



**UNIVERSITY OF LEEDS**

**Multi-Modal Deep Learning for  
Cardiovascular Disease Diagnosis and  
Risk Prediction Using Retinal Imaging**

**Cynthia Lizbeth Maldonado Garcia**

**Submitted in accordance with the requirements for the degree of Doctor  
of Philosophy in Computer Science**

**The University of Leeds**

**Faculty of Engineering and Physical Sciences**

**School of Computer Science**

**September, 2024**

# Intellectual Property and Publication

## Statements

I confirm that the work submitted is my own, except where work that has formed part of jointly authored publications has been included. My contribution and the contributions of the other authors to this work have been explicitly indicated below. I confirm that appropriate credit has been given within the thesis where reference has been made to the work of others.

Details of jointly authored work included in the thesis:

**Chapter 4:** This chapter is based on the following jointly-authored publication:

**Title:** "Predicting Risk of Cardiovascular Disease Using Retinal OCT Imaging"

**Authors:** Cynthia Maldonado-Garcia, Rodrigo Bonazzola, Enzo Ferrante, Thomas H. Julian, Panagiotis I. Sergouniotis, Nishant Ravikumar, Alejandro F. Frangi

**Journal:** arXiv, 2024

**Contribution:** Cynthia Maldonado-Garcia, conceptualization, data curation, develop and implementation of methodology, analysis, investigation, writing original draft; Alejandro F. Frangi, supervision, conceptualization, review and editing of the manuscript; Nishant Ravikumar, supervision, conceptualization, review and editing of the manuscript; Rodrigo Bonazzola, review and editing of the manuscript; Enzo Ferrante, review and editing of the manuscript; Thomas H. Julian, review and editing of the manuscript; Panagiotis I. Sergouniotis, review and editing of the manuscript.

**Chapter 5:** This chapter is based on the following jointly-authored publication:

**Title:** "Integrating Deep Learning with Fundus and Optical Coherence Tomography for Cardiovascular Disease Prediction"

**Authors:** Cynthia Maldonado-Garcia, Arezoo Zakeri, Alejandro F. Frangi, and Nishant Ravikumar

**Journal:** Annual Conference on Medical Image Computing and Computer-Assisted Interventions, 2024

---

**Contribution:** Cynthia Maldonado-Garcia, conceptualization, data curation, develop and implementation of methodology, analysis, investigation, writing original draft; Nishant Ravikumar, supervision, conceptualization, review and editing of the manuscript; Arezoo Zakeri, supervision, review and editing of the manuscript; Alejandro F. Frangi, supervision, review and editing of the manuscript;

# Acknowledgements

Behind this thesis lies the support of many wonderful people, to whom I am deeply grateful.

To begin with, I would like to thank my supervisor, Dr. Nishant Ravikumar, for guiding me throughout this long PhD journey. His experience, patience, and knowledge have greatly improved the quality of my work and motivated me to work harder. I would also like to thank my former supervisor, Professor Alejandro Frangi, for giving me the opportunity to be part of his group. His guidance helped me understand the importance of work independently and provided me with valuable opportunities to collaborate with others in the community.

I am also grateful to my examiners, Dr. Samuel Relton and Professor Emanuele Trucco, for the detailed feedback I received during my viva. Their insights not only improved the quality of my work but also taught me to be more critical in my research. This experience will be incredibly valuable as I continue my research career.

In addition, I would like to thank the software research team, Alejandro and Kattia, for their technical support and for always being willing to help when needed. A special thank you to my CISTIB colleagues who began this PhD journey with me, it was a pleasure to grow and learn alongside you. In particular, I'm deeply grateful to Nurbanu and Rodrigo for their endless brainstorming sessions, and above all, your friendship. The time we shared made my PhD experience much more enjoyable.

Finally, I want to thank my family for always believing in me and giving me the strength to pursue this dream. My deepest gratitude goes to Victor, for being my happy and safe place, and for walking this labyrinth with me. This thesis is dedicated to all of you.

This work was supported by the Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAH-CYT) (scholarship no. 766588).

# Abstract

Cardiovascular diseases (CVDs) remain a major global health concern. Early identification of individuals at risk of CVDs is essential for effective preventive care, reducing healthcare costs, and improving patient outcomes. Retinal imaging has recently emerged as a noninvasive method of detecting microvascular alterations that might enable earlier identification and targeting of at-risk patients. This thesis leverages retinal imaging from the UK Biobank, combined with advanced deep learning techniques, to develop predictive models of CVDs risk and investigate causal links between cardiovascular outcomes and retinal features.

First, we investigated the use of optical coherence tomography (OCT), combined with minimal clinical data, to estimate CVDs risk by developing a convolutional variational autoencoder and a random forest framework. A novel explainability method was proposed to identify clinically interpretable retinal biomarkers. Second, we demonstrated the synergistic value of multimodal retinal imaging through a multi-channel variational autoencoder and a transformer-based classifier architecture that jointly analyses OCT and fundus photographs. An explainability model was applied to highlight the most relevant features from both retinal imaging modalities for classification tasks. Finally, through genome-wide association studies of nnU-Net-derived OCT phenotypes and Mendelian randomization analyses, we identified genetic variants and established causal relationships between cardiovascular traits and specific retinal layer alterations.

These findings collectively advance our understanding of retinal-cardiovascular interactions, providing, computational evidence that multimodal retinal imaging reveals biomarkers linked to systemic vascular health, methodological frameworks for multimodal ophthalmic data integration, and genetic evidence supporting causal links between cardiovascular traits and retinal alterations. The study bridges artificial intelligence and retina imaging data, offering novel insights into the estimation of CVD risk while highlighting the retina's potential as a window to cardiovascular health.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	1
1.2	Aims and objectives . . . . .	3
1.3	Contributions . . . . .	3
1.4	Thesis structure and chapter summaries . . . . .	4
<b>2</b>	<b>Background</b>	<b>7</b>
2.1	UK Biobank data . . . . .	7
2.2	Retina imaging . . . . .	9
2.2.1	Optical coherence tomography . . . . .	9
2.2.2	Fundus photography . . . . .	11
2.3	Single nucleotide polymorphisms microarrays . . . . .	12
2.4	Machine learning methods . . . . .	14
2.4.1	Random forest . . . . .	14
2.4.2	Neural networks . . . . .	15
2.5	Deep learning . . . . .	16
2.5.1	Fully connected networks . . . . .	17
2.5.2	Convolutional neural networks . . . . .	18
2.6	Autoencoders . . . . .	21
2.6.1	Variational autoencoders . . . . .	22
2.6.2	Multi-Channel variational autoencoder. . . . .	23
2.7	Segmentation . . . . .	24
2.7.1	UNet . . . . .	24
2.7.2	nnUNet . . . . .	25

2.8	Genome-wide association studies . . . . .	25
2.8.1	Single-variant models. . . . .	25
2.9	Mendelian randomization . . . . .	27
2.9.1	Inverse variance weighted . . . . .	28
<b>3</b>	<b>Literature review</b>	<b>30</b>
3.1	Cardiovascular diseases . . . . .	30
3.2	Deep learning and retinal imaging . . . . .	31
3.2.1	Retinal imaging for assessing CVD risk factors . . . . .	32
3.2.2	Retinal imaging for assessing CVD events . . . . .	38
3.2.3	Limitations in the current literature review . . . . .	43
3.3	Genetic and causal factors associated to cardiovascular disease risk via retinal imaging . . . . .	44
3.3.1	Limitations in the Current Literature Review . . . . .	47
<b>4</b>	<b>Predicting cardiovascular diseases using optical coherence tomography imaging and demographic and clinical data</b>	<b>48</b>
4.1	Introduction . . . . .	48
4.2	Database . . . . .	49
4.3	Methodology . . . . .	53
4.3.1	Framework of self-supervised feature selection Variational Autoencoder and multimodal random forest classification . . . . .	53
4.3.2	Model explainability . . . . .	56
4.4	Experiments . . . . .	57
4.5	Results . . . . .	59
4.5.1	Classification performance . . . . .	59
4.5.2	Ablation study . . . . .	63
4.5.3	Model explainability . . . . .	63
4.6	Discussion . . . . .	67
<b>5</b>	<b>Integrating deep Learning with fundus and optical coherence tomography for cardiovascular disease prediction</b>	<b>70</b>
5.1	Introduction . . . . .	70
5.2	Dataset . . . . .	71

