A machine learning solution to predict foveal development and visual prognosis in retinal developmental disorders

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

*Master of Science in **Computer Science and Engineering** by Research*

by

GARIMA NISHAD
2019701029
garima.nishad@research.iiit.ac.in

International Institute of Information Technology
Hyderabad - 500 032, INDIA
FEBRUARY 2023

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled " A machine learning solution to predict foveal development and visual prognosis in retinal developmental disorders." by GARIMA NISHAD, has been carried out under my supervision and is not submitted elsewhere for a degree.

27-02-23

Adviser: Dr Girish Varma

To my parents and advisor,

Acknowledgments

I sincerely thank my adviser, Prof. Girish Varma, for accepting me as his graduate student at IIIT Hyderabad. He has been an exceptional mentor and a very compassionate human being. No words can express my profound thanks and admiration for him. I will forever be grateful to him for believing in me and for extending care and unwavering support through difficult times academically and personally. This opportunity led me to a new world of learning and possibilities. As an adviser and a collaborator, he inculcated within me the right mindset to pursue research by following vigour, discipline, and patience.

I also thank my collaborators at the University of Leicester: Dr Mervyn Thomas and Dr Helen Kuht. They have provided me with their sincere efforts to discuss, evaluate and provide critical feedback for my work.

Moreover, I express my gratitude towards the research community, including fellow lab-mates, researchers from other labs within IIIT, my thesis committee members, and the reviewers of my journal and conference proceedings. These people have given me valuable discussions, assessments, and critical comments on my work.

My family, of course, was significant in the completion of this thesis due to their unfailing support and encouragement. My parents have always encouraged me to follow my interests and have supported me throughout the years. In all of my pursuits, my brother has always been a pillar of support.

Abstract

We design a deep learning-based AI algorithm for analysing paediatric patients' optical coherence tomography (OCT) scans. Our system can distinguish normal from abnormal paediatric retinal OCTs in binary and six-group classification, followed by a ten-layer segmentation of retinal OCT scan and nystagmus detection by performing gaze tracking.

Our AI system successfully differentiated normal and abnormal scans with 97.68 % accuracy. Furthermore, the six-point classification system (normal, grade 1-4 FH and atypical FH) achieved a 93.54 % validation accuracy. We have demonstrated the use of AI in classifying paediatric OCT, which can help in the automated diagnosis of abnormal retinal development like Nystagmus at an early age.

To extend our work, we introduce segmentation of the ten retinal layers of the central scan. Since individual layer thinning or thickening may be markers of retinal disorders or precursors to future visual loss; hence segmenting individual layers and location-specific measuring their thicknesses is critical in therapeutic practice. By implementing the state-of-the-art DeepLabV3 model, we aim to achieve fully automated segmentation of each of the ten layers of retinal OCT scan. We use the ten carefully segmented ground truth data of the patients, including left and right eyes. We have achieved 0.834 as our mean IoU and 0.896 as our Dice Score on the best-performing DeepLabV3 ResNet-50 semantic segmentation model. Our model accurately predicts all ten layers of the retinal OCT scan, which would be utilized for calculating layer thinning or thickening in later stages of the project. We can also determine regional thickness, brightness, or texture-based indices of individual layers by retinal layer segmentation. This work on layer segmentation contributes to our understanding of retinal or optic nerve head (ONH) disease processes and is used to assess disease status, treatment responses, and visual function, among many others.

Our Gaze Tracking model helps with nystagmus waveform generation. Nystagmus is characterised by rhythmic, abnormal eye movements that begin with a "slow" movement that moves the eye away from the target and ends with a second movement that returns the eye to the target. Horizontal, vertical, torsional, or a mixture of these movements can be present. Hence to tackle such a problem with machine learning, we design an algorithm that could replace the Gold standard Eyelink II that is used to get the waveform for Nystagmus disease. Our method can properly generate such waveforms with high accuracy. This algorithm might get extended to replace the expensive machinery used in current ophthalmology labs. The video data used to test our model has been taken via mobile phones, which further assists in replacing the need for physical presence at the clinic for the diagnosis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Deep learning has advanced image recognition and analysis, allowing for remarkable performance improvements in nearly every facet. These rapid breakthroughs paved the way for the development of medical diagnostics that are automated, accurate, accessible, and cost-effective. This effectiveness is demonstrated in radiology, pathology and, recently, ophthalmology.

In this thesis, we work on a group of disorders known as Foveal hypoplasia (FH). Foveal hypoplasia is a retinal disorder in which there is a lack of full development of the morphology of the fovea [38]. It is characterised by arrested retinal development and is often associated with infantile vision condition in which the eyes make repetitive, uncontrolled movements(medical term for which is 'nystagmus'). We use optical coherence tomography (OCT) scans to identify the degree of arrested retinal development since this information provides both diagnostic and prognostic value. Recent advancements of high-resolution OCT imaging have unveiled characteristics of foveal hypoplasia that were not detected by conventional imaging methods [38].

We also aimed to design a retinal oct scan segmentation system to detect all ten retinal layer segments. Individual layer thinning or thickening can be signs of retinal disease or a prelude to future vision loss; hence segmenting individual layers and measuring their thickness is critical in therapeutic intervention. As a result, we developed a model that can do retinal layer analysis quickly and accurately, which is critical for detecting and treating retinal illnesses early.

We extend our project to include a gaze tracking algorithm. A study published in 2009 showed the prevalence of nystagmus to be 24.0 per 10,000 population. In the 18 years or younger age group, the prevalence was 16.6 per 10,000 (95% CI, $\pm$1.1) population, with the most common form of nystagmus attributed to INS associated with albinism. In the adult group, the prevalence was estimated to be 26.5 per 10,000 (95% CI, $\pm$ 6.8), with the largest nystagmus group associated with neurological disease [63]. The data show that nystagmus is more widespread than previously considered in the general population. This could have implications for resource allocation and healthcare planning.

Since there are no AI systems for paediatric OCT or childhood nystagmus currently available, we aim to create a highly efficient, automated AI system that could adequately classify, segment, and detect normal foveal structure and grades of FH.

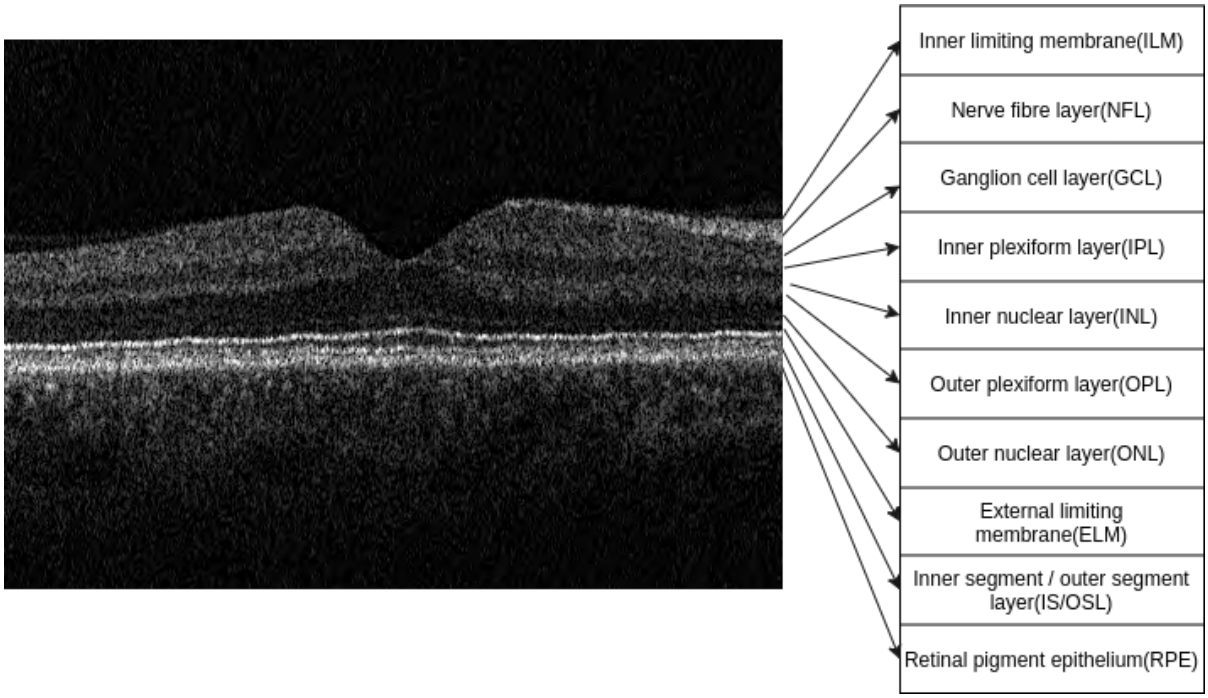


Figure 1.1: A typical OCT scan of a normal subject showing the 10 distinct layers.

1.1 Clinical Background

Human eye is a roughly spherical structure where light enters through the pupil, passes through the lens and is projected onto the back of the eye called the retina. The retina is the light-sensitive innermost layer of tissue in most vertebrates and mollusks' eyes. The optics of the eye form a focused two-dimensional representation of the visual world on the retina, which is translated into electrical nerve impulses and transmitted to the brain to provide visual perception. The retina performs a similar role as the film or image sensor in a camera. The retina of a vertebrate comprises ten layers[55].From the vitreous body's closest point to its furthest point:

1. Inner limiting membrane(ILM) – basement membrane elaborated by Müller cells.
2. Nerve fibre layer(NFL) – ganglion cell body axons (notice that between this layer and the inner limiting membrane lies a thin layer of Müller cell footplates).
3. Ganglion cell layer(GCL) – contains the nuclei of ganglion cells, whose axons become optic nerve fibres, as well as a few misplaced amacrine cells.[11]
4. Inner plexiform layer(IPL) – contains the synapse between the bipolar cell axons and the dendrites of the ganglion and amacrine cells. [11]
5. Inner nuclear layer(INL) – contains the nuclei and surrounding cell bodies (perikarya) of the amacrine cells, bipolar cells, and horizontal cells.[11]

6. Outer plexiform layer(OPL) –Rod and cone projections that culminate in the rod spherule and cone pedicle, respectively. These form synapses with bipolar and horizontal cell dendrites. This is called the Fiber layer of Henle in the macular region.

7. Outer nuclear layer(ONL) – cell bodies of rods and cones. [11]

8. External limiting membrane(ELM) – layer that separates the inner segment portions of the photoreceptors from their cell nuclei.

9. Inner segment / outer segment layer(IS/OSL) – Rods and cones have inner and exterior segments. A very specialised light-sensing system is housed in the outer segments.

10. Retinal pigment epithelium(RPE) – cuboidal epithelial cells in a single layer. The neural retina receives nourishment and support from this layer, which is closest to the choroid. The pigment layer's black pigment, melanin, reduces light reflection across the globe of the eyeball, which is critical for good vision.

The retina of an adult person is 72 percent of a sphere with a diameter of 22 mm. The optic disc, which lacks light receptors and is frequently referred to as "the blind spot," is at the retina's centre. It appears as a three mm2 oval white area. The fovea, a light-sensitive pit at the retina's centre, is responsible for our crisp central vision.

1.2 Retinal Optical Coherence Tomography (OCT) Imaging

Optical Coherence Tomography (OCT) is a comparatively recent non-invasive imaging method that uses the features of infrared light reflectance to offer a high-resolution, 3D cross-sectional picture of the tissues lining the retina. It's an optical version of ultrasonic imaging that creates cross-sectional images of the retina using low-coherence interferometry. It decodes spatial features of tissue microstructures by capturing optical scattering from the tissue. The infrared light from a super-luminescent diode is divided into two parts: one of which is reflected from a reference mirror, and the other is scattered from the biological tissue. The two reflected beams of light are made to produce interference patterns to obtain the echo time delay and their amplitude information that makes up an A-Scan(one-dimensional scans). [7]

Many one-dimensional scans (A-scans) are performed at various depths (B-scan) to get a 2D image. Those B-scans can be transformed into a retina's volumetric image (C-scan) if acquired closely and quickly. The intra-retinal tissue is a multi-layered structure that converts light into neural impulses that the brain may utilise. It is usually separated into seven layers [19] separated by eight boundaries. The boundaries ordered from top to bottom, as depicted in Fig. 1.2, are the:

1. Inner Limiting Membrane (ILM) separating the vitreous and Nerve Fiber Layer(NFL)

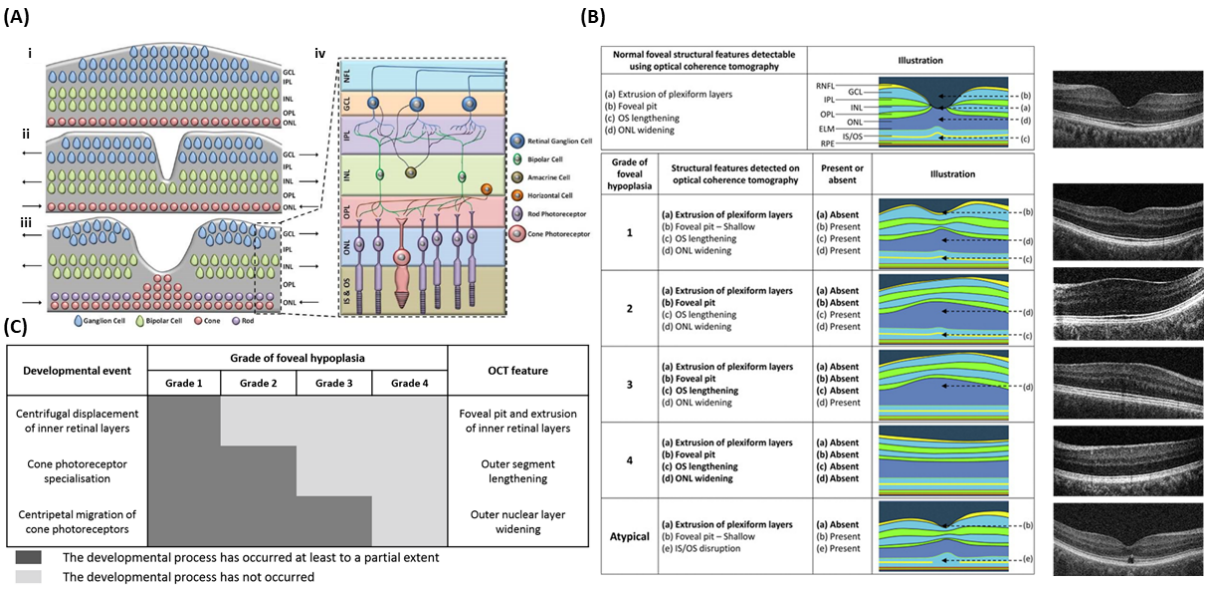


Figure 1.2: (A) and (C) Stages of foveal development and corresponding grades of foveal hypoplasia (B) A diagrammatic representation of the foveal hypoplasia grading scheme, 5 with the corresponding real OCT examples representing the respective foveal hypoplasia grades. Abbreviations: RNFL = retinal nerve fibre layer; GCL = ganglion cell layer; IPL = inner plexiform layer; INL = inner nuclear layer; OPL = outer plexiform layer; ONL = outer nuclear layer; ELM; external limiting membrane; IS = inner segment; OS = outer segment; RPE = retinal pigment epithelium. Adapted and reproduced with permission from Thomas et al. (2011)[70]

2. NFL/GCL boundary separating NFL from the Ganglion Cell and Inner Plexicon layer (GCL-IPL)
3. IPL/INL separating GCL-IPL from the Inner Nuclear Layer (INL)
4. INL/OPL separating INL from the Outer Plexiform Layer(OPL)
5. OPL/ONL separating OPL from the Outer Nuclear and Inner Segment (ONL-IS) the region
6. IS/OS separating ONL-IS from the Outer Segment (OS)
7. The Bruch's Membrane (BM) separating OS from the Retinal Pigment Epithelium Layer (RPE) layer
8. RPE out boundary separating RPE from the choroid.

Optical coherence tomography is a type of tomographic imaging similar to ultrasound. In contrast to sound waves, OCT uses broadband, low-frequency infrared light waves to obtain high-resolution images in the micrometre range in the axial direction (along the A-scans). OCT imaging is done one A-scan at a time by aiming a source laser beam at a precise location on the retinal surface. The light beam is split into two arms, one for the sample arm and the other for the reference arm, directed at the retinal tissue and a moving mirror. The energy of the interference generated by superimposing light back-scattered by the mirror (from the reference arm) and the retinal tissue is stored as the intensity of the OCT image (from the sample arm). By altering the position of the mirror in the sample arm, the whole 1D column of the A-scan can be generated. Due to the mechanical movement of the mirror, this OCT structure is known as the Time Domain OCT, and its acquisition speed is limited to roughly 400 A-scans per second. [37].

There are two types of OCT devices that are used in this project:

1. Hand-Held OCT (HH-OCT) : (Envisu 2300; Leica Microsystems, Germany;2.6-μm axial resolution)
2. Table-Mounted OCT (TM-OCT): (Optopol Technology S.A., Zawiercie, Poland;3-μm axial resolution)

1.2.1 Hand-Held OCT (HH-OCT)

With a traditional table-mounted OCT, retinal imaging in an infant can be problematic. As a result, we used Leica Envisu C2300 HH-OCT equipment to help address this unmet clinical need, transforming paediatric ophthalmology. Hand-held OCT provides unique retinal architecture data. It improves access to paediatric imaging and allows for non-invasive monitoring. It can assess the development of the paediatric eye and can be used as a diagnostic and prognostic tool.

It also avoids the need for more intrusive procedures like MRIs, ERGs (electroretinography), and genetic testing in many circumstances. The adoption of a hand-held OCT as a first-line diagnostic tool

Figure 1.3: Handheld OCT assessment (A) Capturing an image of a child's retina using HH OCT. (B) The HH OCT assessment from a patient's perspective looking into the imaging probe.

is feasible. Hand-held OCT offers a variety of clinical uses. For starters, it aids in the detection of foveal development problems. Hand-held OCT can detect each developmental event, such as the foveal pit and inner retinal layer extrusion. The use of HH-OCT has aided in developing a grading system for foveal hypoplasia and diagnostic algorithms.

Using a hand-held OCT for diagnosis and monitoring of optic nerve illnesses is also effective. In addition, the images might be provided to parents during the appointment as a reassurance scan. It is also a prognostic tool that can assist in forecasting future vision acuity and a surgical planning tool.

