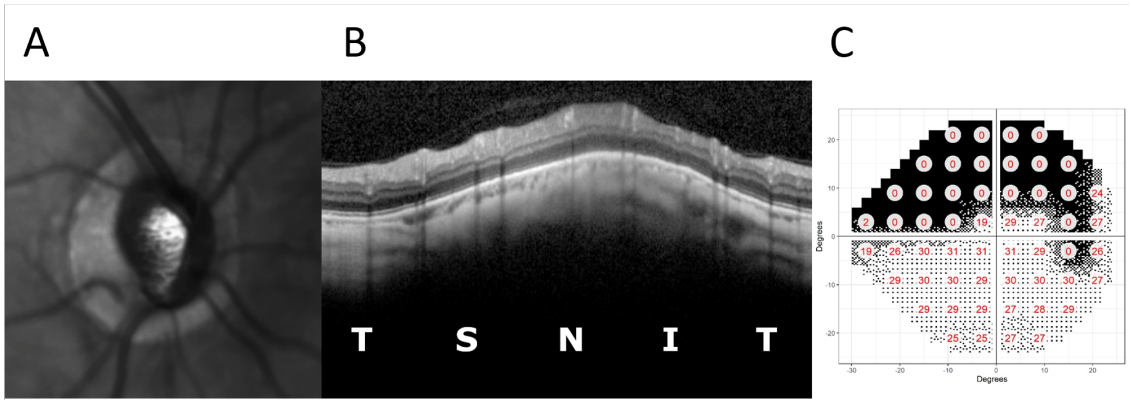

Figure 3.1: An example of a disc (A) and OCT (B) images alongside the corresponding HVF (C). T = temporal; S = superior; N = nasal; I = inferior.

Matching VF testing and structural data in glaucoma has proven problematic for several reasons. The first challenge is posed by the complex non-linear spatial mapping of Standard Automated Perimetry (SAP) locations to the ONH. Although empirical maps exist to describe average trajectories, these are often difficult to customize for individual patients and become unreliable with increasing distance from the ONH. Moreover, structural measurements are often affected by a strong measurement floor effect, whereby little change in structure is observed for large variations in function, especially for more advanced damage. Previous work has shown the potential of machine learning techniques to model such a complex problem. More recently, advancements in the field of Artificial Intelligence (AI) and deep learning (DL), especially for

image analysis with the introduction of Convolutional Neural Networks (CNN), have reignited the interest around structure-function predictions in glaucoma. Recent attempts focused on predicting SAP sensitivity from segmented CpRNFL thickness profiles. Although simple, such an approach relies on segmentations being available and correct. Moreover, it discards valuable information from OCT reflectivity. Finally, little effort has been put into predicting VF data from ONH planar pictures. Despite providing only limited direct quantitative data, these images are rich with contextual information, for example on the general shape of the ONH and the position of the blood vessels, that can be efficiently exploited by CNN architectures. Our work tested the hypothesis that a hybrid DL method combining information from paired ONH and Cp-OCT images can not only improve the prediction of SAP sensitivity directly from structural data without segmentation but also lead to a fully AI-derived structure function mapping. We used a large clinical database acquired from routine glaucoma clinics to develop two separate DL sub-models, each able to provide an estimate of VF sensitivity from either OCT or infrared SLO images. A third policy DL model then combines the two predictions by evaluating feature vectors generated by the two sub-models. Compared to previous work, our work focused on developing a model to predict VF sensitivity values from a combination of different imaging modalities in an agnostic fashion, particularly without any use of segmented OCT data.

3.2 Multi-modal Policy model

We aimed to develop and validate a deep learning (DL) system for predicting each point on VF from disc and OCT imaging and derive a structure-function mapping. We train input specific deep neural networks for each input source, show the potential of forging them together into a multi-modal architecture and train a novel policy network that learns to choose between them. Our approach is considered as a feature level fusion (intermediate fusion) (See Section 2.2). For implementing early fusion, there is no straightforward way to directly align and combine the two input image modalities; one being an en-face fundus image and the other being a peripapillary OCT scan described in a polar coordinate system. Besides, observing the generated feature maps from each modality may include richer context rather than labels, we assume combined feature maps serve as a better indicator for the final policy decision. Our experiments suggest that the OCT information is more informative than the disc images for the task of VF sensitivity prediction. However, for a relatively large number of products ($\sim 8\%$), the disc CNN is correct while the OCT CNN is wrong, indicating a potential gain from using a multi-modal architecture. We show that we can train a deep policy to choose between the two models and give a performance improvement over both state-of-the-art networks.

3.2.1 Single modality DL model

Two separate DL models were trained using an EfficientNet B2 architecture [113], one for the disc and one for the OCT. EfficientNet is one of the state-of-the-art architectures, that designed for a specific input resolution and optimized by a scaling method that uniformly scales each dimension of depth, width and resolution with a fixed set of coefficients. The input resolution for EfficientNet B2 is 260×260 pixels. For training, the batch size was set to 32, NAdam was used as the optimizer, and the initial learning-rate was set to 5×10^{-3}. The last layer of the model had 54 nodes corresponding to the 54 VF sensitivity points on the 24-2.

3.2.2 Multi modality Policy DL model

A policy deep learning model was constructed to take the feature maps from both models and combine using a 1×1 convolutional layer after concatenation of the feature maps. We then used a final fully connected layer which was tasked with learning the relevant connections between the individual values in the 1 dimensional tensor, regardless of strict spatial proximity in the original image space. A network diagram of the complete end-to-end multimodal deep learning model is shown in Figure 3.2, along with a detailed description of the loss function for the policy network.

For each training sample, we generated our target of a binary 54 element vector where each location was labeled as 0 for the disc prediction being closer to the true (observed) HVF and 1 for OCT prediction being closer to true (observed) HVF. The output of the policy network was a 54 dimensional vector where each element was a floating point value from 0 to 1, where 0 represents an absolute preference for the disc model prediction and 1 represents an absolute preference for the OCT model prediction at each VF location. Since this value was generated by a sigmoid function, we chose 0.5 as the threshold to decide between adopting the prediction from the disc or OCT model. The multimodal network also could be trained on VF sensitivities directly, instead of preference between disc and OCT. We purposely chose such an approach considering the importance of interpretability. With our policy model, the prediction indicates "reasoning" of how to select preference among the two sub-models to achieve the best prediction as shown in Figure 3.5 with interpretability plots.

We used the same training settings (learning rate, optimizer, and batch size) as for the base models, with a weighted binary cross-entropy loss function. Considering the limited size of the dataset, we trained policy network while holding all the parameters in the base models "frozen".

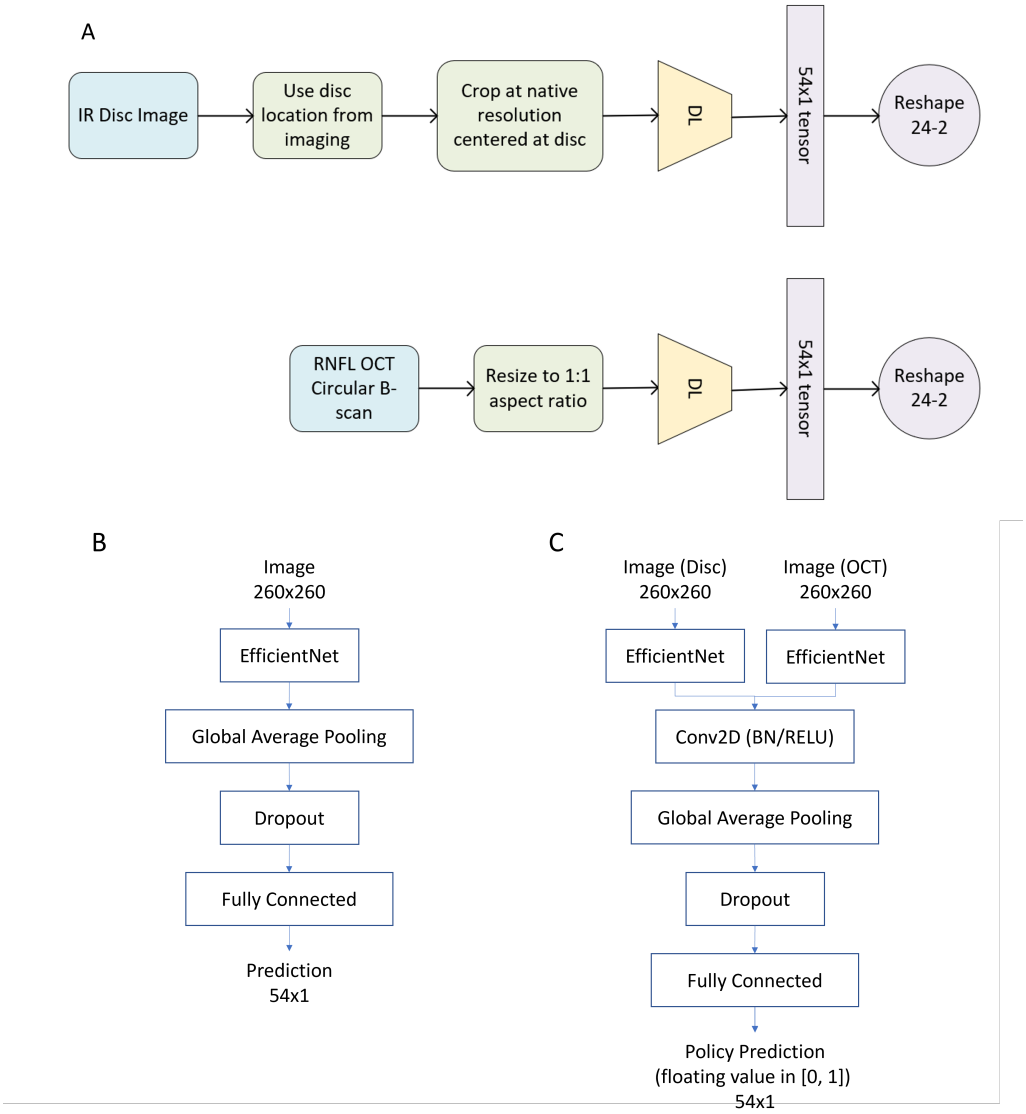

Figure 3.2: A) Process overview of VF prediction from the DL model using the IR SLO image (top) and the OCT image(bottom). B) Model architecture of SLO and OCT models. C) Model architecture of Policy model. The feature maps obtained via EfficientNet from the IR SLO and OCT images were concatenated. We then applied a 1x1 convolution to obtain a multimodal feature representation while holding all the parameters in the base models "frozen" during the training of the policy model. The output of the policy network (54-dimensional vector) represents the policy preference for each VF location ranging from 0 to 1, generated by a fully connected layer. Values closer to 0 indicate a preference for the disc model, values closer to 1 indicate a preference for the OCT model. Since the target labels of the policy model were binary (0 = disc preferred, 1 = OCT preferred), we used a class weighted binary cross-entropy loss function. We also applied a pointwise "risk weight"" which reflects how much the error increases when the wrong choice is made.

3.3 Mitigating data imbalance

3.3.1 Sensitivity-weighted loss function

One important aspect for us was to evaluate the performance of our system in a clinical setting. We used a large clinical database acquired from routine glaucoma clinics as it is, and did not perform any quality assessment on the images. The dataset was highly imbalanced; the majority of VF sensitivity values were around 30 dB, which caused the model to fail to learn patterns of advanced VF loss. To overcome this limitation, we employed the sensitivity-weighted mean square error for the loss function

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\alpha t_i}(t_i - p_i)^2,$$

where t_i is 1×54 true sensitivities and p_i is 1×54 predicted sensitivities and the mean was calculated over samples in the same training batch. In this case α is a hyperparameter that controls how strongly the model weighs lower sensitivities. We used $\alpha = 2$ for our training. The MAE reported for the test dataset was calculated without applying any sensitivity-weights.

3.3.2 Risk-weighted loss function for policy network

For the policy model where output has a binary form (0 =Disc preferred, 1 =OCT preferred), class weight that would cancel out the class imbalance in the dataset has been applied in binary cross-entropy loss. However, the hard labeling still does not reflect how much the error (i.e. the gap between a true and a predicted sensitivity) increases when the wrong choice is made. We therefore applied the "risk weight" for each sample point. Since the target labels of the policy model were binary (0 =Disc preferred, 1 = OCT preferred), we used a class weighted binary cross-entropy loss function. We also applied a pointwise "risk weight" reflects how much the error increases when the wrong choice is made, that is calculated as follows.

$$max\{|p_i^{disc} - t_i|, |p_i^{oct} - t_i|\}$$

3.4 Experiment

3.4.1 Data preparation

Clinical data, imaging data, and HVF data were extracted for a total of 24,248 patients. After applying inclusion and exclusion criteria, the total number of eligible study patients was 6,437 (11,025 eyes) . A flow chart of the patient selection process is shown in Figure 3.3. The training, validation, and held-out test set contained 11,417 HVF-OCT pairs for 3,725 patients, 4,845 pairs for 1,323 patients, and 5,078 pairs for 1,389 patients, respectively. Note that each VF could be paired with multiple OCTs in the training set,

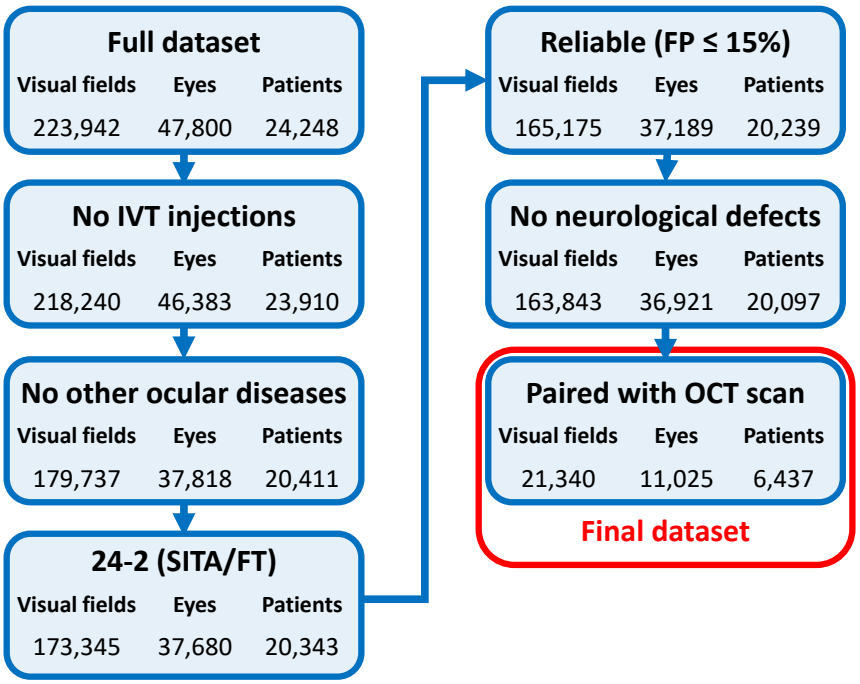

Figure 3.3: Flowchart of the selection process from the full dataset of VF tests. FP = false positive; FT = full-threshold; IVT = intravitreal; SITA = Swedish interactive threshold algorithm; VF = visual field.

but only with the closest OCT for the validation and test sets . Descriptive statistics for the three mutually exclusive sets of patients are shown in Table 3.1.

OCT scan patterns of the optic nerve from Heidelberg Spectralis (either OCT-1 or OCT-2) were identified in the database. Each HVF was linked to OCT scans performed within 7 days. For the training set, we retained all possible pairs, because any variation in the data within such a short time interval could be attributed to test retest variability; this effectively served as a form of data augmentation. For the validation and test sets, we only retained pairs with the OCT scan closest to the HVF, within the selected time-frame. The infrared reflectance (IR) image of the optic disc and the circular B-scan set at 3.5 mm were then preprocessed. For the disc IR image, the manually selected central point of the optic disc during acquisition by the photographer was used to crop a 260×260 pixel image at a native resolution as an eight-bit grayscale image. The circular OCT B-scan was resized to a 1:1 aspect ratio and resized to 260×260 as an eight-bit grayscale image. Each of the paired HVFs was collapsed to a one-dimensional 54 element vector representing each sensitivity point on the 24-2 VF. Two blind spot locations (X = 15; Y = +/ 3 degrees for a right eye) were discarded from the 54 dimensional vector at the time of evaluation. The VF sensitivity values were clipped between 0 and 40 dB and normalized to be between 0 and 1. The upper bound was chosen as it is commonly considered the highest plausible dB value in visually healthy subjects in standard testing conditions. The imaging data was normalized to be between 0 and 1. The data were partitioned into

	Training set	Test set	Validation set
# Patients	4,786	1,585	1,541
# Eyes	8,883	2,902	2,840
# Visual fields	22,866	7,370	7,411
# OCT scans	21,963	7,042	7,110
Quality index (dB)	25 [22, 28]	25 [22, 27]	25 [22, 27.25]
MD (dB)	−1.71 [−4.39, −0.23]	−1.65 [−4.27, −0.18]	−1.63 [−4.25, −0.19]
PSD (dB)	2.07 [1.57, 4.44]	2.06 [1.57, 4.24]	2.1 [1.56, 4.34]
FP (%)	1 [0, 4]	1 [0, 4]	1 [0, 5]
Sex (F / M)	2,492 / 2,294	791 / 794	743 / 798
Baseline age (years)	71 [62, 78]	71 [62, 79]	70 [62, 78]
Average IOP (mmHg)	17.5 [15, 20.25]	17.6 [15, 20.2]	17.33 [14.83, 20]
Cataract surgery (Yes / No)	989 / 7,894	314 / 2,588	324 / 2,516

Table 3.1: Descriptive Statistics of the Selected Sample. Continuous variables are reported as median [interquartile range]. F = female; FP = false positive; IOP = intraocular pressure; M = male; MD = mean deviation; OCT = optical coherence tomography; PSD = pattern standard deviation.

60%, 20%, and 20% for training, validation and held-out test sets at the patient level, respectively.

3.4.2 Results

To visualize the trained model predictions, a mean-occlusion mask was iteratively applied to each pixel position and the change in the predicted VF sensitivities was quantified. Videos of the occlusion affecting the HVF were generated by iterating the occlusion over each position as a separate frame. The heatmap describing magnitude of each difference was projected back onto the 24-2 to determine the occlusion locations that led to the greatest change in the predicted HVF. The final performance of the disc, OCT, and multimodal policy DL models was assessed in the final held-out test set at the end of the study. Pointwise mean absolute error (PMAE) and standard deviation (STD) were assessed by taking the mean of the absolute difference for each sensitivity point between the predicted VF and the true clinical VF.

The final PMAE of the single modality DL model using the IR SLO image of the disc alone was 3.62 dB (95% CI: 3.54 to 3.68 dB). The PMAE of the single modality DL model using the OCT alone was slightly better at 3.19 dB (95% CI: 3.13 to 3.25 dB). EfficientNet has been shown to achieve better performance compared to other well-known CNN architectures such as ResNet and DenseNet with a smaller amount of parameters. For comparison, we report the prediction performance of our model with different architectures in Table 3.4 and 3.3. Other methods of combining outputs from each single modality DL model were

examined and are reported in Table 3.2. The best performing model was the policy fusion DL model that combined the predictions from the OCT and disc sub-models with a PMAE of 3.13 dB (95% CI: 3.07 to 3.19 dB). The details are shown in Table 3.3. Clinical examples of the policy DL model are shown in Figure 3.5 with interpretability plots of the policy model displaying the confidence in choosing the model output from disc or OCT. Failure examples where the model predictions deviated from the clinical HVF are shown in Figure 3.6.

Table 3.2: Prediction performance using other fusion algorithms. Mean: average the two base model outputs, Blending: calculate multiplication by predicted probability to blend model outputs proportionally, Policy (selected approach): use either prediction from disc model or OCT model with threshold=0.5. For each cell, MAE with 95% Confidence Intervals are reported.

Mean	Blending	Policy
3.37 (3.31 to 3.43)	3.21 (3.15 to 3.26)	3.13 (3.07 to 3.19)

Table 3.3: Prediction performance on other EfficientNet types. For each cell, MAE with 95% Confidence Intervals are reported.

Architecture	Disc to HVF	OCT to HVF	Policy
EfficientNet B4	3.81 (3.74 to 3.88)	3.36 (3.30 to 3.42)	3.24 (3.18 to 3.30)
EfficientNet B3	3.48 (3.41 to 3.56)	3.22 (3.16 to 3.28)	3.17 (3.10 to 3.23)
EfficientNet B2	3.62 (3.54 to 3.69)	3.19 (3.13 to 3.25)	3.13 (3.07 to 3.19)
EfficientNet B1	3.55 (3.48 to 3.62)	3.11 (3.05 to 3.17)	3.08 (3.01 to 3.14)
EfficientNet B0	3.89 (3.81 to 3.96)	3.46 (3.39 to 3.52)	3.29 (3.23 to 3.36)

To fairly assess the final model performance across the full range of sensitivities, the pointwise error was plotted for each sensitivity threshold from 0 to 36 dB in the held-out test set (Figure 3.4). In order to probe the structure-function mapping learned by the model, a mean-occlusion mask was iteratively applied to each pixel position on the disc and OCT, thereby measuring the effect on HVF predictions. A remapping of the areas leading to the largest magnitude change for both models is shown as heatmaps in Figure 3.8B (for the disc SLO), Figure 3.8C (for the OCT), and Figure 3.8D (for the Policy). The structure-function mapping for the whole held-out test set is reported in Figure 3.9 for the Cp-RNFL. The map was built by plotting

Table 3.4: Prediction performance using DenseNet. Two different architectures were used to train the disc model and OCT model. The parameter size of DenseNet 201 is comparable with EfficientNet B4, and the parameter size of DenseNet 121 is comparable with EfficientNet B2. EfficientNet achieved a better MAE. For each cell, MAE and 95% Confidence Intervals are reported.

Architecture	Disc to HVF	OCT to HVF
DenseNet201	4.48 (4.39 to 4.56)	3.51 (3.45 to 3.57)
DenseNet121	4.29 (4.21 to 4.37)	3.48 (3.42 to 3.54)

Table 3.5: Prediction performance with different allowed time gaps between imaging and visual field test. For each cell, MAE with 95% Confidence Intervals are reported.

Interval	Disc to HVF	OCT to HVF	Policy
7 days	3.62 (3.54 to 3.68)	3.19 (3.13 to 3.25)	3.13 (3.07 to 3.19)
30 days	3.8 (3.74 to 3.87)	3.38 (3.32 to 3.43)	3.26 (3.20 to 3.32)
60 days	3.81 (3.75 to 3.87)	3.19 (3.13 to 3.25)	3.32 (3.26 to 3.37)
90 days	3.67 (3.61 to 3.73)	3.26 (3.20 to 3.31)	3.18 (3.13 to 3.24)

the distribution of the circumpapillary angles at which the highest peak of the heat-map was located, in polar coordinates. The map was also compared with previously reported structure-function maps based on anatomical data [114, 115, 116, 117, 118]. We demonstrated that the policy network implicitly optimizes the multimodal architecture accuracy, by explicitly learning to choose between single modality networks. However, it is not clear on what basis the policy network makes such a choice. The observation that the preference of the policy network appeared to be influenced by the quality of each image raises an interesting question: is the policy choice linked with the actual superiority on the prediction performance of the two models? To answer this question, we added several levels of noise to either SLO or OCT to degrade its performance on one of these models, and then observed how that affected the policy choice. As the level of noise on SLO increased and as the prediction performance from the SLO network degraded (PMAE: from 3.62 to 4.61), the policy network adopted the prediction from OCT model more (Policy choice: from 63.6% to 67.7%). On the contrary, it trusted the prediction from the SLO model more (Policy choice: 63.6% to 50.4%) as the level of noise on OCT increased and the prediction performance from OCT network degraded (PMAE: from 3.19 to 8.84). Throughout the experiments, our policy network kept a stable performance

(PMAE: from 3.13 to 3.16 with noise on SLO, PMAE: from 3.16 to 3.23 with noise on OCT). The detailed results are shown in Figure 3.7.