5.3	Methodology . . . . .	72
5.3.1	Framework of Task-Aware Multi-Channel Variational Autoencoder (task-aware MCVAE) . . . . .	72
5.4	Experiments . . . . .	74
5.5	Results . . . . .	75
5.5.1	Classification performance . . . . .	75
5.5.2	Ablation study . . . . .	78
5.5.3	Interpretation. . . . .	79
5.6	Discussion . . . . .	81
<b>6</b>	<b>Mendelian randomisation studies of OCT-derived retinal phenotypes and cardiovascular disease</b>	<b>84</b>
6.1	Introduction . . . . .	84
6.2	Dataset . . . . .	85
6.2.1	OCT imaging data from the UK Biobank . . . . .	85
6.2.2	SNP microarrays used in the UK Biobank. . . . .	85
6.2.3	Genotype imputation in the UK Biobank. . . . .	87
6.3	Methodology . . . . .	88
6.3.1	Optical coherence tomography imaging segmentation . . . . .	88
6.3.2	Phenotypes of thickness maps of variational autoencoder . . . . .	89
6.3.3	Mendelian randomization . . . . .	89
6.4	Experiments . . . . .	89
6.5	Results . . . . .	90
6.5.1	Results of qualitative segmentation . . . . .	90
6.5.2	Results from qualitative and quantitative analysis of thickness maps . . . . .	90
6.5.3	Results of the GWAS . . . . .	92
6.5.4	Results of mendelian randomisation . . . . .	99
6.6	Discussion . . . . .	101
<b>7</b>	<b>Conclusions</b>	<b>104</b>
7.1	Summaries and key findings . . . . .	104
7.1.1	Using optical coherence tomography as a predictive tool for cardiovascular diseases. . . . .	104

7.1.2	A multimodal deep learning method using fundus photographs and optical coherence tomography for predicting cardiovascular diseases. . . . .	105
7.1.3	Genetic basis and causal links influencing OCT-derived phenotypes and cardiovascular diseases. . . . .	106
7.2	Implications and significance . . . . .	106
7.3	Limitations, challenges, and opportunities for future research . . . . .	107
7.4	Final remarks . . . . .	109
<b>References</b>		<b>109</b>
<b>A Supplementary information to Chapter 4</b>		<b>130</b>
A.1	Supplementary information to 4.3 . . . . .	130
A.2	Supplementary information to 4.5 . . . . .	135
A.3	Links to Docker image and Github . . . . .	135
<b>B Supplementary information to Chapter 5</b>		<b>136</b>
B.1	Supplementary information to 5.2 . . . . .	136
B.2	Supplementary information to 5.3 . . . . .	139
B.3	Links to Docker image and Github . . . . .	141
<b>C Supplementary information to Chapter 6</b>		<b>142</b>
C.1	Supplementary information to 6.2 . . . . .	142
C.2	Supplementary information to 6.3 . . . . .	142
C.2.1	GWAS databases links used as exposure data in the Mendelian randomization Analysis . . . . .	142
C.3	Links to Docker image and Github . . . . .	143

# List of Figures

2.1	Schematic representation of retinal optical coherence tomography (adapted from Farrah et al. 2020)	10
2.2	Scan types with Optical Coherence Tomography (adapted from Leica-Microsystems 2025)	11
2.3	Examples of three 64th B-scan OCT images, each from a different patient.	11
2.4	Schematic representation of Fundus Photographs (adapted from Zhang et al. 2024)	12
2.5	Examples of three fundus images, each from a different patient.	13
2.6	The schema of Single nucleotide polymorphism (SNP) (adapted from Wikipedia)	13
2.7	A neural network with three input variables, two hidden layers and an output (adapted from Waring et al. 2020)	16
2.8	Example of a fully connected network (adapted from Ramsundar 2019)	18
2.9	Example of a convolutional neural network (adapted from Ramsundar 2019)	19
2.10	Example of a simple autoencoder	21
4.1	Comparison of the distribution of the Quality Index from both eyes.	50
4.2	STROBE flow chart describing participant inclusion and exclusion criteria applied to define the study cohort.	50
4.3	Bar plots showing the methods used to determine the onset date of stroke or MI for the 612 CVD+ subjects involved in our classification task	51
4.4	Distribution of age-sex cohort match. M, Male. F, Female. The left histogram illustrates the total of CVD+ labeled data used solely for the classification task. The middle histogram shows the total of CVD- for the classification task, while the right histogram illustrates CVD- used for the pretraining task.	53

4.5 The workflow diagram illustrates the comprehensive process of training the Variational Autoencoder (VAE) and subsequently using it to acquire the latent vectors (upper section). These latent vectors are then combined with metadata and serve as inputs to the Random Forest (RF) classifier (middle section). Finally, we perform an interpretability analysis by perturbing the most relevant features, reconstructing the corresponding image and computing the vector field between the perturbed reconstructions (lower section).  $z_{left}$  represents the latent vector obtained from the training of the VAE for the left eye.  $Z_{right}$  corresponds to the latent vector acquired from training the VAE for the right eye. . . . . 54

4.6 Comparison of classifier performance in terms of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Values above the bars represent the corresponding percentages for each classifier. . . . . 62

4.7 Calculation of feature importance magnitudes for the seven different classifiers investigated, where each classifier uses different combinations of data channels/modalities. Latent variables starting with 'zr' refer to the right eye, while those starting with 'zl' refer to the left eye. . . . . 64

4.8 Global explanations of features from different data modalities/channels which were considered important by the predictive model for separating the CVD+ and CVD- groups. Bar plot summarises the relative importance of latent variables from left ( $zl$ ) and right ( $zr$ ) eye OCT images and patient metadata, as percentages. . . . . 65

4.9 Vector field maps presented for three different B-scans for the left eye. The top row corresponds to the 1st B-scan, the middle image for the 64th B-scan while the bottom row depicts the final B-scan. The yellow circles represent the regions that the vector field maps highlight when modifying the latent variable  $zl066$ . . . . . 66

5.1 STROBE flow chart outlines the patient selection process for a study using the UK Biobank dataset with OCT and Fundus images of the left eye. . . . . 72