1.2.2 Table-Mounted OCT (TM-OCT)

Table Mounted OCT is a spectral-domain optical coherence tomography (SDOCT) tool. A single tomogram can be improved to the level of an averaged tomogram produced by repeated scanning using advanced AI algorithms. New Extended Depth Retina imaging enables deeper scans for more reliable and convenient monitoring of complex cases based on Full Range technology. This novel imaging modality is ideal for diagnosing even myopic individuals, as scans provide adequate visual detail.

1.3 Thesis focus

Detecting and grading abnormal retinal development has important diagnostic and prognostic implications in children. Hence the scope of this thesis is retinal image analysis in which the OCT is first

classified into a category of disease and non-disease. Once diagnosed with a disease, it ventures into the disease categories (Foveal Hyperplasia has six grades of disease).

Classification gives the grade of the disease; to get each of the ten individual layer thinning or thickening statuses, we perform segmentation. Since each of the ten retinal scans needs to have a thickness profile relatively constant along with the 6-mm scan through the fovea, ranging between 34 – 45 microns [4]. If there are variations in the thickness, it may be a marker of retinal disorders or precursors to future visual loss; hence segmenting individual layers and location-specific measuring of their thicknesses is critical in therapeutic practice.

Moreover, we tackle Nystagmus which is a disease in which the eyes move uncontrollably, repeating patterns. These motions can disturb balance and coordination as well as cause vision and depth perception issues. Involuntary eye motions might be circular, moving from side to side, up and down, or in a zigzag pattern. We track the eye gaze in all nine cardinal directions (up, down, left, right, all four diagonals and center).

1.4 Thesis contributions

This thesis involves contributions from the retinal eye binary and six class classification project, retinal ten layers segmentation project and Nystagmus gaze tracker project.

For the first chapter (Sec. 2.1), we aimed to construct a rapid, automated AI method to grade foveal hypoplasia in paediatric retinal OCTs accurately. We start by explaining how the retinal OCT scan data for the image classification was collected (in Sec. 2.3.1). It includes participants' information (age, gender, ethnicity), varying degrees of retinal development, i.e. grade of eye disorder (Sec. 2.2.2, 2.4.1). We thoroughly explain the tools used to annotate the data collected and how said data was augmented to increase the amount of data by adding slightly modified copies of already existing retinal data (Sec. 2.3.3). Later we provide our transfer learning-based algorithm (Sec. 2.3.4) for both the binary and six class classification with four performance metrics used to evaluate the performance of our algorithm (Sec. 2.3.7). Finally, we state two stages of experiments and results with comparisons of Convolutional Neural Networks(CNNs) and their model performance based on different configurations (Sec. 2.4). Research papers published in this chapter of the thesis are:

1. A machine learning solution to predict foveal development and visual prognosis in retinal developmental disorders.[42]

2. Using Artificial Intelligence (AI) to Classify Retinal Developmental Disorders. [39]

3. Genotypic and Phenotypic Spectrum of Foveal Hypoplasia: A Multi-centre Study. [41]

For the second chapter (Sec. 3.1), we have aimed to create a retinal OCT scan segmentation system that predicts all ten layer segments present on the retina. This chapter explains all ten individual layers of a retinal OCT scan and how thinning or thickening may be markers of retinal disorders or precursors

to the future. We used an ophthalmic image analysis tool known as OCT-Explorer for dataset collection. We explain the workings of the tool and the output obtained by it for each retinal OCT scan (Sec. 3.3.1). We created masks for each patient's data using the OCT tool output. We augmented the segmented points further to correct for possible discontinuities using data augmentation techniques (Sec. 3.3.3). We later explain our semantic segmentation algorithm that addresses the problem of segmenting layers for each retinal layer (Sec. 3.3.4). Finally, we showcased the experiments and results obtained via flattening and non-flattening of the data (Sec. 3.5).

In our final chapter of the thesis (Sec. 4.1), we aim to detect the nystagmus present in the eye by making use of the nystagmus waveform created by our algorithm. We start by collecting vertical and horizontal nystagmus patient videos using a head-mounted video-based eye-tracker known as the EyeLink II (Sec. 4.3.2). We then explain our extensive data preprocessing part, which includes the extraction of 468 face landmarks using MediaPipe after performing a video smoothing operation on the original videos (Sec. 4.3). Finally, we provide our results for each of the experiments conducted. We had three stage experiments in this project, which involved nystagmus waveform creation (Sec. 4.4.2), depth Estimation and ocular motility (Sec. 4.4.3), comparison with "Gold Standard" nystagmus waveform (Sec. 4.5).

Chapter 2

Foveal Hyperplasia classification

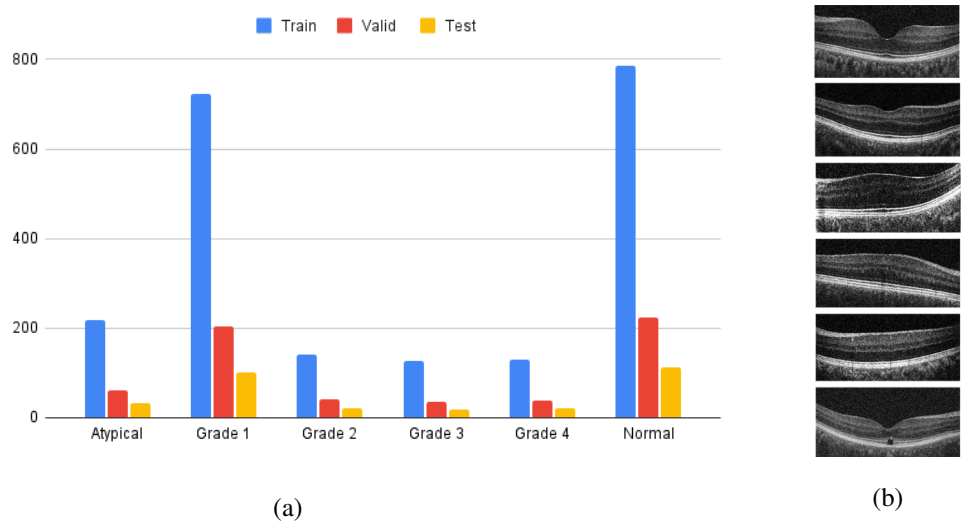
2.1 Introduction

The existence of a highly specialised retinal structure known as the fovea helps humans acquire high-acuity vision. [10] Foveal development is a long and complicated process that begins in the middle of pregnancy and lasts until the child is about 13 years old. [44, 72, 73] There are three distinct steps in normal foveal development:

1. Centrifugal displacement of cells in the inner retinal layers.
2. Centripetal migration of cone photoreceptors towards the incipient fovea.
3. Foveolar cone specialisation as shown in Fig. 1.2 (A).

Foveal hypoplasia refers to the loss of these foveal components. [70] Albinism,[57] prematurity,[30] achromatopsias,[71, 69] PAX6 mutations,[33] AHR mutations,[56] and SLC38A8 mutations are linked to foveal hypoplasia.[72, 70, 40, 58]. The foveal hypoplasia grading system is based on the phase at which retinal development stops. [70] As a result, a higher degree of foveal hypoplasia would indicate retinal developmental arrest earlier in pregnancy, but low-grade foveal hypoplasia could occur later in life. Overall, four degrees of foveal hypoplasia have been identified, and an atypical grade in which photoreceptor apoptosis causes degenerative alterations (Fig. 2.5 b). [70]. The grade of foveal hypoplasia must be correctly identified since it shows the degree of retinal development and hence has diagnostic and prognostic relevance. [70] The grading system has been shown to predict future visual acuity in preverbal infants as a prognosis tool. [61]

The ability to visualise retinal morphology in a high-resolution, non-invasive manner has advanced ocular diagnosis since the introduction of optical coherence tomography (OCT) in the science of ophthalmology. [25, 34] To appropriately diagnose the grade of foveal hypoplasia and differentiate it from normal development. However, substantial clinical experience with OCT interpretation is required. Subtle flaws, such as the subtle distortion of the inner segment ellipsoid seen in atypical foveal hypoplasia (Fig. 2.1 b), may be overlooked by human interpreters, especially by less experienced doctors, resulting in misdiagnoses.



(a) (b)

Figure 2.1: Retinal OCT Dataset. (a) The number of images present for each grade shows (in Copernicus Table Mounted OCT dataset) that the most commonly occurring grade has the highest number of images, i.e. Grade 1 and Normal. (b) Example of each grade; starting from top Atypical, Grade 1, Grade 2, Grade 3, Grade 4, Normal, respectively.

Diabetic retinopathy, macular degeneration (age-related), and optic nerve anomalies have all been treated with OCT. [23, 32]. Over 30 million OCT examinations are performed each year, vastly outnumbering all other ocular imaging equipment. [25] Despite the unsurprising popularity of OCT examinations due to their diagnostic and prognostic value,23 access to clinical professionals for OCT picture interpretation does not keep up with the ever-increasing demand. [25] This demonstrates an unmet clinical need for novel technologies to deal with the growing volume of OCT examinations and interpretations.

There have only been a few AI applications in paediatric illnesses so far. [50] .This could be because there are not large enough annotated paediatric datasets for training and validation. As a result, this group is frequently disregarded, and paediatric populations are frequently denied access to cutting-edge technology. There is an unmet clinical requirement for a system that can detect retinal developmental defects automated, accurate, and timely. Hand-held OCT equipment has made it possible to image paediatric patients' retinas, facilitating the collecting of developmental information [44]. By incorporating Deep Learning into OCT for the interpretation of macular images in paediatric ophthalmology, the clinical pathway would be augmented by minimising inter-examiner variability, improving the diagnosis of retinal developmental abnormalities, and improving time efficiency in busy clinics.

By adopting a multi-layered convolutional neural network (CNN) model to learn and recognise image attributes, we aimed to construct a rapid, automated AI method to grade foveal hypoplasia in paediatric retinal OCTs accurately. This was accomplished by first constructing a model that can distinguish

between aberrant and normal foveal structure, then using a six-point classification system to further split abnormal foveal scans into grades 1-4 and atypical foveal hypoplasia, as well as the normal fovea.

2.2 Background

2.2.1 Study design

The study was divided into two main stages, which included the development of a:

1. Device agnostic binary classification.

2. Device agnostic six-point classification.

The motivation for this was the intended clinical utility of a binary classification as a screening or referral algorithm. However, it would lack a six-point classification's diagnostic and prognostic significance. The binary classification was created to differentiate between normal and abnormal foveal developmental OCT scans. The next step was to create a six-point classification to separate the 'abnormal' dataset into grades 1-4 foveal hypoplasia and atypical foveal hypoplasia. Multiple pre-trained CNNs were evaluated and compared to find the model with the highest validation accuracy during the binary and six-point classifications. The model was trained and verified on data provided by two separate OCT models for both the binary and six-point systems.

2.2.2 Study Participants

Patients with retinal developmental problems (grade classes mentioned in Fig. 2.5 (b)) and healthy controls with no indications of aberrant retinal development were included in the study. We included all diagnoses known to be related to foveal hypoplasia based on past research. We did not sub-categorize these groups based on genetics due to their rarity [70]. Previous research has revealed that the retinal structure, regardless of subgroups of genetic diagnosis, is crucial in determining prognosis [70]. Furthermore, because there is significant evidence of dynamic retinal alterations in achromatopsia [71] and other kinds of foveal hypoplasia, we chose longitudinal datasets. [45]

Poor-quality scans with non-gradable fovea, as judged by experienced clinical graders were excluded, as subjects under one year's age were excluded. We calculated this age cut-off based on our research and existing literature because the outer retinal structures are immature, and foveal specialisation has not yet been attained for reliable grading. [44, 72, 73]

2.3 Method

2.3.1 Dataset Collection

The University of Leicester's paediatric OCT database provided scans with varying degrees of retinal development arrest. The OCT scans were collected for nine years (from 2011 to 2020). The Leicester Grading System for FH was created by Thomas et al. [69, 70, 71]. Arrested retinal development was graded using the Leicester Grading System for FH produced by Thomas et al. Normal foveal hypoplasia (FH) is categorised into four grades (1-4 FH), and atypical foveal hypoplasia [70]. All of the participants came from the University Hospitals of Leicester's paediatric and neuro-ophthalmology clinics and local schools and nurseries in the Leicestershire area. The National Research Ethics Service gave the study full ethical approval (REC reference: 10/H0406/74 and 12/EM/0261).

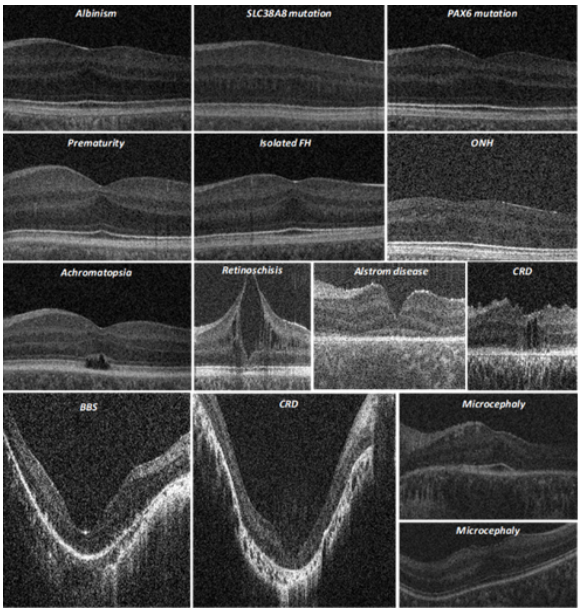


Figure 2.2: Retinal OCT Dataset. Type of mutations are linked to foveal hypoplasia.

The study followed the Declaration of Helsinki's precepts, and all participants or their guardians gave their informed consent. The Leicester OCT database obtained retinal scans representing a representative sample of various degrees of stalled retinal development.

The volumetric OCT scans were obtained using two different OCT models: the Copernicus Table-Mounted OCT (TM-OCT) (Optopol Technology S.A., Zawiercie, Poland; 3-μm axial resolution) and the Leica Hand-Held-OCT (HH-OCT) (Envisu 2300; Leica Microsystems, Germany; 2.6-μm axial resolution), both of which have been extensively used in clinical and research studies previously[44, 70]. Six OCT devices were used to collect data (1 TM-OCT and 5 HH-OCT). Scan dimensions for HH-OCT:

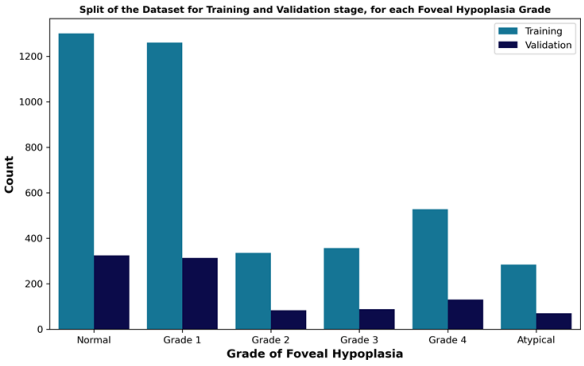
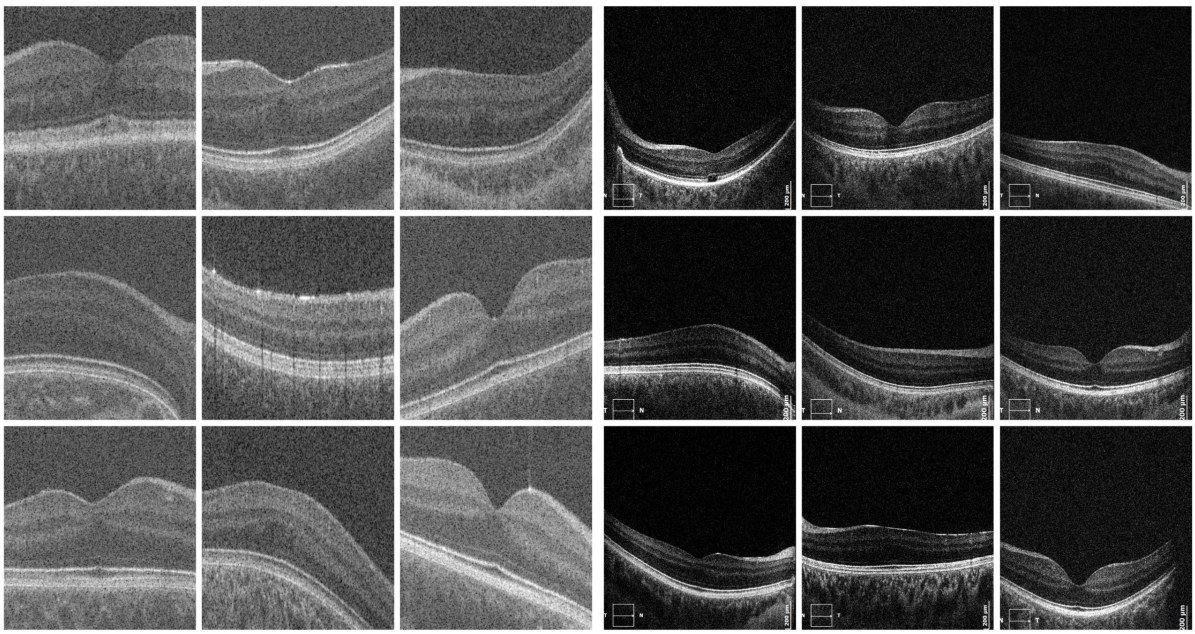


Figure 2.3: Split of the Leica Envisu HH-OCT dataset.

1. 600 A scans x 80 B scans, 12mm x 8mm
2. 500 A scans x 100 B scans, 10mm x 10mm
3. 500 A scans x 100 B scans, 10mm x 5mm

Scan dimensions for TM-OCT:



(a) (b)

Figure 2.4: Examples of HH-OCT and TM-OCT. (a) HH-OCT i.e. Leica Envisu OCT image. (b) TM-OCT image.

1. 743 A scans x 75 B scans, 7mm x 7mm

An aggregate dataset of more than 20,000 volumetric B-scans was used to gather, segment, and annotate the OCT scans.

2.3.2 Data Preprocessing

Suitable OCT scans were selected based on the following:

1. The identification of the deepest foveal pit (where present) or the widest outer nuclear layer (ONL).

2. The absence of artefacts disguising the foveal structure and (3) the absence of a significant tilt of the scan.

In this study, pre-processing was applied to optimize the brightness, contrast, and tilt of the optical coherence tomography (OCT) images, ensuring that the input images were uniform. The data was then normalized by considering the mean and standard deviation. However, it should be noted that the visual characteristics of images acquired using different OCT devices can vary significantly. The examples of the images from both HH-OCT and TM-OCT are shown in Fig.2.4. For example, images acquired

using handheld OCT (HH-OCT) tend to have low contrast, high axial resolution, and large scan size (Fig2.4 a), while images obtained using tabletop OCT (TM-OCT) tend to have high contrast, low axial resolution, and small scan size (Fig.2.4b).