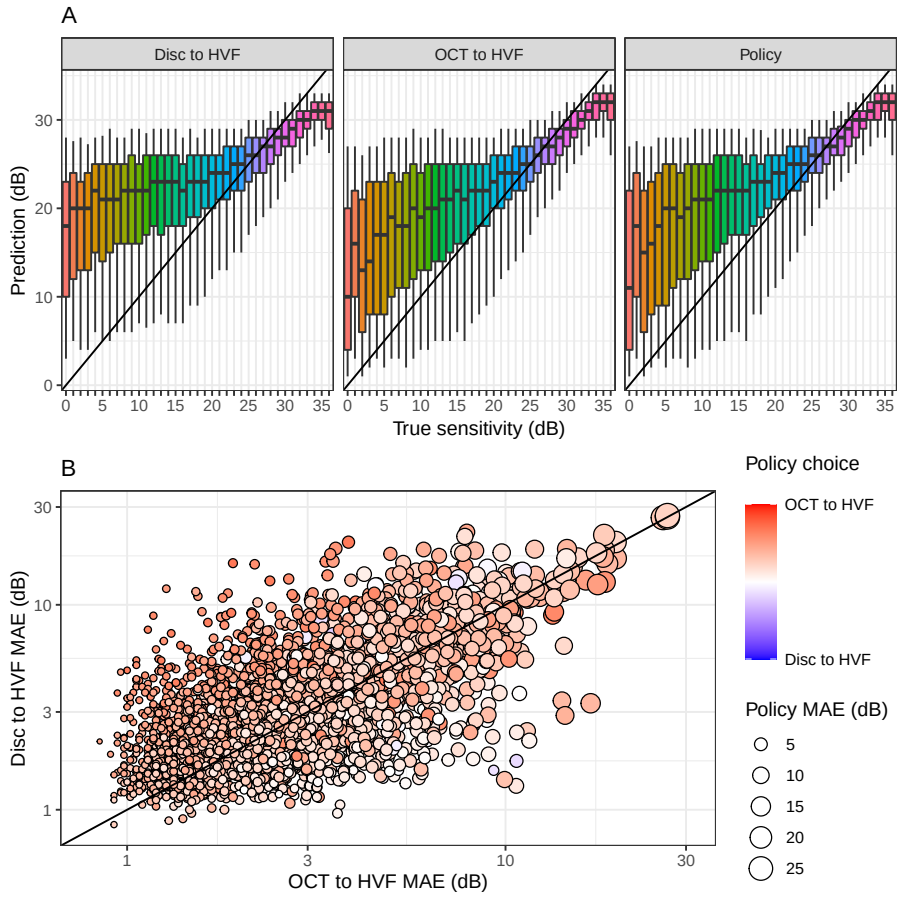

Figure 3.4: A, Predictions from the two sub-models and the policy network for each level of sensitivity. The diagonal line indicates perfect agreement. B, Relationship between the prediction Mean Absolute Error (MAE) from the two sub-models and the final average choice from the policy network. Higher MAE in the disc to HVF model is generally associated with a preference for the OCT to HVF model and vice-versa. The diagonal line indicates equivalent MAE. HVF = Humphrey visual field; MAE = mean absolute error; OCT = optical coherence tomography.

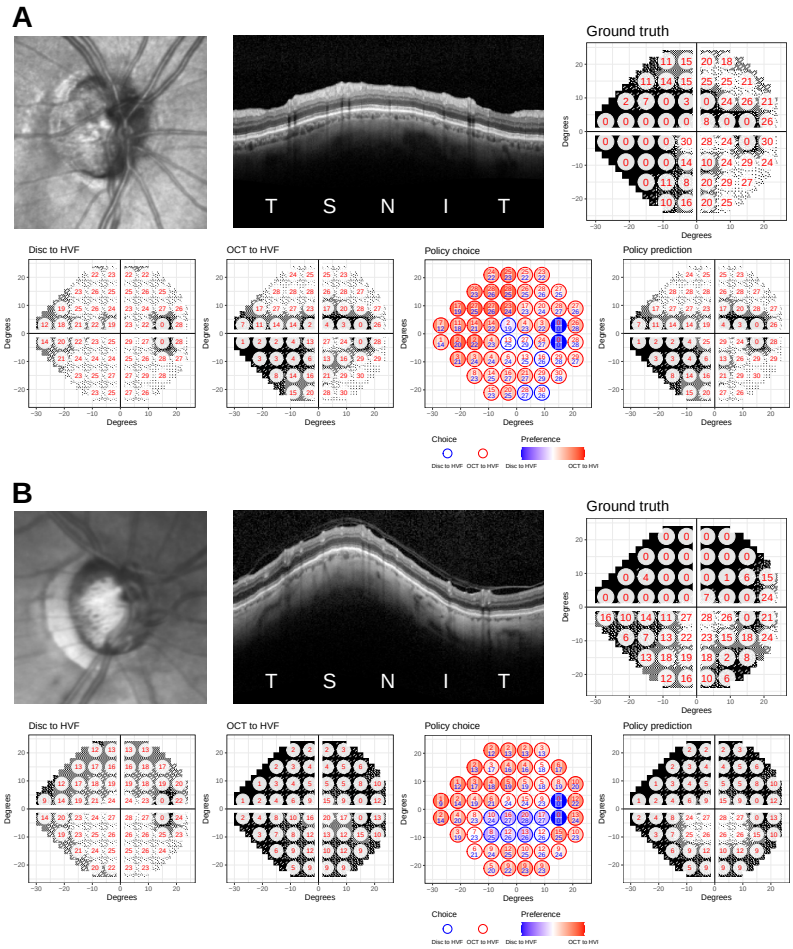

Figure 3.5: Prediction examples from the test set. The top of each panel reports the disc and OCT images used for predictions alongside the ground-truth HVF. The bottom of each panel shows, in order, the predictions from the two sub-models, the choice score of the policy network for each location, and the final prediction of the policy network. A, Both sub-models predict a similar location of the defect, but the OCT to HVF is more accurate in predicting the magnitude of glaucoma damage; the policy network correctly selects this prediction for the superior hemifield. B, the prediction from either individual sub-model is wrong, showing either little damage or diffused advanced loss; however, the policy network correctly selects predictions for each location to obtain a result very close to the ground truth, better characterizing the spared paracentral VF. HVF = Humphrey visual field; I = inferior; N = nasal; OCT = optical coherence tomography; S = superior; T = temporal.

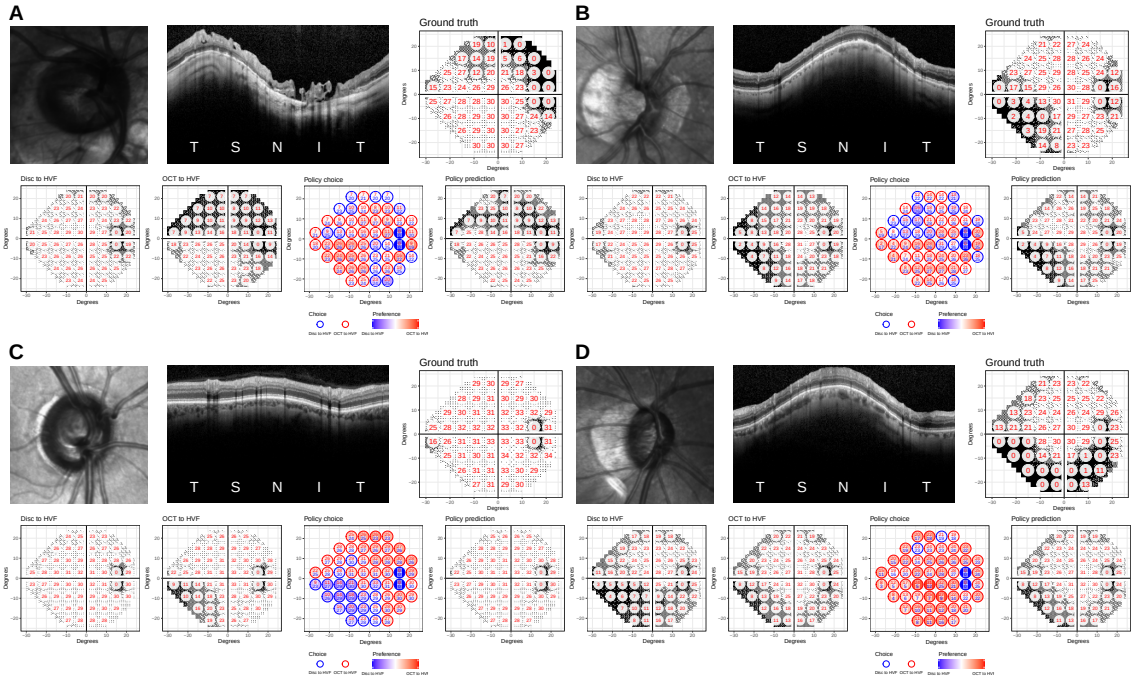

Figure 3.6: Other predictions from the test set. A) Case of myopic disc with extensive peripapillary atrophy and full thickness retinal damage in the inferior sector. Interestingly, the superior loss due to the full thickness defect is captured by the OCT to HVF model and correctly selected by the policy network. B) Synergic interaction, where the policy network is able to select inputs from the two sub-models to achieve a result that would not be possible with either individual sub-model. C) OCT to HVF produces an exaggerated nasal inferior defect; the policy network correctly prefers the result of the Disc to HVF prediction D) Prediction failure: the disc prediction was closer to the ground-truth but the policy network relies more on the OCT prediction. Of notice, both models produced similar results but the Disc to HVF model predicts a shallower HVF defect compared to the ground truth.

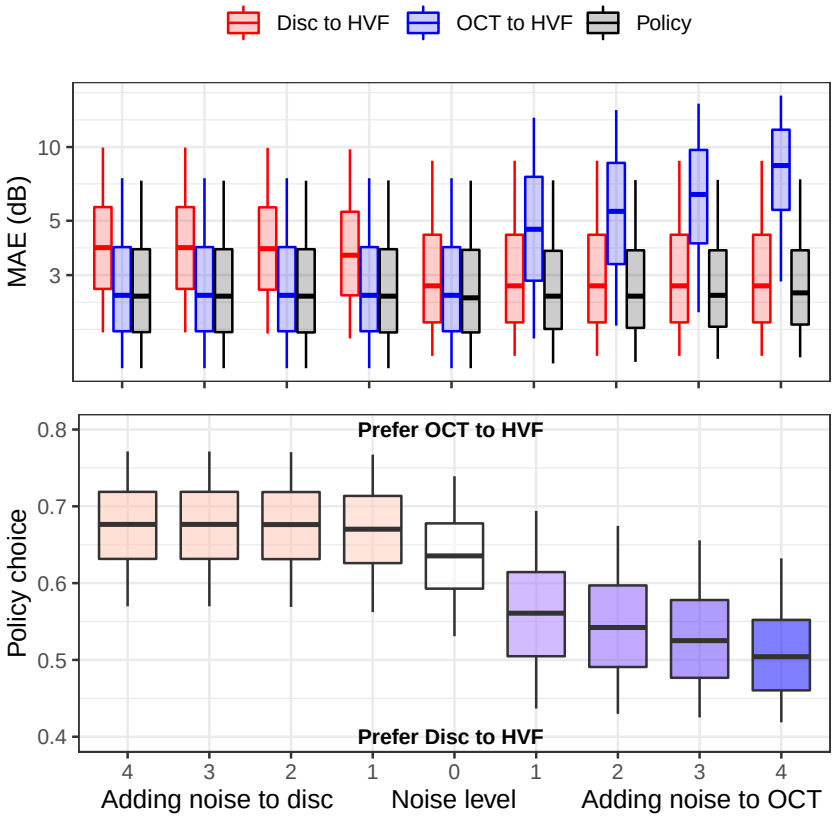

Figure 3.7: Prediction performance (top) and policy choice (bottom) for different noise levels applied to disc and OCT images. As expected, adding noise to OCT images degraded the performance of the OCT prediction (red) while leaving the disc prediction unaltered (blue). The opposite was true when adding noise to the disc images. The policy network was able to maintain a stable performance throughout. This was achieved by changing the preference for inputs from the disc or the OCT prediction according to which provided more reliable results (i.e. contained less noise). In other words, the policy network was able to adapt its choice based on changes in the quality of the input image. The gaussian blur was added to the disc images (kernel size from 0 to 41 pixels from level 0 to 4) while brightness was changed on OCT images (intensities decreased by a value from 0 to 120 from level 0 to 4, over a maximum of 255 and clipped at 0).

3.5 Discussion

Our work demonstrates an end-to-end DL approach for accurately predicting pointwise 24-2 HVF threshold sensitivities from the combination of the IR SLO ONH image and the Cp-OCT B-scan image. In addition, we found that the structure-function relationship that was learned directly from the imaging data in a fully agnostic, data-driven manner falls within prior manually derived efforts.

3.5.1 Explainability

The whole process simultaneously provides three distinct outputs (prediction from OCT, prediction from SLO of the disc, and policy-based fusion of the two results), with each producing directly interpretable results (HVF sensitivity maps). Importantly, the policy model is able to selectively choose between predictions from the two sub-models, thereby improving, for example, the spatial localization of deep defects (Figure 3.5B). The preference of the policy network appeared to be influenced by the quality of each image. Therefore, the policy model seems adept at working with real-world data, which might include a preponderance of artifactual image quality in OCT/SLO scans. The policy network still has the potential to fail if both the SLO and the OCT image are affected by artifacts or poor quality. We explored these aspects more in detail in Figure 3.7 by simulating different amounts of noise for either the OCT or the disc image.

3.5.2 Structure-Function Mapping

Our structure-function mapping was agnostically learned by the two sub-models from the data alone. Remarkably, the OCT-based mapping yielded results in agreement with other mapping schemes based on anatomical features of the RNFL bundles [114, 115, 116, 117, 118]. In Figure 3.9, we represent the structure-function mapping for the Cp-OCT scans in the test set by plotting the angle at which the peak of each heatmap (see Figure 3.8 C) was found around the ONH. Our technique indicated some spread in the distribution of the peaks, suggesting adaptive mapping based on individual features of the Cp-RNFL scans. In many of the locations, this matched the inter-individual variability observed by Jansonius *et al.* [114]. Inevitably, some of the variability is due to measurement noise, as evident by the larger spread in the distribution for the temporal VF locations, where perimetric defects are less commonly detected [119, 120]. One notable difference is the mapping of the macular locations, which in our model are shifted more towards the superior and inferior poles of the ONH (Figure 3.9) compared with Jansonius *et al.* [114]. This is another important line of evidence to support the novelty of our technique because Jansonius *et al.* [114] did not account for the retinal ganglion cell displacement in the macular region [121, 122]. A similar difference was recently reported by Turpin and McKendrick [118], whose model is also reported in Figure 3.9 as a comparison.

From the example in Figure 3.8B, the heatmap for the SLO predictions is mostly focused on the neuroretinal rim for the inferior hemifield; however, it appears more widespread for the location in the superior hemifield. Similarly, the heatmap produced for the Cp-OCT scans mainly focused on the RNFL (example in Figure 3.8C). This suggests that our technique learns the image features relevant for prediction without the need for segmentation, which is noteworthy. The disc heatmap commonly indicates some relevance of image features beyond the optic disc suggesting a role of the reflectivity patterns induced by the loss of RNFL [123]. In the heatmaps, the most frequent location of the blind spot (X = 15; Y = -3 degrees) was rendered inactive during the training, meaning that the network learned that the sensitivity at this point

could not be predicted by any information present in the image. This is inconsequential for our results (both blind spot locations were excluded from our analyses) but provided us with the chance of confirming that the models were able to learn meaningful spatial features from the data.

3.5.3 Comparison with Previous Work

Others have used DL methods to predict global VF metrics [124, 125] but few have attempted pointwise predictions of the VF. Park *et al.*[126] and Hashimoto *et al.* [127] used fully segmented OCT thickness maps. Mariottoni *et al.* also predicted pointwise VF sensitivities agnostically from peripapillary OCT RNFL thickness profiles [128]. However, such an approach not only requires an accurate segmentation to be available in the first place, but also disregards other information contained in the OCT image that might inform prediction, such as the reflectivity of the RNFL [123, 129]. Fewer attempts have instead been made to utilize optic disc images to predict VF data [130]. Interestingly, our data indicate the existence of additional information gained by combining the two imaging modalities. The error profile of our prediction stratified by sensitivity was similar to previous publications [128], in that the error greatly increased below 15 dB. This could be linked to the high level of VF noise at low sensitivities [131], which imposes a lower bound on the achievable predictive performance, and by the previously mentioned floor effect in the structural data [132]. Such a feature is clearly demonstrated by the positive bias at lower sensitivities (Figure 3.4). Notably, non-linear methods have proven more successful than linear models at overcoming this issue [133, 127, 128, 134], but the problem persists in all these attempts.

3.5.4 Practical Implications and Future Directions

VF estimation from imaging allows the conversion of the information contained in OCT/SLO structural data into a more clinically meaningful format. For example, it could allow for structural data to be seamlessly integrated into analyses of VF progression [135] or into the VF test itself [136, 137, 138]. This could expedite perimetry or reduce between-visit measurement variability when following a patient over time. The latter is appealing because this might improve the power to detect disease progression in a trial, and this is the subject of our future work.

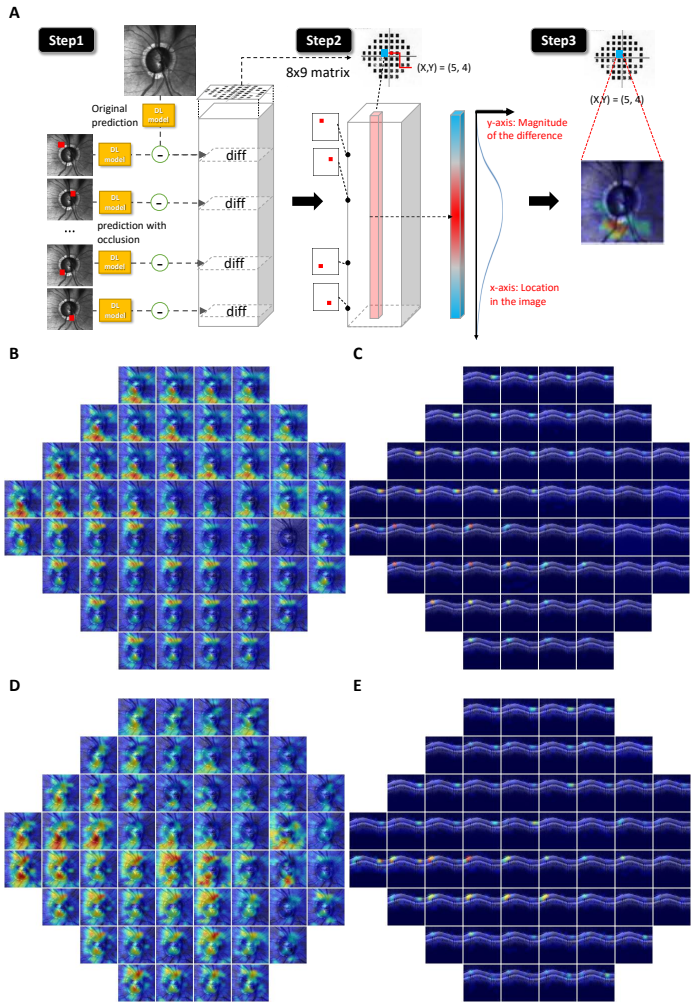

Figure 3.8: A, Process for generating heatmaps for the disc SLO and OCT models. According to the location of the perturbation in the predicted VF, the magnitude of changes reflects where the network focuses on during the predictions. B, Heatmap for one patient-generated example based on our disc SLO. C, Heatmap for one patient-generated example based on the OCT model. D, Heatmap for one patient-generated example from the policy network obtained by occluding the disc SLO. E, Heatmap for one patient-generated example from the policy network obtained by occluding the OCT image. OCT = optical coherence tomography; SLO = scanning laser ophthalmoscopy; VF = visual field.

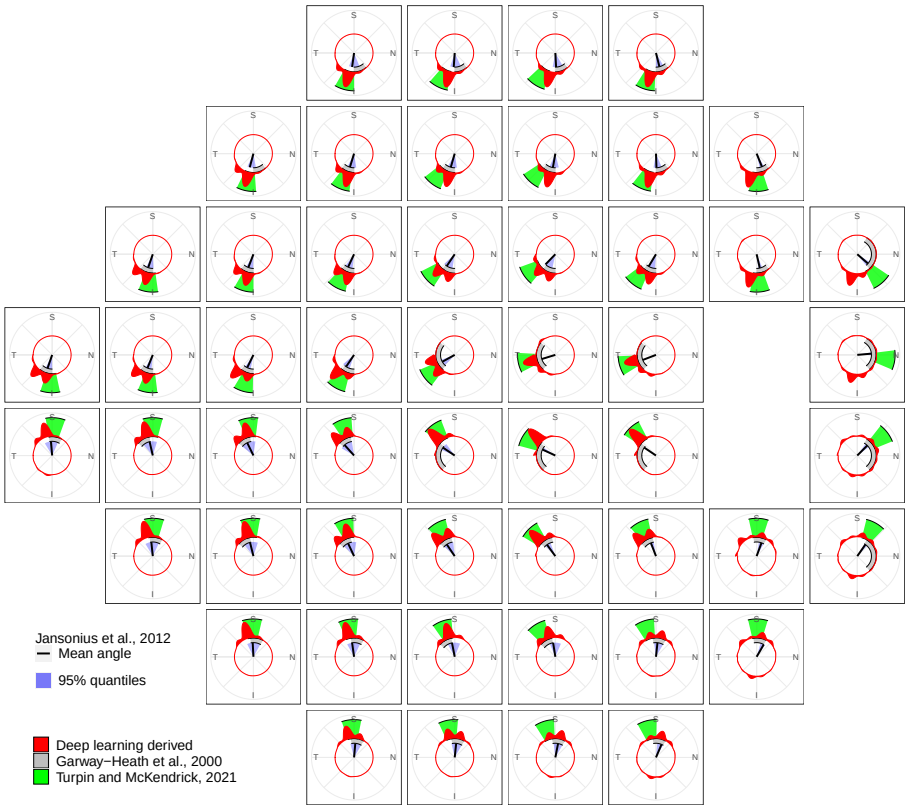

Figure 3.9: Structure-function mapping for the OCT to HVF model generated from the heatmaps for the test set for each testing point in the 24-2 VF pattern shown on polar coordinates. The map was built by plotting the distribution of the circumpapillary angles at which the highest peak of the heatmap was located. Other mapping schemes are reported for comparison. HVF = Humphrey visual field; I = inferior; N = nasal; OCT = optical coherence tomography; S = superior; T = temporal.

3.6 Summary

We demonstrate the ability of DL models to predict VF pointwise sensitivities agnostically from Cp-OCT scans and IR SLO images of the disc, further improving predictions with a policy-based fusion of the two results. The DL models generate structure-function maps compatible with established anatomical features and are able to capture the functional consequence of relevant structural changes in glaucoma. Ubiquitously obtained IR/OCT data may contain information beyond conventional segmentation that could be used to better determine visual function and detect progression in patients with glaucoma.

CHAPTER 4

RETINAL LAYER SEGMENTATION ON LARGE SCALE UNLABELED DATA WITH PSEUDO LABELING

4.1 Introduction

4.1.1 Foveal Curvature and Its Associations

The fovea is a highly specialized retinal region at the center of the macula responsible for driving high visual acuity and color vision [15]. Despite occupying $\sim 2.69\ mm^2$/1100 mm^2 of the retinal area, the fovea maps to half of the primary visual cortex [139, 140]. Absent or poorly formed foveal depressions with presence of inner retinal layers have been associated with poor vision in cases with well characterized diseases (i.e., retinopathy of prematurity, aniridia, ocular albinism, absent or poorly formed foveal depressions with presence of inner retinal layers optic nerve decussation defects and anterior segment dysgenesis syndrome, Stickler syndrome, Alport syndrome, familial exudative vitreoretinopathy, incontinentia pigmenti, nanophthalmos, posterior microphthalmos, and achromatopsia) [141, 142, 143, 144, 145, 146, 147]. Nevertheless, absent foveas or foveas with presence of inner retinal layers have also been described in healthy individuals with good vision [148]. In this context, the detailed noninvasive cross-sectional imaging of the retina with micrometer resolution obtained with optical coherence tomography (OCT) has significantly contributed to the detailed quantitative description of foveal morphology in healthy and diseased individuals and provided insight into postnatal retina development [149, 150, 151]. Our understanding of mechanisms and functional implications of cytoarchitectural and morphological foveal alterations are driven by studies in patients with absent or poorly formed foveal depressions with presence of inner retinal layers in selective settings [152]. What is less understood is the interindividual variation of foveal curvature (FC) in the general population, and what factors may be associated with these differences. Studies analyzing the OCT-derived foveal slope have been limited to using small to moderate sample sizes (typically with less than 400 subjects) [143, 153, 154, 155, 156]. With more than half a million recruited participants and with a subset of about 85,000 patients with enhanced ophthalmological examination, the UK Biobank is one of the world's largest single resources for comprehensive study of health and disease (UK Biobank Eye and Vision Consortium, available at: `https://www.ukbiobankeyeconsortium.org.uk/`). By using comprehensive structured population data and machine learning (ML), we aim to address a gap in our knowledge by exploring the associations of sociodemographic, ocular, and early life factors with OCT-derived FC of healthy individuals.