5.2	The illustration shows the architecture of our model. During the pre-training phase (left), fundus photographs and OCT images are processed through a 2D CNN and a 3D CNN encoder-decoder network (in a multi-channel VAE configuration), respectively. In the fine-tuning phase (right), the images are processed through the pre-trained encoder-decoder networks. The resulting latent vectors are aggregated and input into a transformer architecture, followed by a fully connected layer. The model undergoes end-to-end training in both phases. . . . .	73
5.3	The bar chart displays the classification metrics for Fundus, OCT, and combined Fundus-OCT modalities. The metrics shown include Accuracy, Precision, Sensitivity, Specificity, and AUC (Area Under the Curve) values, with their respective standard deviations. The colors represent different modalities: Fundus (black), OCT (red), and Fundus-OCT (yellow). Each bar is labeled with its corresponding value. . . . .	76
5.4	Comparison of classifier performance in terms of TP, TN, FP, and FN. Values above the bars represent the corresponding percentages for each classifier. . . . .	79
5.5	The top 5 latent vectors $z_{58}$ , $z_{112}$ , $z_{24}$ , $z_{106}$ , and $z_{101}$ are identified based on their SHAP values, which measure the contribution of each feature to the model's prediction of cardiovascular disease risk. . . . .	80
5.6	a) The top row displays a sequence of Optical Coherence Tomography (OCT) images, highlighting the different layers of the retina with yellow markers indicating regions of interest. b) The bottom row shows corresponding fundus images with vascular structures and regions of interest also marked in yellow. These images are utilized for analyzing and predicting cardiovascular disease risks based on retinal biomarkers. . . . .	81
6.1	STROBE flow chart describing participant inclusion and exclusion criteria applied to define the study cohort. . . . .	86
6.2	Example of a 64th B-scan OCT, displaying the boundaries of 10 layers. The corresponding labels for these layers are provided in the figure on the right. . . . .	88
6.3	An example of the nnUNet segmentation results for the ten layers of OCT. . . . .	91
6.4	Thickness maps generated from Optical Coherence Tomography (OCT) data, illustrating the Variational Autoencoder (VAE) model's performance in analyzing 10 layers of tissue. . . . .	91

A.1 Captum images present three different B-scans for the left eye. The top row corresponds to the 1st B-scan, the middle image to the 64th B-scan, and the bottom row depicts the final B-scan. The yellow zones highlight regions of the ablation method when modifying the latent variable  $z_{l066}$ . . . . . 135

B.1 Heatmaps indicating the number of patients with different CVDs and fundus photographs, where the diagonal represents the number of patients with a given diagnosis and the off-diagonal elements indicate co-occurrence of pairs of diagnoses. . . . . 137

B.2 Heatmaps indicating the number of patients with different CVDs and OCT, where the diagonal represents the number of patients with a given diagnosis and the off-diagonal elements indicate co-occurrence of pairs of diagnoses. . . . . 138

# List of Tables

3.1	Overview of studies using AI for predicting CVD risk factors from retinal images . . . . .	35
3.2	Overview of studies using AI for predicting CVD events from retinal images . . . . .	40
4.1	Characteristics of patients in the CVD+ and CVD- sets. N, Never. P, Previous. C, Current. NA, Not answer. . . . .	52
4.2	Predictive analysis of cardiovascular disease (CVD) metrics utilizing UK Biobank data across seven distinct cases. . . . .	60
4.3	Comparison between various classifiers based on different configurations of eye data (both eyes, left eye, right eye) and metadata inclusion. The chi-squared ( $\chi^2$ ) values and associated p-values indicate the statistical significance of differences between these classifiers. . . . .	61
4.4	Comparison of classification metric results between our model employing both ocular data and metadata (BE-MTDT) and the QRISK algorithm. McNemar's Test ( $*p < 0.005$ )	62
4.5	Comparison of classification metric results between our model employing both ocular data and metadata (BE-MTDT-RF) and the MLP algorithm. . . . .	63
5.1	95% CI for classification metrics of the Fundus, OCT, and Fundus-OCT classifiers. Each interval represents the range within which the true metric value is expected to lie with 95% confidence, providing insight into the precision and reliability of the classifiers' performance. . . . .	77
5.2	Comparative analysis of our model (Fundus-OCT) and QRISK algorithm in predictive classification metric. . . . .	78
5.3	Classification metrics for the retina modalities (Ablation Study). This table compares the performance of classifiers trained without the end-to-end approach, highlighting the reduced efficacy across all metrics when the MCVAE is not trained alongside the classifier.	79

6.1	Root Mean Square Error (RMSE) for each retinal layer. . . . .	91
6.2	This table presents the results of our GWAS, including: significant SNPs, their closest genes, prior known associations with retinal phenotypes from existing GWAS literature, and classification as either novel or previously reported in retinal GWAS. . . . .	93
6.3	Summary of statistically significant results from forward MR analyses. The table presents effect size estimates, standard errors, and $p$ -values for associations between exposures (e.g., cardiovascular traits) and outcomes (retinal latent vectors). IVW, inverse-variance weighted. . . . .	100
6.4	This table summaries the top associations (lowest $p$ -values) from analyses testing retinal latent vectors (exposures) against cardiovascular outcomes, including effect estimates ( $\beta$ ), standard errors (SE), and statistical significance ( $p$ values). . . . .	101
A.1	Hyperparameter grid search for the VAE model. Tested values include architectural choices (CNN channels, latent space size), optimization parameters (learning rate, weight decay), and data splits. The optimal configuration was selected via 5 cross-validation. . .	130
A.2	Architectural design of the VAE, showing the encoder and decoder structures with their respective layers, configurations, and dimensional transformations. . . . .	131
A.3	Hyperparameter grid search for RF models. Tested ranges and optimal values are shown for each variant (BE/LE/RE-MTDT-RF, BE/LE/RE-RF, and MTDT-RF). Step sizes are indicated in parentheses for ranged parameters. . . . .	132
A.4	Hyperparameter grid search for the MLP model. Tested values include architectural choices (CNN channels), optimization parameters (learning rate, weight decay), and data splits. The optimal configuration was selected via 5 cross-validation. . . . .	133
A.5	List of Variables used to calculate QRISK3 . . . . .	134
B.1	Hyperparameter grid search for the MCVAE model. Tested values include architectural choices (CNN channels, latent space size), optimization parameters (learning rate, weight decay), and data splits. The optimal configuration was selected via 5-fold cross-validation. . . . .	139
B.2	Architectural design of the VAE for fundus images, showing the encoder and decoder structures with their respective layers, configurations, and dimensional transformations. .	140
B.3	Architectural design of the VAE for OCT images, showing the encoder and decoder structures with their respective layers, configurations, and dimensional transformations. .	141

C.1	Hyperparameter grid search for the VAE model. Tested values include architectural choices (CNN channels, latent space size), optimization parameters (learning rate, weight decay), and data splits. The optimal configuration was selected via 5-fold cross-validation.	142
C.2	Architectural design of the VAE for thickness maps of OCT retinal layers, showing the encoder and decoder structures with their respective layers, configurations, and dimensional transformations. . . . .	143

# Publications

1. **Maldonado-García, C.**, Zakeri, A., Frangi, A.F., Ravikumar, N. (2025). Integrating Deep Learning with Fundus and Optical Coherence Tomography for Cardiovascular Disease Prediction. In: Rekik, I., Adeli, E., Park, S.H., Cintas, C. (eds) Predictive Intelligence in Medicine. PRIME 2024. Lecture Notes in Computer Science, vol 15155. Springer, Cham. [https://doi.org/10.1007/978-3-031-74561-4\\_14](https://doi.org/10.1007/978-3-031-74561-4_14)
2. **Maldonado-García, C.**, Bonazzola, R., Ferrante, E., Julian, T.H., Sergouniotis, P.I., Ravikumar, N., and Frangi, A.F. (2024). Predicting risk of cardiovascular disease using retinal OCT imaging. ArXiv, abs/2403.18873. (Under review in Frontiers in Artificial Intelligence). <https://arxiv.org/abs/2403.18873>
3. Girach, Z., Sarian, A., **Maldonado-García, C.**, Ravikumar, N., Sergouniotis, P.I., Rothwell, P.M., Frangi, A.F., and Julian, T.H. (2024). Retinal imaging for the assessment of stroke risk: a systematic review. Journal of neurology. <https://link.springer.com/article/10.1007/s00415-023-12171-6>
4. **Maldonado-García, C.**, Bonazzola, R., Ravikumar, N., and Frangi, A.F. (2022). Predicting Myocardial Infarction Using Retinal OCT Imaging. In: Yang, G., Aviles-Rivero, A., Roberts, M., Schönlieb, CB. (eds) Medical Image Understanding and Analysis. MIUA 2022. Lecture Notes in Computer Science, vol 13413. Springer, Cham. [https://doi.org/10.1007/978-3-031-12053-4\\_58](https://doi.org/10.1007/978-3-031-12053-4_58)