To address these differences and mitigate the potential for high error rates when combining data from two unique devices, domain adaptation was utilised in this study. As the images obtained from the two datasets were distinctly different, domain adaptation was deemed an appropriate method for adjusting for these differences.

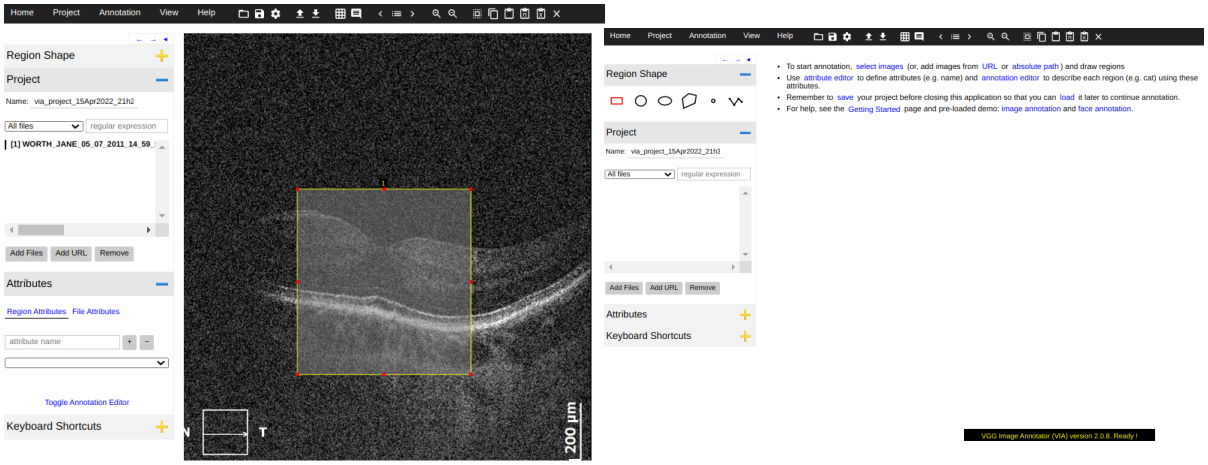
In this classification model, we employed transfer learning to make our model robust and capable of handling data from different optical coherence tomography (OCT) devices. Specifically, the model was first trained on images obtained using tabletop OCT (TM-OCT) and then fine-tuned on images acquired using handheld OCT (HH-OCT). This approach allowed us to leverage the knowledge gained from the TM-OCT data and apply it to the HH-OCT data, ultimately resulting in a model that was able to generalise well to both datasets. By using transfer learning in this manner, we were able to achieve a high level of performance on both types of OCT data, demonstrating the effectiveness of this approach in handling distinct and potentially challenging datasets.

2.3.2.1 VGG Image Annotator tool

This lightweight, independent, and offline software programme operates entirely in a web browser and requires no installation or setup. Human annotators can define and characterise spatial regions in pictures or video frames and temporal segments in audio or video using the VIA software. These handwritten annotations can be exported to plain text data formats like JSON and CSV, making them accessible to other software tools for additional processing. VIA also allows a group of human annotators to collaborate on annotating a vast dataset.

A VGG Image Annotator tool [21] was implemented to mark a bounding box around the foveal area of the HH-OCT scans during training. This was due to the standard HH-OCT scan window size often encompassing optic nerve structures, which may present confusion for the algorithm. Six JSON files were obtained for each class once a suitable bounding box had been finalised. The dataset was then cropped, using the pre-determined JSON files encompassing the annotation boxes, and subsequently inputted for training of the model.

During stage 1, we evaluated the impact of implementing a further preprocessing step of flattening the OCT pictures before inserting them into the CNN. Using custom scripts in ImageJ (version 1.48 (National Institutes of Health, Bethesda, Maryland, USA [60]), spline segmentation of Bruch's membrane was performed, followed by flattening the image (supplementary figure 1). Model accuracy using flattened images was compared to model accuracy using original images to see if the flattening preprocessing step impacts model performance. As a result, the level of image preprocessing needed to create the final binary and six-class classification systems was led by this.



(a) (b)

Figure 2.5: Screenshots of VIA running as an offline and standalone application in a web browser. (a) The main page of VIA annotation editor.(b) Spatial region extraction i.e. bounding box of region of interest in an OCT scan.

2.3.3 Data Augmentation

To ensure that the model was trained on a balanced dataset, we used data augmentation techniques to increase the dataset's variability. The following six data enhancement procedures were used: (1) brightness modifications, (2) contrast adjustments, (3) jitter adjustments, (4) zoom function (10.00 per cent and 20.00 per cent), (5) vertical image flipping, and (6) random rotation between -10° and +10° are all available. Similar data augmentation strategies have been developed and deployed effectively using AI, and OCT. Fig. 2.6 depicts some of the data augmentation strategies that have been used. During the model's training stage, we implemented data augmentation techniques.

2.3.4 Algorithm

We used a transfer learning-based technique (Fig. 2.7). Transfer learning takes a previously trained deep learning model and tweaks it to fit a new dataset [43]. The pre-existing model we chose was ImageNet. We replaced the 1000-neuron fully connected (FC) with 2-neuron FC for normal and abnormal classification and a 6-neuron FC for the other grades to fine-tune ImageNet to suit our dataset (normal, grade 1, grade 2, grade 3, grade 4 and atypical). We tweaked the transfer learning process to improve the algorithm's performance by allocating different layers to distinct learning parameters (frozen and newly developed).

We tested different CNNs on the same dataset to see which one produced the best final validation accuracy. We examined five different CNN models to get the highest performance for binary and six-point classification:

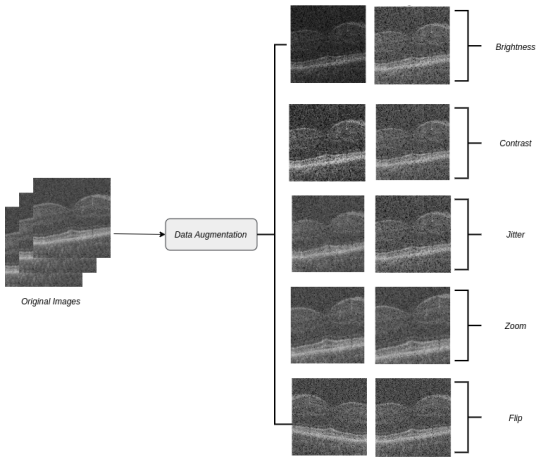


Figure 2.6: Data Augmentation. Input images were augmented via (starting from top) brightness adjustments, contrast adjustments, jitter adjustments, zoom function (10.00% and 20.00%), flipping of the image.

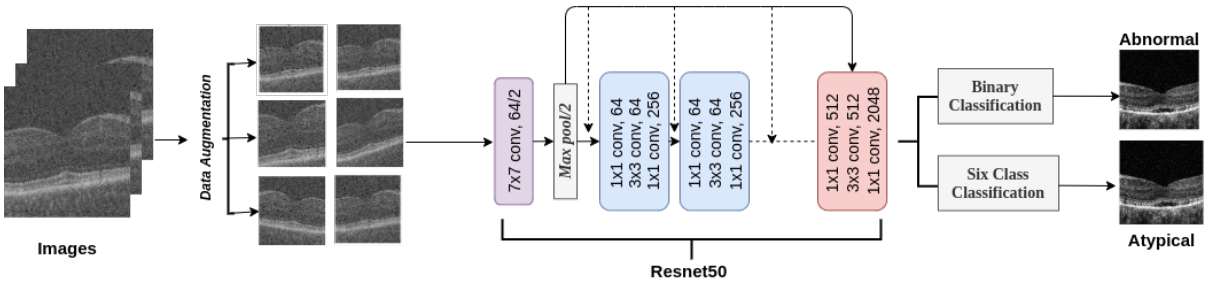


Figure 2.7: Outline of the proposed system.

1. ResNet-18
2. ResNet-34
3. ResNet-50
4. ResNet-101
5. DenseNet-121

A five-fold cross-validation process was utilised to reduce variation, and the validation accuracy of each CNN was averaged. A re-sampling approach called cross-validation is used to test machine learning models on a small sample of data. The highest-performing model was further optimised by adjusting the configurable hyperparameters (such as the learning rate) to get the highest possible accuracy. This final model would serve as the foundation for the binary and six-point models.

2.3.5 Training and Validation

We split our dataset into two groups for binary and six-point classification: (1) training (80.00 %), and (2) validation (20.00 %). This dataset split was done for each specific grade of foveal hypoplasia to ensure that each grade was given the exact weighting in the training and validation groups (figure 6).

To create a device-agnostic classification system, we ran tests evaluating the model's performance when trained and verified on various dataset setups. We looked at the model's performance in various training and validation dataset setups. The following nine configurations were tested:

1. Trained on HH-OCT, validated on HH-OCT
2. Trained on HH-OCT, validated on TM-OCT
3. Trained on HH-OCT, validated on combined (TM-OCT + HH-OCT)
4. Trained on TM-OCT, validated on TM-OCT
5. Trained on TM-OCT, validated on HH-OCT
6. Trained on TM-OCT, validated on combined(Representing the final model for binary and six-point classification)

2.3.6 AI vs Human grading of Retinal OCT scan

We compared the grading performance of clinical graders with varied years of experience in paediatric OCT interpretation (10 years to 1 year) to the grading performance of the unique AI system. This research used seven graders who were not aware of the clinical diagnosis. The true positive rate and false-positive rate for each grader were calculated using the validation dataset from configuration 9 (above) and plotted against the receiver operating characteristic curve for the AI-systems diagnostic accuracy.

2.3.7 Statistical Analysis/Performance Metrics

To measure the accuracy of the developed model, we derived various performance metrics. These included: (1) sensitivity, (2) specificity and (3) accuracy as shown in the equation below, where TP = number of true positives, FP = number of false positives, TN = number of true negatives and FN = number of false negatives.

$$Sensitivity = \frac{TP}{TP + FN} \tag{2.1}$$

$$Specificity = \frac{TN}{TN + FP} \tag{2.2}$$

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{2.3}$$

The metrics above are empirically computed on a test set. We also report 95% binomial confidence intervals using the following formula:

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p} - \hat{p}^2}{n}} \tag{2.4}$$

where p is the empirical value of the metric and n is the size of the test set.

$$HitRate = \frac{TP}{P} \tag{2.5}$$

$$FalseAlarmRate = \frac{FP}{N} \tag{2.6}$$

2.3.8 Grad-CAM

We used Gradient-weighted Class Activation Mapping (Grad-CAM) to test model interpretability [17]. Grad-CAM visually checks where the CNN is looking, ensuring that the model was built to classify significant patterns in the image [17]. Grad-CAM was used to:

1. Detect and identify the network's last convolution layer.

2. Assess the gradient information flowing into this layer after our model has been constructed and modified. Grad-CAM generates a heat map that may be placed on the foveal tomogram to show which portions of the picture the CNN used for categorization.

2.3.9 Area under the ROC curve

A receiver operating characteristic (ROC curve) is a graph that shows how well a classification model performs across all classification thresholds."Area under the ROC Curve" is the abbreviation for "Area under the ROC Curve." AUC, in other words, assesses the full two-dimensional area beneath the entire ROC curve. AUC is desirable for the following two reasons:

1. AUC is scale-invariant. It measures how well predictions are ranked rather than their absolute values.

2. AUC is classification-threshold-invariant. It measures the quality of the model's predictions irrespective of what classification threshold is chosen.

As mentioned in Sec. 2.3.6, we calculated the ROC curve for each of the seven specialists based on how well they graded the OCT scans.

2.4 Experiments and Results

2.4.1 Dataset characteristics

OCT data was collected from a total of 707 individuals (female: 50.50%, male: 48.23% and undisclosed: 1.27%). The age range of the participants was between 1 year and 80 years (mean ± standard deviation: 31.13 years ± 17.50 years). Individuals were from a range of ethnic backgrounds: White (79.92%), Asian or Asian British (15.28%), Black, African, Caribbean or Black British (0.71%), Other (0.57%) and undisclosed (1.41%). The dataset included a representative sample of varying degrees of retinal development, with a total of 5078 OCT images (HH-OCT: 2041 (40.19%), TM-OCT: 3037 (59.81%)). Within the total dataset, 1626 (32.02%) images demonstrated normal fovea, 1575 (31.02%) grade 1 foveal hypoplasia, 419 (8.26%) grade 2 foveal hypoplasia, 445 (8.76%) grade 3 foveal hypoplasia, 659 (12.98%) grade 4 foveal hypoplasia and 354 (6.97%) atypical foveal hypoplasia.

2.4.2 Stage 1: Flattened vs Non-flattened/original dataset

In Stage 1, we trained with both flattened and original dataset images. The findings stated that the validation accuracy of the system, training with flattened and original data, was 96.70% and 96.44%, respectively, on the test dataset. This experiment demonstrated no significant improvement in performance associated with flattening OCT scans as a pre-processing step. We, therefore, continued to train and validate the algorithms on original data without the pre-processing step of flattening.

2.4.3 Stage 2: Binary and Six-Point Classification System

We achieved a device-agnostic, binary classification to differentiate between normal and abnormal fovea structure with a final validation accuracy of 98.10% with ResNet-50, with 99.13% sensitivity and 95.69% specificity. Following the effectiveness of the binary classification, we expanded it from two to six categories, which included normal fovea, grade 1-4 foveal hypoplasia, and atypical foveal hypoplasia. With ResNet-50, this device-independent, six-point classification achieved a final validation accuracy of 95.00 %. Furthermore, the method distinguished normal fovea from all other grades of foveal hypoplasia with a sensitivity of 97.85 % and a specificity of 98.84 % in the validation dataset.

For all degrees of foveal hypoplasia, sensitivity values were consistently high (range between 88.10 % - 100.00 %). The method was least sensitive to grade 2 and grade 3 foveal hypoplasia, with a sensitivity of 88.09 % and 88.76 %, respectively. All categories were quite specific, ranging from 98.49 % to 100.00 %. We plotted the GRAD CAM to find out the ROI present in the image. These GRAD CAM images were manually checked by senior specialists and verified to be true in it's accuracy for finding out the region of interest.

2.4.4 Binomial proportion confidence interval

A binomial proportion confidence interval is a range of values that is likely to contain the true probability of success based on a given number of success-failure experiments (also known as Bernoulli trials). It is calculated using the number of experiments performed (n) and the number of successful outcomes (nS). This type of confidence interval is used to estimate the probability of success in a situation where only the number of experiments and the number of successes are known.

It is a range of values within which the true probability is likely to fall based on a given level of confidence. This interval is calculated using the number of times the event occurred in a series of trials and the total number of trials. The confidence level specifies how likely it is that the interval contains the true proportion. For example, a 95% confidence interval means that there is a 95% probability that the interval contains the true proportion.

Section 2.4.4 in the thesis explains in detail how to calculate the confidence interval (C.I.). To do this, we plotted a hit rate versus false alarm plot. For the hit rate confidence interval, we set n = p and $\hat{p}$ = TP/P. For the false alarm rate C.I., we considered n = N and $\hat{p}$= TP/N. Here, P is the number of positive ground truth images, and N is the number of negative ground truth images. These correspond to the blue-shaded region shown in Figure 2.10(B).

2.4.5 Comparisons of Convolutional Neural Networks

Compared to the other CNNs tested, ResNet-50 achieved the highest validation accuracy (93.54 %). ResNet-18 had the lowest validation accuracy (88.90 %). Configuring the learning rate to 1e-2 increased the ResNet-50 validation accuracy even more. This was considered the best-configured model for our dataset (Refer to Table. 2.1). ResNet has been shown to yield good validation accuracy values in image recognition tasks in the past, which influenced our decision to test multiple models of this CNN [31].

2.4.6 Comparing Model Performance Based on Different Configurations

Validation accuracy results demonstrated that each six-point classification model performed better when using the device-specific dataset that it was trained on to validate the model (Table 2.2 and 2.3). For example, training the model using HH-OCT data and validating with HH-OCT data achieved higher accuracy (97.05%) than training with HH-OCT and validating on the TM-OCT dataset (8.11%) or com-

Model	Accuracy
Resnet18	88.9
Resnet34	90.2
Resnet50	93.54
Resnet101	91.00
Densenet121	93.38

Table 2.1: Comparisons of CNN models.

Dataset	TM-OCT	HH-OCT	Combined
TM-OCT	97.35	93.08	-
HH-OCT	24.42	99.26	43
Combined	97.68	98.28	98.12

Table 2.2: Comparing Binary Classification Model Performance at the validation stage. The validation accuracy findings showed that each six-point classification model performed better when validated using the device-specific dataset that it was trained on. Drop-down values are the base models; we can use them to get exact accuracy while validating the model.

Dataset	TM-OCT	HH-OCT	Combined
TM-OCT	93.54	33.94	-
HH-OCT	8.11	97.05	22.6
Combined	94.53	95.09	94.96

Table 2.3: Comparing Six-point Classification Model Performance at the validation stage.

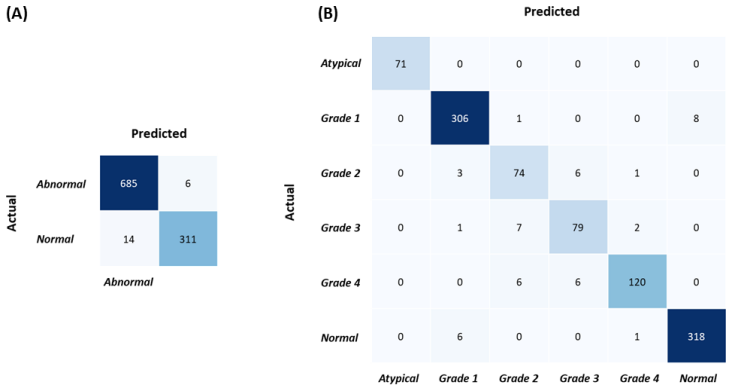


Figure 2.8: Confusion Matrices. (a)Binary classification.(b) Six-point classification.

bined dataset (22.60%). Interestingly, the model that was trained on the combined dataset generates consistently high validation accuracies when validated on all datasets; TM-OCT only (94.53%), HH-OCT only (95.09%) and combined (94.96%). Moreover, the model trained using the combined dataset produces higher accuracy when validating on TM-OCT data (94.53%) than the model trained and validated on TM-OCT only (93.54%). The highest validation accuracy for a six-point classification system was observed on the algorithm trained and tested with HH-OCT scans (97.05%). A table of comparisons for each configuration can be found in Table 2.2 and Table 2.3.