4.1.2 Macular Curvature and Dome-Shaped Configuration

It is commonly believed that the retina follows the roundness of the globe and shows a slight outward directed curve. Nevertheless, divergent macular configurations such as staphyloma or a dome-shaped macular configuration have been described [157, 158, 159, 160]. These divergent configurations are reported to be mainly associated with such features as thickened sclera and choroid in the context of or secondary to high myopia (apart from space-occupying processes such as choroidal tumors), 5 defined by refractive error

exceeding -6 diopters (D) or axial length of ≥ 26 mm [157, 158, 159, 160, 161]. In the cohort with high myopia, it was reported with a frequency of 4% to 15%. Given the increasing myopia prevalence [162], understanding macula malformation in the context of myopia is vital, as it is a strong driver of maculopathy development is the presence of a posterior staphyloma with anomalous macula curvature [162, 163, 164]. However, myopia research is ominated by exploring associations of refractive error or occasionally axial length, but not macula curve, due to difficulty measuring it. Therefore, the pathophysiology underlying macular malformations, especially dome-shaped maculopathy, is not completely understood nor has the prevalence of such malformations been systematically investigated. In the last decade, high-definition optical coherence tomography (OCT) has become widely available that allows for three-dimensional imaging of the macula [165]. In this context, objective quantification of the macular curvature has recently been introduced [166]. It might offer the potential for an effective comprehensive evaluation. Although macular curvature has been reported to be associated with choroidal thickness and axial length, as well as retinitis pigmentosa and associated genes [166, 167, 168, 169], large epidemiologic studies are still pending.

4.2 Curvature Analysis on UKBB

The dataset include eye data from more than 133,000 participants underwent an enhanced ophthalmic assessment. A subset of these (87,624 participants) had undilated macular spectral-domain OCT (SD-OCT, Topcon 3D OCT-1000; Topcon Optical Company, Tokyo, Japan) imaging. These OCT data were stored as FDS files for each eye, a proprietary image storage file format that contains the raw fundus image and OCT B-scan images, on the UKBB supercomputers in Oxford, UK, without prior analysis. Each eye data contains 128 B-scans with 512 A-scans. With the dataset, our aim was to extract the area between the inner limiting membrane and retinal pigment epithelium (RPE), and the area between the internal limiting membrane (ILM) and RPE from OCT B-scans, and thus detect the boundary of the ILM and RPE without any human annotations. The curve fitting was then executed by fitting a two-dimensional polynomial curve on the extracted ILM boundary for foveal, and RPE boundary for macular curvature analysis. The process overview for foveal curvature fitting is shown in Figure 4.1.

For this purpose, we took an iterative pseudo-labeling technique, a semi-supervised algorithm which efficiently performs multiple iterations of pseudo-labeling on unlabeled data as the model evolves. In particular, we extracted "hard samples" to promote generalization of the model rather than overfit to incorrect samples.

4.2.1 Retinal Layer Segmentation with Pseudo Labels

We developed a machine learning model to extract the area between the internal limiting membrane (ILM) and retinal pigment epithelium (RPE) from OCT B-scans and detect ILM and RPE boundaries

without need for human annotations. The process overview is shown in Figure 4.1. We first employed the $A^\star$ algorithm to obtain initial segmentation masks of the retinal layers [170], and we used them as pseudo labels, which are the initial training target for the machine learning model (Figure 4.1 Step1). The $A^\star$ algorithm is an extension of Edsger Dijkstra's algorithm, which is widely used in pathfinding and graph traversal [170], and thus can be adopted to distinguish different parts of boundaries in graph-based image segmentation problems. However, this traditional image segmentation algorithm is not robust enough to handle images with poor contrast or abnormal brightness filled with random noise; in fact, the output segmentation masks included many errors. In order to filter out these errors, we calculated the minimum retinal thickness and standard deviation (SD) for each B-scan from predicted segmentation masks. In this study, only the samples with a minimal retinal thickness between 30 and 80 pixels (which converts to 180 to 480 μ m at an axial resolution of 6 μ m/pixel) and $SD \leq 10$ pixels (60 μ m) were accepted. Through the process described above, we collected a total of 6,409 input–output pairs to train a segmentation network in Figure 4.1 Step2. We used the Pyramid Parsing Network (PSPNet) with the ResNet18 backbone as our segmentation architecture [171, 172]. The model was trained with a binary cross-entropy loss function using the Adam optimizer. We chose a batch size of 8. The learning rate was initially set to 1×10^{-3}, and decay over each update was set to initial learning ratio divided by epochs. All inputs and outputs were cropped to 512×512 by setting y-coordinates whose sum of pixel values had a maximum intensity to the center and then resized to 256×256. All images were normalized to a range between 0 and 1. We assessed the performance of the neural network using cross-validation with the validation set and evaluated the generalizability with an independent dataset. Samples were divided into training sets at 80%, validation sets at 20%. The training and validation sets contained images from mutually exclusive groups of subjects. During the training, shadows and Gaussian/speckle noises were randomly applied to the training data, as well as basic data augmentations such as shift, flip, and rotation. When we filtered out segmentation errors from the $A^\star$ algorithm, many hard examples were also excluded from the dataset; therefore, the trained model was not robust enough at this point. In order to bolster this weakness, we retrained our model with additional hard examples. For this, we first processed 1,250 FDS data and excluded segmentation error. Next, we computed a sum of entropy for each obtained segmentation mask where each pixel had a probability value of being foreground. We then regarded examples with high entropy value as "hard examples", which were sorted in descending order, and the top 200 images were added to the original dataset for the training. The trained model achieved a mean intersection over union value of 0.97 on the validation set. In the end, a total of 170,079 eye images were processed and $128 \times 170,079$ segmentation masks were generated for each B-scan by the segmentation network. Because the B-scans often include several vertically flipped images, we built a binary classification network with the LeNet architecture that can identify flipped B-scans and unflip them automatically. RPE/C boundaries were then extracted simply by tracking the bottom boundary for each segmentation mask.

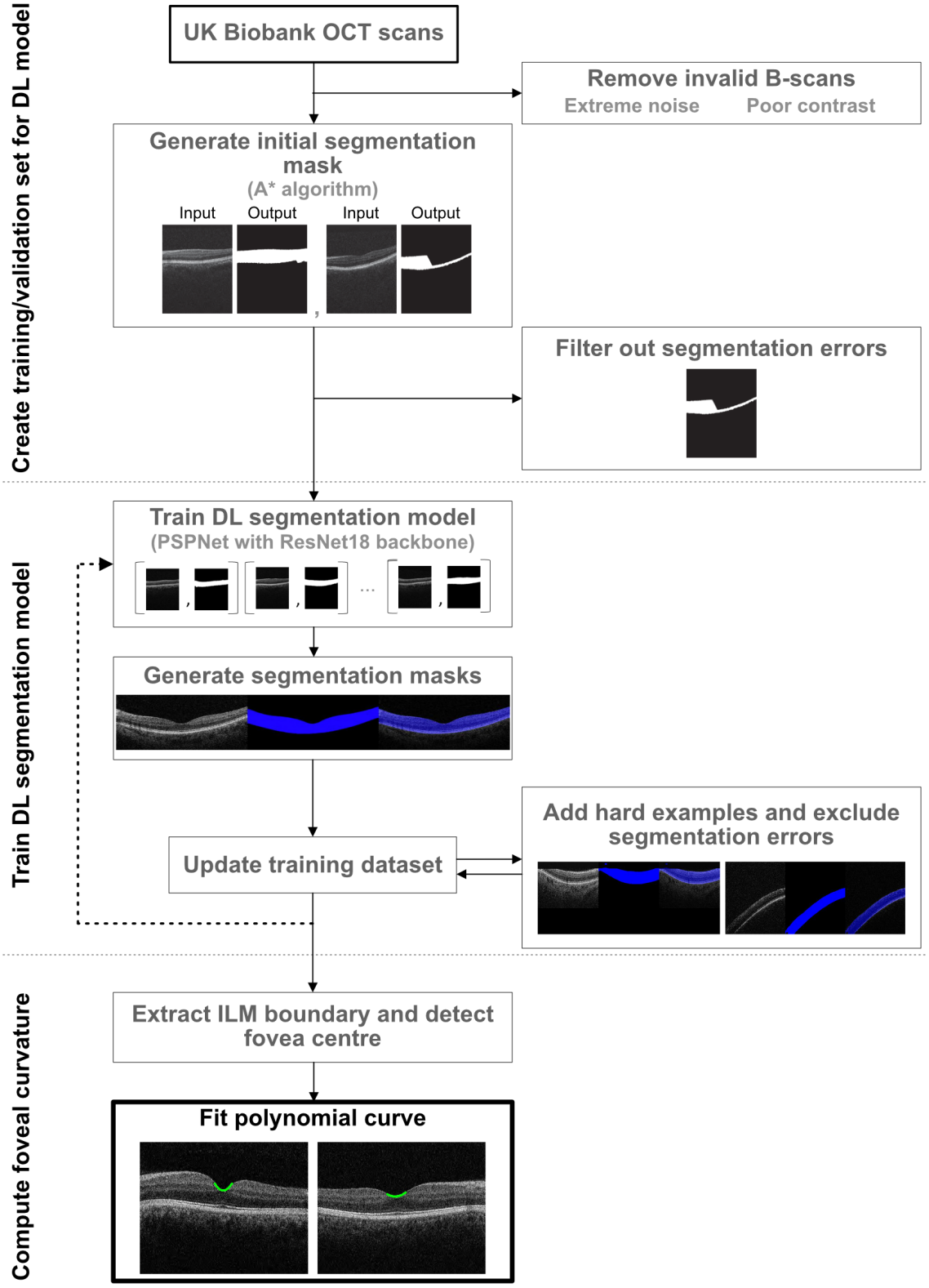

Figure 4.1: Diagram of machine learning metodology implemented to generate our automated OCT-derived FC quantification.

4.2.2 Center Point Retinal Thickness Analysis.

The retinal thickness was computed as the distance between the ILM and RPE boundary, and center point retinal thickness (CPRT) was determined to be the center of the area with the thinnest retina for each OCT volume scan. By mapping the thickness of ILM and RPE boundaries for each location on the 128 B-scans, we obtained a 128×256 height map where each pixel represents the height of ILM-RPE. Next, we applied gaussian blur to the thickness map (Figure4.2 a). We then binarized the filtered height map using Otsu's algorithm [173] (Figure4.2 b). The resulting donut-shaped blob was then detected by a minimum circularity threshold (Figure4.2 c). The location of the center point was then given as a tuple of slice number and B-scan x-coordinate.



Figure 4.2: Process overview of Center Point Retinal Thickness Analysis.

4.2.3 Foveal Curvature Analysis

The curve fitting was executed on the pixel coordinates. Given the center point for each OCT volume scan, a two-dimensional polynomial curve was fitted on the extracted ILM boundary with a range of pixels left and right from the center point. The degree of the polynomial used was two dimensional, and the coefficient of highest degree (leading term) was used to describe the curvature (Figure 4.3).

■ Validation of Curvature Analysis

In order to validate the automated foveal curvature analysis, two retina specialists with wide experience in OCT grading (A.T., A.O-B.) were asked to classify 10 different image sets, composed of 3 B-scans from each FC tertile (see Figure 4.4), from flattest to steepest scan in each set. The reference standard was a tertile classification based on ML-derived FC quantification. Human graders correctly classified each FC tertile in all 10 image sets.

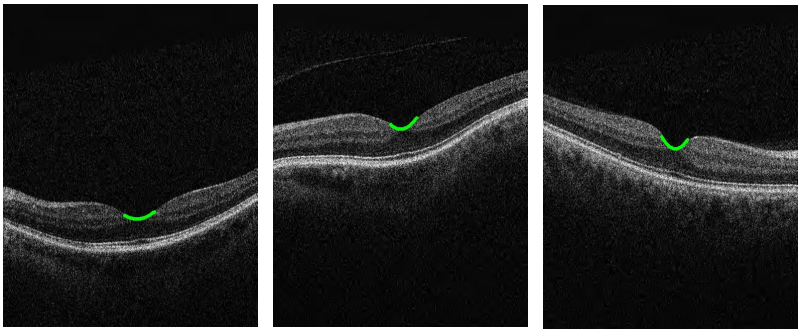

Figure 4.3: Examples of detected curve of fovea.

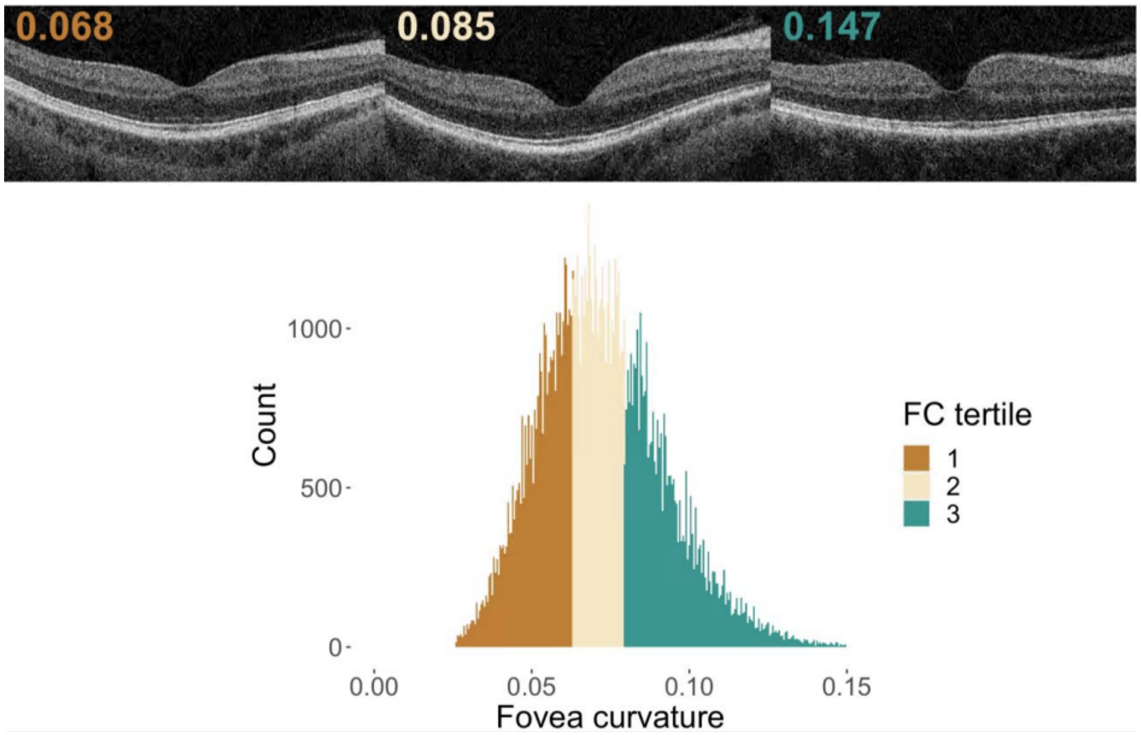

Figure 4.4: Top: foveal curvature quantification from central b-scans for each curvature tertile (from right to left, flattest to steepest quantified curvature measurements). Bottom: histogram of foveal curvature distribution by FC tertile.

■ Statistical analysis

R version 4.0.2 was used to analyze the data [174]. The "lme4" (version 1.1-28) package was used for linear multilevel regression models fitted by restricted maximum likelihood. P values were calculated

via Satterthwaite's degrees of freedom method with the "lmerTest" (version 3.1-3) package. In view of systematic differences in FC between men and women, and associated covariates related to growth, all analyses were stratified by sex. Multilevel linear regression models adjusting for age, ethnicity, height, and UK Biobank assessment center as fixed effects, with a random effect for person to allow for the right- and left-eye data from the same participant to contribute to the analysis (Model 1), were used to examine associations with FC. Model 2 extended model1 with further adjustment for VA, SE, corneal astigmatism, IOP, MC, and CPRT. Model 3 extended model 2 allowing for deprivation, higher education, fluid intelligence score, annual income, and birth order. Fovea curvature measures were modeled as z scores. Coefficients represent the standard deviation (SD) change in FC per specified increase in covariates or the standardized difference between groups (Tables 4.1 and 4.2, show estimates for raw $FC \times 100$ as dependent variable). Data missing on categorical variables were included as an additional category for each variable to minimize data loss. In sensitivity analyses model 3 was extended by allowing for birth weight, maternal age at birth, maternal smoking around birth, and breastfeeding status as a baby to examine FC associations with early life factors. Additionally, multilevel models were fitted again after exclusion of individuals with SE < -6 D and > 6 D and vision < 80 Early Treatment Diabetic Retinopathy Study letters (worse than 6/7.5 Snellen, or worse than 0.1 logMAR equivalent).

Table 4.1: Regression table for females with fovea curvature$\times 100$ as dependent variable.

	Females model 1*	Females model 2 †	Females model 3 ‡
Characteristic			
Age (per decade)	0.13 (0.11, 0.16); 6.8e-28	0.16 (0.14, 0.19); 1.8e-34	0.16 (0.14, 0.19); 9.4e-30
Ethnicity			
White	1	1	1
Black	-1.1 (-1.2, -1.0); 5.1e-75	-1.5 (-1.6, -1.3); 1.8e-120	-1.4 (-1.5, -1.3); 2.3e-93
Asian	-0.93 (-1.1, -0.80); 2.2e-44	-1.2 (-1.3, -1.0); 1.4e-63	-1.1 (-1.3, -1.0); 9.9e-48
Other	-0.71 (-0.87, -0.55); 3.8e-18	-0.90 (-1.1, -0.73); 4.4e-25	-0.80 (-1.0, -0.62); 8.2e-18
Mixed	-0.30 (-0.49, -0.11); 0.002	-0.51 (-0.71, -0.31); 3.7e-07	-0.47 (-0.68, -0.26); 7.5e-06
Chinese	-0.31 (-0.59, -0.03); 0.032	-0.63 (-0.92, -0.34); 2.5e-05	-0.59 (-0.90, -0.28); 1.7e-04
Prefer not to say	-0.84 (-1.2, -0.51); 6.9e-07	-1.0 (-1.4, -0.67); 1.7e-08	-1.0 (-1.4, -0.66); 9.5e-08
Missing	-0.81 (-1.3, -0.37); 3.5e-04	-0.88 (-1.4, -0.40); 3.1e-04	§
Height (per 5cm)	0.08 (0.06, 0.09); 5.5e-24	0.07 (0.06, 0.09); 2.4e-19	0.07 (0.05, 0.08); 6.9e-16
Visual acuity (per 5 letters)		0.03 (0.02, 0.03); 1.6e-16	0.03 (0.02, 0.03); 1.4e-15
Spherical equivalent (per diopter)		-0.08 (-0.09, -0.08); 6.4e-105	-0.08 (-0.09, -0.07); 1.3e-96
Corneal astigmatism (per diopter)		-0.14 (-0.16, -0.12); 3.2e-33	-0.14 (-0.16, -0.12); 4.0e-32
Macula curvature (per 0.01)		0.58 (0.46, 0.69); 1.4e-21	0.55 (0.43, 0.67); 1.7e-19
Center point retinal thickness (per 10 μ m)		-0.26 (-0.27, -0.25); 0.0e+00	-0.26 (-0.27, -0.25); 0.0e+00
Fluid intelligence			0.02 (0.01, 0.03); 9.8e-04
Annual income (Great British Pound)			
Less than 18,000			1
18,000 to 30,999			0.01 (-0.05, 0.08); 0.701
31,000 to 51,999			0.02 (-0.05, 0.09); 0.589
52,000 to 100,000			0.02 (-0.05, 0.09); 0.605
Greater than 100,000			0.07 (-0.03, 0.17); 0.181
Prefer not to say			0.07 (0.00, 0.15); 0.055
Missing			-0.04 (-0.53, 0.45); 0.872
Birth order			
1			1
2			0.02 (-0.02, 0.07); 0.311
3			-0.05 (-0.13, 0.03); 0.259
4			-0.03 (-0.11, 0.06); 0.569
Missing			0.38 (-0.45, 1.2); 0.371

Bold p-values represent statistically significant results.

* Model 1: multilevel model adjusts for age, ethnicity, and height as fixed effects, and a random effect for person to allow for within person eye measurements (59,642 eyes of 35,097 patients).

† Model 2 adjusts as model 1 plus visual acuity, spherical equivalent, corneal astigmatism, macular curvature, and center point foveal thickness as fixed effects (54,489 eyes of 32,564 patients).

‡ Model 3 adjusts as model 2 plus deprivation, higher education, fluid intelligence score, annual income, and birth order as fixed effects (53,135 eyes of 31,727 patients).

§ No missing data on ethnicity on this model.

Table 4.2: Regression table for males with fovea curvature x 100 as dependent variable.

	Males model 1*	Males model 2 †	Males model 3 ‡
Characteristic			
Age (per decade)	-0.06 (-0.08, -0.03); 8.4e-05	-0.02 (-0.05, 0.01); 0.168	-0.01 (-0.04, 0.02); 0.466
Ethnicity			
White	1	1	1
Black	-1.2 (-1.3, -1.0); 6.0e-49	-1.7 (-1.8, -1.5); 2.1e-88	-1.6 (-1.8, -1.4); 5.2e-68
Asian	-0.71 (-0.86, -0.56); 7.4e-21	-1.1 (-1.2, -0.92); 1.7e-41	-1.1 (-1.2, -0.90); 2.3e-34
Other	-0.57 (-0.78, -0.36); 1.5e-07	-1.0 (-1.2, -0.74); 1.5e-17	-0.94 (-1.2, -0.71); 7.2e-15
Mixed	-0.41 (-0.69, -0.12); 0.005	-0.77 (-1.1, -0.47); 3.8e-07	-0.73 (-1.0, -0.43); 3.0e-06
Chinese	-0.68 (-1.1, -0.29); 6.5e-04	-1.1 (-1.5, -0.65); 4.5e-07	-1.1 (-1.5, -0.65); 8.9e-07
Prefer not to say	-0.26 (-0.61, 0.09); 0.150	-0.25 (-0.62, 0.11); 0.168	-0.18 (-0.57, 0.21); 0.372
Missing	0.00 (-0.51, 0.52); 0.986	-0.09 (-0.61, 0.44); 0.749	-1.1 (-3.8, 1.6); 0.437
Height (per 5cm)	0.08 (0.06, 0.10); 4.3e-19	0.07 (0.05, 0.09); 3.4e-15	0.06 (0.05, 0.08); 4.2e-12
Visual acuity (per 5 letters)		0.02 (0.01, 0.03); 3.5e-07	0.02 (0.01, 0.03); 8.4e-07
Spherical equivalent (per diopter)		-0.07 (-0.08, -0.07); 5.3e-53	-0.07 (-0.08, -0.06); 5.6e-50
Corneal astigmatism (per diopter)		-0.12 (-0.15, -0.10); 3.6e-20	-0.13 (-0.15, -0.10); 5.3e-20
Macula curvature (per 0.01)		0.27 (0.14, 0.41); 1.0e-04	0.28 (0.14, 0.42); 1.2e-04
Center point retinal thickness (per 10 μ m)		-0.29 (-0.30, -0.28); 0.0e+00	-0.29 (-0.30, -0.28); 0.0e+00
Fluid intelligence			0.00 (-0.02, 0.01); 0.559
Annual income (Great British Pound)			
Less than 18,000			1
18,000 to 30,999			0.11 (0.03, 0.19); 0.010
31,000 to 51,999			0.10 (0.02, 0.18); 0.021
52,000 to 100,000			0.13 (0.04, 0.22); 0.004
Greater than 100,000			0.16 (0.05, 0.28); 0.005
Prefer not to say			0.01 (-0.10, 0.12); 0.857
Missing			0.22 (-0.37, 0.81); 0.459
Birth order			
1			1
2			0.01 (-0.04, 0.06); 0.728
3			0.06 (-0.04, 0.16); 0.216
4			-0.05 (-0.15, 0.06); 0.375
Missing			-0.39 (-1.3, 0.50); 0.389

Bold p-values represent statistically significant results.