# Chapter 1

## Introduction

### 1.1 Background

Cardiovascular diseases (CVD) are a major global health concern and the leading cause of death worldwide. In 2021, 20.5 million people died from cardiovascular conditions, accounting for one-third of all global deaths (World-Heart et al., 2023). Ischaemic heart disease is the leading cause of premature death in 146 countries for males and 98 countries for females, with up to 80% of premature cardiovascular conditions being potentially preventable. Furthermore, a significant health disparity is evident, with approximately 80% of CVD-related fatalities occurring in low- and middle-income countries, while advancements in cardiovascular health are predominantly observed in high-income nations (Lindstrom et al. 2023; D'Agostino et al. 2013). This inequity underscores an urgent need for targeted interventions and global health initiatives. Notably, early identification of individuals at risk is crucial since premature CVD is highly preventable. Effective primary prevention strategies can lead to a decrease in CVD mortality and morbidity, as demonstrated in several previous clinical studies (Littlejohns et al. 2019; Wong et al. 2022).

Emerging research suggests that retinal examination, direct imaging of the light-sensitive tissue lining the back of the eye, may offer insights into certain systemic diseases, including microvascular abnormalities and central nervous system disorders. However, such findings remain investigational and are not yet used in clinical practice. Advances in imaging technology have enabled the cost-effective assessment of retinal parameters on a large scale, both in hospital settings and within community optician practices (Farrah et al., 2019; Girach et al. 2024). This is facilitated by imaging techniques such as fundus photography and optical coherence tomography (OCT). Fundus photography captures images of

the retina and optic nerve (Zhang et al. 2024), while OCT provides a more detailed, non-invasive, in vivo cross-sectional view of retinal tissues (Farrah et al. 2020).

The retinal and choroidal microvasculature have been identified as sensitive indicators of systemic vascular conditions, including those affecting the cerebral and coronary vasculature (Farrah et al. 2020). Thus, examining the microvasculature at the back of the eye through retinal imaging offers a valuable opportunity to identify individuals at risk of common yet serious cardiovascular diseases, such as stroke and myocardial infarction. The non-invasive assessment of the microvasculature via retinal imaging provides a practical means of detecting microvascular dysfunction in the peripheral vasculature, which may aid in the early identification of individuals at heightened risk of cardiovascular disease (Anderson et al. 1995; Farrah et al. 2020; Rudnicka et al. 2022; Diaz-Pinto et al. 2022).

Taking advantage of its non-invasive nature and cost-effectiveness, retinal imaging is increasingly recognised in research as a valuable tool for the early detection of cardiovascular disease, which is essential for delivering timely preventive care and formulating effective treatment strategies (Wagner et al. 2020). Retinal imaging techniques are routinely employed in hospital eye clinics and most optometric practices, highlighting their potential as a widely accessible method for assessing cardiovascular disease risk. The role of the retina in systemic diseases is well established in research (Flammer et al. 2013; Hanssen et al. 2022; Wong et al. 2022), with eye care professionals frequently diagnosing various systemic conditions, such as diabetes, hypertension, and atherosclerosis, based on retinal appearance. As advancements in imaging technology enhance the observation of retinal structure, a growing body of research is focused on exploring the associations between retinal features and cardiovascular disease risk factors, as well as developing predictive models for the early identification of individuals at risk of cardiovascular disease.

The intricate nature of biological tissues, organs, and disease processes necessitates the use of multi-modal imaging techniques within the medical field to accurately characterise disease phenotypes and extract clinically relevant quantitative information (Rajiah et al. 2019; Govindarajan et al. 2005). Recent advancements in artificial intelligence (AI) methodologies for computational image analysis and phenotype extraction across multiple imaging modalities have shown considerable promise in elucidating the underlying anatomical and physiological alterations associated with CVD (Amal et al. 2022; Milosevic et al. 2024). As a result, the development of advanced statistical and machine learning techniques for the effective analysis and integration of these diverse data sources is crucial for improving patient care outcomes.

## 1.2 Aims and objectives

The general aim of this thesis is to investigate and demonstrate how retinal medical imaging, in conjunction with advanced artificial intelligence techniques, can serve as an approach for the early prediction and prevention of cardiovascular diseases. This research is guided by the following objectives:

- To implement deep convolutional autoencoder-like algorithms for extracting clinically relevant features from OCT, which are predictive of cardiovascular diseases, and apply it to the population-scale OCT dataset from UK Biobank of over 60,000 subjects. To learn a rich representation of the dataset we aim to pretrain our algorithm with around 20,000 of participants.
- To implement interpretability algorithms which highlight the most relevant retinal features for the task of predicting CVD, thereby enhancing clinical understanding and decision-making processes.
- To employ state-of-the-art DL-based architectures to perform segmentation of the different retinal layers to obtain a finer-grained understanding of them.
- To perform genome-wide association studies on features obtained of retinal layer-wise thickness maps by using an unsupervised autoencoder-like phenotyping algorithm, to understand the genetic basis of these features.
- To employ Mendelian randomization techniques for exploring the biological mechanisms underlying the association between identified retinal biomarkers and CVD and establishing potential causal relationships.
- To assess the potential of retinal multi-modal phenotyping, using fundus photographs and OCT simultaneously, in order to improve the predictive performance for CVDs, by using multi-channel variational autoencoders to perform unsupervised phenotypes and a transformer-based head for carrying out the final prediction.

## 1.3 Contributions

This thesis makes different contributions, which are outlined below:

### **Methodological contributions:**

- A two-stage framework: First, a deep convolutional autoencoder is used to learn rich representa-

tions of OCT scans from 20,000 images (both eyes). Second, a random forest classifier estimates the risk of CVD using data from patients at increased risk and a matched control group.

- An explainable algorithm based on vector fields that highlights relevant features in retinal images contributing to the estimation of CVD risk in patients.
- An end-to-end multimodal retinal framework that combines a multi-channel variational autoencoder for unsupervised phenotype extraction from OCT and fundus images, with a transformer-based classification head for cardiovascular risk prediction.
- A segmentation framework utilising the nnU-Net model to segment 10 retinal OCT layers using limited annotated data (50 ground truth images).

#### **Applications in retinal–cardiovascular interaction:**

- Demonstrated the utility of OCT imaging in estimating cardiovascular risk, having highest predictive results that using metadata. The choroidal layer identified as a particularly relevant feature.
- Developed a multimodal imaging approach using both fundus photographs and 3D OCT volumes, showing that their combined use provides complementary information for predicting CVD.

#### **Genetic and causal insights:**

- Expanded the number of genetic loci associated with retinal OCT layer thickness through genome-wide association studies.
- Provided evidence of causal links between cardiovascular traits and retinal thickness using forward-direction Mendelian randomization analyses.

## **1.4 Thesis structure and chapter summaries**

**Chapter 1** provides a comprehensive introduction to the thesis. It outlines the significance of cardiovascular diseases, the potential of retinal imaging in disease prediction, and presents the aim and objectives of the research.

**Chapter 2** establishes the theoretical and contextual groundwork for the research. It covers fundamental concepts underpinning the study, including:

- Description of the UK Biobank dataset, retina imaging, and genetic data used in the thesis experiments
- Deep learning architectures utilised in this work, such as convolutional neural networks, auto-encoder models, and segmentation models (e.g., nn-UNet)
- Concepts related to genetic association studies and Mendelian randomisation

The goal of this chapter is to provide a comprehensive background with the necessary theory to facilitate understanding of the methodologies proposed in subsequent chapters.