2.4.7 Actual vs. Predicted

Confusion matrices were generated to demonstrate the predicted grades of foveal hypoplasia produced from the AI system compared with the actual grades interpreted by experienced clinicians. Following the training of the algorithm and adjustments of parameters, the final validation accuracies we achieved surpassed the validation accuracies achieved by the human interpreters.

An example of heatmaps generated from Grad-CAM experimentation can be viewed in Fig. 2.9.

2.5 Conclusion

We have created an intelligent diagnostic system that uses OCT scans to distinguish between normal and pathological retinal development for the first time. Our proof-of-concept study with a ResNet-50 model demonstrated high sensitivity and specificity for a binary and six-point classification after testing and comparing different CNNs. Our deep learning system was trained and verified on imaging data from various OCT devices, resulting in a device-independent system. Subclassifying the stage of aberrant retinal development depending on the grade of foveal hypoplasia (six-point classification) also gives mechanistic insight into the retinal developmental event that has been stopped, whether it happened in utero or after birth. As a result of our system's excellent sensitivity and specificity in identifying degrees of foveal hypoplasia, it has the potential to provide a significant diagnostic utility.

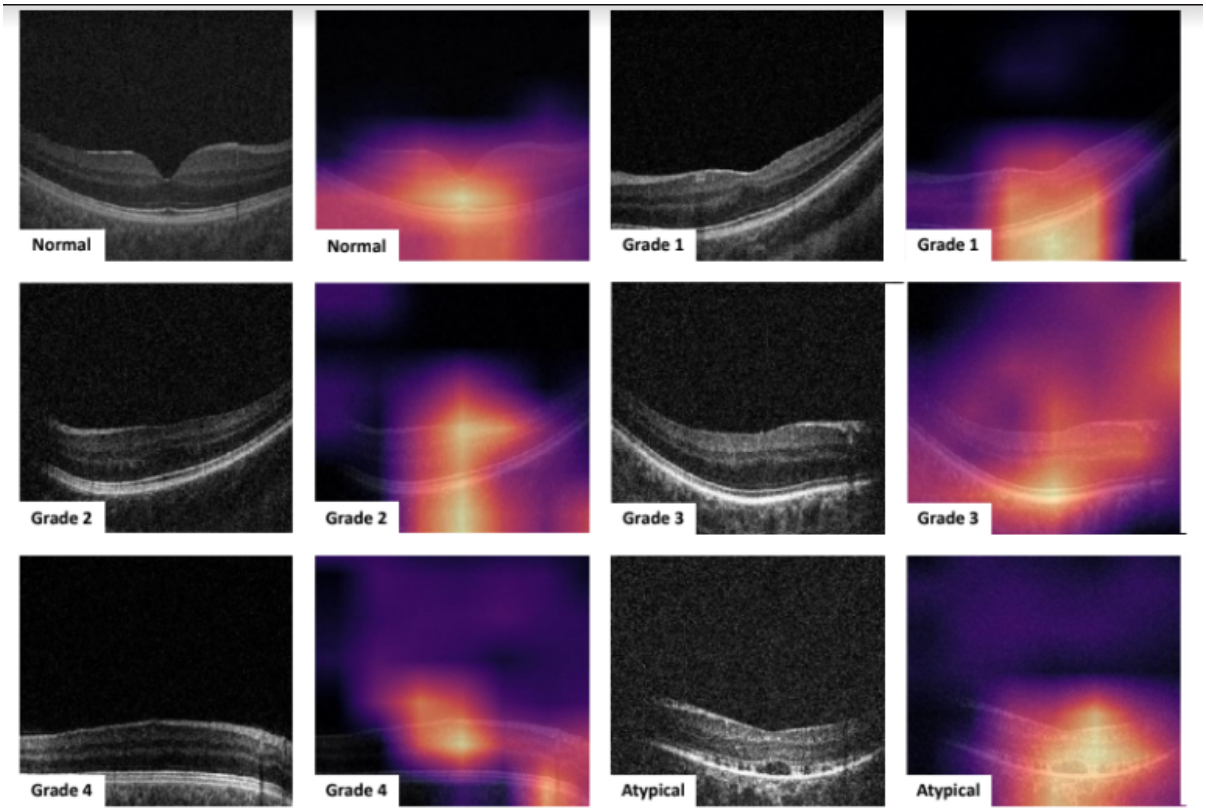


Figure 2.9: Gradient-weighted Class Activation Mapping (Grad-CAM). As shown in image it produces a coarse localization map highlighting the ROI(regions of interest). Resulting heatmaps generated from Grad-CAM superimposed onto the original foveal tomograms. Examples of all grades of foveal hypoplasia and normal foveal morphology.

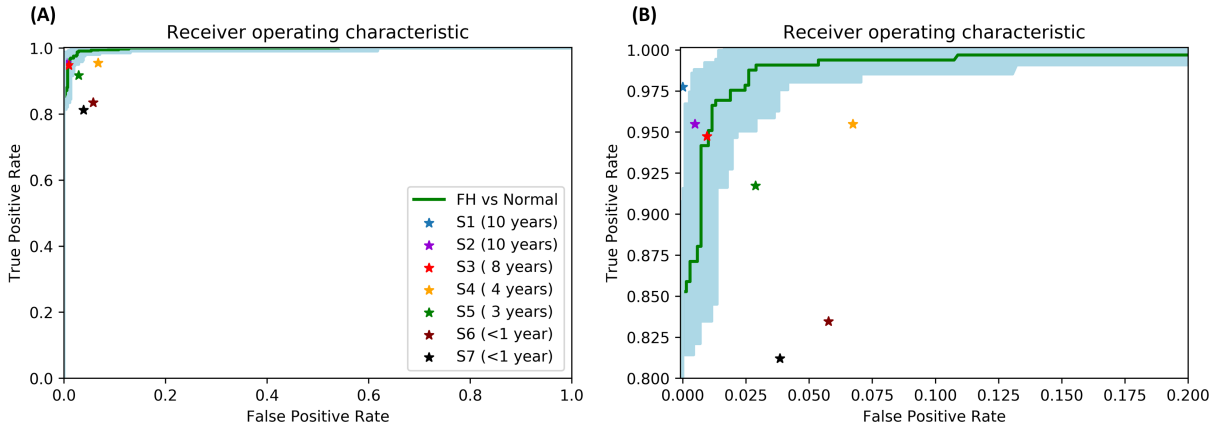


Figure 2.10: ROC curve. (A) Performance of the AI-system demonstrated with receiver operating characteristic (ROC) curve for foveal hypoplasia versus normal classification. 95% confidence interval shown with the light blue shaded area. Stars represent the performance of clinician graders of varying years of experience (grading was performed blinded to the clinical diagnosis). (B) zoomed-in version of ROC curve. .

This could eliminate the need for comprehensive testing and offer vital information for patients to be referred to ocular genetics for specific genetic testing, resulting in an earlier diagnosis. Furthermore, we believe that our system's capacity to identify the grade of foveal hypoplasia correctly will be helpful in the future in predicting future visual acuity in preverbal children [61].

Only table-mounted versions of OCT were available when it was initially debuted in 1991, making it generally only accessible for the examination of an adult or a cooperative youngster. [7] This was the case until 2010 when a hand-held spectral-domain OCT suitable for assessing newborns and young children was created and optimised. [46, 54] The gap between the development of adult DL systems and corresponding paediatric intelligent systems could partly explain the time it took for TM-OCT and HH-OCT to be released, resulting in smaller paediatric image datasets.

Furthermore, compared to the relatively static character of adult settings, paediatric retinal developing conditions are dynamic. As a result, establishing a paediatric intelligent diagnostic system to manage the spectrum of illnesses depending on the stage of stopped development poses considerable difficulty. Similar restrictions are connected with the lack of adequate adult child datasets in other specialities, such as cancer, where similar limits are associated with the availability of AI systems for adult populations compared to juvenile populations. [20] Due to a young child's or infant's incapacity to communicate effectively, paediatric patient's investigations often rely on more objective measurements than adult patients.

This highlights the critical need to prioritising the development of intelligent paediatric systems to supplement the clinical pathway of objective assessments in children. The Leicester OCT database includes a sizeable paediatric dataset (¿20,000 foveal tomograms) and adult data, allowing for the con-

struction of a system that can distinguish between retinal development problems in both children and adults.

Despite our model's demonstrated outstanding performance, our research has significant drawbacks. Only retrospective data was used to train and evaluate our model. Normative datasets were collected from schools, nurseries, and hospitals in the Midlands region of the United Kingdom. The abnormal scans also included data from paediatric and neuro-ophthalmology clinics in the same region. OCT collection from clinics across nine years resulted in an imbalance in class distribution among the grades. As a result, more validation studies with external datasets and prospective algorithm testing are vital future studies to consider prior to adoption. The algorithm's performance measures mostly showed good sensitivity and specificity; however, the sensitivity for grade 2 and grade 3 foveal hypoplasia was lower than for the other classes (supplementary figure 2 and supplementary table 1). Even though our DL system misclassified some grade 3 foveal hypoplasia images, it consistently identified these scans as abnormal and categorised them as grade 2 most of the time. One probable explanation is that grade 2 (n=419) and grade 3 (n=445) foveal hypoplasia scans are given a lower weighting than other categories, resulting in higher mistake rates.

The anatomical changes between grade 2 and grade 3 fovea hypoplasia are often modest, with only the existence of OS lengthening serving as a distinguishing factor. As a result, grade 2 and 3 foveal hypoplasia are challenging to distinguish and are frequently misinterpreted by human interpreters. As a result, the results of our built model represent the difficulties that human interpreters face when identifying grade 2 and 3 foveal hypoplasia. Misclassifying grade 3 foveal hypoplasia as grade 2 could lead to false assurances about future visual acuity. This emphasizes the significance of prospective human-interpreter reinforcement alongside the intelligent system, especially in grade 3 foveal hypoplasia, where an expert physician may be necessary to validate the prediction.

Further research should be considered to develop a more quantitative measure of foveal morphological features to eliminate or reduce misclassification errors associated with grade 2 and 3 foveal hypoplasia. Furthermore, the binary classification system had a high sensitivity (99.13%) and specificity (95.69%), indicating that our intelligent system might be employed as a screening tool to distinguish between normal and pathological fovea.

It is not easy to determine how DL approaches to arrive at a judgement or categorization [67]. Because physicians, medical policymakers, and patients value evidence-based management, the black-box nature of DL methods continues to be a source of worry. Techniques, such as Grad-CAM ways to generate visual explanations with feature maps of how the DL system learns, are being proposed to make neural network models easier to read. We used the Grad-CAM method, which provided comforting proof of which areas of the image the DL system was used to classify it.

Finally, this proof-of-concept study reveals the first use of AI in distinguishing between normal and pathological retinal development, with an application in paediatric ophthalmology to detect the presence and severity of foveal hypoplasia from OCT images. We showcase that this has the potential to be a more time-efficient and accurate method of detecting foveal developmental abnormalities. As

the number of doctors with skills to interpret paediatric retinal imaging scans continues to grow, this approach could facilitate virtual clinics led by junior allied health professionals and trainee clinicians to meet the growing demand. Our model has a considerable impact as a front-line diagnostic tool in paediatric ophthalmology, displaying the ability to empower physicians and determine downstream investigations.

2.5.1 Code

The GitHub link to the code can be found at https://github.com/Garima13a/Foveal-Hyperplasia-classification. The repository includes a detailed README file that provides step-by-step instructions on how to run the code, as well as any necessary dependencies and configurations.

Chapter 3

Retinal OCT Scan segmentation

3.1 Introduction

Ophthalmologists frequently identify fundus retinal disorders, with retinopathy being the most common cause. Fundus illnesses such as age-related macular degeneration (AMD), diabetic retinopathy (DR), and central serous chorioretinopathy affect more than 300 million people globally. Age-related macular degeneration (AMD) is the most common cause of visual loss in the U.S. and is a growing public health problem. Currently, almost 7.3 million Americans (6.12% of Americans aged 40 years and older) have some form of AMD, and AMD is the cause of blindness for 54% of all legally blind Americans [24]. India has about 77 million people at or above the age of 60 years representing a large group vulnerable to vision-related disorders, and the number is estimated to reach 180 million by 2026. As reported in population-based studies, the prevalence of AMD in India ranges from 39.5% to 0.3% [66].

Human eye is a roughly spherical structure where light enters through the pupil, passes through the lens and is projected onto the back of the eye called the retina. The retina of a vertebrate comprises ten layers[55] and the thickness of a layer is measured by the distance between the target layer and the layer above or below it. From the vitreous body's closest point to its furthest point:

1. Inner limiting membrane(ILM)
2. Nerve fibre layer(NFL)
3. Ganglion cell layer(GCL)
4. Inner plexiform layer(IPL)
5. Inner nuclear layer(INL)
6. Outer plexiform layer(OPL)
7. Outer nuclear layer(ONL)
8. External limiting membrane(ELM)
9. Retinal pigment epithelium(RPE)

Individual layer thinning or thickening may be markers of retinal disorders or precursors to future visual loss; hence segmenting individual layers and location-specific measuring their thicknesses is critical in therapeutic practice. According to studies, in most fundus illnesses, retinal morphological

alterations appear before visual changes. The use of precise and sensitive technologies to analyse the retinal morphological structure will aid in the early detection of fundus retinal disorders. Different retina layers are affected differently by various ophthalmic and neurological disorders. Diabetic macular edoema (DME), for example, is characterised by fluid accumulation within the retina; neovascular AMD is characterised by fluid beneath the retina; and multiple sclerosis is linked to changes in the RNFL thickness. The ability to accurately evaluate retinal structure in retinal OCT, particularly the need for accurate OCT segmentation, is highlighted by these numerous disorders and their impact on the retina.

Manual OCT segmentation takes a lot of time and effort. As a result, autonomous, dependable, and precise OCT segmentation is critical for extending the utility of OCT technology. Several methods for measuring retinal thickness from OCT images have recently been described, but they are not extracting the thickness of all the layer segments. [15] The analysis of the optic nerve head geometry was described using a method based on an extension of the Markov model. [8]

As a result, we require a model that can quickly and accurately perform retinal layer analysis, which is critical for early identification and treatment of retinal illnesses. Since retinal fundus pictures have been increasingly popular in diagnosing, screening, and treating retinal disorders in recent years, we aim to automate retinal layer segmentation using retinal optical coherence tomography (OCT) images. Our model can aid in the diagnosis and monitoring of retinal disorders by extracting the retinal layer thickness.

3.2 Clinical Background

Ophthalmic Coherence Tomography (OCT) is a widely used non-invasive medical imaging technique for high-resolution scanning of ocular tissue. In particular, it is useful for examining the fundus nerve tissue, which is important for the diagnosis and treatment of retinal diseases. However, the analysis of retinal OCT images can be challenging due to the presence of speckle noise and low contrast between adjacent structures.

In recent years, there has been an increasing interest in using computer-aided methods for the segmentation of retinal OCT images. This is because automatic segmentation of the retina is considered a critical and challenging step in the development of computer-aided diagnostic systems for ocular diseases. The difficulty in accurately segmenting retinal layers is due to the complexity of retinal OCT images and the limited resolution of the OCT scanning system [65]. Additionally, retinal diseases can cause severe deformation of the retinal layer, further complicating the segmentation process. [29, 12]

Although many algorithms for segmenting the retinal layer have been developed, they still have limitations, and some layers may not be segmented. The existing methods for retinal OCT image segmentation mainly include active contour [26], classifiers, three-dimensional map search and deep learning techniques. Despite the progress that has been made, the problem of accurately and completely segmenting retinal layers in OCT images remains an ongoing challenge.

In one study, authors proposed the use of active contours for the segmentation of retinal OCT images. They used a combination of continuous curves and energy functions to convert the segmentation process into finding the minimum value of the energy function. The contour that corresponds to the minimum energy value is considered the boundary of the layer. However, this method may be computationally expensive and sensitive to the starting position of the contour. [26]

Another research paper [2] presented a classifier-based method for the segmentation of retinal OCT images that uses a combination of Support Vector Machine (SVM) [27] and Fuzzy C-Means clustering [47]. The SVM-based segmentation classifies image pixels in a feature space, and the Fuzzy C-Means clustering-based segmentation groups pixels into distinct regions, thus achieving automatic segmentation of the retinal OCT images. With the ongoing optimization of the algorithm, [3] the segmentation accuracy of this method has been improved to within 2 pixels.

There is also a graph theory-based method for the segmentation of retinal OCT images, which involves dividing the image into nine layers by removing specific edges from the graph [35]. Graph-based methods such as GraphCut [9], GrabCut [59], and RandomWalk [64] were used in this approach. While this method can achieve a high level of accuracy (1 pixel), it is sensitive to noise and image degradation as it relies on treating pixels as nodes. [68, 49]

Fang et al. [22] proposed a deep learning-based method that combines a convolutional neural network (CNN) and graph search for the segmentation of retinal layers in OCT images. CNN is used to extract features of specific layer boundaries, and graph search methods are applied to the resulting probability maps to obtain the final boundaries. This method can automate the segmentation of nine-layer boundaries of the retina to a certain extent, but the computational load is high, and the method can be sensitive to the position of the boundaries.

Another proposed method combines a Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) to extract layers of interest, extract edges, and trace the layer boundary, respectively. The model was trained on a dataset that includes both normal and AMD cases with minimal data. Evaluation of the proposed model on three public datasets showed that the error is lower than the inter-marker error of existing methods, and the performance is comparable to the current state-of-the-art techniques. [28]

In another paper, authors present a new approach to using atrous convolution for semantic image segmentation. [16] Atrous convolution, also known as dilated convolution, is a technique that allows for increased resolution in image feature maps without increasing the number of parameters. The authors propose a modified version of atrous convolution that uses multiple atrous rates in a single layer and a global context module to aggregate contextual information from the entire image. The authors evaluate their proposed method on several popular semantic segmentation benchmarks and show that it achieves competitive results compared to state-of-the-art methods. The paper also shows the advantage of using Atrous convolution for semantic image segmentation, as it helps in increasing the resolution of feature maps by using multiple atrous rates and also aggregating contextual information through a global context

module. Based on this model, we carried out experiments and proposed a new method for the automatic segmentation of retinal layers in retinal OCT images.