* Model 1: multilevel model adjusts for age, ethnicity, and height as fixed effects, and a random effect for person to allow for within person eye measurements (59,642 eyes of 35,097 patients).

† Model 2 adjusts as model 1 plus visual acuity, spherical equivalent, corneal astigmatism, macular curvature, and center point foveal thickness as fixed effects (54,489 eyes of 32,564 patients).

‡ Model 3 adjusts as model 2 plus deprivation, higher education, fluid intelligence score, annual income, and birth order as fixed effects (53,135 eyes of 31,727 patients).

4.2.4 Macular Curvature Analysis

We used 32 central Bscans for macular curvature fitting. For each B-scan, we fitted a quadratic function to the extracted RPE boundary, then took the coefficient of the leading term as curvature value. After collecting the 32 curvature values, we took median as the final macular curvature of the volume. Again, the curve fitting was executed on the pixel coordinates.

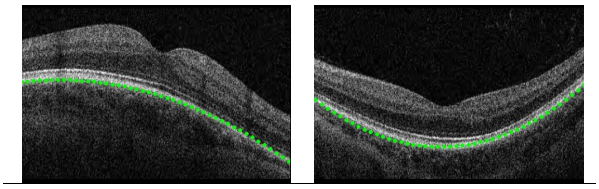

Figure 4.5: Examples of detected curve of macular.

■ Validation of Curvature Analysis

In order to validate the macular curvature analysis, the eyes were divided into quartiles based on macular curvature values. Randomly (using the True Random Number Generator, `https://www.random.org/`), one eye of each quartile was chosen and the central OCT scan was included in an Adobe PhotoShop PSD file (Figure 4.6). Ten PSD files consisting of one OCT scan per quartile were generated. The files were saved twice: once with layers in random order for validation and once with the correct order of layers as control. Two independent experienced retinal specialists (P.L.M. and A.T.), masked to the results of the other, were then asked to order the four OCT scans in each PSD file (random order) according to the appearance of macular curvature. Finally, the results were compared to the control files, which revealed perfect agreement. Furthermore, the 100 cases with the most extensive curvature values were manually checked.

■ Statistical analysis

Statistical analysis was performed using R 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria) and Python 3.7 (Python Software Foundation, Wilmington, DE, USA). The distribution of the macular curvature and its correlation with ethnicity (white, comprised of English/Irish or other white background; Asian, comprised of British Asian, Indian, Pakistani, Bangladeshi, or other Asian background; black, comprised of black British, Caribbean, African, or other black background; Chinese; mixed, comprised of white and black Caribbean or African, white and Asian, or other mixed background; and other, comprised of undefined ethnicity), demographic, ocular, and functional parameters were investigated. These parameters included refractive error, as spherical equivalent (D), measured by autorefractor and calculated as sphere + [cylinder/2]); VA (in logarithm of the minimum angle of resolution [logMAR]); corneal-corrected IOP (mmHg),measured using the Ocular Response Analyzer (Reichert Corp., Buffalo, USA); age; sex; corneal

curvature, measured as maximal curvature (KMAX); childhood environment (birth weight, maternal smoking); and fluid intelligence (FI) score based on a baseline touchscreen questionnaire with 13 questions. Factors were chosen according to literature evaluations, as associations with eye development or effects on the pathogenesis of dome-shaped macular curvature have been hypothesized [157, 160, 175, 176, 177, 178, 179, 180]. Multilevel linear regression models adjusting for age, gender, and ethnicity (demographic parameters) as fixed effects, with a random effect for person to allow for the right- and left-eye data from the same person to contribute to the analysis (model 1), were used to examine associations with macular curvature. Model 2 extended model 1 with further adjustment for ocular measures (refractive error, IOP, corneal curvature) and functional data (VA, FI). Model 3 extended model 2 with further adjustment for childhood environment (birth weight, maternal smoking). Macular curvature measures were modeled as z-scores in the regression models, and coefficients represent the fraction of SD change per unit increase or per group change (for categorical values) in covariates. Age represents the changes per decade increase, VA per 0.1 logMAR change, and IOP per 5-mmHg rise; other continuous variables are expressed in their units. References for categorical variables are female versus male, white versus all other ethnicities, and maternal smoking versus no maternal smoking. For inter-eye comparisons, paired t-tests were used. P values with $\alpha < 0.05$ were considered statistically significant.

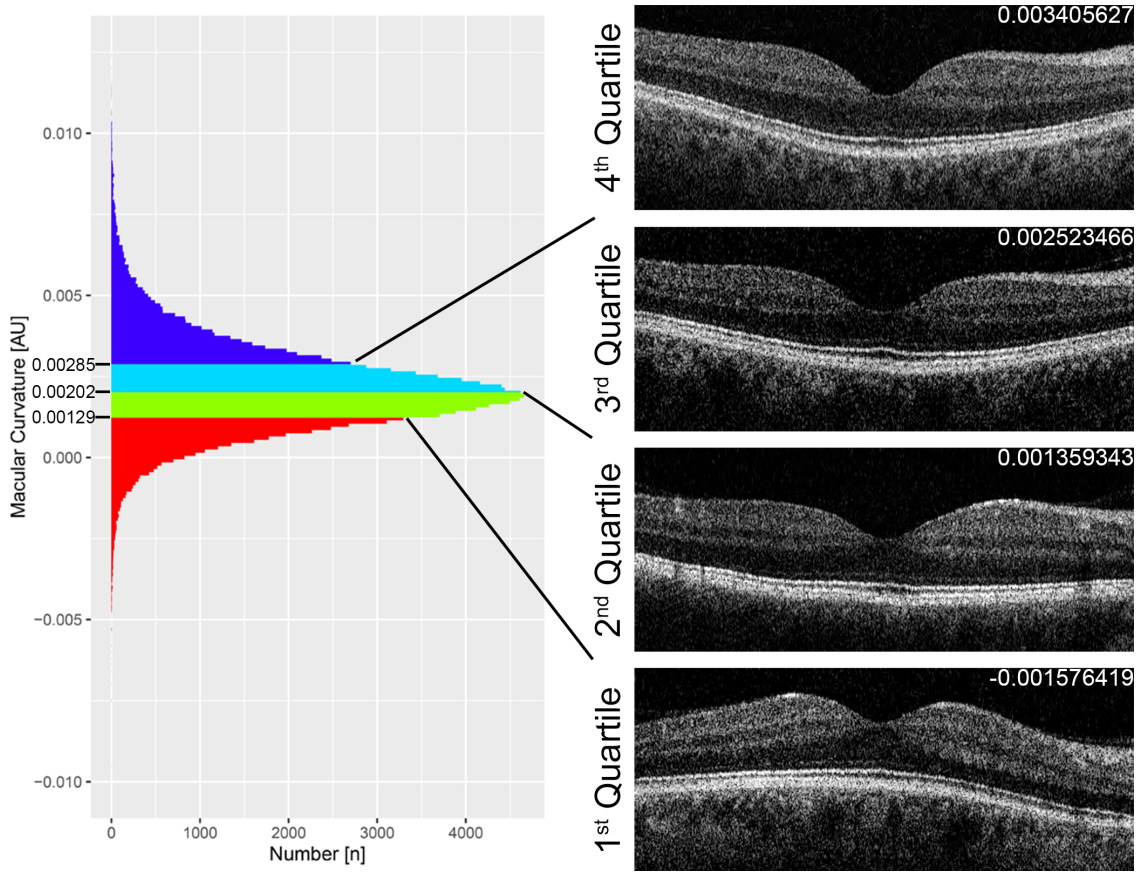

Figure 4.6: The spectrum of macular curvature. The histogram (left) shows the distribution of macular curvature values; the colors highlight the four different quartiles. The optical coherence tomography B-scans (right) represent exemplary images for all four quartiles as used for validation. For the demonstrated exemplary set, the individual curvature value of the respective eyes is displayed. The bottom image shows an inward-directed macular curve associated with a negative curvature value, indicating a dome-shaped configuration.

4.3 Results

4.3.1 Results from Foveal Curvature Analysis

A total of 109,160 eyes (54,055 right eyes, and 55,105 left eyes) of 63,939 participants (45.1% male) were included in the analysis. Mean age (SD) was 56 years (±8.0), and 92% of the participants were White. The FC followed a normal distribution (Figure 4.4) and had a mean of 0.072 (±0.02). Table 4.3 shows the overall patient characteristics of our study cohort. A summary of eye-level characteristics is found in Table 4.4. Figure 4.7 shows the association of FC with each covariate (deciles of continuous variables), adjusted for age, height, and UK Biobank assessment center. Foveal curvature showed an inverse association with

each decile increase in SE, CPRT, VA, and corneal astigmatism. A positive linear association of FC was found for each decile increase in MC. Associations in different directions between males and females were observed with age. Associations were less clear for height, IOP, higher education, fluid intelligence, annual income, and deprivation.

■ Sex Differences

Males had on average steeper FC and greater CPRT. Crude difference in FC between men and women was 0.47 SDs, and after adjustment for age, ethnicity, height, and UK biobank assessment center (i.e., as in model 1), this difference was 0.36 SDs (95% CI 0.34–0.38; p 4.8×10^{-247}). After additional adjustment for covariates included in model 3 (SE, VA, IOP, MC, CPRT, corneal astigmatism, higher education, annual income, fluid intelligence, deprivation, and birth order) the sex-difference was 0.46SD (95% CI 0.44–0.48; p 4.3×10^{-360}). Formal tests for interaction with sex (females vs. males from multilevel models adjusting for age, sex, height, and UK biobank center as fixed effects, and a random effect per person) were significant for age, MC and CPRT only (all instances $P < 0.0001$). As a result of the observed systematic sex differences, associations are presented for females (Table 4.5) and males (Table 4.6) separately.

■ Age, Ethnicity, and Height (Models 1 to 3)

Differences in FC by ethnicity showed the greatest effect overall, and these held after adjustment, and after exclusion of extreme refractive status and low VA. Black, Asian, mixed, Chinese, and other ethnic groups showed flatter FC when compared with whites. Black participants showed the biggest difference (In models 3, males 0.80SD decrease [p 2.1×10^{-88}]; females 0.70SD decrease when compared to whites [p 1.8×10^{-120}]). In females, every decade increase in age was associated with a 0.07 SD rise in FC (Model 1; p 6.8×10^{-28}), and additional adjustments, or exclusion of those on basis of high SE, did not materially alter the strength of this association (Table 4.1). In males, a 0.03 SD decline in FC per decade rise in age was observed in Model, 1 but this was attenuated to the null with further adjustment (Models 2 and 3, Table 4.6). Every 5 cm increase in height was associated with steeper foveas in males and females, with equal effect sizes across all models in both sexes. Formal test for interaction between ethnicity with age, height, and SE showed that patterns were consistent across ethnicity (data available on request) except for the age in females. Analyses showed that for female non-white ethnicities there was no association with age but the rise in FC with age was present in white females only (0.10 SD increase in FC per decade rise in age; 95% CI 0.09–0.11).

■ Functional and Ocular Factors (Models 2 and 3)

Every five Early Treatment Diabetic Retinopathy Study letter increase in VA (better vision) was associated with a 0.01 SD rise in FC in both males and females in all models (p in all instances $\leq 3.5 \times 10^{-7}$). Every D increase in corneal astigmatism and in SE was associated with a flatter FC in both males and females. An inverse association was observed with foveal thickness in both sexes, per 10 μ m increase in foveal thickness FC decreased by approximately 0.1 SD. Steeper MC were found to be associated with steeper FC measurements and the effect sizes were double in females when compared with males in all models. After exclusion of extreme refractive status and VA worse than 6/7.5, the association with MC was attenuated to the null in males (Supplementary Table 4.2) but did not materially change in females. IOP did not show an association with FC.

■ Additional Sociodemographic Factors (Model 3)

Fovea curvature showed an increasing trend in curvature steepness with increasing annual income in males (p for linear trend 0.005). When compared to annual income $<$ 18,000 GBP, earning $> 100,000$ GBP per year was associated with a 0.08 SD rise in FC in males (Model 3, p for linear trend 0.005). The FC associations with income were not observed in females. Fluid intelligence showed a significant 0.01 SD rise in FC per score unit increase (95% CI 0.002–0.099; p 9.8×10^{-4}) in females (model 3). Fluid intelligence did not show associations with FC in males. Townsend deprivation indices, level of education and birth order did not show associations with FC. In sensitivity analysis (exclusion of cases with high refractive errors and poor VA), the coefficients presented in Tables 4.5 and 4.6 remained remarkably stable. An additional model did not show clear associations with other early life factors (including self-recalled birth weight, maternal age at birth, maternal smoking around birth, and breastfeeding status as a baby), and significant coefficients shown in the results section remained remarkably stable (data not shown, available on request).

Table 4.3: Patient Level Characteristics Stratified by Sex

Characteristic	Overall, N = 63,939*	Female, N = 35,097*	Male, N = 28,842*
Age	56 (8.0)	56 (7.9)	56 (8.2)
Ethnicity			
White	58,915 (92.1%)	32,204 (91.8%)	26,711 (92.6%)
Black	1,614 (2.5%)	979 (2.8%)	635 (2.2%)
Asian	1,477 (2.3%)	767 (2.2%)	710 (2.5%)
Other	822 (1.3%)	486 (1.4%)	336 (1.2%)
Mixed	521 (0.8%)	339 (1.0%)	182 (0.6%)
Chinese	248 (0.4%)	153 (0.4%)	95 (0.3%)
Prefer not to say	226 (0.4%)	109 (0.3%)	117 (0.4%)
Missing	116 (0.2%)	60 (0.2%)	56 (0.2%)
Height (in cm)	169 (9.2)	163 (6.3)	176 (6.7)
Townsend deprivation quintiles			
1	20,388 (31.9%)	10,995 (31.3%)	9,393 (32.6%)
2	13,277 (20.8%)	7,266 (20.7%)	6,011 (20.8%)
3	11,828 (18.5%)	6,618 (18.9%)	5,210 (18.1%)
4	11,170 (17.5%)	6,304 (18.0%)	4,866 (16.9%)
5	7,204 (11.3%)	3,877 (11.0%)	3,327 (11.5%)
Missing	72 (0.1%)	37 (0.1%)	35 (0.1%)
Income (in GBP)			
Less than 18,000	9,802 (15.3%)	5,696 (16.2%)	4,106 (14.2%)
18,000 to 30,999	13,569 (21.2%)	7,584 (21.6%)	5,985 (20.8%)
31,000 to 51,999	15,261 (23.9%)	8,061 (23.0%)	7,200 (25.0%)
52,000 to 100,000	13,530 (21.2%)	6,745 (19.2%)	6,785 (23.5%)
Greater than 100,000	4,431 (6.9%)	2,138 (6.1%)	2,293 (8.0%)
Prefer not to say	7,104 (11.1%)	4,741 (13.5%)	2,363 (8.2%)
Missing	242 (0.4%)	132 (0.4%)	110 (0.4%)
Education			
Degree	23,875 (37.3%)	12,729 (36.3%)	11,146 (38.6%)
O levels, CSEs, or equivalent	15,840 (24.8%)	9,227 (26.3%)	6,613 (22.9%)
A levels, professional, or equivalent	13,730 (21.5%)	7,344 (20.9%)	6,386 (22.1%)
None	8,256 (12.9%)	4,350 (12.4%)	3,906 (13.5%)
Missing	2,238 (3.5%)	1,447 (4.1%)	791 (2.7%)
Fluid intelligence†	6 (2.1)	6 (2.1)	6 (2.2)
Birth order			
1	35,661 (55.8%)	19,803 (56.4%)	15,858 (55.0%)
2	20,216 (31.6%)	10,855 (30.9%)	9,361 (32.5%)
3	4,034 (6.3%)	2,237 (6.4%)	1,797 (6.2%)
4	3,989 (6.2%)	2,182 (6.2%)	1,807 (6.3%)
Missing	39 (0.1%)	20 (0.1%)	19 (0.1%)

*Mean (SD) for continuous variables; n (%) for categorical variables; †Continuous variable with missing data (2.7%).

GBP; pound sterling, CSE; certificate of secondary education.

Table 4.4: Eye Level Characteristics Stratified by Sex

	Overall*	Sex*	
Characteristic	**N = 108,871†**	**Female, N = 59,642†**	**Male, N = 49,229†**
Fovea curvature x 100	7.25 (7.24, 7.27)	6.83 (6.82, 6.85)	7.77 (7.75, 7.78)
Visual acuity (in ETDRS letters)	85.09 (85.03, 85.14)	84.78 (84.70, 84.85)	85.46 (85.37, 85.55)
Spherical equivalent (Diopter)	-0.69 (-0.70, -0.67)	-0.71 (-0.73, -0.69)	-0.66 (-0.69, -0.64)
Corneal astigmatism (Diopter)	0.85 (0.85, 0.85)	0.88 (0.88, 0.89)	0.81 (0.80, 0.81)
Macula curvature x 100	0.21 (0.21, 0.22)	0.21 (0.21, 0.22)	0.22 (0.21, 0.22)
Center point retinal thickness (in µm)	226.18 (226.06, 226.30)	223.11 (222.96, 223.26)	229.90 (229.72, 230.08)

*N eyes of 63,939 participants; †Mean (95% CI).

ETDRS; Early Treatment Diabetic Retinopathy Study.

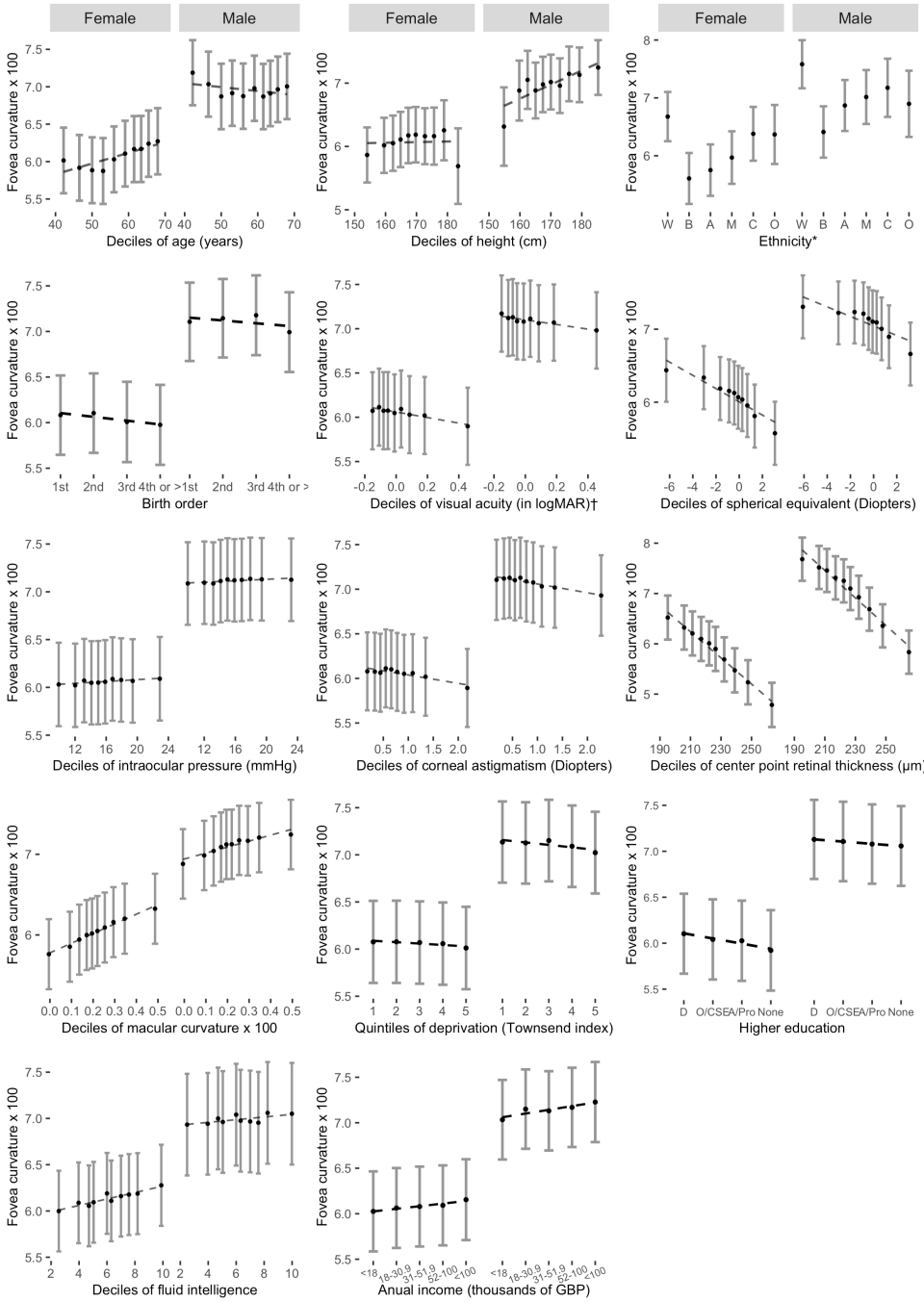

Figure 4.7: Adjusted mean foveal curvature by deciles of covariates stratified by sex (annual income and Townsend index of deprivation shown in quintiles). Adjusted means (solid black dots), 95% confidence intervals (vertical solid lines), and regression line (dotted line) are from a multilevel model allowing for age, height, ethnicity, and UK Biobank center as fixed effects, and repeated foveal curvatur measurement for each person. *Ethnicity codes: W, white; B, black; A, Asian; M, mixed; C, Chinese; O, other. † Visual acuity shown in logMAR for visualization purposes. GBP, pound sterling; O/CSE, O levels, certificate of secondary education or equivalent.

Table 4.5: Standardized difference in fovea curvature per specified differences in covariates ([95% CI]; p-value) for females.

Characteristic			
Age (per decade)	0.07 (0.06, 0.08); **6.8e-28**	0.08 (0.07, 0.09); **1.8e-34**	0.08 (0.07, 0.10); **9.4e-30**
Ethnicity			
White	1.00	1.00	1.00
Black	-0.54 (-0.59, -0.48); **5.1e-75**	-0.73 (-0.79, -0.67); **1.8e-120**	-0.70 (-0.77, -0.63); **2.3e-93**
Asian	-0.46 (-0.53, -0.40); **2.2e-44**	-0.59 (-0.66, -0.52); **1.4e-63**	-0.56 (-0.63, -0.48); **9.9e-48**
Other	-0.35 (-0.43, -0.27); **3.8e-18**	-0.45 (-0.53, -0.36); **4.4e-25**	-0.40 (-0.49, -0.31); **8.2e-18**
Mixed	-0.15 (-0.24, -0.05); **0.002**	-0.26 (-0.35, -0.16); **3.7e-07**	-0.24 (-0.34, -0.13); **7.5e-06**
Chinese	-0.15 (-0.30, -0.01); **0.032**	-0.32 (-0.46, -0.17); **2.5e-05**	-0.30 (-0.45, -0.14); **1.7e-04**
Prefer not to say	-0.42 (-0.59, -0.26); **6.9e-07**	-0.51 (-0.69, -0.33); **1.7e-08**	-0.52 (-0.72, -0.33); **9.5e-08**
Missing	-0.41 (-0.63, -0.18); **3.5e-04**	-0.44 (-0.68, -0.20); **3.1e-04**	§
Height (per 5cm)	0.04 (0.03, 0.05); **5.5e-24**	0.04 (0.03, 0.04); **2.4e-19**	0.03 (0.03, 0.04); **6.9e-16**
Visual acuity (per 5 letters)		0.01 (0.01, 0.02); **1.6e-16**	0.01 (0.01, 0.02); **1.4e-15**
Spherical equivalent (per diopter)		-0.04 (-0.05, -0.04); **6.4e-105**	-0.04 (-0.04, -0.04); **1.3e-96**
Corneal astigmatism (per diopter)		-0.07 (-0.08, -0.06); **3.2e-33**	-0.07 (-0.08, -0.06); **4.0e-32**
Macula curvature (per 0.01)		0.29 (0.23, 0.35); **1.4e-21**	0.28 (0.22, 0.34); **1.7e-19**
Center point retinal thickness (per 10µm)		-0.13 (-0.13, -0.13); **0.0e+00**	-0.13 (-0.14, -0.13); **0.0e+00**
Fluid intelligence			0.01 (0.00, 0.01); **9.8e-04**
Annual income (Great British Pound)			
Less than 18,000			1.00
18,000 to 30,999			0.01 (-0.03, 0.04); 0.701
31,000 to 51,999			0.01 (-0.02, 0.04); 0.589
52,000 to 100,000			0.01 (-0.03, 0.05); 0.605
Greater than 100,000			0.03 (-0.02, 0.08); 0.181
Prefer not to say			0.04 (0.00, 0.07); 0.055
Missing			-0.02 (-0.27, 0.23); 0.872
Per increase in income category			0.02 (-0.01, 0.06); 0.230
Birth order			
1			1.00
2			0.01 (-0.01, 0.03); 0.311
3			-0.02 (-0.06, 0.02); 0.259
4			-0.01 (-0.06, 0.03); 0.569
Missing			0.19 (-0.23, 0.60); 0.371

Bold p-values represent statistically significant results.
* Model 1: multilevel model adjusts for age, ethnicity, and height as fixed effects, and a random effect for person to allow for within person eye measurements (59,642 eyes of 35,097 patients).
† Model 2 adjusts as model 1 plus visual acuity, spherical equivalent, corneal astigmatism, macular curvature, and center point foveal thickness as fixed effects (54,489 eyes of 32,564 patients).
‡ Model 3 adjusts as model 2 plus deprivation, higher education, fluid intelligence score, annual income, and birth order as fixed effects (53,135 eyes of 31,727 patients).
§ No missing data on ethnicity on this model.