**Chapter 3** presents a critical review of recent literature, focusing on:

- The application of artificial intelligence methodologies with retinal medical images for two main objectives: assessing CVD risk factors and biomarkers, and CVD events.
- Studies investigating causal relationships between CVD risk factors or events and retinal phenotypes, with emphasis on deep learning methodologies

This review highlights the importance of retinal images, AI methods, and their intersection with cardiovascular disease research

**Chapter 4** details our published work, proposing a framework that utilises OCT features in an unsupervised learning approach to extract relevant features and then combine them with clinical and demographic data to predict patients likely to experience a cardiac event. The key contributions in this chapter include:

- Development of a predictive model that merges multi-modal patient data (e.g., OCT imaging, demographic, and clinical information).
- An innovative method for enhancing model interpretability, offering detailed localisation of the features in retinal layers that most impact accurately identifying patients at risk of adverse CVD events.
- A pioneering investigation into the application of 3D OCT imaging and artificial intelligence for automatically forecasting patients vulnerable to adverse CVD incidents.
- Systematic evaluation of the contribution when combining OCT images from both eyes, a single-eye OCT, or integrating OCT data with clinical and demographic information.

**Chapter 5** presents work from our published paper, where we develop an innovative end-to-end mul-

timodal deep generative learning model and transformer classifier. This model extracts retina-derived phenotypes from both fundus and OCT images to classify patients at risk of cardiovascular disease. Key contributions include:

- A novel predictive model that integrates multimodal retinal imaging (OCT and fundus retinal data) by employing a multichannel variational autoencoder and a transformer network classifier to identify individuals at risk of developing cardiovascular diseases within five years of image acquisition.
- A pioneering study into the application of fused 3D OCT and fundus imaging to enhance the prediction of patients at risk of cardiac events.
- An ablation study that reveals the impact of training our feature extraction model simultaneously with classification, compared to performing these tasks separately.
- Systematic analysis quantifying the performance contributions of individual retinal modalities, demonstrating that the combined use of both retinal imaging modalities enhances model performance.

**Chapter 6** describes a genome-wide association study to identify associations between genetic variations and retina-derived phenotypes. Additionally, it employs Mendelian randomisation techniques to investigate the biological factors underlying the causal relationship between cardiovascular disease and retinal morphology. The key contributions in this chapter include:

- First-time application of state-of-the-art nnUNet model for ten OCT layers segmentation.
- First-time application of genome-wide association study to autoencoder phenotypes involving ten OCT layers.
- Identification of new genetic loci implicated in OCT-derived phenotypes.
- Establishment of a causal relationship between cardiovascular incidents and OCT features using Mendelian randomization analysis.

**Chapter 7** The final chapter presents a comprehensive overview of the thesis, synthesising contributions, key findings and deriving conclusions. It also addresses limitations of the study and proposes avenues for future research in this field.

## Chapter 2

# Background

This chapter aims to establish a robust foundation for the subsequent chapters by presenting key concepts, contextual information, data details, and essential background necessary to support the study.

### 2.1 UK Biobank data

The UK Biobank is a large-scale, community-based prospective cohort study conducted in the United Kingdom, recognised for its extensive and detailed data collection (UK Biobank 2024; Sudlow et al. 2015). The study recruited 500,000 participants aged 40–69 years at the time of enrolment between 2006–2010, who will be followed for many years through linkage to their health-related records (UK Biobank 2024). Comprehensive baseline data were gathered from multiple sources, including detailed questionnaires addressing general health, disabilities, socio-demographic factors, and lifestyle behaviours such as smoking, alcohol consumption, and diet (Bycroft et al. 2018).

In addition, a wide range of physical measurements were recorded, including electrocardiography, exercise tolerance tests, spirometry, and bone density assessments. Medical imaging data, such as magnetic resonance imaging (MRI) and retinal imaging, were also acquired. Biological samples—blood, urine, and saliva—were collected for further analysis, with DNA extracted from blood samples used for high-throughput genotyping across all participants, generating an extensive genetic dataset (UK Biobank 2024).

The UK Biobank’s rich array of data offers an unparalleled resource for advancing the understanding of disease risk factors and their interplay with genetic and environmental variables. All participants provided written, informed consent prior to their inclusion in the study (UK Biobank 2024; Sudlow et

al. 2015).

The UK Biobank is a valuable resource for investigating the complex interplay between genetic, environmental, and lifestyle factors in relation to health outcomes (UK Biobank 2024; Sudlow et al. 2015). Its extensive data collection, including biological samples, imaging, and detailed participant information, enables robust analyses that contribute to the advancement of medical and epidemiological research. However, despite its strengths, the UK Biobank has notable limitations in terms of generalizability.

The cohort is not fully representative of the broader UK population, with differences observed in ethnicity, socioeconomic status, health behaviors, and disease prevalence. The overrepresentation of healthier, wealthier, and more health-conscious individuals introduces selection bias, limiting its use for deriving accurate population-level disease prevalence and incidence rates. For instance, findings indicate that 94.6% of UK Biobank participants were of white ethnicity, a proportion higher than the 91.3% reported in the 2011 UK Census. Regarding sociodemographic factors, UK Biobank participants were more likely to own their property outright (56.7%) compared to 40.6% in the 2001 UK Census. Conversely, 39.6% of individuals in the 2001 UK Census had a mortgage or loan, compared to 33.9% of UK Biobank participants. Moreover, UK Biobank participants were less likely to live in rental accommodations than the general population within the same age range (50–64 years) (Batty et al. 2020; Fry et al. 2017).

Additionally, lifestyle and health-related differences compared to national survey data suggest caution in extrapolating absolute risk estimates from the UK Biobank to the general public. UK Biobank participants, both men and women, were less likely to be current smokers across all age groups compared to the Health Surveys for England (HSE) 2008. Furthermore, while they were less likely to be lifelong abstainers from alcohol, they were also less likely to consume alcohol daily compared to the general population included in HSE 2008. Regarding health conditions, UK Biobank participants exhibited a lower prevalence of cardiovascular disease, stroke, hypertension, diabetes, chronic kidney disease, and respiratory diseases compared to the general population, based on various HSE datasets (Fry et al. 2017).

Nevertheless, the large sample size, high-quality data collection, and long-term follow-up make the UK Biobank an exceptional platform for studying associations between exposures and health outcomes. While absolute disease frequencies may not be directly applicable to other populations, the relationships identified between risk factors and diseases remain highly informative and generalizable when interpreted within the appropriate context. Despite its limitations, the UK Biobank remains a powerful tool for advancing scientific knowledge on disease etiology and public health.

## 2.2 Retina imaging

Optical Coherence Tomography (OCT) scans and fundus photography are non-invasive techniques that enable high-resolution imaging of the retina, making them highly effective for detecting subtle retinal changes. These methods offer a safe, cost-effective approach for diagnosing and monitoring the progression of eye diseases, and also play a crucial role in identifying individuals at risk of developing ocular conditions, facilitating early intervention. Furthermore, OCT and retinal photography provide insights into ageing processes and systemic conditions affecting the cardiovascular and neurological systems, by allowing direct visualisation of a highly vascularized part of the central nervous system (CNS) tissue and its vasculature. As a result, these imaging techniques serve as a valuable tool for assessing retinal health, which may reflect broader systemic health (Warwick et al. 2022).

### 2.2.1 Optical coherence tomography

OCT is a non-invasive, high-resolution imaging technique that uses light waves to capture cross-sectional images of biological tissues. By scanning vertically into tissue, OCT reveals internal structures, such as retinal layers (nerve fibers, photoreceptors) or skin strata, in micrometer-scale detail. OCT relies on reflected light, making it ideal for semi-transparent tissues (e.g., retina, blood vessels, or cornea). Often called 'optical ultrasound,' it provides real-time, subsurface visualization with exceptional precision (Huang et al. 1991).