3.3 Method

3.3.1 OCT-Explorer

OCT Explorer is an ophthalmic image analysis tool by The Iowa Institute for Biomedical Imaging Research for segmenting the layers of an OCT scan of the retina. It uses LOGISMOS (Layed Optimal Graph Image Segmentation for Multiple Objects, and Surfaces [48] multi-surface multi-object algorithm is very well suited to the task. LOGISMOS simultaneously detects multiple interacting surfaces and allows for creating surface- and layer-specific cost functions that reflect both surfaces- and regional (layer-specific) information. The resulting surface shapes can be influenced by regionally varying surface smoothness and shape-preference constraints, and both the cost functions and constraints can be learned automatically from segmentation examples. The LOGISMOS method is utilised to segment 11 intra-retinal layers simultaneously and computationally efficiently (12 surfaces). Fig. 3.1 shows the OCT-Explorer environment.

3.3.2 Dataset collection

Our initial experimentation started with Farsiu Dataset [18]. We took the raw images provided with the masks as our base dataset to conduct initial experimentation for the segmentation. The dataset contained images from ten patients, with each having 61 images of the retinal OCT scan and their corresponding feature masks. This dataset lacked proper grade information for each image. Hence, it was discarded.

We hence obtained the annotated dataset from the OCT-Explorer tool (Sec. 3.3.1) with automated segmentation of retinal layers (RNFL retinal nerve fiber layer, GCL ganglion cell layer, IPL inner plexiform layer, INL inner nuclear layer, OPL outer plexiform layer, ONL outer nuclear layer, ELM external limiting membrane below which lie the photoreceptor layers). This dataset was provided to us by the University of Leicester's Ophthalmology department. We had ten patients' retinal scans, which included both left and right eye data separately.

The data consisted of 75 b-scans/eye per patient. For each patient, ten ".dat" format files were present. Each .dat file consists of the y coordinate of the segmented layer, assuming that the x-axis is image width in pixels. In our case, the image width was fixed to 743 pixels. Hence, for each image with the segmented layer on top; we created masks for each of the scans.

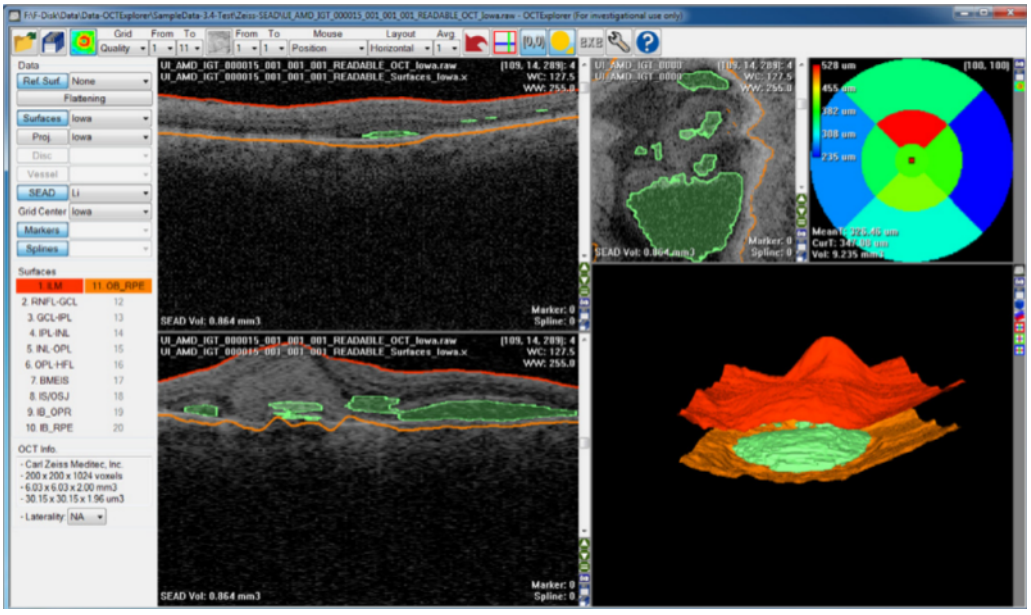


Figure 3.1: Screenshot of OCT-Explorer environment–11 layer surfaces segmented (A 3-D fluid-associated abnormalities in the retina, called as symptomatic exudate-associated derangements (SEAD) is shown in green). The upper right panels show an x-y (horizontal) slice cutting through the fluid region and the ability to perform regional analyses. The lower-right image shows a 3D visualization of the SEADs bounded by the ILM and RPE layer surfaces.

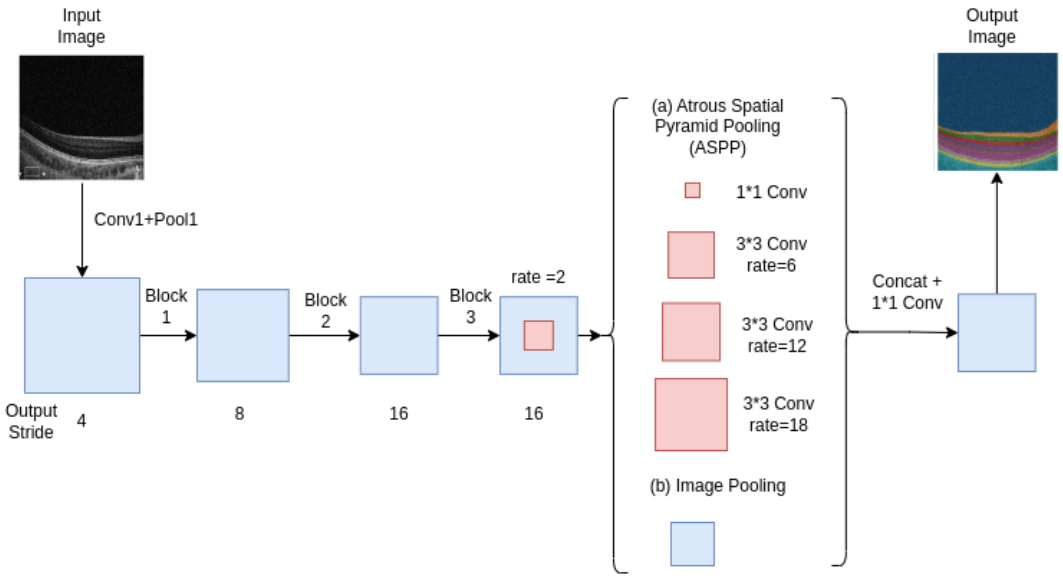


Figure 3.2: Diagram of the Deeplab v3 architecture, featuring an encoder-decoder structure with atrous spatial pyramid pooling and multi-scale feature fusion for semantic image segmentation.

3.3.3 Data Augmentation

In a typical 2-D approach, we first create masks from the original image. We do this process in the following manner:

1. Preprocess the image, i.e. plot XY coordinates obtained from the segmentation machine on each B-scan of the image to find points on each border of interest.

2. Process the points further to correct for possible discontinuities in the 1-D border detection approaches (e.g., spline-fit)

3. Fill up the area between two consecutive layers and the area outside the layers.

Once we have the masks, we augment them via:

1. Random cropping of the masks was performed at 500* 500 height and width.

2. Then the image is resized at (100 * 100), i.e. height and width.

3. A random horizontal flip is applied, which flips the image. Flipping the image horizontally would mean that the right eye scan is now the left eye scan and vice versa. Vertical flipping is not added since retina cross-section (i.e. a retinal OCT scan) will always be parallel to the ground.

4. A random salt and pepper noise was also added to enhance the efficiency of the model whenever it predicts the images that do not belong to central scan $\pm$ 15.

3.3.4 Algorithm

One of the difficulties in implementing deep convolutional neural networks (DCNNs) to separate objects in images is that the input feature map shrinks as the network traverses, information about objects of a smaller scale can be lost. To overcome this issue, we use DeepLabv3 Resnet50 model as our base segmentation model. DeepLabv3 is a semantic segmentation architecture that addresses the problem of segmenting objects at many scales. Modules that use atrous convolution in cascade or parallel to capture multi-scale context using multiple atrous rates have been built. In addition, DeepLabv2's Atrous Spatial Pyramid Pooling module was enhanced with image-level features that encode global context and improve efficiency. [16]

A fully connected(FC) layer at the end of the base model was added, in which the number of classes is equal to the curve length * number of curves present on the retinal OCT scan. Since we wanted to segment all the layers present, the number of classes was ten * curve length (width of the image in pixels), corresponding to each of the ten layers in the scan. The input image shape for the DeepLabv3 model is (N, 3, H, W), where N is the number of images, H and W are the height and width of the image, and 3 is the RGB channel. The images are loaded into a [0, 1] range and then normalised using mean = [0.485, 0.456, 0.406] and std = [0.229, 0.224, 0.225]. The model returns an OrderedDict

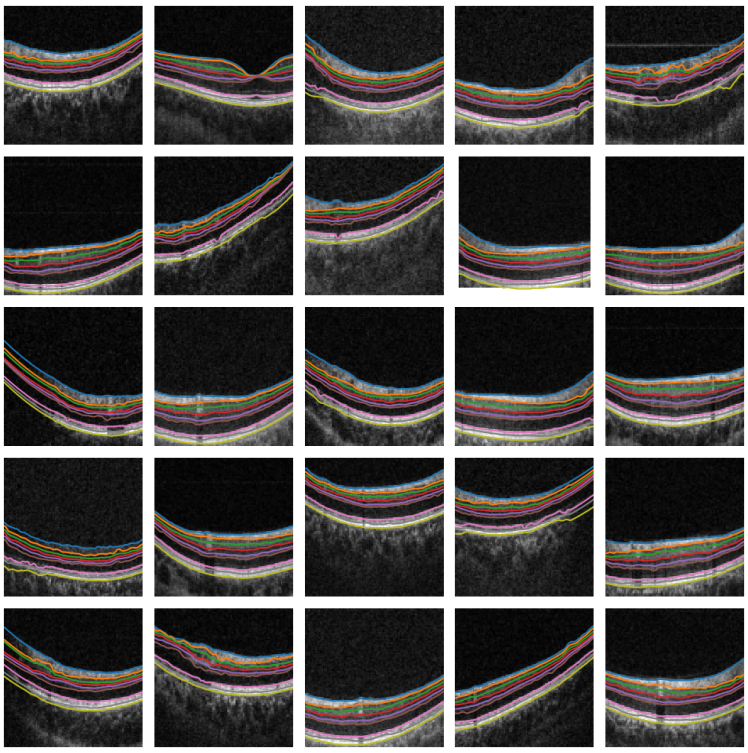


Figure 3.3: Segmented lines on central scans $\pm$ 12 Retinal OCT scans. The central scans are usually of good quality and hence are clearly segmented by the OCT-Explorer as mentioned in Sec. 3.3.1

with two tensors with the same height and width as the input Tensor, but with ten classes each. The semantic masks are stored in output['out'], while the auxiliary loss values are stored in output['aux']. output['aux'] is useless in inference mode. As a result, output['out'] is in shape (N, 10, H, W) We use mean squared error (MSE) Loss, which quantifies the average of the squares of the errors, i.e., the average squared difference between the estimated and real values. We used SGD optimizer with 0.9 momentum and 0.01 as the learning rate to get the best performing model.

3.4 Performance Metrics

3.4.1 Intersection Over Union(IoU)

The region of overlap between the expected segmentation and the ground truth is divided by the area of union between the predicted segmentation and the ground truth to arrive at the IoU. This statistic ranges from 0–1 (0–100%), with 0 indicating no overlap and 1 indicating perfect overlap.

$$\text{IoU} = \frac{\text{Area of overlap}}{\text{Area of union}} \tag{3.1}$$

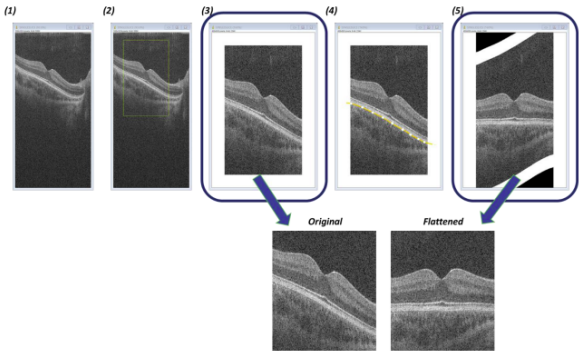


Figure 3.4: An overview of Retinal image flattening.

Model	Mean IoU	Dice Score
FCN ResNet-50	0.736	0.823
FCN ResNet-101	0.733	0.820
DeepLabV3 ResNet-50	0.834	0.896
DeepLabV3 ResNet-101	0.829	0.893

Table 3.1: Comparing Segmentation Model Performance at validation stage. The validation accuracy findings showed that DeepLabV3 ResNet-50 model performed the best.

3.4.2 Dice score

The Dice score, which ranges from 0 to 1, is used to evaluate model performance, i.e. it measures how much two masks overlap. A pixel perfect match between the deep learning model output and ground truth annotation equates to a value of 1. A perfect overlap is indicated by a 1; no overlap is indicated by a 0.

$$\text{Dice score} = \frac{2 * \text{Area of Overlap}}{\text{Total number of pixels in both images}} \tag{3.2}$$

3.5 Results

We calculated the train and test loss for four different model combinations with variations in learning rate as mentioned in Table.3.1. We used FCN ResNet50 ,FCN ResNet101, DeepLabV3 ResNet50, DeepLabV3 ResNet101. Our dataset was divided into a 70-30 train-test split. We used the test split to calculate the mean IoU and dice score.

Learning Rate	Mean IoU	Dice Score
0.01	0.834	0.896
0.07	0.830	0.890
0.001	0.771	0.850

Table 3.2: DeepLabV3 ResNet-50 learning rate comparison. Optimal learning is 0.01.

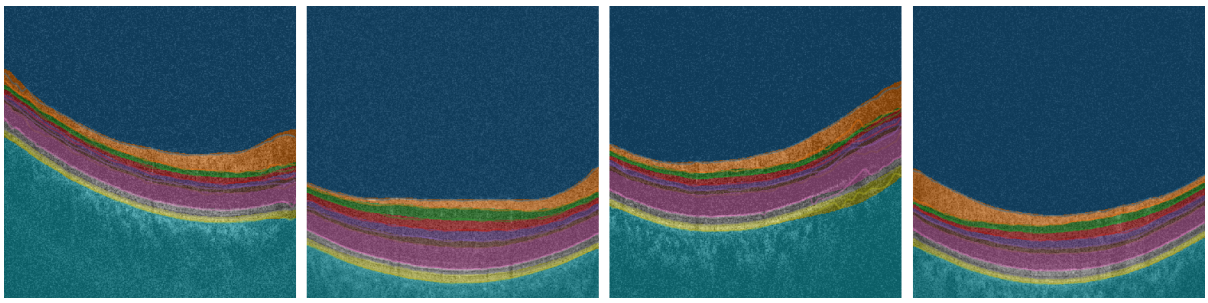


Figure 3.5: DeepLabV3 ResNet-50 predictions. Best performing model's semantic segmentation output on the retinal OCT scans.

DeepLabV3 ResNet50 was the top-performing model based on the mean IoU and Dice score. With momentum of 0.9, we employed SGD as our optimization algorithm. We found the optimal learning rate to be 0.01 as mentioned in Table. 3.2.

3.6 Conclusion

We have aimed to create a retinal OCT scan segmentation system that predicts all ten layer segments present on retina. Our proof-of-concept study with a DeepLabV3 ResNet50 demonstrated convergence of the model without overfitting/underfiting the data. We compared FCN ResNet50 ,FCN ResNet101, DeepLabV3 ResNet50, DeepLabV3 ResNet101 (all with both Adam and SGD optimizers) to get the best possible segmentation on OCT scan. Our deep learning system was trained and verified on imaging data from two OCT devices (Handheld OCT and Table mounted OCT device), resulting in a device-agnostic system.

We have achieved 0.834 as our mean IoU and 0.896 as our Dice Score on the best performing DeepLabV3 ResNet-50 semantic segmentation model. Our model accurately predicts all ten layers of the retinal OCT scan, which would be utilized for calculating layer thinning or thickening in later stages of the project.

Individual layer thinning or thickening may be indications of retinal illnesses or precursors to future vision loss; hence segmenting individual layers and quantifying their thicknesses location-specifically

is crucial in therapeutic intervention. According to studies, retinal morphological abnormalities develop in most fundus disorders before visual changes. Early detection of fundus retinal abnormalities will be aided by adopting accurate and sensitive technologies to analyse the retinal morphological structure.

As a result, we created a model that can do retinal layer analysis rapidly and reliably, which is crucial for the early detection and treatment of retinal diseases. We aim to automate retinal layer segmentation using retinal optical coherence tomography (OCT) images since retinal fundus pictures have been increasingly popular in diagnosing, screening, and treating retinal problems. By extracting the retinal layer thickness, our model can aid in diagnosing and monitoring retinal illnesses.

3.6.1 Future Scope

Despite our model's demonstrated good performance, our research has significant improvement areas. We need to extract exact values of the layer segments in μm units for both the ground truth and predicted OCT layer segments to get a better understanding of the model prediction. The extracted μm units would then aid in classifying the type of disease present in the eye.

We have started our work on flattening the retinal OCT scans, which might help in improving the segmentation of the retinal layers (shown in Fig. 3.4). Although we augment the data after masking, the current pipeline uses the raw OCT scan; therefore, layers retain their original curved shape.

We also aim to add 'Active Learning' to our project. This might help select a subset of images from a sizeable unlabeled pool of data so that obtaining annotations for those images will result in the most significant possible improvement in model accuracy.

Once the image has been segmented, the human in the loop would be given the option to correct the segmentation if it is wrong. Obtaining annotations of those images will result in a maximal increase in model accuracy since we can retrain our model using newly occupied data.

3.6.2 Code

The GitHub link to the code can be found at https://github.com/Garima13a/Retinal-OCT-Scan-segmentation. The repository includes a detailed README file that provides step-by-step instructions on how to run the code, as well as any necessary dependencies and configurations.