Table 4.6: Standardized difference in fovea curvature per specified differences in covariates ([95% CI]; p-value) for males.

Characteristic	**Males model 1***	**Males model 2†**	**Males model 3‡**
Age (per decade)	-0.03 (-0.04, -0.01); **8.4e-05**	-0.01 (-0.03, 0.00); 0.168	-0.01 (-0.02, 0.01); 0.466
Ethnicity			
White	1.00	1.00	1.00
Black	-0.59 (-0.67, -0.51); **6.0e-49**	-0.84 (-0.92, -0.76); **2.1e-88**	-0.80 (-0.89, -0.71); **5.2e-68**
Asian	-0.36 (-0.43, -0.28); **7.4e-21**	-0.54 (-0.62, -0.46); **1.7e-41**	-0.54 (-0.62, -0.45); **2.3e-34**
Other	-0.28 (-0.39, -0.18); **1.5e-07**	-0.48 (-0.60, -0.37); **1.5e-17**	-0.47 (-0.59, -0.35); **7.2e-15**
Mixed	-0.20 (-0.35, -0.06); **0.005**	-0.38 (-0.53, -0.24); **3.8e-07**	-0.37 (-0.52, -0.21); **3.0e-06**
Chinese	-0.34 (-0.54, -0.15); **6.5e-04**	-0.53 (-0.73, -0.32); **4.5e-07**	-0.54 (-0.76, -0.33); **8.9e-07**
Prefer not to say	-0.13 (-0.31, 0.05); 0.150	-0.13 (-0.31, 0.05); 0.168	-0.09 (-0.29, 0.11); 0.372
Missing	0.00 (-0.25, 0.26); 0.986	-0.04 (-0.31, 0.22); 0.749	-0.54 (-1.9, 0.82); 0.437
Height (per 5cm)	0.04 (0.03, 0.05); **4.3e-19**	0.04 (0.03, 0.04); **3.4e-15**	0.03 (0.02, 0.04); **4.2e-12**
Visual acuity (per 5 letters)		0.01 (0.01, 0.01); **3.5e-07**	0.01 (0.01, 0.01); **8.4e-07**
Spherical equivalent (per diopter)		-0.04 (-0.04, -0.03); **5.3e-53**	-0.04 (-0.04, -0.03); **5.6e-50**
Corneal astigmatism (per diopter)		-0.06 (-0.08, -0.05); **3.6e-20**	-0.06 (-0.08, -0.05); **5.3e-20**
Macula curvature (per 0.01)		0.14 (0.07, 0.21); **1.0e-04**	0.14 (0.07, 0.21); **1.2e-04**
Center point retinal thickness (per 10µm)		-0.14 (-0.15, -0.14); **0.0e+00**	-0.15 (-0.15, -0.14); **0.0e+00**
Fluid intelligence			0.00 (-0.01, 0.00); 0.559
Annual income (Great British Pound)			
Less than 18,000			1.00
18,000 to 30,999			0.05 (0.01, 0.10); **0.010**
31,000 to 51,999			0.05 (0.01, 0.09); **0.021**
52,000 to 100,000			0.07 (0.02, 0.11); **0.004**
Greater than 100,000			0.08 (0.02, 0.14); **0.005**
Prefer not to say			0.00 (-0.05, 0.06); 0.857
Missing			0.11 (-0.18, 0.41); 0.459
Per increase in income category			0.06 (0.02, 0.10); **0.005**
Birth order			
1			1.00
2			0.00 (-0.02, 0.03); 0.728
3			0.03 (-0.02, 0.08); 0.216
4			-0.02 (-0.08, 0.03); 0.375
Missing			-0.20 (-0.65, 0.25); 0.389

Bold p-values represent statistically significant results.
* Model 1: multilevel model adjusts for age, ethnicity, and height as fixed effects, and a random effect for person to allow for within person eye measurements (49,229 eyes of 28,842 patients).
† Model 2 adjusts as model 1 plus visual acuity, spherical equivalent, corneal astigmatism, macular curvature, and center point foveal thickness as fixed effects (45,296 eyes of 26,982 patients).
‡ Model 3 adjusts as model 2 plus deprivation, higher education, fluid intelligence score, income, and birth order as fixed effects (44,118 eyes of 26,252 patients).

4.3.2 Results from Macular Curvature Analysis

A total of 126,291 eyes of 65,023 subjects (35,176 female) with a mean age $\pm$ SD of 57.3 $\pm$ 8.11 years were included. The vast majority were assigned to white ethnicity (90.4% of included participants). In terms of refractive error, 58,432 eyes of 34,813 subjects (46.3%) were emmetropic; 33,283 eyes of 20,007 subjects (26.4%) were hypermetropic, with 25,843 eyes (16,769 subjects), 6317 eyes (4494 subjects), and 1123 eyes (828 subjects) assigned to mild, moderate, and high hypermetropia, respectively. Also, 34,576 eyes (27.4%) of 20,205 subjects were myopic with 17,278 eyes (11,986 subjects), 12,355 eyes (8139 subjects), and 4943 eyes (3322 subjects) assigned to mild, moderate, and high myopia, respectively (Figure 4.6).

■ Association with Macular Curvature

The overall macular curvature described a Gaussian distribution with a mean ($\pm$ SD) of 0.00213 $\pm$ 0.00145 (Figure 4.6). Multiple putative associated features of macular curvature were initially considered, including demographic characteristics (age, ethnicity, sex), ocular measures (refractive error, IOP, corneal curvature), functional data (VA, FI), and childhood environment (birth weight, maternal smoking). Figure 3 shows the association of macular curvature with each explored covariate. Here, as well as in a linear regression analysis, refractive error indicated the most distinct correlation (r = $-$ 0.391, P $>$ 0.001), revealing lower macular curve values (i.e., flatter curve) with increasing spherical equivalent. Associations of other parameters with macular curvature were lower but evident (Figure 4.8). Maternal smoking during pregnancy (mean = 0.00212) and female sex (mean = 0.00212) were associated with a slightly lower macular curvature (maternal non-smoking, mean = 0.00214; males, mean = 0.00214). Concerning ethnicity, white subjects (mean = 0.00212 $\pm$ 0.00145) and Asian subjects (mean = 0.00215 $\pm$ 0.00132) showed the lowest macular curvature, whereas Chinese subjects (mean = 0.00272 $\pm$ 0.00192) and black subjects (mean = 0.00235 $\pm$ 0.00127) revealed the highest overall values (Figure 4.8). Of note, white participants (mean = $-$ 0.30 $\pm$ 2.71 D) and Chinese participants (mean = $-$ 2.26 $\pm$ 3.19 D) also represented the subgroups in terms of lowest and highest refractive errors, respectively. However, the black subgroup also revealed low refractive error (mean = $-$ 0.39 $\pm$ 2.25 D), in the range of white subjects. To minimize association biases and because features might exhibit a significance in a multilinear model despite not showing any real correlation in a bivariate analysis, we fitted three multilinear models (Table 4.7). Model 1 focused on demographic parameters only, and there was a significant effect of ethnicity such that black and Chinese participants revealed a positive correlation to macular curvature scores. Age and sex revealed a significant effect, as well, which changed distinctively in the models 2 and 3, which included functional and ocular measures and infancy factors, respectively. In addition to ethnicity, refractive error consistently revealed the most significant effect on macular curvature scores. The impact of VA, corneal curvature, IOP, FI, and birthweight was low, but still significant. Maternal smoking, however, did not reveal any significance in the model.

Table 4.7: Standard deviation differences in macular curvature by change in independent variables Model 1 stands for a multilevel model that adjusts for demographical parameters (age, sex, and ethnicity) as fixed effects and a random effect for person to allow for within person eye measurements. Model 2 adjusts as model 1 plus functional (visual acuity and fluid intelligence) as well as ocular measures (spherical equivalent, corneal curvature and intraocular pressure). Model 3 adjusts as model 2 plus infancy factors (maternal smoking and birthweight). The marginal R^2 (i.e., variance explained by the models) and the conditional R^2 were 0.004 and 0.747 for model 1, 0.161 and 0.794 for model 2, and 0.167 and 0.797 for model 3, respectively. CI = confidence interval, LogMAR = Logarithm of the Minimum Angle of Resolution; mmHg = millimeters of mercury

	Model 1			**Model 2**			**Model 3**		
Independent Variables	*Estimates*	*CI*	*P-Value*	*Estimates*	*CI*	*P-Value*	*Estimates*	*CI*	*P-Value*
Age (per decade)	-0.05	-0.06 – -0.05	**1.3E-32**	0.04	0.04 – 0.05	**5.6E-20**	0.02	0.01 – 0.04	**1.1E-04**
Sex									
Female (Ref)	---	---	---	---	---	---	---	---	---
Male	0.02	0.01 – 0.04	**5.1E-03**	-0.01	-0.02 – 0.01	3.3E-01	-0.03	-0.05 – -0.01	**1.1E-02**
Ethnicity									
White (Ref)	---	---	---	---	---	---	---	---	---
Black	0.14	0.10 – 0.18	**1.6E-10**	0.23	0.18 – 0.27	**7.7E-21**	0.14	0.06 – 0.22	**1.1E-03**
Asian	0.002	-0.04 – 0.04	9.4E-01	0.07	0.03 – 0.12	**2.3E-03**	0.07	-0.01 – 0.15	7.5E-02
Other	0.002	-0.05 – 0.05	9.4E-01	0.11	0.05 – 0.18	**2.0E-04**	0.12	0.03 – 0.22	**1.4E-02**
Mixed	0.04	-0.03 – 0.12	2.8E-01	0.05	-0.04 – 0.13	2.8E-01	0.03	-0.09 – 0.14	6.6E-01
Chinese	0.40	0.29 – 0.50	**1.1E-12**	0.21	0.09 – 0.33	**7.6E-04**	0.08	-0.10 – 0.26	3.6E-01
Visual acuity (per 0.1 LogMAR)				-0.02	-0.02 – -0.01	**8.4E-36**	-0.02	-0.02 – -0.01	**1.8E-21**
Refractive error (per diopter)				-0.15	-0.15 – -0.15	**0.0E+00**	-0.15	-0.16 – -0.15	**0.0E+00**
Corneal curvature (per diopter)				-0.02	-0.03 – -0.02	**2.2E-19**	-0.02	-0.03 – -0.02	**1.7E-13**
Intraocular pressure (per mmHg)				-0.02	-0.03 – -0.02	**3.8E-14**	-0.02	-0.03 – -0.02	**8.2E-09**
Fluid Intelligence				0.03	0.03 – 0.03	**4.6E-63**	0.03	0.03 – 0.04	**1.9E-40**
Maternal Smoking							0.02	-0.00 – 0.004	7.3E-02
Birthweight (per kg)							-0.03	-0.05 – -0.02	**4.8E-05**

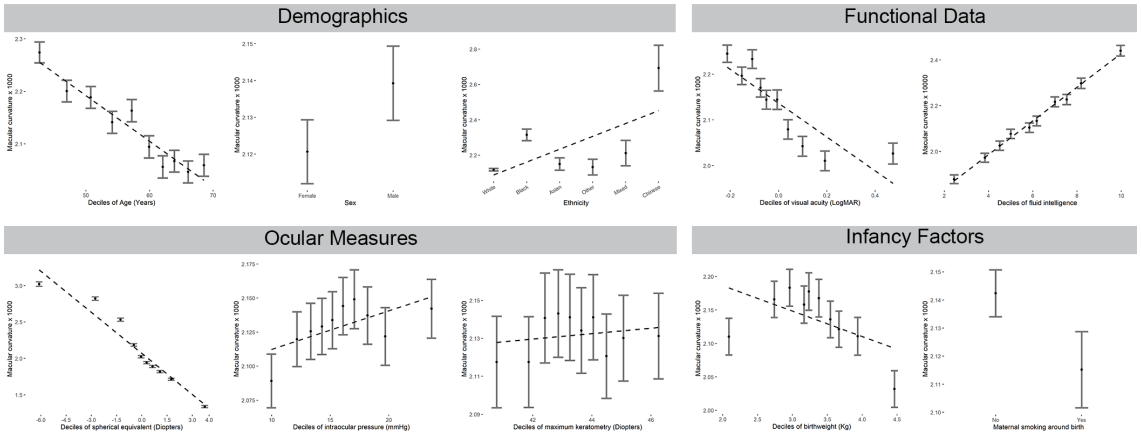

Figure 4.8: Macular curvature and associated parameters. The panel reveals adjusted mean macular curvature by deciles of variables of interest. Adjusted means (solid black dots), 95% confidence intervals (vertical solid lines), and the regression line (dotted line) are from a multilevel model allowing for repeated macular curvature measurement for each person.

■ Dome-Shaped Configuration

Defined as inverted macular curvature (negative values) (Figure 4.8), the prevalence of a macular dome-shaped configuration was overall 4.78% (6040 eyes of 4725 subjects) and was more common in hypermetropic than in emmetropic or myopic eyes (Table 4.8). Of note, the prevalence increased with more extreme refractive errors into both directions. In contrast, an extensive dome-shaped macular configuration (fourth quartile of eyes with negative macular curvature values; macular curvature, < -0.00105589; 1.26%; 1599 eyes of 1346 subjects) was more common in myopic than in emmetropic or hypermetropic eyes. The highest prevalence of this particular extensive macular shape was found in the subgroup of high myopia (Table 4.8), in particular in the subset of those eyes with very high refractive error (< -9 diopters spherical equivalent), which revealed a prevalence of 6.49% (79 of 1217 eyes). The odds ratios for the extensive macular dome-shaped configuration were 3.23 and 5.34 for high and very high myopia, respectively. In the set of the 100 eyes with the most negative macular curvature (i.e., most extensive macular dome-shaped configuration), subretinal fluid was present in 29% of cases, whereas it was absent in all other assessed OCT scans (used for validation), including those with negative macular curvature. The 100 eyes with the most negative macular curvature revealed significantly impaired VA compared to other eyes (mean $\pm$ SD, 0.140 ± 0.212 vs. 0.020 ± 0.202; $P < 0.0001$), whereas those with and without subretinal fluid did not show a significant difference in VA (P = 0.344). In terms of ethnicity, a macular dome-shaped and an xtensive dome-shaped configuration were most common in subjects of Chinese origin (5.38% and 2.15%, respectively) followed by white origin (4.99% and 1.33%, respectively) and other origin (3.86% and 0.96%, respectively). Subjects with Asian (2.74% and 0.60%, respectively), mixed (2.44% and 0.72%, respectively), or black (1.90%

and 0.41%, respectively) origin showed the particular concave macular configuration less frequently. These findings were partly independent from the impact of refraction on macular curvature, as Chinese subjects had a significantly higher proportion of dome-shaped macular presentation in each refraction subgroup excluding moderate and high hypermetropia (which was absent in our Chinese cohort) compared to other ethnicities; for example, in the high myopia subgroup, Chinese represented 8.54% and white 8.39%. This difference became even more obvious in the group of those with extensive dome-shaped macular configurations; for example, in the high myopia subgroup, Chinese represented 6.10% and white 3.94%. For this phenotype, the odds ratios for Chinese subjects with high or very high myopia were 4.83 and 7.91, respectively.

Table 4.8: Dome-Shaped Macular Configuration and Refractive Error

Refractive Subgroup		**Dome-Shape Configuration (n = 6040)**		**Extensive Dome-Shape Configuration (n = 1599)**	
Emmetropia	(n = 58432)	3.17 %	(n = 1855)	0.74%	(n = 435)
Hypermetropia	(n = 33283)	7.37 %	(n = 2453)	1.43 %	(n = 475)
Mild	(n = 25843)	5.43 %	(n = 1404)	1.19 %	(n = 308)
Moderate	(n = 6317)	12.70 %	(n = 802)	2.09 %	(n = 132)
High	(n = 1123)	21.99 %	(n = 247)	3.12 %	(n = 35)
Myopia	(n = 34576)	5.01 %	(n = 1732)	1.99 %	(n = 689)
Mild	(n = 17279)	4.06 %	(n = 702)	1.52 %	(n = 262)
Moderate	(n = 12355)	5.14 %	(n = 635)	1.95 %	(n = 241)
High	(n = 4943)	7.99 %	(n = 395)	3.76 %	(n = 186)

■ Inter-Eye Comparison

In 61,319 subjects (33,158 female) with a mean age of 57.2 $\pm$ 8.11 years (range, 39.2–70.5) both eyes were included in this study. Inter-eye comparison revealed small but significant differences in macular curvature (OD, 0.00214 $\pm$ 0.00144; OS, 0.00213 $\pm$ 0.00143; P = 0.012), refractive error (OD, $-$ 0.339 $\pm$ 2.640 D; OS, $-$ 0.280 $\pm$ 2.670 D; P $<$ 0.001), VA (OD, 0.021 $\pm$ 0.199 logMAR; OS, 0.018 $\pm$ 0.204 logMAR; P = 0.003), and IOP (OD, 16.0 $\pm$ 4.18; OS, 15.9 $\pm$ 4.22; P $<$ 0.001). Nevertheless, all parameters showed a high intereye correlation (Supplementary Figure 4.9). In this cohort, the prevalence of macular dome-shaped and extensive domeshaped configurations was similar to the aforementioned overall study population with values of 4.67% (5723 eyes of 4327 participants) and 1.20% (1467 eyes of 1187 partici-

pants), including a binocular manifestation in 32.26% and 23.59% of subjects, respectively. In the majority of participants with monocular manifestation, the affected eye was more hypermetropic or less myopic (65.50% and 54.10% for dome-shaped and extensive dome-shaped macular configurations, respectively).

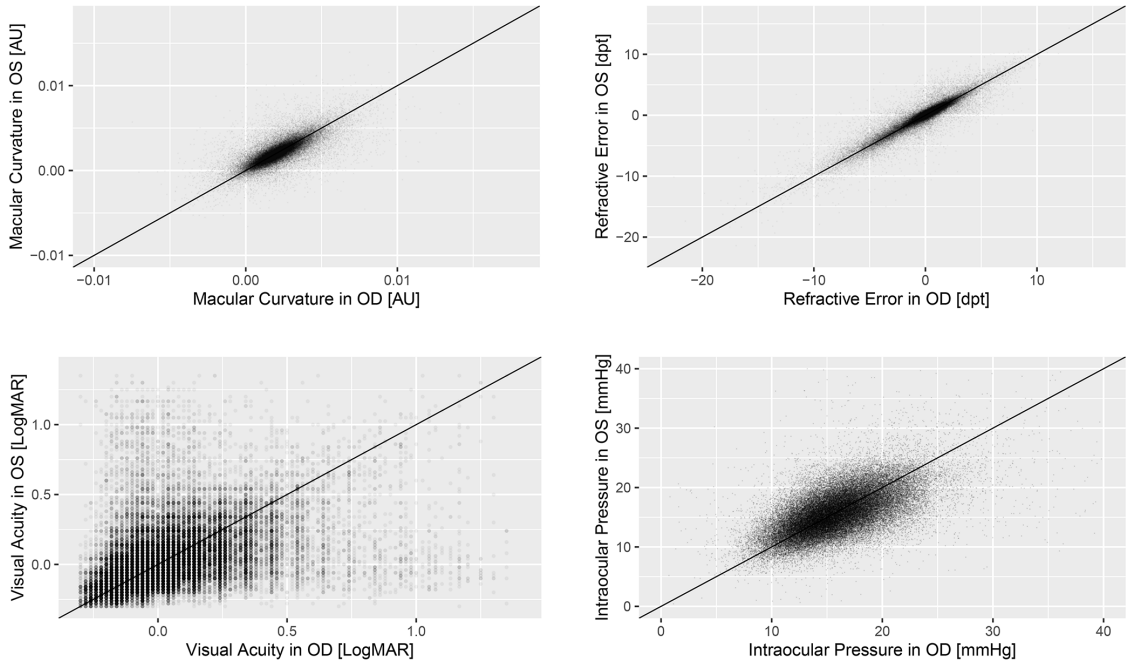

Figure 4.9: Inter-eye correlation. The scatter plots visualize the high inter-eye correlation for macular curvature, refractive error (as spherical equivalent; in diopters, dpt), visual acuity (in Logarithm of the Minimum Angle of Resolution, LogMAR), and intraocular pressure between right and left eyes of participants with both eyes included. There was a high inter-eye correlation for all parameters (macular curvature, r = 0.746, $P < 0.001$; refractive error, r = - 0.451, $P < 0.001$; visual acuity, r = 0.368, $P < 0.001$; intraocular pressure, r = 0.595, $P < 0.001$). The diagonal symbolizes perfect inter-eye agreement. OD = Right eye; OS = Left eye.

4.4 Summary

This project investigated FC and MC based on OCT images using the unique imaging dataset for the UKBB study, the world's largest cohort study of adults. A deep learning model was used to segment the retinal pigment epithelium without human supervision.

Our findings in FC curvature analysis highlight novel associations between OCT-derived FC and sociodemographic, VA, and ocular factors, with the greatest effect sizes in ethnicity. The findings could represent the result of independent maturation or development of inner and outer retinal layers during development and suggest FC as a candidate marker to comprehensively assess the fovea in health and disease.

MC revealed associations with demographic, functional, ocular, and infancy factors, as well as increasing

prevalence of a dome-shaped macular configuration in high refractive error including high myopia and hypermetropia. These findings imply different pathophysiologic processes that lead to macular development and might open new fields to future myopia and macula research.

CHAPTER 5

SEMANTIC SEGMENTATION FOR DETECTION OF NONEXUDATIVE MACULAR NEOVASCULARIZATION

5.1 Introduction

Type 1 macular neovascularization (MNV) is the most common form of MNV seen in eyes with age-related macular degeneration (AMD)[181, 182]. This neovascularization arises from the choroid and grows under the retinal pigment epithelium (RPE) and resides between Bruch's membrane (BM) and the RPE [25]. In AMD, these nonexudative neovascular lesions have an increased risk of progressing to exudation and vision loss[26, 27, 28]. Since early detection and treatment of exudative AMD has been shown to result in better visual acuity outcomes, it is important to identify and closely follow nonexudative MNV (neMNV) even before exudation develops so that treatment can be initiated once symptomatic exudation arises. Optical coherence tomography angiography (OCTA) is the non-invasive imaging strategy of choice for the detection of neMNV [183, 184].

While OCTA is able to detect neMNV, not all clinical practices are equipped with an OCT instrument capable of angiographic imaging. However, most practices are equipped with standard OCT instruments that provide structural B-scan images. The presence of type 1 neMNV has been associated with the presence of a double layer sign (DLS), also known as a shallow irregular RPE elevation (SIRE), on structural OCT B-scan images [29, 185]. A cost-effective strategy to detect these nonexudative neovascular lesions without requiring the use of the more expensive OCTA technology would be to train clinicians to detect neMNV using structural OCT B-scans.

Shi *et al*. [29] investigated whether graders could accurately identify the presence of a DLS in eyes with type 1 neMNV. After training on eyes with known subclinical neovascular lesions, the graders assessed a total of 100 eyes with AMD in which 20 eyes had both drusen and type 1 neMNV, 13 eyes had geographic atrophy (GA) along with type 1 neMNV, 44 eyes only had drusen, and 23 eyes only had GA. While a statistically significant association was found between the presence of the DLS and type 1 neMNV, the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) for the junior graders were 73%, 84%, 69%, and 86%, respectively, while the sensitivity, specificity, PPV, and NPV for the senior grader was 88%, 87%, 76%, and 94%, respectively. These results suggested that training and experience should yield improved results, and the grading of these structural OCT images for the presence of a DLS could be approached by developing machine learning algorithms to detect these lesions.