The technical principles underlying OCT are akin to those of ultrasound; however, instead of sound waves, OCT measures light reflections. A low-coherence light source, typically in the near-infrared range, is employed. OCT operates on the principle of interferometry, comparing reflected light from the sample with a reference beam. The light beam is split into two paths: one directed at the tissue sample, while the other serves as the reference. The sample beam is scanned across the tissue, penetrating to various depths. Light reflections from different retinal layers are then recombined with the reference beam. This recombination produces interference patterns containing detailed information about the depth and reflectivity of the tissue structures (see Figure 2.1) (Aumann et al. 2019; Farrah et al. 2020).

A photodetector captures the interference patterns, and advanced signal processing techniques, including Fourier transformation, are employed to analyse these patterns, whether through a spectrometer (in spectral-domain OCT) or a tunable laser (in swept-source OCT). The analysis results in high-resolution, two-dimensional cross-sectional images or three-dimensional volumetric images of the tissue. Given its non-invasive nature and exceptional resolution, OCT has become an widely adopted tool in ophthalmol-

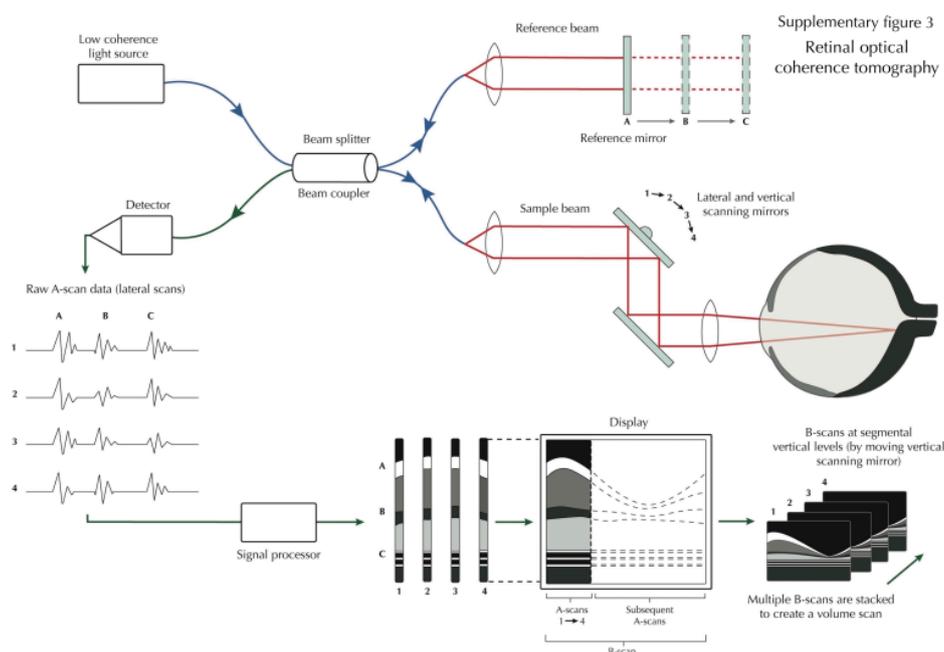
Supplementary figure 3  
Retinal optical coherence tomography

Figure 2.1: Schematic representation of retinal optical coherence tomography (adapted from Farrah et al. 2020)

ogy and numerous other medical fields (Aumann et al. 2019; Farrah et al., 2019).

These cross-sectional images, or tomograms, are composed of neighboring A-scans, each capturing the reflective properties of the subject at a specific depth along the scanned beam's path. By systematically sweeping the beam across the sample and compiling adjacent A-scans, a B-scan (cross-sectional image) is produced, providing a detailed slice of the tissue. A volumetric image is then constructed from multiple B-scans (see Figure 2.2).

Due to its non-invasive nature and exceptional resolution, OCT has become an essential imaging tool in ophthalmology and various other medical fields (Aumann et al. 2019; Farrah et al. 2020).

One of the key advantages of OCT is its exceptionally high axial resolution, typically ranging from 3 to 8  $\mu\text{m}$  when imaging the retina. Furthermore, OCT enables comprehensive retinal scanning in just a few seconds, often capturing over 100 scans. Owing to its precision and non-invasiveness, OCT has been described as an “in vivo clinical biopsy,” providing high-resolution imaging of retinal structures. It is particularly well-suited for accurately measuring neurosensory retinal thickness and visualising the multilayered architecture of the retina, thus enabling the assessment of individual retinal sublayers (Warwick et al. 2022; Keane et al. 2016).

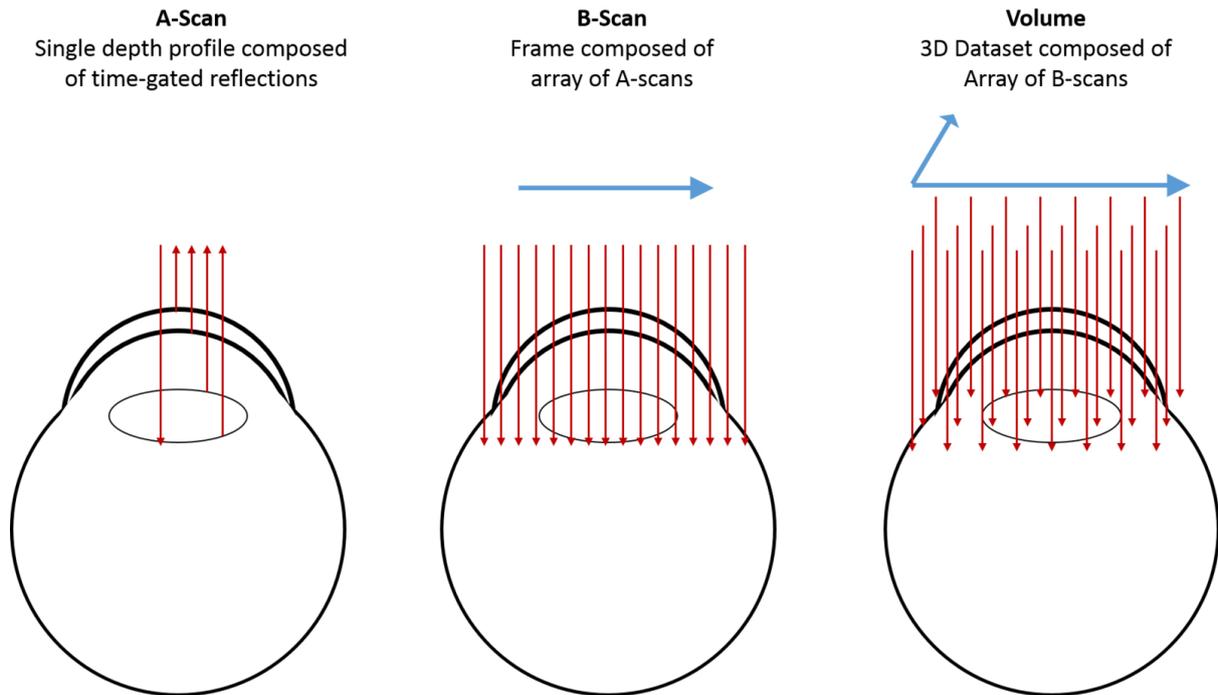


Figure 2.2: Scan types with Optical Coherence Tomography (adapted from Leica-Microsystems 2025)

#### Examples of three 64th B-scan OCT images

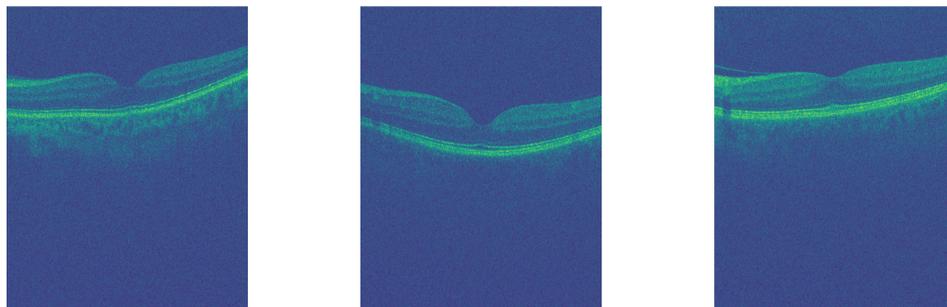


Figure 2.3: Examples of three 64th B-scan OCT images, each from a different patient.