Chapter 4

Gaze tracker

4.1 Introduction

The term gaze is frequently used in physiology to describe coordinated motion of the eyes and neck. The horizontal gaze center is a collection of local circuit neurons near the midline in the pons responsible for generating horizontal eye movements. The vertical gaze center is located in the rostral part of the midbrain reticular formation and is responsible for vertical movements. Activation of each gaze center separately results in movements of the eyes along a single axis, either horizontal or vertical. Activation of the gaze centers in concert results in oblique movements whose trajectories are specified by the relative contribution of each center. [14] The conjugate gaze is the movement of both eyes in the same order simultaneously, whereas conjugate gaze palsy is the lack of coordination between both eyes. The conjugate gaze is controlled by four different mechanisms:

1. Saccadic system: A saccade is a simultaneous movement of both eyes between two or more phases of focus in the same direction. A saccadic system allows for voluntary direction of the gaze. [13]

2. Pursuit system: A pursuit system describes the ability of the eye to move in a way where the eyes remain fixated on a moving object, i.e. it allows the subject to follow a moving object.

3. Nystagmus: Nystagmus is a condition in which the eyes move involuntarily (or, in some situations, voluntarily [74]). It can be passed down the generations, but it is more typically acquired during infancy or later in life. In many situations, it can lead to eyesight loss or impairment. It has been dubbed ”dancing eyes” because of involuntary eye movement.

4. Vestibulo-ocular reflex system (VOR system): This system controls head movements to maintain a consistent visual representation of the world. It is a reaction that maintains a steady gaze during head movement and eye movement caused by vestibular system activation.

This project aims to detect the nystagmus present in the eye by making use of the nystagmus waveform created by our algorithm.

4.2 Background

4.2.1 Study Design

The study was divided into three main stages:

1. First stage, we experimented with videos of 12 patients to test horizontal/vertical nystagmus.
2. Second stage, we included diagonal movements of the eye.
3. Third stage, we gathered data from the gold standard algorithm and compared it to our algorithm.

4.3 Method

4.3.1 Dataset Collection

We used twelve patients' eye videos for the first stage of the project, which was provided by the Department of Ophthalmology (at the University of Leicester). Out of 12 patients, six had the vertical nystagmus eye condition, whereas the other six had horizontal nystagmus. These videos were recorded at 30 fps.

For later stages, we manually recorded the video for gaze tracking using a tripod and camera at 50 fps. This dataset consisted of two types of videos:

1. Human eye movement with the gold standard equipment on the bare face.
2. Human eye movement with the gold standard equipment on face with a mask on.

4.3.2 Eyelink II

For determining the nystagmus waveforms, we use a head-mounted video-based eye tracker known as the EyeLink II, as shown in Figure 4.1. It has the fastest data rate and highest resolution, which results in very low-velocity noise, making it ideal for a wide range of eye-tracking research. The EyeLink II has Pupil + Corneal Reflection and Pupil-only tracking modes. It has binocular eye-tracking at 500 Hz, 0.5o average accuracy, and 0.01o resolution, as well as 3.0 msec delay eye position data. It has high-quality computer-based or scene camera eye tracking.

4.3.3 MediaPipe Iris

This project makes use of the MediaPipe Iris: Real-time Iris Tracking & Depth Estimation by Google [53]. With just a single RGB camera and no additional technology, this MediaPipe Iris model can monitor landmarks affecting the iris, pupil, and eye shapes. With a relative inaccuracy of less than 10%, the solution can also compute the metric distance between the subject and the camera using iris

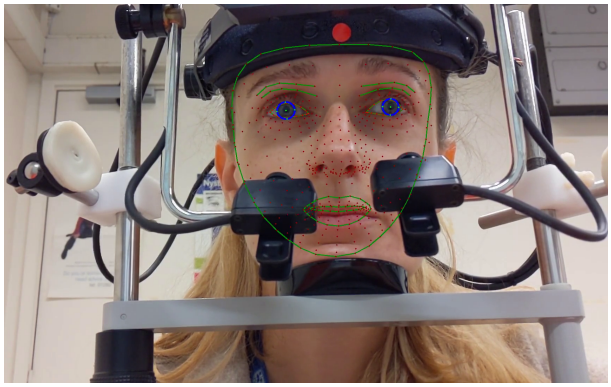


Figure 4.1: A participant wearing Eyelink II. Video-based eye-tracker has a headrest to ensure less eye movement and a one cm diameter red sticker on the forehead panel to extract relative eye movement distance.

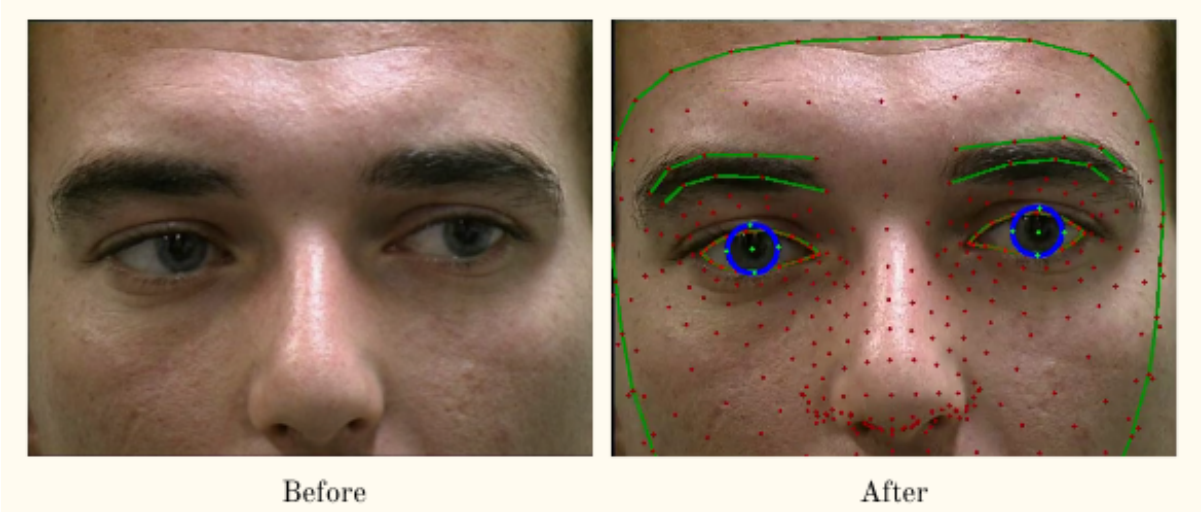


Figure 4.2: Mediapipe facemesh. The red dots represent the 468 landmarks on the image, the blue lines connecting landmarks illustrate the contours around the eyes and the green lines connecting landmarks show eyebrows, lips and the entire face.

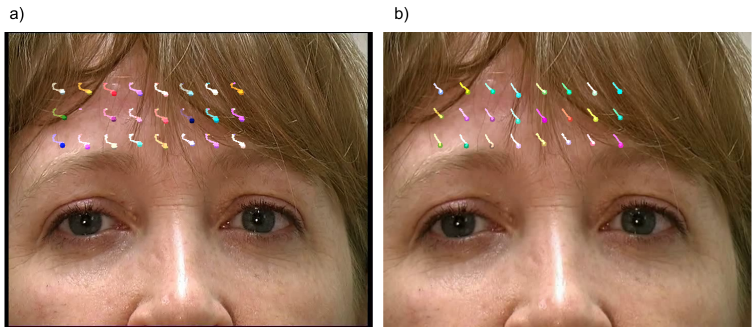


Figure 4.3: Optical Flow: 24 keypoints on forehead representing the movement of first frame throughout video. **a)** Last video frame before smoothing. **b)** Last video frame after smoothing.

landmarks. It's worth noting that iris tracking doesn't infer where people are looking or provide any kind of identification recognition.[53]

MediaPipe Facial Mesh is used in the first step of the pipeline to create a mesh of the approximate face geometry as shown in Figure 4.2. We isolate the eye region in the original image using this face mesh for iris tracking phase that follows. The iris model takes an image patch of the eye region and estimates both the eye landmarks (along the eyelid) and iris landmarks (along the iris contour). [1]

4.3.4 Data Prepossessing

Each video is first smoothed using optical flow, which is the pattern of the apparent movement of image objects between two consecutive frames caused by the object's movement or camera. According to OpenCV documentation Optical flow works on several assumptions:

1. The pixel intensities of an object do not change between consecutive frames.

2. Neighbouring pixels have similar motion.

Hence, to ensure that the pixel accuracy does not change during the video, we made sure to select the center part of the forehead to track throughout the video. In the later stages of the project, we make use of a red sticker to use as a tracker as well as for depth estimation. Figure 4.3 shows 24 points on the forehead that were tracked to receive the optical flow vectors of those points.

4.3.4.1 Video Smoothing

This method involves tracking a few feature points between two consecutive frames. The tracked features allow us to calculate and adjust for motion between frames. As shown in Figure 4.4,we will go through all of the frames, looking for the motion in the current and former frames. We apply the Euclidean motion model, which requires only two points in each frame to move. Since tracking algorithms use a small patch around a point to track it, we used 24 points to track the forehead.

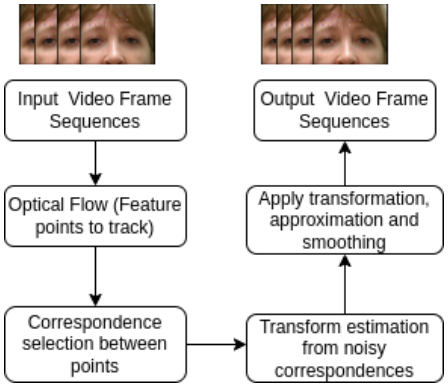


Figure 4.4: Block diagram for the smoothing of the video

We use the Lucas-Kanade Optical Flow algorithm to track good features in the previous frame in the next frame.[52][51] We use the Lucas-Kanade Optical Flow algorithm to track good features in the previous frame in the next frame. We can utilize these two sets of points to identify the rigid (Euclidean) transformation that maps the last frame to the current frame because we know the features in the current and previous frames.

Once we have estimated the motion, we can divide it into x and y translation and rotation angle. These values are stored in an array to be changed effortlessly. The motion trajectory is then determined by putting all the calculated differential motions together. So we have three curves that show how the motion (x, y, and angle) varies over time, and we use a moving average filter to smooth them out. Moving average filter replaces the value of a function at the point by its neighbours' average defined by a window. Once we have obtained a smooth trajectory, we will use it to obtain smooth transforms that can be applied to frames of the videos to stabilize it. This is done by finding the difference between the smooth and original trajectories and adding this difference back to the original transforms. Then we loop over the frames and apply the transforms.

4.4 Experiments and Results

4.4.1 Stage 1: Initial Eye tracking

We used Mediapipe iris to get the face mesh on the patient's video. The eye region was extracted using a circular template resembling an eye; to create a mask of the iris. We implemented template matching with k means clustering of the masked area to extract the x-y coordinates of the centre of the eye. This technique was implemented in each frame for both eyes.

As mentioned in 4.3.4.1, we first smoothed patients' video. Then we tracked vertical and horizontal movements of both the left and right eye (Here, the left and right eye is addressed in terms of the first

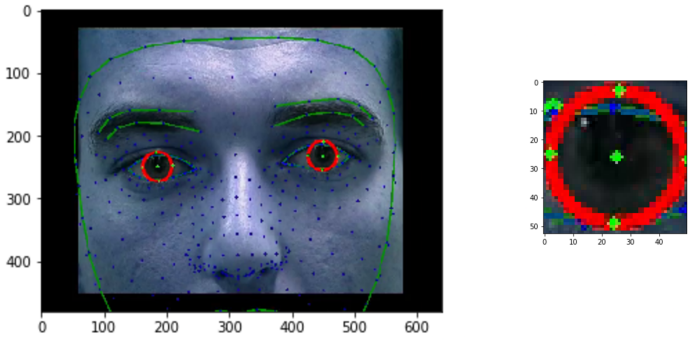


Figure 4.5: Extraction of the iris

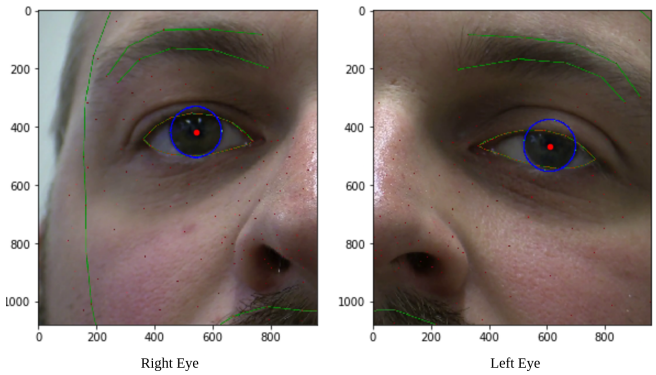


Figure 4.6: Right and Left Eye: From the perspective of patient

person's perspective.) This gave us insight into the coordination of the left and right eyes on the x-axis (horizontal direction) and the y-axis (vertical direction).

4.4.1.1 Nystagmus

Nystagmus is defined as involuntary, rhythmical, recurrent oscillations of one or both eyes in any or all fields of view. It can be pendular (with undulating motions of equal speed, amplitude, and duration in each direction) or jerky (with jerky movements of equal speed, amplitude, and duration). Movement can be slower in one direction, followed by a faster return to the original position. Horizontal, vertical, oblique, rotational, circular, or any combination of these movements are possible. The smaller the amplitude, the faster the pace (and vice versa). The defect is characterised based on where the eyes are when it occurs:

1. Grade I: Only when the eyes are oriented toward the rapid component.

2. Grade II: When the eyes are in their primary/basic posture.

3. Grade III: When the eyes are oriented toward the slow component.

and also on how the eyes move:

1. side to side (horizontal nystagmus)

2. up and down (vertical nystagmus)

3. in a circle (rotary nystagmus)

The failure to keep a constant fixation causes reduced acuity. Nystagmus can be reduced by head tilting, which is usually involuntary (toward the fast component in jerky nystagmus or in such a position to minimise pendular nystagmus). Congenital nystagmus is frequently accompanied by head nodding. Dizziness or vertigo may develop if oscillopsia (illusory movement of things) happens.

Nystagmus can be a symptom of a variety of neurological illnesses and a side effect of certain medications (including barbiturates). Nystagmus is a condition that has no recognised cure. Certain varieties of jerky nystagmus (usually Grade I types) do, however, improve spontaneously during development (up to about age 10). Muscle surgery may be an option for this type (essentially, repositioning muscles to take advantage of the point of least nystagmus or relative rest position). Except for brief oscillopsia episodes, most people with nystagmus consider objects to be stationary. The brain is thought to be responsible for perceptual adjustment. Children with nystagmus (who may lose their place in early reading instruction) may benefit from using a typoscope (a card with a rectangular hole that allows you to see one word or line at a time) or an underliner (a card or strip of paper that allows you to ”underline” the line you are reading). Children with nystagmus appear to require these assistance devices less frequently as they grow older.

4.4.2 Stage 1 Results

Figure 4.7 shows a person diagnosed with horizontal nystagmus i.e involuntary side-to-side eye movements. We faced challenges in the fps of the video data. Since the videos were at 30 FPS, it was hard for us to identify the pattern in the form of graphs. Moreover, we used template matching to extract the iris; hence only the centre eye data were available. To better understand the movement of the eye w.r.t. translation movement of the head and its accuracy with Mediapipe, we required data for the corner of the eye as well. Since there was a need for x-y coordinates for all eye corners, including iris points and centre, we tweaked the Mediapipe iris code to extract iris points with corners.

4.4.3 Stage 2: Depth Estimation and Ocular motility

4.4.3.1 Depth Estimation

Depth Estimation means the metric distance of a subject to the camera. In our case, we are looking at the position of the eye in absolute measurements, i.e. millimeters (mm) To make use of the depth

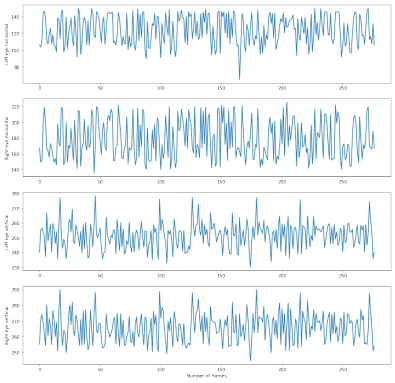


Figure 4.7: Horizontal Nystagmus: Showing no co-ordination between 'Right eye horizontal' and 'Left Eye Horizontal'

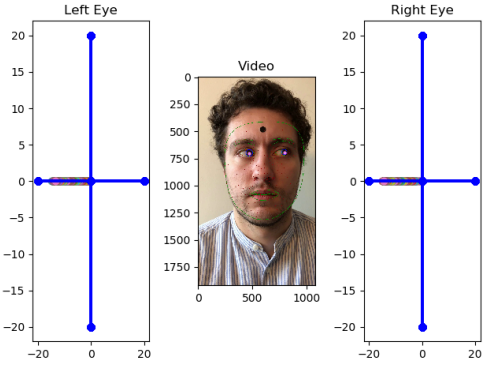


Figure 4.8: Sticker on the head (1cm in diameter): Graph showing a healthy participant's eye movement. Each dot represents where the eye moved in each frame.

estimation in the project, we added a 1cm diameter sticker on the forehead of the participants. This sticker is used as a feature tracker in optical flow as well as it is used to get absolute values of the movement of the eye, i.e. how many millimeters (mm) the eye moved in a certain time limit.

To get the absolute distance of the eye movement, we rely on the fact that the horizontal iris diameter of the human eye remains roughly constant at 11.7±0.5 mm across a wide population, along with some simple geometric arguments.[6, 62, 5] Hence, each experiment's data will be converted into absolute value using this method.

4.4.3.2 Ocular Motility

The study of the twelve extraocular muscles and their effects on eye movement is called ocular motility. Each eye consists of six muscles, four rectus and two obliques, that allow the eyes to operate together in a wide range of gazes when they are working correctly. Patients are frequently examined

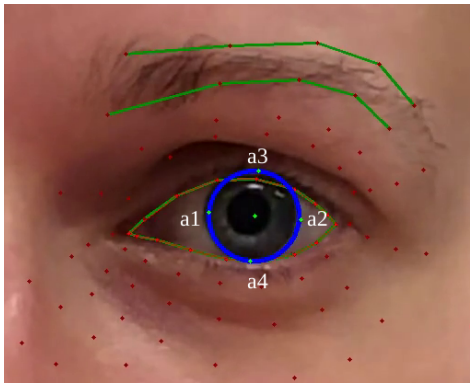


Figure 4.9: Eye diameter: a1-a2 = a3-a4 = 11.7 ± 0.5 mm

for eye position and movement during a standard eye examination. The patient is advised to look in the nine diagnostic cardinal positions to monitor the functioning of all twelve muscles. While the physician closely examines their eye movements, the patient follows a target to various sight places. Any restriction or misalignment of the eyes should be investigated further since it could suggest muscle weakness or paralysis.