In this study, we aimed to develop a deep learning algorithm to detect a DLS based on cross-sectional structural OCT B-scans.

5.2 Detection of Nonexudative Macular Neovascularization with ViT-based Segmentation model

Vision Transformer (ViT) [186] is now considered the state of the art in many computer vision tasks. We trained a ViT segmentation model using eyes with and without type 1 neMNV that was confirmed on

swept-source OCTA (SS-OCTA) imaging. A large dataset of annotated structural B-scans was fed into the algorithm to train the model. After training and preliminary testing, the machine learning model was finally applied on the same dataset of 100 eyes that were evaluated by human graders in our previous study [29]. The sensitivity, specificity, NPV, and PPV of the machine learning algorithm were then compared with the performance of the human graders.

5.2.1 ViT-based Segmentation model

The segmentation model was built using a fully ViT-based encoder-decoder architecture [187] mapping a sequence of patch embeddings to pixel-level class labels. An overview of the model is shown in Figure 5.1. Since ViT is relatively novel and has limited application in ophthalmology, we briefly describe the architecture of our model. First, an input image was split into a sequence of patches. Each patch was flattened into a 1D vector and then fed into a linear projection layer that would produce a sequence of patch embeddings. To retain positional information, learnable position embeddings are added to the patch embeddings to get the resulting input sequence of tokens. A transformer layer consists of a multi-headed self-attention (MSA) block followed by a point-wise multi-layer perceptrons (MLP) block of two layers with layer norm (LN) applied before every block and residual connections added after every block. The transformer encoder was applied to the sequence of tokens to generate a contextualized encoded sequence. The decoder learns to map patch-level encodings coming from the encoder to patch-level class labels. A pointwise linear layer was applied to the patch-level encodings to produce patch-level class logits. The sequence was then reshaped into a 2D feature map and upsampled using bilinear interpolation to the original image size. A softmax was then applied on the class dimension to obtain the final segmentation map.

The resolution of the original B-scan image was 500 pixels in width, and 1536 pixels in height. The B-scan was then systematically cropped into 500×768 accordingly so that the retinal layer comes to the center of the cropped image. During the training, we applied mean subtraction, random resizing of the image to a ratio between 0.5 and 2.0 and random left-right flipping for data augmentation purposes. We randomly cropped large images and pad small images to a fixed input size of 512×512. The output of the network was the same size as the input but had 3 channels corresponding to the background, DLS, and drusen. Our model was trained end-to-end with a per-pixel cross-entropy loss with SGD optimizer. At inference time, argmax was applied after upsampling to obtain a single class label per pixel.

Our backbone ViT model had 12 layers, 768 token sizes, and 12 heads, and was pre-trained on ImageNet-21k which was publicly available provided by the image classification library timm (available at `http://github.com/rwightman/pytorch-image-models/`).

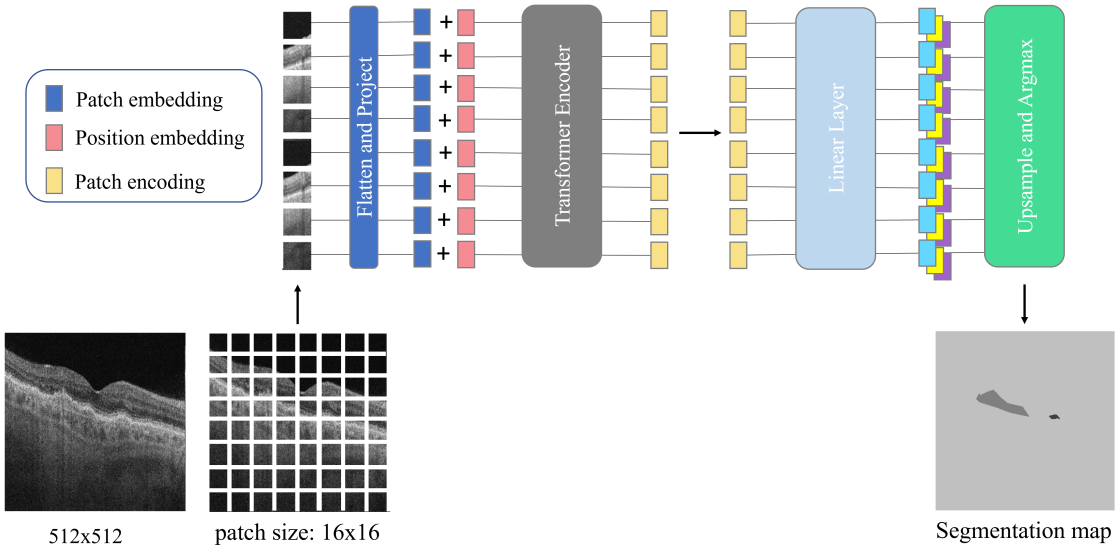

Figure 5.1: Framework of ViT segmentation model. 512×512 images were used as input. 16×16 image patches were projected to a sequence of embeddings and then encoded with a transformer and reshaped into a segmentation map.

5.2.2 Classification criteria

We built a binary classification algorithm that identifies eyes with MNV from an en-face prediction map generated by processing all 500 slices of B-scans and calculating prediction masks using the segmentation model as shown in Figure 5.2. The en-face images had $500(width) \times 500(height)$ that corresponded to the original width and depth of the B-scans, respectively. From the en-face images, we first rescaled them to 128×128 to reduce noise, then extracted all connected blobs of DLS lesions. The size of the largest component for each eye was measured, then DLS labels were assigned to the eye if the largest component size was larger than 65 pixels. In order to set the threshold value, we processed the above criteria to an internal validation set and observed how the number of classification errors changed. We then selected the center point of component size with the lowest error as our threshold. See Figure 5.3 for details.

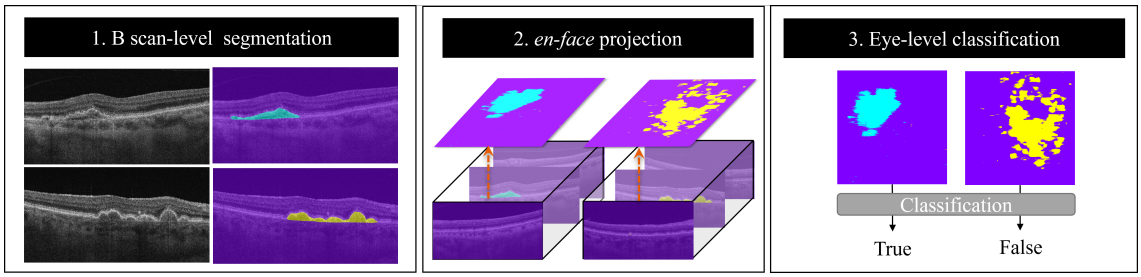

Figure 5.2: Process overview. A. DLSs and drusen were extracted from B scans by a segmentation model. The detected DLSs and drusen were depicted in blue and yellow, respectively. B. After processing predictions for all B-scans, we generated en-face maps. C. eyes with MNV were then identified based on the en-face projection map.

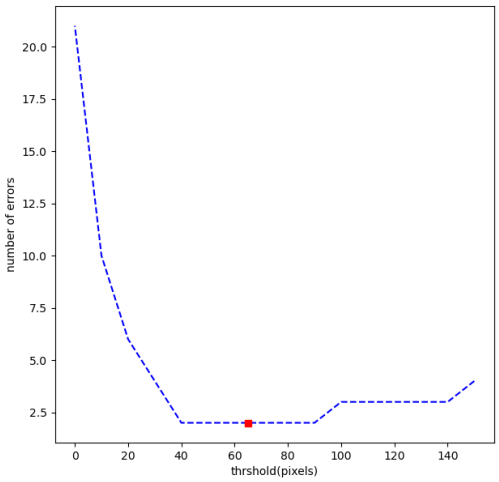

Figure 5.3: Threshold vs. Classification errors. On each 128×128 en-face projection map on the validation set, we observed how the number of classification errors changed with regard to the threshold. Between threshold values were 40 to 60, the errors got minimum, and 65 was the middle point.

5.2.3 Results

A total 251 eyes from 210 patients with 182 eyes with DLS and 115 eyes with drusen were used for model training. Out of 125,500 B-scans, 6,879 B-scans were manually annotated. The data were partitioned into 70% for training and 30% for validation sets at the patient level. Out of a total of 125,500 B-scans, 5,256 B-scans and 1,623 B-scans were used for manual labeling of training and validation, respectively, and provided for segmentation model training. A vision Transformer segmentation model was then built to extract DLS and drusen from B-scans.

As a point of comparison, we also trained a U-Net [188] model using a traditional fully convolutional approach on the same dataset as our ViT segmentation model. Just like ViT, systematically cropped 500×768 B-scans were used. The images were then rescaled to 512×512 and normalized to a range between 0 and 1. The model was trained with a cross-entropy loss function using Adam optimizer. We chose a batch size of 8. The learning rate was initially set to 1×10^{-4}, and decay over each update was set to the initial learning ratio divided by epochs. Basic data augmentations such as shift, flip, and rotation were applied. Furthermore, Spatial Dropout which drops entire feature maps instead of individual elements was additionally applied to regularize the activations and reduce overfitting.

Intersection over union (IoU) was DLS: 58.80%, drusen: 61.11% for the Transformer model, and DLS: 55.39%, drusen: 54.30% for the Unet model on the validation set. Some examples of predicted masks are shown in Figure 5.4.

ViT showed significantly better performance than the U-Net model on the segmentation task. Although U-Net and ViT are just one example of CNN and Transformer models, respectively, the performance gap could have arisen from the fundamental difference in their model architectures. ViT is quite different from CNN. CNNs start with a feature of large spatial sizes and a small channel size and gradually increase the channel size while decreasing the spatial size. In ViT, input images are divided into 16×16 patches and fed to the transformer network; except for the first embedding layer, there is no convolution operation in ViT, and the position interactions occur through the self-attention layers. While CNNs have restricted spatial interactions, ViT uses multi-head self-attention that allows all the positions in an image to interact. These characteristics of ViT might have helped with improving segmentation performance, especially on eyes with GA.

■ Test set description

A completely separate set of patients were used as a test set and these patients were not used for training and internal validation. The test-set consisted of 100 AMD eyes with drusen and GA, with or without nonexudative type 1 MNV, and this is the same test-set that was used by human graders in our previous study [29]. This test-set is shown in Table 5.1. The consensus grading results of these 100 eyes were used as the ground truth to be compared with the output from the current model. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the AI algorithm for identifying DLS associated with neMNV were calculated. The extracted prediction masks from all B-scans in a volume were projected to an en face image and an eye level projection map was obtained for each eye. A binary classification algorithm was established to identify eyes with neMNV from the projection map.

Table 5.1: Presence of Subclinical Macular Neovascularization in Test Eyes with Nonexudative Age-Related Macular Degeneration.

Group	Intermediate Age-Related Macular Degeneration (n = 64)	Late Age-Related Macular Degeneration (n = 36)
Presence of subclinical MNV (n = 33)	20	13
Absence of subclinical MNV (n = 67)	44	23

MNV = macular neovascularization.

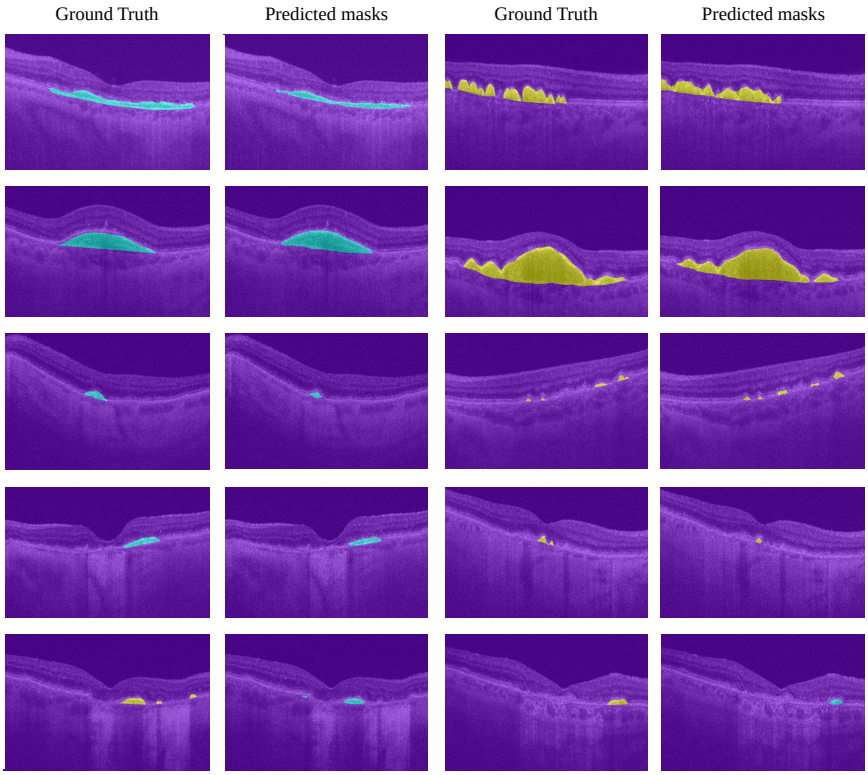

Figure 5.4: Prediction examples from the validation set. Grounded Truth(left) and predicted masks(right) overlaid on input B scan. 1 to 4th rows: The detected DLSs and drusen were depicted in blue and yellow, respectively. Our segmentation model mostly was able to extract DLSs and drusen at different locations/volumes properly. 5-th row: Failure case examples. Our segmentation model tends to mislabel drusen to DLSs on eyes with geographic atrophy.

■ **Comparison with human graders**

Two junior graders read the scans separately then reached a consensus grading. They detected DLSs in 24 of 33 eyes with subclinical MNV and did not detect DLSs in 56 of 67 eyes without MNV. Their sensitivity, specificity, PPV, and NPV were 73%, 84%, 69%, and 86%, respectively. The senior grader detected DLSs in 29 of 33 eyes with subclinical MNV and did not detect DLSs in 58 of 67 eyes without MNV, achieving a sensitivity, specificity, PPV, and NPV of 88%, 87%, 76%, and 94%, respectively. For all graders, there were statistically significant associations between type 1 MNV and presence of the double-layer sign ($P < 0.001$). Compared to human grader, AI grader consistently performs better than junior graders and as good as senior graders. Above all, AI grader shows robust performance on eyes with late AMD while human graders had more difficulty in identifying a DLSs in eyes with GA.

All the results are summarized in Table 5.2. Our model, referred to as AI grader, detected DLSs in 27 of 33 eyes with subclinical MNV and did not detect a double-layer sign in 60 of 67 eyes without MNV, and achieved 82%, 90%, 79%, and 91% sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), respectively on a separate test set of 100 eyes that were evaluated by human graders in a previous study. The area under the curve (AUC) value was calculated as 0.91 (95% CI: 0.85 to 0.98). The results of the algorithm showed excellent agreement with the senior human grader ($kappa = 0.83$, $p < 0.001$) and moderate agreement with the junior grader consensus ($kappa = 0.54$, $p < 0.001$).

Table 5.2: Sensitivity, Specificity, and Predictive Values of Double-Layer sign for Identifying Subclinical Macular Neovascularization in Nonexudative Age-Related Macular Degeneration

	Age-Related Macular Degeneration (n = 100)			Intermediate Age-Related Macular Degeneration (n = 64)			Late Age-Related Macular Degeneration (n = 36)		
Values	AI Grader	Consensus from Junior Grader	Results from Senior Grader	AI Grader	Consensus from Junior Grader	Results from Senior Grader	AI Grader	Consensus from Junior Grader	Results from Senior Grader
Sensitivity	0.82	0.73	**0.88**	0.75	0.85	**0.90**	**0.92**	0.54	0.85
Specificity	**0.90**	0.84	0.87	**0.96**	0.84	0.93	**0.78**	0.83	0.74
PPV	**0.79**	0.69	0.76	**0.88**	0.71	0.86	**0.71**	0.64	0.65
NPV	0.91	0.86	**0.94**	0.89	0.93	**0.95**	**0.95**	0.76	0.89
P value	≤ 0.001	≤ 0.001	≤ 0.001	≤ 0.001	≤ 0.001	≤ 0.001	≤ 0.001	0.056	≤ 0.001

NPV = negative predictive value; PPV = positive predictive value.

5.3 Detection of Nonexudative Macular Neovascularization with Semi-supervised Segmentation model

In section 5.2, we have developed a ViT segmentation model that detects the presence of neMNV from the SS-OCT scans. Swept source has a longer wavelength and it penetrates better to deeper layers, however, SS-OCT is more expensive and came out later. Many available clinical dataset consists of SD-OCT scans. Our motivation is to expand our scope to SD-OCT scans. Our SD-OCT dataset is challenged by the difficulty in acquiring a large set of annotated volumetric images for training. For the development of a machine learning algorithm to identify subclinical MNV based on SD-OCT imaging, we conducted a preliminary experiment with semi-supervised setting. To overcome the problems, we studied the semi-supervised semantic segmentation problem via exploring both labeled data and extra unlabeled data. We employed a consistent regularization approach, called cross pseudo supervision (CPS) [108].

5.3.1 Cross Pseudo Super Vision

Cross Pseudo Super Vision (CPS) imposes the consistency on two segmentation networks $P_1 = f(X; \Theta_1)$, and $P_2 = f(X; \Theta_2)$ that share the same structure but are initialized differently.

$$X \to f(\Theta_1) \to P_1 \to Y_1 \tag{5.1}$$

$$\searrow f(\Theta_2) \to P_2 \to Y_2 \tag{5.2}$$

P_1 (P_2) is the segmentation confidence map, which is the network output after softmax normalization. The labeled and unlabeled images are fed into the two networks. The outputs of the two networks on the labeled data are supervised separately by the corresponding ground-truth segmentation map. On the unlabeled images, each segmentation network for an input image estimates a pseudo segmentation map, and output from one perturbed segmentation network is used to supervise the other segmentation network, and vice versa.

The training objective contains two losses: supervision loss L_s and cross pseudo supervision loss L_{cps}.

Given a set D_l of N labeled images, The supervision loss L_s is formulated using the standard pixel-wise cross-entropy loss on the labeled images over the two parallel segmentation networks:

$$L_s = \frac{1}{|D^l|} \sum_{X \in D^l} \frac{1}{W \times H} \sum_{i=0}^{W \times H} (l_{ce}(p_{1i}, y^\star_{1i}) + l_{ce}(p_{2i}, y^\star_{2i})).$$

where l_{ce} is the cross-entropy loss function and $y^\star_{1i}$ and $y^\star_{2i}$ are the ground truths. W and H represent the width and height of the input image.

Given a set D_u of M unlabeled images, The cross pseudo supervision loss is defined as follows.

$$L_{cps}^{u} = \frac{1}{|D^{u}|} \sum_{X \in D^{u}} \frac{1}{W \times H} \sum_{i=0}^{W \times H} (l_{ce}(p_{1i}, y_{2i}) + l_{ce}(p_{2i}, y_{1i})).$$

The cross pseudo supervision loss L_{cps}^{l} on the labeled data is defined in the same way. The whole cross pseudo supervision loss is the combination of the losses on both the labeled and unlabeled data: $L_{cps} = L_{cps}^{l} + L_{cps}^{u}$. The whole training objective is written as:

$$L = L_s + \lambda L_{cps}.$$

5.3.2 Preliminary validation of CPS

Using the SS-OCT dataset, we studied the performance of the semi-supervised segmentation model for 3 different training sizes. We divided the whole annotated B-scans in the training set to two groups via randomly sub-sampling $\frac{1}{8}$ and $\frac{1}{64}$ of the whole set as the labeled set and regarded the remaining images, including B-scans without annotations, as the unlabeled set. The number of labeled and unlabeled data in the training set are summarized in Table 5.3 and Table 5.4. Only one network branch was used to generate results for evaluation.

Table 5.3: The number of the labeled and unlabeled data for each training size on SS-OCT dataset.

ratio of labeled set	1/64	1/8	1
# of labeled data	82	657	5256
# of unlabeled data	90918	90343	85744

Table 5.4: The number of the labeled and unlabeled data for each training size on SD-OCT dataset.

ratio of labeled set	1/64	1/8	1
# of labeled data	36	294	2352
# of unlabeled data	27364	17106	15048

All the methods are based on DeepLabv3+ [189] with ResNet-50 [172]. We initialize the weights of two backbones in the two segmentation networks with the same weights pre-trained on ImageNet and the weights of two segmentation heads of DeepLabv3+ randomly. We adopt mini-batch SGD with momentum to train our model with Sync-BN [16]. The momentum is fixed as 0.9 and the weight decay is set to 0.0005. We employ a poly learning rate policy where the initial learning rate is multiplied by $(1 - itermaxiter) \times 0.9$. For the supervised baseline trained on the full training set, we use random horizontal flipping and multi-scale as data augmentation if not specified.

We illustrate the improvements of our method compared with the supervised baseline under all partition protocols in Figure 5.5. Sub-sampling rate=1 stands for a setting where all the annotated images are used as labeled data and the remaining images are used as unlabeled data. Figure 5.5 show CPS consistently outperforms the supervised baseline. Specifically, the improvements of our method over the baseline method are 3.55% and 2.03% for DLS class, and 3.43% and 6.66% for drusen class under $\frac{1}{64}$ and $\frac{1}{8}$ partition protocols separately. When all the annotated images are used as labeled data and the remaining images are used as unlabeled data, the improvement was 1.93%. Mean IOU of CPS model with a full annotation was 60.30%, which slightly outperforms the ViT model in Section 5.2.3.

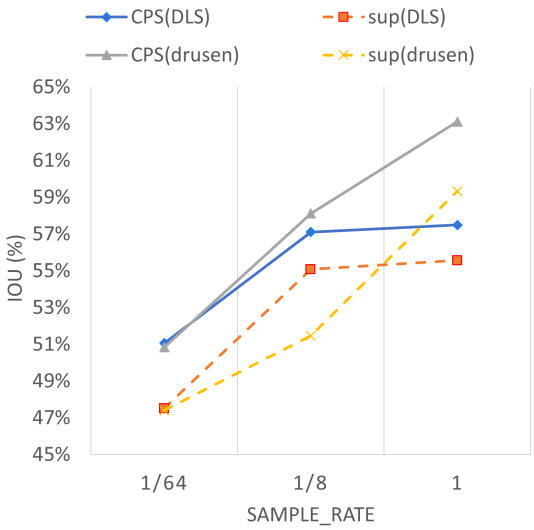

Figure 5.5: Improvements over the supervised baseline.

5.4 Summary

Our network was able to detect the presence of neMNV from structural B-scans alone by applying a purely transformer-based model. In this study, we developed a deep learning algorithm to detect a DLS based on cross-sectional structural OCT B-scans. We built a ViT segmentation model using eyes with and without type 1 neMNV that was confirmed on swept-source OCTA (SS-OCTA) imaging. The presented frameworks using Vision Transformers were able to detect the presence of neMNV from the SS-OCT scans. The evaluation demonstrated that segmentation performance can be consistent and reliable, which can substantially improve sensitivity and specificity in the final classification task. For expanding our algorithm to SD-OCT imaging where we have smaller number of labeled data, we conducted experiments using cross pseudo supervision to leverage both labeled and unlabeled data for learning. The experiments show the effectiveness of the semi-supervised approach on this task.

CHAPTER 6

CONCLUSIONS

In this dissertation I presented machine learning algorithms using three real world medical datasets. The first dataset consists of clinical data, imaging data, and VF data curated from three regionally different National Health Service glaucoma clinics in the United Kingdom including a total of 24,248 patients. The second dataset is UK Biobank (UKBB), one of the world's largest single resources for comprehensive study of health and disease, which contains OCT data from more than 60,000 participants. The last dataset is for an AMD study that includes one hundred eyes with nonexudative AMD from 94 patients enrolled in a prospective OCT imaging study at the Bascom Palmer Eye Institute (BPEI). In order to gain profound medical knowledge through deep learning from these datasets, I addressed three major issues; imbalanced data, having limited or lacking target domain annotations, and noisy labels and images. The conclusions of this paper are described below.