### 2.2.2 Fundus photography

Retinal fundus photography allows for non-invasive, in vivo visualisation of the vascular system within the superficial inner retina, including central and branch veins, arteries, venules, and arterioles. These retinal blood vessels are composed of tightly sealed endothelial cells (ECs), forming the inner blood-retina barrier, and are encased by smooth muscle cells (SMCs) that support the vessel walls (Bharadwaj et al. 2013).

The process of acquiring fundus photographs begins with the illumination of the retina. A bright light source, typically a xenon flash lamp, generates a brief, intense burst of light, which passes through the pupil to illuminate the interior of the eye. The fundus camera utilises a sophisticated system of lenses

and mirrors, creating an optical pathway for both illumination and imaging. The outer portion of this pathway is used for illumination, while the central portion captures the reflected light from the fundus. As the light reflects off the retina, it travels back through the pupil and passes through the camera's optical system, where it is focused to form an image. A digital sensor then captures the image, converting the focused light into a digital signal. The separation of the illumination and imaging pathways within the camera design reduces reflections from the cornea and lens. Illumination light enters through the periphery of the pupil, while the retinal image is captured through the centre. The resulting image can be viewed, stored, and analysed immediately (see Figure 2.4) (Mishra 2024; Saine 2024).

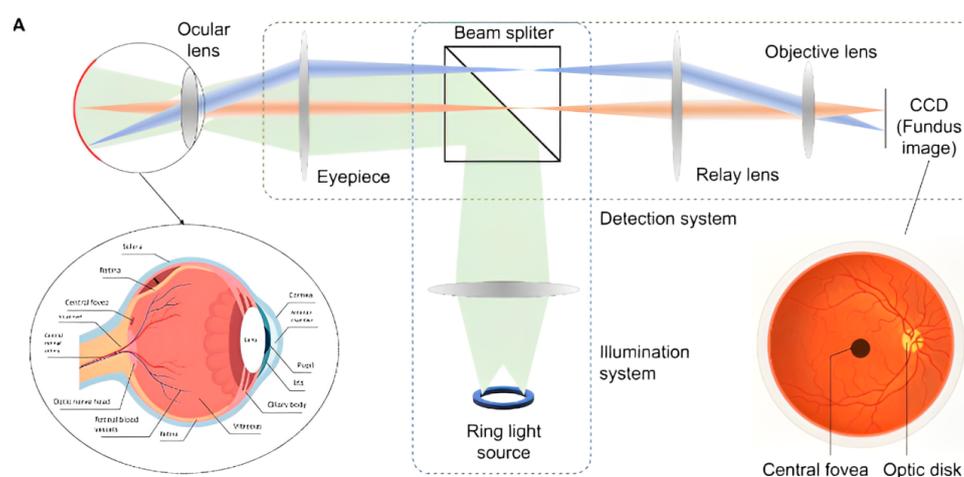


Figure 2.4: Schematic representation of Fundus Photographs (adapted from Zhang et al. 2024)

Advanced fundus cameras, such as those employing scanning laser ophthalmoscopy, enable ultra-widefield imaging. These systems use lasers to rapidly scan the retina, capturing a much wider view of the fundus in a single image. By incorporating these advanced optical and imaging techniques, fundus photography provides detailed visualisation of the retina, optic disc, blood vessels, and other structures at the back of the eye, facilitating the diagnosis and monitoring of a wide range of ocular and systemic conditions (Mishra 2024; Saine 2024).

### 2.3 Single nucleotide polymorphisms microarrays

Single Nucleotide Polymorphisms (SNPs) are variations at a single nucleotide position within the DNA sequence that occur among individuals (see Figure 2.6). These variations can be located in both coding and non-coding regions of the genome, potentially influencing gene expression or protein function. SNPs represent the most prevalent form of genetic variation in humans (Visscher et al. 2017).

Genetic analysis has been revolutionized by the advent of SNP microarrays, which enable the detection

## Examples of three fundus images

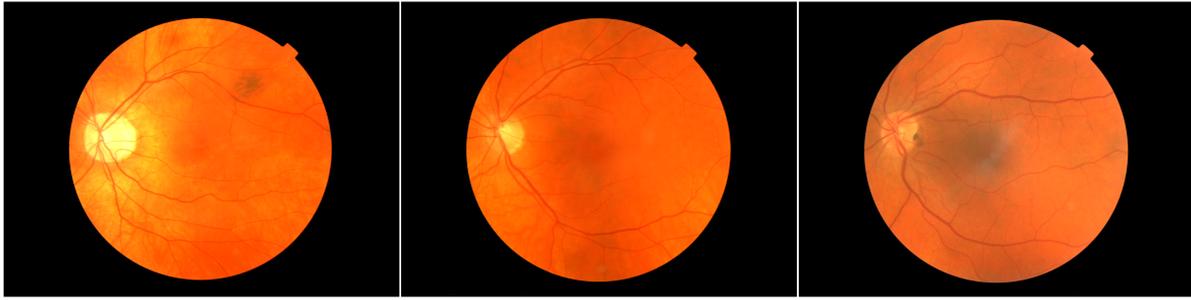


Figure 2.5: Examples of three fundus images, each from a different patient.

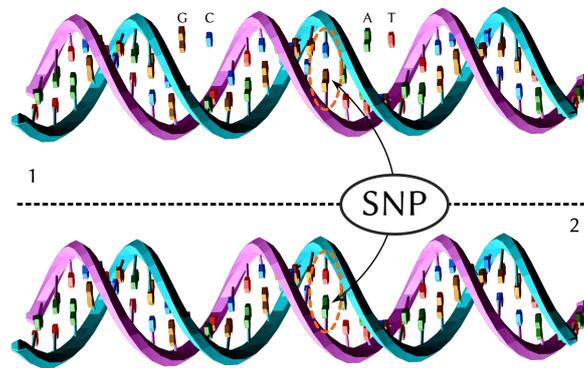


Figure 2.6: The schema of Single nucleotide polymorphism (SNP) (adapted from Wikipedia)

of minute variations in DNA sequences (Wang et al. 2009). These sophisticated tools employ a vast array of DNA probes, ranging from thousands to millions, each carefully engineered to bind with specific genomic regions. SNP arrays are pairing of nucleotide bases: adenine (A) with thymine (T), and cytosine (C) with guanine (G).

The process involves the application of single-stranded DNA from a sample to the array, which is populated with an extensive collection of distinct nucleotide probe sequences. These probes are designed to be either perfectly complementary or nearly so to the DNA segments containing SNP sites. The strength of the hybridization signal that results from this interaction allows for the identification of various genetic alterations, including SNPs, insertions, deletions, and changes in copy number (Mccarroll et al. 2008).

This technology has significantly enhanced our ability to map genetic variations across the genome, providing invaluable insights into human genetic diversity and disease susceptibility (Stranger et al. 2007).