The "cardinal positions" are nine positions of gaze which allow comparisons of the horizontal, vertical, and diagonal ocular movements produced by the six extra-ocular muscles when both eyes and multiple muscles are working together. These are the nine cardinal positions:

1. up/right
2. up/left
3. right
4. left
5. down/right
6. down/left
7. front
8. up
9. down

A muscle of one eye is paired with a muscle of the other eye in each gaze position to move the eyes together in a specific direction. Each of the nine cardinal gaze positions is represented in Fig. 4.11.

4.4.4 Stage 2 Results

We took into consideration two participants' videos for calculating ocular motility. The equipment list and method for making participants' videos are as follows:

1. Mobile camera at 60 FPS
2. Flash unit (camera mounted)
3. Tripod stand
4. A target for following gaze (in our case we used a pen to guide across the target board).

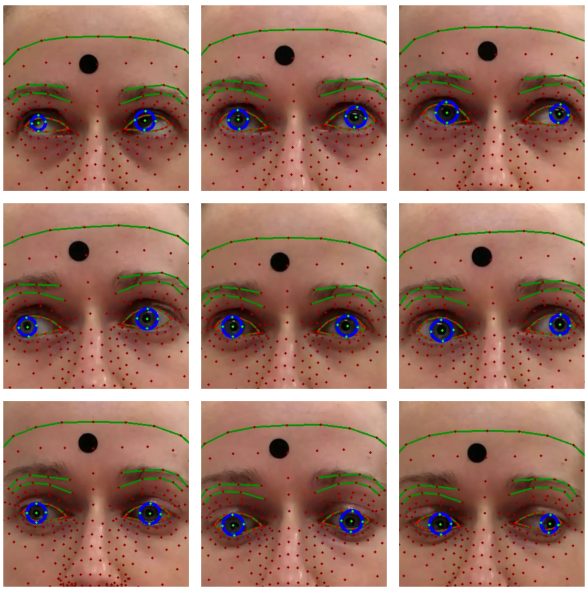
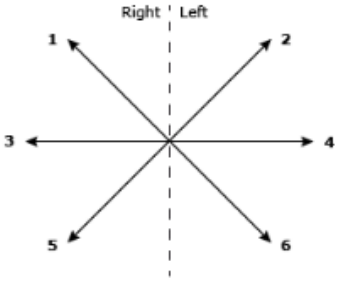


Figure 4.10: Cardinal Positions: The six cardinal positions of eye gaze isolate the individual extraocular muscles. In order to determine which muscle is abnormal, the six positions of gaze are observed to determine if the eye moves in each direction.



Eye position	Extraocular muscles
1) Up and to the right	Right superior rectus Left inferior oblique
2) Up and to the left	Left superior rectus Right inferior oblique
3) Right	Right lateral rectus Left medial rectus
4) Left	Left lateral rectus Right medial rectus
5) Down and to the right	Right inferior rectus Left superior oblique
6) Down and to the left	Left inferior rectus Right superior oblique

Figure 4.11: Nine Cardinal Positions of the eye

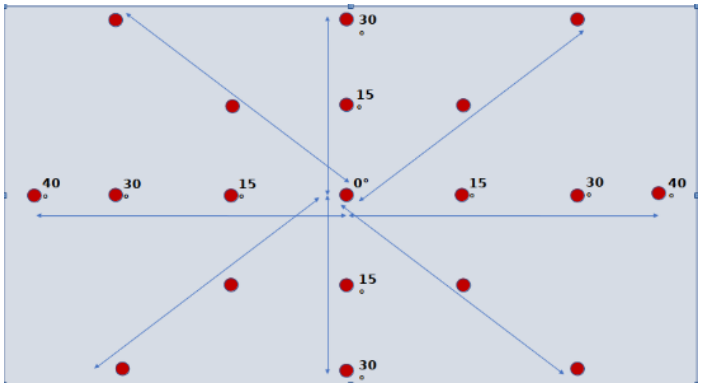


Figure 4.12: Ocular Motility target board: Nine cardinal positions with degrees mentioned on them were used to track the eye movement starting from 0 degrees.

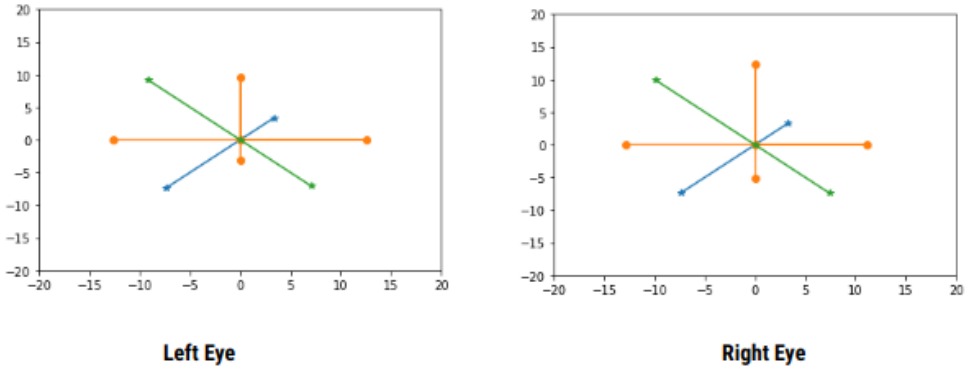


Figure 4.13: Ocular Motility for participant 1: Since eyes do not have any nystagmus, there is a correlation between left and right eye's nine cardinal positions.

5. A target image with degrees mentioned for eye movement. Refer Fig. 4.12

The target started from the origin, i.e. the primary front position of the eye, before moving onto the nine cardinal points. The participant follows the target with his/her gaze without moving their head. To account for the error in involuntary head movements during the video, we first perform video smoothing mentioned in Sec. 4.3.4.1.

Graphs of ocular motility may be required to document any abnormal muscle movement. These graphs can be utilised to back up a doctor's clinical judgment for medical-legal reasons or as a starting point for muscle surgery. Patients with strabismus (A term for an imbalance of the extraocular muscles causing misalignment of the eyes) and cranial nerve palsies are excellent candidates for motility documentation.

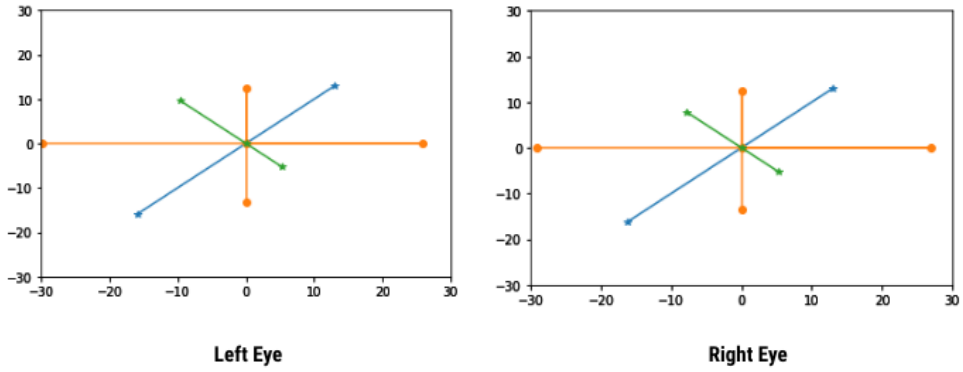


Figure 4.14: Ocular Motility for participant 2: Since eyes do not have any nystagmus, there is a correlation between left and right eye's nine cardinal positions.

4.5 Stage 3: Gold Standard Nystagmus waveform

4.5.1 Nystagmus waveform

Nystagmus waveforms get their name because they have a slow phase velocity profile (See Fig. 4.15). The pendular form has no fast phase and is best represented by Figure 2's first wave. Congenital nystagmus is linked to the exponentially increasing velocity type. The exponential declining velocity waveform is prevalent in gaze-evoked nystagmus, which can be a pathologic finding. A linear waveform characterises vestibular nystagmus. Dissociated nystagmus occurs when nystagmus occurs in the same direction but with different amplitudes in the two eyes. Sea-saw nystagmus is an example of disconjugate nystagmus when the two eyes oscillate in opposite directions.

Nystagmus has been divided into jerk nystagmus, which exhibits a quick and slow phase, and pendular nystagmus, which is a sinusoidal-like oscillation without any obvious quick phase. In jerk nystagmus, the direction of nystagmus is defined by the quick phase of the jerk (e.g., downbeat). The slow phase can have accelerating and decelerating velocities. [36]

4.5.2 Stage 3 Results

Since it is helpful to differentiate between infantile and acquired nystagmus, this can be done by considering the age of onset and waveform characteristics of nystagmus. The waveform can be formally obtained by eye movement recordings. Techniques used are electrooculography, scleral search coil and video eye-tracking devices [36], i.e. Eyelink II. We compared the waveform generated by Eyelink II(gold standard) and our gaze tracking model.

We tracked the centre of the iris throughout the video and plotted the x and y coordinates, which correspond to horizontal and vertical eye movement, respectively. As shown in Fig 4.17,

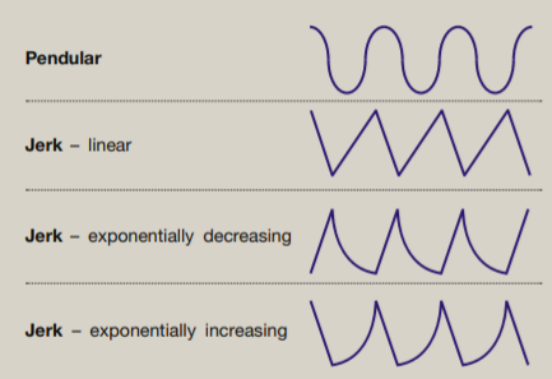
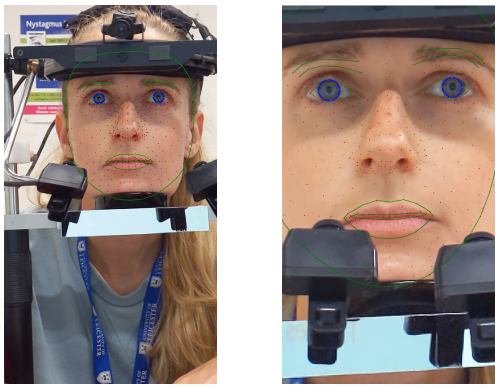
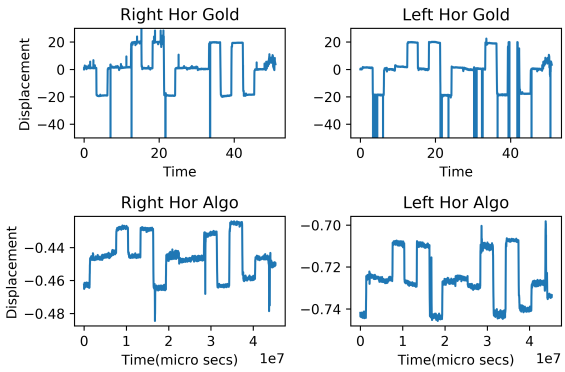


Figure 4.15: Waveform characteristics of different types of nystagmus.[36]

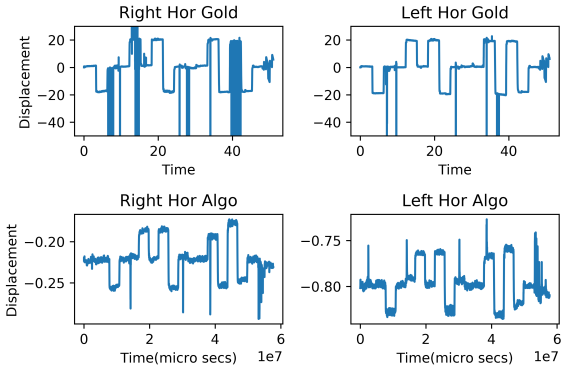


(a) a (b) b

Figure 4.16: Participant with mounted Eyelink II: [a] Zoomed out. [b] Zoomed in

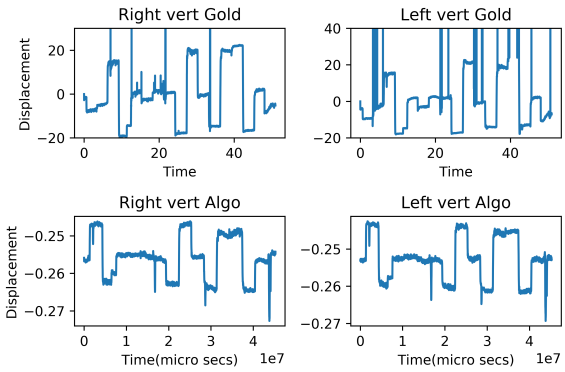


(a) a

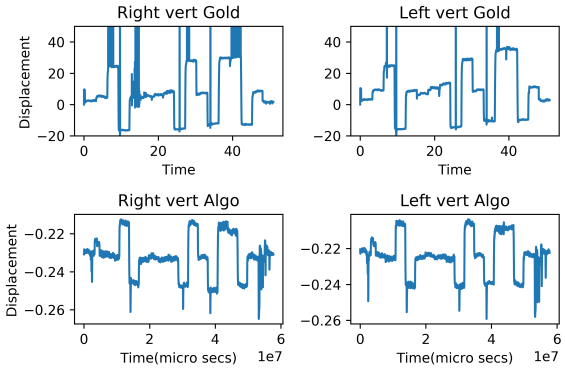


(b) b

Figure 4.17: Horizontal eye movements compared with the Gold standard Eyelink II waveform. Here, Right/Left Hor Gold refers to Right/Left Horizontal Gold Standard and Right/Left Hor Algo refers to Right/Left Horizontal algorithm. [a] Zoomed out version [b] Zoomed in version



(a) a



(b) b

Figure 4.18: Vertical eye movements compared with the Gold standard Eyelink II waveform. Here, Right/Left Ver Gold refers to Right/Left Vertical Gold Standard and Right/Left Ver Algo refers to Right/Left Vertical algorithm. [a] Zoomed out version [b] Zoomed in version

4.6 Conclusion

We have created an algorithm that detects the nystagmus present in the eye by creating a nystagmus waveform via a video (shot by any type of camera). We have successfully extracted data using eye tracker algorithm [53] and predicted both horizontal and vertical nystagmus present in the patient.

We have made our algorithm flexible enough to incorporate video at any frame rate. Although it was observed that higher frame rate gives more clear nystagmus waveform output and hence making the diagnosis much easier.

We have extracted three different informations using one single patient video having all nine cardinal positions (Fig. 4.10 , 4.12). First, we find out ROI (region of interest) in the video using Mediapipe iris (Refer to Sec. 4.3.3). Iris is tracked throughout the video and using the position of the eye in each frame; we draw graph to locate nystagmus waveform. Based on the curve in the graph (Fig. 4.15), we find out the type of nystagmus present in the patient.

To enhance the utility of our algorithm, we calculate Ocular Motility (Sec. 4.4.3.2. Each eye is separately tracked throughout the video to check the coordination between the two eyes (Fig. 4.13). If left and right eye graphs are not similar for all the cardinal positions then it indicates presence of an eye disorder.

Since our algorithm gave an outstanding performance in generation of nystagmus waveforms, we took it further and compared the waveform generated by Gold standard Eyelink II (Sec. 4.3.2) with our algorithm. It's safe to say that our model gave accurate results for each patient video (Fig. 4.17 , 4.18).

A neurological issue at birth or develops in early childhood is the most common cause of nystagmus. Later-life acquired nystagmus might indicate another ailment or disease, such as stroke, multiple sclerosis, or trauma. Lack of development of normal eye movement control early in childhood is another reason for nystagmus. Causes of nystagmus can be:

1. Albinism.
2. Refractive error: Nearsightedness (myopia) or astigmatism.
3. Cataracts that are present at birth.
4. The inner ear is inflamed.
5. Diseases of the central nervous system.
6. Anti-epilepsy medications

To tackle early detection of the disorders mentioned above, we developed our gaze tracking algorithm to detect nystagmus even when the patient is not physically present at the ophthalmology clinic.

Since our algorithm can properly track gaze and detect nystagmus, it is much faster and easier for the patient to send a mobile shot video now and still get the correct diagnosis.

4.6.1 Code

The GitHub link to the code can be found at https://github.com/Garima13a/Gaze-tracker. The repository includes a detailed README file that provides step-by-step instructions on how to run the code, as well as any necessary dependencies and configurations.

Chapter 5

Conclusions

In this thesis, we worked on a group of disorders known as Foveal hypoplasia (FH), retinal layer segment-related disorders and nystagmus. [38]

In the first stage, we use optical coherence tomography (OCT) scans to identify the degree of arrested retinal development i.e. the grades of Foveal hypoplasia (FH), since this information provides both diagnostic and prognostic value. It is characterised by arrested retinal development and often associated with infantile vision condition in which the eyes make repetitive, uncontrolled movements(medical term for which is 'nystagmus'). We use optical coherence tomography (OCT) scans to identify the degree of arrested retinal development since this information provides both diagnostic and prognostic value. Recent advancements of high-resolution OCT imaging have unveiled characteristics of foveal hypoplasia that were not detected by conventional imaging methods [38]. Our classification diagnostic system uses OCT images to differentiate between normal and diseased retinal development for the first time. After testing and comparing multiple CNNs, our proof-of-concept study with a ResNet-50 model provides good sensitivity and specificity for binary and six-point classification. Our deep learning system was trained and validated on imaging data from various OCT devices, resulting in a device-agnostic system. Subclassifying the stage of abnormal retinal development based on the severity of foveal hypoplasia (six-point classification) provides mechanistic insight into the retinal developmental event that was stopped, whether it occurred in utero or after birth. Our classification system exhibited high sensitivity (99.13%) and specificity (95.69%), implying that our intelligent system might be used as a screening tool to distinguish between normal and abnormal fovea.

We also aimed to design a retinal OCT scan segmentation system to detect all ten retinal layer segments. Individual layer thinning or thickening can be signs of retinal disease or a prelude to future vision loss; hence segmenting individual layers and measuring their thickness is critical in therapeutic intervention. As a result, we developed a model that can do retinal layer analysis quickly and accurately, which is critical for detecting and treating retinal illnesses early. We have achieved 0.834 as our mean IoU and 0.896 as our Dice Score on the best-performing DeepLabV3 ResNet-50 semantic segmentation model. Our segmentation technique predicts all ten layer segments found on the retina. Our proof-of-concept experiment with a DeepLabV3 ResNet50 showed that the model converged without

overfitting the data. To acquire the best possible segmentation on an OCT scan, we experimented on FCN ResNet50, FCN ResNet101, DeepLabV3 ResNet50, and DeepLabV3 ResNet101 segmentation models with various hyperparameter values. Our deep learning system was trained and validated using imaging data from various OCT devices, resulting in a device-independent solution.