Chapter 3 demonstrated the ability of DL multimodal models with the glaucoma dataset which consists of VF data and two different modality imaging resources; Cp-OCT scans and IR SLO images of the disc. One of the goals of this study is to build a DL model that works robustly in a clinical setting. Therefore, quality assessments of the images in the dataset were not performed. As a result, the dataset inevitably includes a certain amount of noisy or poor quality images. To address this issue, we built a policy-based fusion network to maximally exploit the available information from the two imaging resources, including cases when either one was noisy or of poor quality. The dataset also has skewed VF sensitivity distributions, which caused the model to fail to learn patterns of advanced VF loss. To overcome this limitation, I used loss reweight approaches to alleviate the effects of imbalance. Our policy model was able to selectively choose between predictions from the two sub-models, and outperformed single modality models. Occlusion masking of our developed model shows that the DL models learned the correct structure-function mapping in a data-driven, feature agnostic fashion.

Chapter 4 reported the OCT derived FC and MC curvature analysis based on the UKBB data. For the calculation of FC and MC, the boundaries of RPE and ILM, and RPE and RPE choroid (RPE/C) needed to be detected. I present a methodology implemented to generate OCT-derived FC and MC quantification that does not require time-consuming annotations but relies on a fast and fully automated framework. A deep learning model was built to segment the retinal pigment epithelium. The model was trained based on pseudo labels that were generated in an unsupervised manner. FC and MC curvature were then systematically calculated by extracting the center fovea from the segmentation masks. Our findings in FC curvature

analysis may represent the result of independent maturation or development of inner and outer retinal layers during development and suggest that FC is a candidate marker for comprehensively assessing the fovea in health and disease. Our findings in MC curvature imply different pathophysiologic processes that lead to macular development and might open new fields to future myopia and macula research.

Chapter 5 presented a deep learning system to detect DLS based on cross-sectional structural OCT B-scans in the AMD dataset. The dataset consists of SS-OCT scans of a total of 251 eyes from 210 patients. Previous study shows experienced human graders could identify DLS in eyes with type 1 neMNV. That suggested that the grading of these structural OCT images for the presence of a DLS could be approached by developing machine learning algorithms to detect these lesions. I thus developed a semantic segmentation algorithm to detect DLS based on cross-sectional structural OCT B-scans with ViT. Compared with human graders, the presented model consistently performed better than junior graders and as well as the senior grader. The results of the algorithm also showed excellent agreement with the senior human grader and moderate agreement with the junior grader consensus. I further explored possible machine-learning algorithms to detect DLS from SD-OCT scans. Due to the expensive cost for the annotations, I conducted experiments using semi-supervised method called CPS to leverage both labeled and unlabeled data for learning. The experiments show the effectiveness of the semi-supervised approach on this task.

Several areas for future work are apparent. The problem of noisy labels and images, data imbalance, and having limited or lacking target domain annotations is naturally widespread in real-world applications. Handling the situation where these challenges occur simultaneously; the limited annotations for training is not uniformly distributed or with noise, requires further research. Applying the presented methods to a larger variety of datasets will help to identify the preferred deep learning method for future application. More work is required with non-convolutional deep learning to determine if the methods presented will generalize well to alternative architectures, e.g. multi-layer perceptrons and recurrent neural networks.

REFERENCES

[1] B. Medical, "Medical gallery of blausen medical 2014", WikiJournal of Medicine, vol.1, no.2, pp.1–79, 2014.

[2] E. A. Swanson, and J. G. Fujimoto, "The ecosystem that powered the translation of oct from fundamental research to clinical and commercial impact", Biomedical optics express, vol.8, no.3, pp.1638–1664, 2017.

[3] M. Adhi, and J. S. Duker, "Optical coherence tomography–current and future applications", Current opinion in ophthalmology, vol.24, no.3, p.213, 2013.

[4] B. Karamata, K. Hassler, M. Laubscher, and T. Lasser, "Speckle statistics in optical coherence tomography", JOSA A, vol.22, no.4, pp.593–596, 2005.

[5] J. M. Schmitt, S. Xiang, and K. M. Yung, "Speckle in optical coherence tomography", Journal of biomedical optics, vol.4, no.1, pp.95–105, 1999.

[6] T. T. Hormel, D. Huang, and Y. Jia, "Artifacts and artifact removal in optical coherence tomographic angiography", Quantitative Imaging in Medicine and Surgery, vol.11, no.3, p.1120, 2021.

[7] R. F. Spaide, J. G. Fujimoto, and N. K. Waheed, "Image artifacts in optical coherence angiography", Retina (Philadelphia, Pa.), vol.35, no.11, p.2163, 2015.

[8] S. Yun, G. Tearney, J. De Boer, and B. Bouma, "Motion artifacts in optical coherence tomography with frequency-domain ranging", Optics Express, vol.12, no.13, pp.2977–2998, 2004.

[9] N. Venkateswaran, A. Galor, J. Wang, and C. L. Karp, "Optical coherence tomography for ocular surface and corneal diseases: a review", Eye and Vision, vol.5, no.1, pp.1–11, 2018.

[10] R. Murthy, S. Haji, K. Sambhav, S. Grover, and K. Chalam, "Clinical applications of spectral domain optical coherence tomography in retinal diseases", Biomedical journal, vol.39, no.2, pp.107–120, 2016.

[11] K. Fort, G. Adda, and K. B. Cohen, "Amazon mechanical turk: Gold mine or coal mine?", Computational Linguistics, pp.413–420, 2011.

[12] S. Guo, W. Huang, H. Zhang, C. Zhuang, D. Dong, M. R. Scott, and D. Huang, "Curriculumnet: Weakly supervised learning from large-scale web images", Proceedings of the European Conference on Computer Vision, pp.135–150, 2018.

[13] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.2097–2106, 2017.

[14] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, et al., "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison", Proceedings of the AAAI conference on artificial intelligence, vol.33, no.01, pp.590–597, 2019.

[15] D. A. Dartt, Encyclopedia of the Eye, vol.1Academic Press, 2010.

[16] S. R. Flaxman, R. R. Bourne, S. Resnikoff, P. Ackland, T. Braithwaite, M. V. Cicinelli, A. Das, J. B. Jonas, J. Keeffe, J. H. Kempen, et al., "Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis", The Lancet Global Health, vol.5, no.12, pp.e1221–e1234, 2017.

[17] S. L. James, D. Abate, K. H. Abate, S. M. Abay, C. Abbafati, N. Abbasi, H. Abbastabar, F. Abd-Allah, J. Abdela, A. Abdelalim, et al., "Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017", The Lancet, vol.392, no.10159, pp.1789–1858, 2018.

[18] T. J. Heesterbeek, L. Lorés-Motta, C. B. Hoyng, Y. T. Lechanteur, and A. I. den Hollander, "Risk factors for progression of age-related macular degeneration", Ophthalmic and Physiological Optics, vol.40, no.2, pp.140–170, 2020.

[19] W. L. Wong, X. Su, X. Li, C. M. G. Cheung, R. Klein, C. Y. Cheng, and T. Y. Wong, "Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis", The Lancet Global Health, vol.2, no.2, pp.e106–e116, 2014.

[20] N. M. Schultz, S. Bhardwaj, C. Barclay, L. Gaspar, and J. Schwartz, "Global burden of dry age-related macular degeneration: A targeted literature review", Clinical Therapeutics, vol.43, no.10, pp.1792–1818, 2021.

[21] A. Stahl, "The diagnosis and treatment of age-related macular degeneration", Deutsches Ärzteblatt International, vol.117, no.29-30, p.513, 2020.

[22] M. D. Davis, R. E. Gangnon, L. Y. Lee, L. D. Hubbard, B. Klein, R. Klein, F. L. Ferris, S. B. Bressler, R. C. Milton, et al., "The age-related eye disease study severity scale for age-related macular degeneration: Areds report no. 17.", Archives of ophthalmology (Chicago, Ill.: 1960), vol.123, no.11, pp.1484–1498, 2005.

[23] A. Gheorghe, L. Mahdi, and O. Musat, "Age-related macular degeneration", Romanian journal of ophthalmology, vol.59, no.2, p.74, 2015.

[24] P. Mitchell, G. Liew, B. Gopinath, and T. Y. Wong, "Age-related macular degeneration", The Lancet, vol.392, no.10153, pp.1147–1159, 2018.

[25] M. Shen, P. J. Rosenfeld, G. Gregori, and R. K. Wang, "Predicting the onset of exudation in treatment-naïve eyes with nonexudative age-related macular degeneration", Ophthalmology Retina, vol.6, no.1, pp.1–3, 2022.

[26] J. R. de Oliveira Dias, Q. Zhang, J. M. Garcia, F. Zheng, E. H. Motulsky, L. Roisman, A. Miller, C. L. Chen, S. Kubach, L. de Sisternes, et al., "Natural history of subclinical neovascularization in nonexudative age-related macular degeneration using swept-source oct angiography", Ophthalmology, vol.125, no.2, pp.255–266, 2018.

[27] Y. Yanagi, A. Mohla, S. Y. Lee, R. Mathur, C. M. Chan, I. Yeo, T. Y. Wong, and C. M. G. Cheung, "Incidence of fellow eye involvement in patients with unilateral exudative age-related macular degeneration", JAMA ophthalmology, vol.136, no.8, pp.905–911, 2018.

[28] J. Yang, Q. Zhang, E. H. Motulsky, M. Thulliez, Y. Shi, C. Lyu, L. De Sisternes, M. K. Durbin, W. Feuer, R. K. Wang, et al., "Two-year risk of exudation in eyes with nonexudative age-related macular degeneration and subclinical neovascularization detected with swept source optical coherence tomography angiography", American journal of ophthalmology, vol.208, pp.1–11, 2019.

[29] Y. Shi, E. H. Motulsky, R. Goldhardt, Y. Zohar, M. Thulliez, W. Feuer, G. Gregori, and P. J. Rosenfeld, "Predictive value of the oct double-layer sign for identifying subclinical neovascularization in age-related macular degeneration", Ophthalmology retina, vol.3, no.3, pp.211–219, 2019.

[30] D. Banik, and D. Bhattacharjee, in Data Preprocessing, Active Learning, and Cost Perceptive Approaches for Resolving Data Imbalance, pp.66–89, IGI Global, 2021.

[31] Y. Cui, M. Jia, T. Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.9268–9277, 2019.

[32] S. Park, J. Lim, Y. Jeon, and J. Y. Choi, “Influence-balanced loss for imbalanced visual classification”, Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.735–744, 2021.

[33] S. Zhang, Z. Li, S. Yan, X. He, and J. Sun, “Distribution alignment: A unified framework for long-tail visual recognition”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.2361–2370, 2021.

[34] Y. Hong, S. Han, K. Choi, S. Seo, B. Kim, and B. Chang, “Disentangling label distribution for long-tailed visual recognition”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.6626–6636, 2021.

[35] J. Shu, Q. Xie, L. Yi, Q. Zhao, S. Zhou, Z. Xu, and D. Meng, “Meta-weight-net: Learning an explicit mapping for sample weighting”, Advances in neural information processing systems, vol.32, 2019.

[36] J. Byrd, and Z. Lipton, “What is the effect of importance weighting in deep learning?”, International Conference on Machine Learning, pp.872–881PMLR, 2019.

[37] A. K. Menon, S. Jayasumana, A. S. Rawat, H. Jain, A. Veit, and S. Kumar, “Long-tail learning via logit adjustment”, arXiv preprint arXiv:2007.07314, 2020.

[38] K. Cao, C. Wei, A. Gaidon, N. Arechiga, and T. Ma, “Learning imbalanced datasets with label-distribution-aware margin loss”, Advances in neural information processing systems, vol.32, 2019.

[39] J. Tan, C. Wang, B. Li, Q. Li, W. Ouyang, C. Yin, and J. Yan, “Equalization loss for long-tailed object recognition”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.11662–11671, 2020.

[40] J. Wang, W. Zhang, Y. Zang, Y. Cao, J. Pang, T. Gong, K. Chen, Z. Liu, C. C. Loy, and D. Lin, “Seesaw loss for long-tailed instance segmentation”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.9695–9704, 2021.

[41] A. More, “Survey of resampling techniques for improving classification performance in unbalanced datasets”, arXiv preprint arXiv:1608.06048, 2016.

[42] L. Shen, Z. Lin, and Q. Huang, “Relay backpropagation for effective learning of deep convolutional neural networks”, Proceedings of the European Conference on Computer Vision, pp.467–482Springer, 2016.

[43] Y. Gao, X. Bu, Y. Hu, H. Shen, T. Bai, X. Li, and S. Wen, “Solution for large-scale hierarchical object detection datasets with incomplete annotation and data imbalance”, arXiv preprint arXiv:1810.06208, 2018.

[44] B. Zhou, Q. Cui, X. S. Wei, and Z. M. Chen, “Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.9719–9728, 2020.

[45] D. Mahajan, R. Girshick, V. Ramanathan, K. He, M. Paluri, Y. Li, A. Bharambe, and L. Van Der Maaten, “Exploring the limits of weakly supervised pretraining”, Proceedings of the European Conference on Computer Vision, pp.181–196, 2018.

[46] B. Kang, S. Xie, M. Rohrbach, Z. Yan, A. Gordo, J. Feng, and Y. Kalantidis, “Decoupling representation and classifier for long-tailed recognition”, arXiv preprint arXiv:1910.09217, 2019.

[47] Y. Wang, W. Gan, J. Yang, W. Wu, and J. Yan, “Dynamic curriculum learning for imbalanced data classification”, Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.5017–5026, 2019.

[48] J. Ren, C. Yu, X. Ma, H. Zhao, S. Yi, et al., “Balanced meta-softmax for long-tailed visual recognition”, Advances in neural information processing systems, vol.33, pp.4175–4186, 2020.

[49] Y. Zang, C. Huang, and C. C. Loy, “Fasa: Feature augmentation and sampling adaptation for long-tailed instance segmentation”, Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.3457–3466, 2021.

[50] Y. Li, T. Wang, B. Kang, S. Tang, C. Wang, J. Li, and J. Feng, “Overcoming classifier imbalance for long-tail object detection with balanced group softmax”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.10991–11000, 2020.

[51] J. Cai, Y. Wang, and J. N. Hwang, “Ace: Ally complementary experts for solving long-tailed recognition in one-shot”, Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.112–121, 2021.

[52] S. Ando, and C. Y. Huang, “Deep over-sampling framework for classifying imbalanced data”, Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp.770–785Springer, 2017.

[53] P. Chu, X. Bian, S. Liu, and H. Ling, “Feature space augmentation for long-tailed data”, Proceedings of the European Conference on Computer Vision, pp.694–710Springer, 2020.

[54] S. Li, K. Gong, C. H. Liu, Y. Wang, F. Qiao, and X. Cheng, “Metasaug: Meta semantic augmentation for long-tailed visual recognition”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.5212–5221, 2021.

[55] Y. Wang, X. Pan, S. Song, H. Zhang, G. Huang, and C. Wu, "Implicit semantic data augmentation for deep networks", Advances in Neural Information Processing Systems, vol.32, 2019.

[56] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority oversampling technique", Journal of artificial intelligence research, vol.16, pp.321–357, 2002.

[57] J. Kim, J. Jeong, and J. Shin, "M2m: Imbalanced classification via major-to-minor translation", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.13896–13905, 2020.

[58] J. Kozerawski, V. Fragoso, N. Karianakis, G. Mittal, M. Turk, and M. Chen, "Blt: Balancing long-tailed datasets with adversarially-perturbed images", Proceedings of the Asian Conference on Computer Vision, 2020.

[59] Z. Zhong, J. Cui, S. Liu, and J. Jia, "Improving calibration for long-tailed recognition", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.16489–16498, 2021.

[60] H. P. Chou, S. C. Chang, J. Y. Pan, W. Wei, and D. C. Juan, "Remix: rebalanced mixup", Proceedings of the European Conference on Computer Vision, pp.95–110Springer, 2020.

[61] Z. Xu, Z. Chai, and C. Yuan, "Towards calibrated model for long-tailed visual recognition from prior perspective", Advances in Neural Information Processing Systems, vol.34, pp.7139–7152, 2021.

[62] S. Park, Y. Hong, B. Heo, S. Yun, and J. Y. Choi, "The majority can help the minority: Context-rich minority oversampling for long-tailed classification", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.6887–6896, 2022.

[63] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features", Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.6023–6032, 2019.

[64] T. Wang, Y. Li, B. Kang, J. Li, J. Liew, S. Tang, S. Hoi, and J. Feng, "The devil is in classification: A simple framework for long-tail instance segmentation", Proceedings of the European Conference on Computer Vision, pp.728–744Springer, 2020.

[65] X. Wang, L. Lian, Z. Miao, Z. Liu, and S. X. Yu, "Long-tailed recognition by routing diverse distribution-aware experts", arXiv preprint arXiv:2010.01809, 2020.

[66] Y. Zhang, B. Hooi, L. Hong, and J. Feng, "Test-agnostic long-tailed recognition by test-time aggregating diverse experts with self-supervision", arXiv preprint arXiv:2107.09249, 2021.

[67] B. Kang, Y. Li, S. Xie, Z. Yuan, and J. Feng, "Exploring balanced feature spaces for representation learning", International Conference on Learning Representations, 2020.

[68] P. Wang, K. Han, X. S. Wei, L. Zhang, and L. Wang, “Contrastive learning based hybrid networks for long-tailed image classification”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.943–952, 2021.

[69] J. Cui, Z. Zhong, S. Liu, B. Yu, and J. Jia, “Parametric contrastive learning”, Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.715–724, 2021.

[70] D. Samuel, and G. Chechik, “Distributional robustness loss for long-tail learning”, Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.9495–9504, 2021.

[71] J. Goh, and M. Sim, “Distributionally robust optimization and its tractable approximations”, Operations research, vol.58, no.4-part-1, pp.902–917, 2010.

[72] S. Y. Boulahia, A. Amamra, M. R. Madi, and S. Daikh, “Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition”, Machine Vision and Applications, vol.32, no.6, pp.1–18, 2021.

[73] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, “Multisensor data fusion: A review of the state-of-the-art”, Information fusion, vol.14, no.1, pp.28–44, 2013.

[74] D. Ramachandram, and G. W. Taylor, “Deep multimodal learning: A survey on recent advances and trends”, IEEE signal processing magazine, vol.34, no.6, pp.96–108, 2017.

[75] D. Kiela, E. Grave, A. Joulin, and T. Mikolov, “Efficient large-scale multi-modal classification”, Proceedings of the AAAI Conference on Artificial Intelligence, vol.32, no.1, 2018.

[76] A. Owens, J. Wu, J. H. McDermott, W. T. Freeman, and A. Torralba, “Ambient sound provides supervision for visual learning”, Proceedings of the European Conference on Computer Vision, pp.801–816Springer, 2016.

[77] S. Poria, E. Cambria, and A. Gelbukh, “Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis”, Proceedings of the 2015 conference on empirical methods in natural language processing, pp.2539–2544, 2015.

[78] D. Yi, Z. Lei, and S. Z. Li, “Shared representation learning for heterogenous face recognition”, 2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG), vol.1, pp.1–7IEEE, 2015.

[79] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, “Large-scale video classification with convolutional neural networks”, Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp.1725–1732, 2014.

[80] C. Ding, and D. Tao, “Robust face recognition via multimodal deep face representation”, IEEE transactions on Multimedia, vol.17, no.11, pp.2049–2058, 2015.

[81] L. I. Kuncheva, Combining pattern classifiers: methods and algorithms, John Wiley & Sons, 2014.

[82] V. Fernandez Arguedas, Q. Zhang, and E. Izquierdo, “Bayesian multimodal fusion in forensic applications”, Proceedings of the European Conference on Computer Vision, pp.466–475Springer, 2012.

[83] M. S. Kankanhalli, J. Wang, and R. Jain, “Experiential sampling in multimedia systems”, IEEE Transactions on Multimedia, vol.8, no.5, pp.937–946, 2006.

[84] D. E. Nirmala, B. S. Paul, and V. Vaidehi, “A novel multimodal image fusion method using shift invariant discrete wavelet transform and support vector machines”, 2011 International Conference on Recent Trends in Information Technology (ICRTIT), pp.932–937IEEE, 2011.

[85] D. Arsic, B. Schuller, and G. Rigoll, “Suspicious behavior detection in public transport by fusion of low-level video descriptors”, 2007 IEEE International Conference on Multimedia and Expo, pp.2018–2021IEEE, 2007.

[86] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler, “A system for traffic sign detection, tracking, and recognition using color, shape, and motion information”, IEEE Proceedings of the Intelligent Vehicles Symposium, 2005., pp.255–260IEEE, 2005.

[87] M. Meuter, C. Nunn, S. M. Gormer, S. Muller-Schneiders, and A. Kummert, “A decision fusion and reasoning module for a traffic sign recognition system”, IEEE Transactions on Intelligent Transportation Systems, vol.12, no.4, pp.1126–1134, 2011.

[88] G. Wang, S. Zhai, G. Lasio, B. Zhang, B. Yi, S. Chen, T. J. Macvittie, D. Metaxas, J. Zhou, and S. Zhang, “Semi-supervised segmentation of radiation-induced pulmonary fibrosis from lung ct scans with multi-scale guided dense attention”, IEEE transactions on medical imaging, vol.41, no.3, pp.531–542, 2021.

[89] W. Bai, O. Oktay, M. Sinclair, H. Suzuki, M. Rajchl, G. Tarroni, B. Glocker, A. King, P. M. Matthews, and D. Rueckert, “Semi-supervised learning for network-based cardiac mr image segmentation”, International Conference on Medical Image Computing and Computer-Assisted Intervention, pp.253–260Springer, 2017.

[90] Y. Zhang, L. Yang, J. Chen, M. Fredericksen, D. P. Hughes, and D. Z. Chen, “Deep adversarial networks for biomedical image segmentation utilizing unannotated images”, International conference on medical image computing and computer-assisted intervention, pp.408–416Springer, 2017.

[91] H. Tang, X. Liu, S. Sun, X. Yan, and X. Xie, "Recurrent mask refinement for few-shot medical image segmentation", Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.3918–3928, 2021.

[92] L. Yu, S. Wang, X. Li, C. W. Fu, and P. A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation", International Conference on Medical Image Computing and Computer-Assisted Intervention, pp.605–613Springer, 2019.

[93] P. Cascante-Bonilla, F. Tan, Y. Qi, and V. Ordonez, "Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning", Proceedings of the AAAI Conference on Artificial Intelligence, vol.35, no.8, pp.6912–6920, 2021.

[94] K. Sohn, Z. Zhang, C. L. Li, H. Zhang, C. Y. Lee, and T. Pfister, "A simple semi-supervised learning framework for object detection", arXiv preprint arXiv:2005.04757, 2020.

[95] J. Peng, G. Estrada, M. Pedersoli, and C. Desrosiers, "Deep co-training for semi-supervised image segmentation", Pattern Recognition, vol.107, p.107269, 2020.

[96] Z. Feng, Q. Zhou, Q. Gu, X. Tan, G. Cheng, X. Lu, J. Shi, and L. Ma, "Dmt: Dynamic mutual training for semi-supervised learning", Pattern Recognition, p.108777, 2022.

[97] J. Yuan, Y. Liu, C. Shen, Z. Wang, and H. Li, "A simple baseline for semi-supervised semantic segmentation with strong data augmentation", Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.8229–8238, 2021.

[98] L. Yang, W. Zhuo, L. Qi, Y. Shi, and Y. Gao, "St++: Make self-training work better for semi-supervised semantic segmentation", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.4268–4277, 2022.

[99] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks", Communications of the ACM, vol.63, no.11, pp.139–144, 2020.

[100] N. Souly, C. Spampinato, and M. Shah, "Semi supervised semantic segmentation using generative adversarial network", Proceedings of the IEEE international conference on computer vision, pp.5688–5696, 2017.

[101] W. C. Hung, Y. H. Tsai, Y. T. Liou, Y. Y. Lin, and M. H. Yang, "Adversarial learning for semi-supervised semantic segmentation", arXiv preprint arXiv:1802.07934, 2018.

[102] S. Mittal, M. Tatarchenko, and T. Brox, "Semi-supervised semantic segmentation with high-and low-level consistency", IEEE transactions on pattern analysis and machine intelligence, vol.43, no.4, pp.1369–1379, 2019.

[103] R. Mendel, L. A. d. Souza, D. Rauber, J. P. Papa, and C. Palm, "Semi-supervised segmentation based on error-correcting supervision", Proceedings of the European Conference on Computer Vision, pp.141–157Springer, 2020.

[104] G. French, T. Aila, S. Laine, M. Mackiewicz, and G. Finlayson, "Semi-supervised semantic segmentation needs strong, high-dimensional perturbations", arXiv preprint arXiv:1906.01916.

[105] J. Kim, J. Jang, and H. Park, "Structured consistency loss for semi-supervised semantic segmentation", arXiv preprint arXiv:2001.04647, 2020.