We extend our project to include a gaze tracking algorithm. We accurately identify the video's ROI (region of interest) (Sec.4.3.3). Iris is tracked throughout the video, and we create a graph based on the eye's position in each frame to find the nystagmus waveform. The curve determines the type of nystagmus present in the patient in the graph (Fig. 4.15). We calculate Ocular Motility (Sec. 4.4.3.2) to improve the usability of our approach. Each eye is tracked separately throughout the movie to assess the coordination between the two eyes (Fig. **??**). If the left and right eye graphs for all cardinal points are not identical, it suggests the presence of an eye condition.

This work has laid a solid platform for future testing with our AI algorithm, bringing us one step closer to implementing a real-time intelligent diagnosis solution for paediatric OCT.

Related Publications

Parts of the work presented in this thesis have been previously reported in following research papers:

1. H. Kuht, S. Wang, G. Nishad, S. George, G. Maconachie, V. Sheth, Z. Tu, M. Hisaund, R. McLean, S. Teli,et al. Using artificial intelligence (AI) to classify retinal developmental disorders. Investigative Ophthalmology & Visual Science, 61(7):4030–4030, 2020. [39]

2. H. J. Kuht, G. Nishad, S. S. Wang, G. Maconachie, V. Sheth, Z. Tu, M. Hisaund, R. J. McLean, R. Purohit, S. Teli, et al. A machine learning solution to predict foveal development and visual prognosis in retinal developmental disorders. Investigative Ophthalmology & Visual Science, 62(8):2739–2739, 2021. [42]

Bibliography

[1] A. Ablavatski, A. Vakunov, I. Grishchenko, K. Raveendran, and M. Zhdanovich. Real-time pupil tracking from monocular video for digital puppetry. *CoRR*, abs/2006.11341, 2020.

[2] M. D. Abràmoff, K. Lee, M. Niemeijer, W. L. Alward, E. C. Greenlee, M. K. Garvin, M. Sonka, and Y. H. Kwon. Automated segmentation of the cup and rim from spectral domain oct of the optic nerve head. *Investigative ophthalmology & visual science*, 50(12):5778–5784, 2009.

[3] P. Agarwal, S. Kumar, R. Singh, P. Agarwal, and M. Bhattacharya. A combination of bias-field corrected fuzzy c-means and level set approach for brain mri image segmentation. In *2015 Second International Conference on Soft Computing and Machine Intelligence (ISCMI)*, pages 84–87. IEEE, 2015.

[4] A. M. Bagci, M. Shahidi, R. Ansari, M. Blair, N. P. Blair, and R. Zelkha. Thickness profiles of retinal layers by optical coherence tomography image segmentation. *American journal of ophthalmology*, 146(5):679–687, 2008.

[5] M. Baumeister, E. Terzi, Y. Ekici, and T. Kohnen. Comparison of manual and automated methods to determine horizontal corneal diameter. *Journal of Cataract & Refractive Surgery*, 30(2):374–380, 2004.

[6] J. P. Bergmanson and J. G. Martinez. Size does matter: what is the corneo-limbal diameter? *Clinical and Experimental Optometry*, 100(5):522–528, 2017.

[7] M. Bhende, S. Shetty, M. K. Parthasarathy, and S. Ramya. Optical coherence tomography: A guide to interpretation of common macular diseases. *Indian journal of ophthalmology*, 66(1):20, 2018.

[8] K. L. Boyer, A. Herzog, and C. Roberts. Automatic recovery of the optic nervehead geometry in optical coherence tomography. *IEEE transactions on medical imaging*, 25(5):553–570, 2006.

[9] Y. Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In *Proceedings eighth IEEE international conference on computer vision. ICCV 2001*, volume 1, pages 105–112. IEEE, 2001.

[10] 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*, 66:49–84, 2018.

[11] E. Britannica. Sensory reception: human vision: structure and function of the human eye. *Encyclopedia Brittanica*, 27:179, 1987.

[12] T. R. Burke, D. W. Rhee, R. T. Smith, S. H. Tsang, R. Allikmets, S. Chang, M. A. Lazow, D. C. Hood, and V. C. Greenstein. Quantification of peripapillary sparing and macular involvement in stargardt disease (stgd1). *Investigative ophthalmology & visual science*, 52(11):8006–8015, 2011.

[13] B. Cassin, S. Solomon, and M. Rubin. Dictionary of eye terminology, gainesville. *Florida: Triad Publishing*, 1990.

[14] N. Catz and P. Thier. Neural control of saccadic eye movements. *Neuro-Ophthalmology*, 40:52–75, 2007.

[15] A. Chan, J. S. Duker, H. Ishikawa, T. H. Ko, J. S. Schuman, and J. G. Fujimoto. Quantification of photoreceptor layer thickness in normal eyes using optical coherence tomography. *Retina (Philadelphia, Pa.)*, 26(6):655, 2006.

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

[17] V. Cheplygina, M. de Bruijne, and J. P. Pluim. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical image analysis*, 54:280–296, 2019.

[18] S. J. Chiu, M. J. Allingham, P. S. Mettu, S. W. Cousins, J. A. Izatt, and S. Farsiu. Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. *Biomed. Opt. Express*, 6(4):1172–1194, Apr 2015.

[19] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu. Automatic segmentation of seven retinal layers in sdoct images congruent with expert manual segmentation. *Optics express*, 18(18):19413–19428, 2010.

[20] H. Daldrup-Link. Artificial intelligence applications for pediatric oncology imaging. *Pediatric radiology*, 49(11):1384–1390, 2019.

[21] A. Dutta and A. Zisserman. The via annotation software for images, audio and video. In *Proceedings of the 27th ACM international conference on multimedia*, pages 2276–2279, 2019.

[22] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu. Automatic segmentation of nine retinal layer boundaries in oct images of non-exudative amd patients using deep learning and graph search. *Biomedical optics express*, 8(5):2732–2744, 2017.

[23] M. Fleckenstein, P. C. Issa, H.-M. Helb, S. Schmitz-Valckenberg, R. P. Finger, H. P. Scholl, K. U. Loeffler, and F. G. Holz. High-resolution spectral domain-oct imaging in geographic atrophy associated with age-related macular degeneration. *Investigative ophthalmology & visual science*, 49(9):4137–4144, 2008.

[24] D. S. Friedman, B. J. O'Colmain, B. Munoz, S. C. Tomany, C. McCarty, P. De Jong, B. Nemesure, P. Mitchell, J. Kempen, et al. Prevalence of age-related macular degeneration in the united states. *Arch ophthalmol*, 122(4):564–572, 2004.

[25] J. Fujimoto and E. Swanson. The development, commercialization, and impact of optical coherence tomography. *Investigative ophthalmology & visual science*, 57(9):OCT1–OCT13, 2016.

[26] A. Gawish, P. Fieguth, S. Marschall, and K. Bizheva. Undecimated hierarchical active contours for oct image segmentation. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 882–886. IEEE, 2014.

[27] E. P. Giri and A. M. Arymurthy. Quantitative evaluation for simple segmentation svm in landscape image. In *2014 International Conference on Advanced Computer Science and Information System*, pages 369–374. IEEE, 2014.

[28] K. Gopinath, S. B. Rangrej, and J. Sivaswamy. A deep learning framework for segmentation of retinal layers from oct images. In *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 888–893. IEEE, 2017.

[29] A. Hagiwara, Y. Mitamura, K. Kumagai, T. Baba, and S. Yamamoto. Photoreceptor impairment on optical coherence tomographic images in patients with retinitis pigmentosa. *British Journal of Ophthalmology*, 97(2):237–238, 2013.

[30] D. X. Hammer, N. V. Iftimia, R. D. Ferguson, C. E. Bigelow, T. E. Ustun, A. M. Barnaby, and A. B. Fulton. Foveal fine structure in retinopathy of prematurity: an adaptive optics fourier domain optical coherence tomography study. *Investigative ophthalmology & visual science*, 49(5):2061–2070, 2008.

[31] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[32] M. R. Hee, C. R. Baumal, C. A. Puliafito, J. S. Duker, E. Reichel, J. R. Wilkins, J. G. Coker, J. S. Schuman, E. A. Swanson, and J. G. Fujimoto. Optical coherence tomography of age-related macular degeneration and choroidal neovascularization. *Ophthalmology*, 103(8):1260–1270, 1996.

[33] M. Hingorani, K. A. Williamson, A. T. Moore, and V. van Heyningen. Detailed ophthalmologic evaluation of 43 individuals with pax6 mutations. *Investigative ophthalmology & visual science*, 50(6):2581–2590, 2009.

[34] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, et al. Optical coherence tomography. *science*, 254(5035):1178–1181, 1991.

[35] M. A. Hussain, A. Bhuiyan, and K. Ramamohanarao. Disc segmentation and bmo-mrw measurement from sd-oct image using graph search and tracing of three bench mark reference layers of retina. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 4087–4091. IEEE, 2015.

[36] N. Hussain. Diagnosis, assessment and management of nystagmus in childhood. *Paediatrics and Child Health*, 26(1):31–36, 2016.

[37] R. Kafieh, H. Rabbani, and S. Kermani. A review of algorithms for segmentation of optical coherence tomography from retina. *Journal of medical signals and sensors*, 3(1):45, 2013.

[38] H. Kondo. Foveal hypoplasia and optical coherence tomographic imaging. *Taiwan journal of ophthalmology*, 8(4):181, 2018.

[39] H. Kuht, S. Wang, G. Nishad, S. George, G. Maconachie, V. Sheth, Z. Tu, M. Hisaund, R. McLean, S. Teli, et al. Using artificial intelligence (ai) to classify retinal developmental disorders. *Investigative Ophthalmology & Visual Science*, 61(7):4030–4030, 2020.

[40] 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*, 29(18):2989–3002, 2020.

[41] H. J. Kuht, G. D. Maconachie, J. Han, L. Kessel, M. M. Van Genderen, R. J. McLean, M. Hisaund, Z. Tu, R. W. Hertle, K. Gronskov, et al. Genotypic and phenotypic spectrum of foveal hypoplasia: A multi-centre study. *Ophthalmology*, 2022.

[42] H. J. Kuht, G. Nishad, S. S. Wang, G. Maconachie, V. Sheth, Z. Tu, M. Hisaund, R. J. McLean, R. Purohit, S. Teli, et al. A machine learning solution to predict foveal development and visual prognosis in retinal developmental disorders. *Investigative Ophthalmology & Visual Science*, 62(8):2739–2739, 2021.

[43] S. Kuwayama, Y. Ayatsuka, D. Yanagisono, T. Uta, H. Usui, A. Kato, N. Takase, Y. Ogura, and T. Yasukawa. Automated detection of macular diseases by optical coherence tomography and artificial intelligence machine learning of optical coherence tomography images. *Journal of ophthalmology*, 2019, 2019.

[44] H. Lee, R. Purohit, A. Patel, E. Papageorgiou, V. Sheth, G. Maconachie, A. Pilat, R. J. McLean, F. A. Proudlock, and I. Gottlob. In vivo foveal development using optical coherence tomography. *Investigative ophthalmology & visual science*, 56(8):4537–4545, 2015.

[45] H. Lee, R. Purohit, V. Sheth, E. Papageorgiou, G. Maconachie, R. J. McLean, A. Patel, A. Pilat, S. Anwar, N. Sarvanathan, et al. Retinal development in albinism: a prospective study using optical coherence tomography in infants and young children. *The Lancet*, 385:S14, 2015.

[46] H. Lee, V. Sheth, M. Bibi, G. Maconachie, A. Patel, R. J. McLean, M. Michaelides, M. G. Thomas, F. A. Proudlock, and I. Gottlob. Potential of handheld optical coherence tomography to determine cause of infantile nystagmus in children by using foveal morphology. *Ophthalmology*, 120(12):2714–2724, 2013.

[47] C. Li, Y. Li, and X. Wu. Novel fuzzy c-means segmentation algorithm for image with the spatial neighborhoods. In *2012 2nd International Conference on Remote Sensing, Environment and Transportation Engineering*, pages 1–4. IEEE, 2012.

[48] K. Li, X. Wu, D. Z. Chen, and M. Sonka. Optimal surface segmentation in volumetric images-a graph-theoretic approach. *IEEE transactions on pattern analysis and machine intelligence*, 28(1):119–134, 2005.

[49] X. Li, L. Jin, E. Song, and L. Li. Full-range affinities for graph-based segmentation. In *2013 IEEE International Conference on Image Processing*, pages 4084–4087. IEEE, 2013.

[50] H. Lin, E. Long, X. Ding, H. Diao, Z. Chen, R. Liu, J. Huang, J. Cai, S. Xu, X. Zhang, et al. Prediction of myopia development among chinese school-aged children using refraction data from electronic medical records: a retrospective, multicentre machine learning study. *PLoS medicine*, 15(11):e1002674, 2018.

[51] B. D. Lucas. *Generalized image atching by the method of differences*. Carnegie Mellon University, 1985.

[52] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI'81: 7th international joint conference on Artificial intelligence*, volume 2, pages 674–679, 1981.

[53] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C. Chang, M. G. Yong, J. Lee, W. Chang, W. Hua, M. Georg, and M. Grundmann. Mediapipe: A framework for building perception pipelines. *CoRR*, abs/1906.08172, 2019.

[54] R. S. Maldonado, J. A. Izatt, N. Sarin, D. K. Wallace, S. Freedman, C. M. Cotten, and C. A. Toth. Optimizing hand-held spectral domain optical coherence tomography imaging for neonates, infants, and children. *Investigative ophthalmology & visual science*, 51(5):2678–2685, 2010.

[55] R. H. Masland. The neuronal organization of the retina. *Neuron*, 76(2):266–280, 2012.

[56] A. K. Mayer, M. Mahajnah, M. G. Thomas, Y. Cohen, A. Habib, M. Schulze, G. D. Maconachie, B. Al-Moallem, E. De Baere, B. Lorenz, et al. Homozygous stop mutation in ahr causes autosomal recessive foveal hypoplasia and infantile nystagmus. *Brain*, 142(6):1528–1534, 2019.

[57] J. T. McAllister, A. M. Dubis, D. M. Tait, S. Ostler, J. Rha, K. E. Stepien, C. G. Summers, and J. Carroll. Arrested development: high-resolution imaging of foveal morphology in albinism. *Vision research*, 50(8):810–817, 2010.

[58] J. A. Poulter, M. Al-Araimi, I. Conte, M. M. Van Genderen, E. Sheridan, I. M. Carr, D. A. Parry, M. Shires, S. Carrella, J. Bradbury, et al. Recessive mutations in slc38a8 cause foveal hypoplasia and optic nerve misrouting without albinism. *The American Journal of Human Genetics*, 93(6):1143–1150, 2013.

[59] C. Rother, V. Kolmogorov, and A. Blake. " grabcut" interactive foreground extraction using iterated graph cuts. *ACM transactions on graphics (TOG)*, 23(3):309–314, 2004.

[60] C. T. Rueden, J. Schindelin, M. C. Hiner, B. E. DeZonia, A. E. Walter, E. T. Arena, and K. W. Eliceiri. Imagej2: Imagej for the next generation of scientific image data. *BMC bioinformatics*, 18(1):1–26, 2017.

[61] S. R. Rufai, M. G. Thomas, R. Purohit, C. Bunce, H. Lee, F. A. Proudlock, and I. Gottlob. Can structural grading of foveal hypoplasia predict future vision in infantile nystagmus?: a longitudinal study. *Ophthalmology*, 127(4):492–500, 2020.

[62] F. Rüfer, A. Schröder, and C. Erb. White-to-white corneal diameter: normal values in healthy humans obtained with the orbscan ii topography system. *Cornea*, 24(3):259–261, 2005.

[63] N. Sarvananthan, M. Surendran, E. O. Roberts, S. Jain, S. Thomas, N. Shah, F. A. Proudlock, J. R. Thompson, R. J. McLean, C. Degg, et al. The prevalence of nystagmus: the leicestershire nystagmus survey. *Investigative ophthalmology & visual science*, 50(11):5201–5206, 2009.

[64] S. Shang, S. R. Kulkarni, P. W. Cuff, and P. Hui. A randomwalk based model incorporating social information for recommendations. In *2012 ieee international workshop on machine learning for signal processing*, pages 1–6. IEEE, 2012.

[65] R. F. Spaide and C. A. Curcio. Anatomical correlates to the bands seen in the outer retina by optical coherence tomography: literature review and model. *Retina*, 31(8):1609–1619, 2011.

[66] S. Srinivasan, G. Swaminathan, V. Kulothungan, S. Ganesan, T. Sharma, and R. Raman. Age-related macular degeneration in a south indian population, with and without diabetes. *Eye*, 31(8):1176–1183, 2017.

[67] M. Stocker. Be wary of'ethical'artificial intelligence. *Nature*, 540(7634):525–525, 2016.

[68] S. Sun, M. Sonka, and R. R. Beichel. Graph-based 4d lung segmentation in ct images with expert-guided computer-aided refinement. In *2013 IEEE 10th International Symposium on Biomedical Imaging*, pages 1312–1315. IEEE, 2013.

[69] M. G. Thomas, A. Kumar, S. Kohl, F. A. Proudlock, and I. Gottlob. High-resolution in vivo imaging in achromatopsia. *Ophthalmology*, 118(5):882–887, 2011.

[70] 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*, 118(8):1653–1660, 2011.

[71] M. G. Thomas, R. J. McLean, S. Kohl, V. Sheth, and I. Gottlob. Early signs of longitudinal progressive cone photoreceptor degeneration in achromatopsia. *British journal of ophthalmology*, 96(9):1232–1236, 2012.

[72] M. G. Thomas, E. Papageorgiou, H. J. Kuht, and I. Gottlob. Normal and abnormal foveal development. *British Journal of Ophthalmology*, 2020.

[73] L. Vajzovic, A. E. Hendrickson, R. V. O'Connell, L. A. Clark, D. Tran-Viet, D. Possin, S. J. Chiu, S. Farsiu, and C. A. Toth. Maturation of the human fovea: correlation of spectral-domain optical coherence tomography findings with histology. *American journal of ophthalmology*, 154(5):779–789, 2012.

[74] J. R. Zahn. Incidence and characteristics of voluntary nystagmus. *Journal of Neurology, Neurosurgery & Psychiatry*, 41(7):617–623, 1978.