[106] Z. Ke, D. Qiu, K. Li, Q. Yan, and R. W. Lau, "Guided collaborative training for pixel-wise semi-supervised learning", Proceedings of the European Conference on Computer Vision, pp.429–445Springer, 2020.

[107] T. DeVries, and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout", arXiv preprint arXiv:1708.04552, 2017.

[108] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.2613–2622, 2021.

[109] Y. Zou, Z. Zhang, H. Zhang, C. L. Li, X. Bian, J. B. Huang, and T. Pfister, "Pseudoseg: Designing pseudo labels for semantic segmentation", arXiv preprint arXiv:2010.09713, 2020.

[110] Y. Liu, Y. Tian, Y. Chen, F. Liu, V. Belagiannis, and G. Carneiro, "Perturbed and strict mean teachers for semi-supervised semantic segmentation", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.4258–4267, 2022.

[111] Y. Ouali, C. Hudelot, and M. Tami, "Semi-supervised semantic segmentation with cross-consistency training", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.12674–12684, 2020.

[112] X. Lai, Z. Tian, L. Jiang, S. Liu, H. Zhao, L. Wang, and J. Jia, "Semi-supervised semantic segmentation with directional context-aware consistency", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.1205–1214, 2021.

[113] M. Tan, and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks", International conference on machine learning, pp.6105–6114PMLR, 2019.

[114] N. M. Jansonius, J. Schiefer, J. Nevalainen, J. Paetzold, and U. Schiefer, "A mathematical model for describing the retinal nerve fiber bundle trajectories in the human eye: average course, variability, and influence of refraction, optic disc size and optic disc position", Experimental eye research, vol.105, pp.70–78, 2012.

[115] N. M. Jansonius, J. Nevalainen, B. Selig, L. Zangwill, P. Sample, W. Budde, J. Jonas, W. A. Lagrèze, P. Airaksinen, R. Vonthein, et al., "A mathematical description of nerve fiber bundle trajectories and their variability in the human retina", Vision research, vol.49, no.17, pp.2157–2163, 2009.

[116] D. F. Garway-Heath, D. Poinoosawmy, F. W. Fitzke, and R. A. Hitchings, "Mapping the visual field to the optic disc in normal tension glaucoma eyes", Ophthalmology, vol.107, no.10, pp.1809–1815, 2000.

[117] A. Turpin, G. P. Sampson, and A. M. McKendrick, "Combining ganglion cell topology and data of patients with glaucoma to determine a structure–function map", Investigative ophthalmology & visual science, vol.50, no.7, pp.3249–3256, 2009.

[118] A. Turpin, and A. M. McKendrick, "Improving personalized structure to function mapping from optic nerve head to visual field", Translational Vision Science & Technology, vol.10, no.1, pp.19–19, 2021.

[119] G. E. Pennebaker, and W. C. Stewart, "Temporal visual field in glaucoma: a re-evaluation in the automated perimetry era", Graefe's archive for clinical and experimental ophthalmology, vol.230, no.2, pp.111–114, 1992.

[120] M. Wall, E. J. Lee, R. J. Wanzek, L. X. Chong, and A. Turpin, "Temporal wedge defects in glaucoma: structure/function correlation with threshold automated perimetry of the full visual field", Journal of glaucoma, vol.29, no.3, pp.191–197, 2020.

[121] N. Drasdo, C. L. Millican, C. R. Katholi, and C. A. Curcio, "The length of henle fibers in the human retina and a model of ganglion receptive field density in the visual field", Vision research, vol.47, no.22, pp.2901–2911, 2007.

[122] G. Montesano, G. Ometto, R. E. Hogg, L. M. Rossetti, D. F. Garway-Heath, and D. P. Crabb, "Revisiting the drasdo model: implications for structure-function analysis of the macular region", Translational Vision Science & Technology, vol.9, no.10, pp.15–15, 2020.

[123] J. Van Der Schoot, K. A. Vermeer, J. F. de Boer, and H. G. Lemij, "The effect of glaucoma on the optical attenuation coefficient of the retinal nerve fiber layer in spectral domain optical coherence tomography images", Investigative ophthalmology & visual science, vol.53, no.4, pp.2424–2430, 2012.

[124] M. Christopher, C. Bowd, A. Belghith, M. H. Goldbaum, R. N. Weinreb, M. A. Fazio, C. A. Girkin, J. M. Liebmann, and L. M. Zangwill, "Deep learning approaches predict glaucomatous visual field damage from oct optic nerve head en face images and retinal nerve fiber layer thickness maps", Ophthalmology, vol.127, no.3, pp.346–356, 2020.

[125] H. H. Yu, S. R. Maetschke, B. J. Antony, H. Ishikawa, G. Wollstein, J. S. Schuman, and R. Garnavi, "Estimating global visual field indices in glaucoma by combining macula and optic disc oct scans using 3-dimensional convolutional neural networks", Ophthalmology Glaucoma, vol.4, no.1, pp.102–112, 2021.

[126] K. Park, J. Kim, and J. Lee, "A deep learning approach to predict visual field using optical coherence tomography", PloS one, vol.15, no.7, p.e0234902, 2020.

[127] Y. Hashimoto, R. Asaoka, T. Kiwaki, H. Sugiura, S. Asano, H. Murata, Y. Fujino, M. Matsuura, A. Miki, K. Mori, et al., "Deep learning model to predict visual field in central 10 from optical coherence tomography measurement in glaucoma", British Journal of Ophthalmology, vol.105, no.4, pp.507–513, 2021.

[128] E. B. Mariottoni, S. Datta, D. Dov, A. A. Jammal, S. I. Berchuck, I. M. Tavares, L. Carin, and F. A. Medeiros, "Artificial intelligence mapping of structure to function in glaucoma", Translational vision science & technology, vol.9, no.2, pp.19–19, 2020.

[129] D. C. Hood, B. Fortune, M. A. Mavrommatis, J. Reynaud, R. Ramachandran, R. Ritch, R. B. Rosen, H. Muhammad, A. Dubra, and T. Y. Chui, "Details of glaucomatous damage are better seen on oct en face images than on oct retinal nerve fiber layer thickness maps", Investigative ophthalmology & visual science, vol.56, no.11, pp.6208–6216, 2015.

[130] J. Lee, Y. W. Kim, A. Ha, Y. K. Kim, K. H. Park, H. J. Choi, and J. W. Jeoung, "Estimating visual field loss from monoscopic optic disc photography using deep learning model", Scientific reports, vol.10, no.1, pp.1–10, 2020.

[131] P. H. Artes, A. Iwase, Y. Ohno, Y. Kitazawa, and B. C. Chauhan, "Properties of perimetric threshold estimates from full threshold, sita standard, and sita fast strategies", Investigative ophthalmology & visual science, vol.43, no.8, pp.2654–2659, 2002.

[132] D. C. Hood, and R. H. Kardon, "A framework for comparing structural and functional measures of glaucomatous damage", Progress in retinal and eye research, vol.26, no.6, pp.688–710, 2007.

[133] H. Zhu, D. P. Crabb, P. G. Schlottmann, H. G. Lemij, N. J. Reus, P. R. Healey, P. Mitchell, T. Ho, and D. F. Garway-Heath, "Predicting visual function from the measurements of retinal nerve fiber layer structure", Investigative ophthalmology & visual science, vol.51, no.11, pp.5657–5666, 2010.

[134] Z. Guo, Y. H. Kwon, K. Lee, K. Wang, A. Wahle, W. L. Alward, J. H. Fingert, D. I. Bettis, C. A. Johnson, M. K. Garvin, et al., "Optical coherence tomography analysis based prediction of humphrey 24-2 visual field thresholds in patients with glaucoma", Investigative ophthalmology & visual science, vol.58, no.10, pp.3975–3985, 2017.

[135] R. A. Russell, R. Malik, B. C. Chauhan, D. P. Crabb, and D. F. Garway-Heath, "Improved estimates of visual field progression using bayesian linear regression to integrate structural information in patients with ocular hypertension", Investigative ophthalmology & visual science, vol.53, no.6, pp.2760–2769, 2012.

[136] J. Denniss, A. M. McKendrick, and A. Turpin, "Towards patient-tailored perimetry: automated perimetry can be improved by seeding procedures with patient-specific structural information", Translational vision science & technology, vol.2, no.4, pp.3–3, 2013.

[137] G. Montesano, L. M. Rossetti, D. Allegrini, M. R. Romano, and D. P. Crabb, "Improving visual field examination of the macula using structural information", Translational vision science & technology, vol.7, no.6, pp.36–36, 2018.

[138] S. B. Ganeshrao, A. M. McKendrick, J. Denniss, and A. Turpin, "A perimetric test procedure that uses structural information", Optometry and Vision Science, vol.92, no.1, pp.70–82, 2015.

[139] H. Wässle, U. Grünert, J. Röhrenbeck, and B. B. Boycott, "Cortical magnification factor and the ganglion cell density of the primate retina", Nature, vol.341, no.6243, pp.643–646, 1989.

[140] A. Bringmann, S. Syrbe, K. Görner, J. Kacza, M. Francke, P. Wiedemann, and A. Reichenbach, "The primate fovea: structure, function and development", Progress in retinal and eye research, vol.66, pp.49–84, 2018.

[141] R. S. Maldonado, R. V. O'Connell, N. Sarin, S. F. Freedman, D. K. Wallace, C. M. Cotten, K. P. Winter, S. Stinnett, S. J. Chiu, J. A. Izatt, et al., "Dynamics of human foveal development after premature birth", Ophthalmology, vol.118, no.12, pp.2315–2325, 2011.

[142] H. J. Kuht, J. Han, G. D. Maconachie, S. E. Park, S. T. Lee, R. McLean, V. Sheth, M. Hisaund, B. Dawar, N. Sylvius, et al., "Slc38a8 mutations result in arrested retinal development with loss of cone photoreceptor specialization", Human molecular genetics, vol.29, no.18, pp.2989–3002, 2020.

[143] J. D. Akula, I. A. Arellano, E. A. Swanson, T. L. Favazza, T. S. Bowe, R. J. Munro, R. D. Ferguson, R. M. Hansen, A. Moskowitz, and A. B. Fulton, "The fovea in retinopathy of prematurity", Investigative Ophthalmology & Visual Science, vol.61, no.11, pp.28–28, 2020.

[144] V. M. Villegas, S. G. Schwartz, T. D. Hamet, C. A. McKeown, H. Capó, and H. W. Flynn Jr, "Variable clinical profile of fovea plana in normal children", Ophthalmic Surgery, Lasers and Imaging Retina, vol.49, no.4, pp.251–257, 2018.

[145] P. L. Müller, T. Treis, A. Alsaedi, A. R. Webster, P. Khaw, M. Michaelides, L. Wickham, D. Siriwardena, P. Foster, M. Moosajee, et al., "Foveal structure and visual function in nanophthalmos and posterior microphthalmos", British Journal of Ophthalmology, vol.106, no.8, pp.1164–1170, 2022.

[146] I. Matsushita, T. Nagata, T. Hayashi, K. Kimoto, T. Kubota, M. Ohji, S. Kusaka, and H. Kondo, "Foveal hypoplasia in patients with stickler syndrome", Ophthalmology, vol.124, no.6, pp.896–902, 2017.

[147] K. Hess, M. Pfau, M. W. Wintergerst, K. U. Loeffler, F. G. Holz, and P. Herrmann, "Phenotypic spectrum of the foveal configuration and foveal avascular zone in patients with alport syndrome", Investigative Ophthalmology & Visual Science, vol.61, no.2, pp.5–5, 2020.

[148] K. Pakzad-Vaezi, P. A. Keane, J. N. Cardoso, C. Egan, and A. Tufail, "Optical coherence tomography angiography of foveal hypoplasia", British Journal of Ophthalmology, vol.101, no.7, pp.985–988, 2017.

[149] E. A. Swanson, J. A. Izatt, M. R. Hee, D. Huang, C. Lin, J. Schuman, C. Puliafito, and J. G. Fujimoto, "In vivo retinal imaging by optical coherence tomography", Optics letters, vol.18, no.21, pp.1864–1866, 1993.

[150] M. R. Hee, C. A. Puliafito, C. Wong, J. S. Duker, E. Reichel, J. S. Schuman, E. A. Swanson, and J. G. Fujimoto, "Optical coherence tomography of macular holes", Ophthalmology, vol.102, no.5, pp.748–756, 1995.

[151] A. Hendrickson, D. Possin, L. Vajzovic, and C. A. Toth, "Histologic development of the human fovea from midgestation to maturity", American journal of ophthalmology, vol.154, no.5, pp.767–778, 2012.

[152] M. G. Thomas, A. Kumar, S. Mohammad, F. A. Proudlock, E. C. Engle, C. Andrews, W. M. Chan, S. Thomas, and I. Gottlob, "Structural grading of foveal hypoplasia using spectral-domain optical coherence tomography: a predictor of visual acuity?", Ophthalmology, vol.118, no.8, pp.1653–1660, 2011.

[153] A. M. Dubis, J. T. McAllister, and J. Carroll, "Reconstructing foveal pit morphology from optical coherence tomography imaging", British Journal of Ophthalmology, vol.93, no.9, pp.1223–1227, 2009.

[154] B. Nesmith, A. Gupta, T. Strange, Y. Schaal, and S. Schaal, "Mathematical analysis of the normal anatomy of the aging fovea", Investigative Ophthalmology & Visual Science, vol.55, no.9, pp.5962–5966, 2014.

[155] M. A. Zouache, G. Silvestri, W. M. Amoaku, V. Silvestri, W. C. Hubbard, C. Pappas, S. Akafo, S. Lartey, R. R. Mastey, J. Carroll, et al., "Comparison of the morphology of the foveal pit between african and caucasian populations", Translational Vision Science & Technology, vol.9, no.5, pp.24–24, 2020.

[156] M. Wagner-Schuman, A. M. Dubis, R. N. Nordgren, Y. Lei, D. Odell, H. Chiao, E. Weh, W. Fischer, Y. Sulai, A. Dubra, et al., "Race-and sex-related differences in retinal thickness and foveal pit morphology", Investigative ophthalmology & visual science, vol.52, no.1, pp.625–634, 2011.

[157] F. Dai, S. Li, Y. Wang, S. Li, J. Han, M. Li, Z. Zhang, X. Jin, and S. Dou, "Correlation between posterior staphyloma and dome-shaped macula in high myopic eyes", Retina (Philadelphia, Pa.), vol.40, no.11, p.2119, 2020.

[158] O. Lichtwitz, M. Boissonnot, M. Mercié, P. Ingrand, and N. Leveziel, "Prevalence of macular complications associated with high myopia by multimodal imaging", Journal Français d'Ophtalmologie, vol.39, no.4, pp.355–363, 2016.

[159] M. H. Errera, M. Michaelides, P. A. Keane, M. Restori, M. Paques, A. T. Moore, J. Yeoh, D. Chan, C. A. Egan, P. J. Patel, et al., "The extended clinical phenotype of dome-shaped macula", Graefe's archive for clinical and experimental ophthalmology, vol.252, no.3, pp.499–508, 2014.

[160] D. Gaucher, A. Erginay, A. Lecleire-Collet, B. Haouchine, M. Puech, S. Y. Cohen, P. Massin, and A. Gaudric, "Dome-shaped macula in eyes with myopic posterior staphyloma", American journal of ophthalmology, vol.145, no.5, pp.909–914, 2008.

[161] Y. Imamura, T. Iida, I. Maruko, S. A. Zweifel, and R. F. Spaide, "Enhanced depth imaging optical coherence tomography of the sclera in dome-shaped macula", American journal of ophthalmology, vol.151, no.2, pp.297–302, 2011.

[162] E. Ueda, M. Yasuda, K. Fujiwara, S. Hashimoto, K. Ohno-Matsui, J. Hata, T. Ishibashi, T. Ninomiya, and K. H. Sonoda, "Five-year incidence of myopic maculopathy in a general japanese population: the hisayama study", JAMA ophthalmology, vol.138, no.8, pp.887–893, 2020.

[163] C. L. Jan, and N. Congdon, "Chinese national policy initiative for the management of childhood myopia", The Lancet. Child & adolescent health, vol.2, no.12, pp.845–846, 2018.

[164] Y. N. Yan, Y. X. Wang, Y. Yang, L. Xu, J. Xu, Q. Wang, J. Y. Yang, X. Yang, W. J. Zhou, K. Ohno-Matsui, et al., "Ten-year progression of myopic maculopathy: the beijing eye study 2001–2011", Ophthalmology, vol.125, no.8, pp.1253–1263, 2018.

[165] P. L. Müller, S. Wolf, R. Dolz-Marco, A. Tafreshi, S. Schmitz-Valckenberg, and F. G. Holz, "Ophthalmic diagnostic imaging: retina", High Resolution Imaging in Microscopy and Ophthalmology, pp.87–106, 2019.

[166] S. Minami, Y. Ito, S. Ueno, K. Kataoka, J. Takeuchi, H. Ito, Y. Nakano, M. Kitagawa, C. Leahy, J. Straub, et al., "Analysis of macular curvature in normal eyes using swept-source optical coherence tomography", Japanese Journal of Ophthalmology, vol.64, no.2, pp.180–186, 2020.

[167] Y. Koyanagi, S. Ueno, Y. Ito, T. Kominami, S. Komori, M. Akiyama, Y. Murakami, Y. Ikeda, K. H. Sonoda, and H. Terasaki, "Relationship between macular curvature and common causative genes of retinitis pigmentosa in japanese patients", Investigative ophthalmology & visual science, vol.61, no.10, pp.6–6, 2020.

[168] M. Meinert, S. Ueno, S. Komori, Y. Koyanagi, A. Sayo, S. Andreasson, T. Kominami, Y. Ito, and H. Terasaki, "Longitudinal changes of macular curvature in patients with retinitis pigmentosa", Translational Vision Science & Technology, vol.9, no.10, pp.11–11, 2020.

[169] S. Komori, S. Ueno, Y. Ito, A. Sayo, M. Meinert, T. Kominami, D. Inooka, M. Kitagawa, K. Nishida, K. Takahashi, et al., "Steeper macular curvature in eyes with non-highly myopic retinitis pigmentosa", Investigative Ophthalmology & Visual Science, vol.60, no.8, pp.3135–3141, 2019.

[170] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths", IEEE transactions on Systems Science and Cybernetics, vol.4, no.2, pp.100–107, 1968.

[171] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.2881–2890, 2017.

[172] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.770–778, 2016.

[173] N. Otsu, "A threshold selection method from gray-level histograms", IEEE transactions on systems, man, and cybernetics, vol.9, no.1, pp.62–66, 1979.

[174] R. C. Team, et al., "R: A language and environment for statistical computing", , 2013.

[175] E. Mountjoy, N. M. Davies, D. Plotnikov, G. D. Smith, S. Rodriguez, C. E. Williams, J. A. Guggenheim, and D. Atan, "Education and myopia: assessing the direction of causality by mendelian randomisation", bmj, vol.361, 2018.

[176] S. Lindqvist, T. Vik, M. S. Indredavik, and A. M. Brubakk, "Visual acuity, contrast sensitivity, peripheral vision and refraction in low birthweight teenagers", Acta Ophthalmologica Scandinavica, vol.85, no.2, pp.157–164, 2007.

[177] M. Mehdizadeh, and M. H. Nowroozzadeh, "Dome-shaped macula in eyes with myopic posterior staphyloma", American journal of ophthalmology, vol.146, no.3, p.478, 2008.

[178] G. Soudier, A. Gaudric, V. Gualino, P. Massin, M. Nardin, R. Tadayoni, C. Speeg-Schatz, and D. Gaucher, "Macular choroidal thickness in myopic eyes with and without a dome-shaped macula: a case-control study", Ophthalmologica, vol.236, no.3, pp.148–153, 2016.

[179] A. A. Ellabban, A. Tsujikawa, Y. Muraoka, K. Yamashiro, A. Oishi, S. Ooto, H. Nakanishi, Y. Kuroda, M. Hata, A. Takahashi, et al., "Dome-shaped macular configuration: longitudinal changes in the sclera and choroid by swept-source optical coherence tomography over two years", American journal of ophthalmology, vol.158, no.5, pp.1062–1070, 2014.

[180] A. A. Ellabban, A. Tsujikawa, A. Matsumoto, K. Yamashiro, A. Oishi, S. Ooto, I. Nakata, Y. Akagi-Kurashige, M. Miyake, H. S. Elnahas, et al., "Three-dimensional tomographic features of dome-shaped macula by swept-source optical coherence tomography", American journal of ophthalmology, vol.155, no.2, pp.320–328, 2013.

[181] R. F. Spaide, G. J. Jaffe, D. Sarraf, K. B. Freund, S. R. Sadda, G. Staurenghi, N. K. Waheed, U. Chakravarthy, P. J. Rosenfeld, F. G. Holz, et al., "Consensus nomenclature for reporting neovascular age-related macular degeneration data: consensus on neovascular age-related macular degeneration nomenclature study group", Ophthalmology, vol.127, no.5, pp.616–636, 2020.

[182] R. Laiginhas, J. Yang, P. J. Rosenfeld, and M. Falcão, "Nonexudative macular neovascularization–a systematic review of prevalence, natural history, and recent insights from oct angiography", Ophthalmology retina, vol.4, no.7, pp.651–661, 2020.

[183] M. Shen, Q. Zhang, J. Yang, H. Zhou, Z. Chu, X. Zhou, W. Feuer, X. Jiang, Y. Shi, L. De Sisternes, et al., "Swept-source oct angiographic characteristics of treatment-naive nonexudative macular neovascularization in amd prior to exudation", Investigative ophthalmology & visual science, vol.62, no.6, pp.14–14, 2021.

[184] S. T. Bailey, O. Thaware, J. Wang, A. M. Hagag, X. Zhang, C. J. Flaxel, A. K. Lauer, T. S. Hwang, P. Lin, D. Huang, et al., "Detection of nonexudative choroidal neovascularization and progression to exudative choroidal neovascularization using oct angiography", Ophthalmology Retina, vol.3, no.8, pp.629–636, 2019.

[185] C. Narita, Z. Wu, P. J. Rosenfeld, J. Yang, C. Lyu, E. Caruso, M. McGuinness, and R. H. Guymer, "Structural oct signs suggestive of subclinical nonexudative macular neovascularization in eyes with large drusen", Ophthalmology, vol.127, no.5, pp.637–647, 2020.

[186] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., "An image is worth 16x16 words: Transformers for image recognition at scale", arXiv preprint arXiv:2010.11929, 2020.

[187] R. Strudel, R. Garcia, I. Laptev, and C. Schmid, "Segmenter: Transformer for semantic segmentation", Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.7262–7272, 2021.

[188] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation”, International Conference on Medical image computing and computer-assisted intervention, pp.234–241Springer, 2015.

[189] L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation”, arXiv preprint arXiv:1706.05587, 2017.

PUBLICATIONS AND AWARDS

Journal

[1] Kihara, Y., Montesano, G., Chen, A., Amerasinghe, N., Dimitriou, C., Jacob, A., Chabi, A., Crabb, D.P. and Lee, A.Y., "Policy-Driven, Multimodal Deep Learning for Predicting Visual Fields from the Optic Disc and OCT Imaging". Ophthalmology, vol. 129, no. 7, pp.781-791. 2022.

[2] Olvera-Barrios, A., Kihara, Y., Wu, Y., Warwick, A.N., Müller, P.L., Williams, K.M., Rudnicka, A.R., Owen, C.G., Lee, A.Y., Egan, C. and Tufail, A., "Foveal Curvature and Its Associations in UK Biobank Participants". Investigative Ophthalmology & Visual Science, vol. 63, no. 8, pp.26-26. 2022.

[3] Müller, P.L., Kihara, Y., Olvera-Barrios, A., Warwick, A.N., Egan, C., Williams, K.M., Lee, A.Y. and Tufail, A., "Quantification and Predictors of OCT-Based Macular Curvature and Dome-Shaped Configuration: Results From the UK Biobank". Investigative Ophthalmology & Visual Science, vol. 63, no. 9, pp.28-28. 2022.

[4] Kihara, Y., Shen, M., Shi, Y., Jiang, X., Wang, L., Laiginhas, R., Lyu, C., Yang, J., Liu, J., Morin, R. Lu, R., Fujiyoshi, F., Feuer, W. Gregori, G. Rosenfeld, P. Lee, A. "Detection of Nonexudative Macular Neovascularization on Structural OCT Images Using Vision Transformers". Ophthalmology Science, vol. 2, no. 4, p.100197. 2022.