## 2.4 Machine learning methods

Machine Learning (ML) is a subfield of artificial intelligence (AI) that enables systems to automatically learn patterns from data and make predictions without being explicitly programmed (Jordan et al. 2015). ML techniques have been widely applied across various domains, including healthcare, finance, and engineering, due to their ability to model complex relationships in data (Jordan et al. 2015; Waring et al. 2020).

ML models are broadly categorized into three learning paradigms: supervised learning, where the model learns from labeled data; unsupervised learning, which uncovers hidden patterns in unlabeled data; and reinforcement learning, which optimizes decision-making through trial and error (Bishop 2006).

There are several classes of ML algorithms, each with distinct characteristics and suitability for different types of problems. These include linear models, tree-based models, and kernel-based models, among others. In this thesis, we will focus specifically on random forests, as they are the ML method employed in some of our work.

### 2.4.1 Random forest

Random forests (RF) are an ensemble machine learning method that involves multiple decision trees, each of which is trained on a randomly selected subset of training data (Breiman 2001). Each decision tree in the RF is trained on a bootstrap sample, a randomly selected subset of the training data with replacement, while the remaining data, known as out-of-bag (OOB) samples, are used for internal validation. Additionally, at each node of a decision tree, a random subset of features is considered for splitting, rather than the full feature set, which reduces overfitting and enhances model generalizability.

Formally, given a training dataset  $D = \{(x_i, y_i)\}_{i=1}^N$ , where  $x_i$  represents the feature vector and  $y_i$  the corresponding label, RF constructs  $T$  decision trees  $\{h_t(x)\}_{t=1}^T$ , each trained on a bootstrap sample of  $D$ . The function  $h_t(x)$  represents the prediction of the  $t$ -th decision tree in the forest when given input  $x$ . So, if we have  $T$  trees in the forest, we get  $T$  predictions:

$$h_1(x), h_2(x), h_3(x), \dots, h_T(x) \tag{2.1}$$

Each  $h_t(x)$  is an output from a different tree, and they may not always be the same because each tree learns slightly different patterns from the data.

The final prediction for classification is determined by majority voting:

$$\hat{y} = \text{mode}\{h_t(x)\}_{t=1}^T \quad (2.2)$$

Since each tree gives a separate classification, we need a way to combine their results into a final decision. The function mode simply finds the most frequently occurring class label among all  $T$  trees.

For regression tasks, the prediction is obtained by averaging the outputs of individual trees:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(x) \quad (2.3)$$

In a regression problem, the RF predicts a continuous numerical value instead of a class label. Instead of using "mode," RF take the average of all tree predictions to get the final output.

RF have been extensively used in medical applications, including for CVD diagnosis, due to their robustness to noisy data and ability to handle high-dimensional feature spaces (Khozeimeh et al. 2022; Yang et al. 2020). A key advantage of RF over other classification methods, such as support vector machines (SVMs) or neural networks, is their ability to natively handle heterogeneous data types, including categorical, ordinal, and continuous variables, without requiring extensive preprocessing.

The performance of an RF model is influenced by several hyperparameters, including the number of trees ( $T$ ), the maximum depth of individual trees, and the number of features considered at each split ( $m_{try}$ ). Increasing  $T$  generally improves stability but may lead to higher computational costs, while selecting an optimal  $m_{try}$  balances variance reduction and model interpretability.

## 2.4.2 Neural networks

Neural networks are a subset of machine learning models inspired by the structure and functioning of the human brain. These models consist of interconnected layers of neurons (also called nodes) that process and transmit information. A neural network typically contains an input layer, one or more hidden layers, and an output layer (see Figure 2.7). Each neuron in a given layer receives input from the previous layer, applies a mathematical operation (e.g., a weighted sum followed by an activation function), and passes the result to the next layer. Neural networks learn to perform tasks by adjusting the weights of the connections between neurons based on the error of their predictions, a process typically guided by the backpropagation algorithm (Schmidhuber 2014).

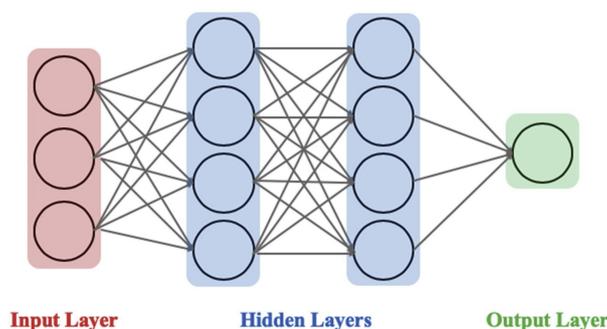


Figure 2.7: A neural network with three input variables, two hidden layers and an output (adapted from Waring et al. 2020)

Neural networks are capable of recognizing patterns in data and have been widely applied to tasks such as classification, regression, and time-series forecasting. However, traditional neural networks with only one or two hidden layers are limited in their ability to model highly complex relationships within data.

## 2.5 Deep learning

Deep learning (DL) is a subfield of Artificial Intelligence (AI) and ML that focuses on training models using deep neural networks. Unlike traditional ML methods as RF, SVM, logistic regression, which often require manual feature engineering, DL models autonomously learn relevant features from large datasets, enabling them to intuitively discern patterns and insights. This ability to extract hierarchical features has made DL particularly effective in domains such as image recognition, natural language processing, and biomedical data analysis.

In traditional ML, feature selection is a crucial preprocessing step aimed at improving model performance, reducing dimensionality, and enhancing interpretability. Common feature selection techniques including filter methods (mutual information, correlation-based selection), wrapper methods (recursive feature elimination) and embedded methods (regression models, tree-based feature importance). In contrast, DL models perform implicit feature selection through multiple layers of transformations. DL facilitates the creation of computational models composed of multiple processing layers, each capable of learning data representations with varying levels of abstraction. By employing the backpropagation algorithm, deep learning models can uncover intricate structures within extensive datasets. This algorithm guides the adjustment of internal parameters, determining how each layer's representation should evolve from the previous layer's output (Voulodimos et al. 2018; Bishop 2006).

Essentially, deep learning operates by comparing the output error to the desired outcome and optimizing

this discrepancy through weight adjustments across the architecture's layers. This optimization is performed automatically using methods such as Adam (Kingma et al. 2015) or Stochastic Gradient Descent (SGD) (Ruder 2016), which iteratively modify weights based on gradient calculations over small data batches, aiming to reach a local minimum (backpropagation).

The backpropagation procedure for computing the gradient of an objective function with respect to the weights of a multilayered stack of modules is a practical application of the chain rule for derivatives. The critical insight is that the derivative (or gradient) of the objective function concerning a module's input can be determined by tracing backward from the gradient concerning the module's output (or the input of the subsequent module) (Bishop 2006).

Despite its advantages, the lack of explicit feature selection in DL can lead to challenges, such as overfitting, computational inefficiency, and reduced interpretability. Besides, DL models often require large amounts of data to learn meaningful features effectively. To mitigate these issues, researchers have proposed hybrid approaches that combine traditional feature selection techniques with deep learning architectures, such as Autoencoders for feature extraction.

### 2.5.1 Fully connected networks

Fully Connected Networks (FCNs), are a type of deep neural network where each neuron in one layer is connected to every neuron in the next layer. This dense connectivity allows the network to learn complex relationships between inputs and outputs. FCNs are typically composed of an input layer, one or more hidden layers, and an output layer. Each layer is fully connected to the next layer, meaning every neuron from the previous layer contributes to every neuron in the next layer. This type of network is powerful but can become computationally expensive and prone to overfitting with very large datasets (Schmidhuber 2014).

A fully connected layer is a function from  $R^n$  to  $R^m$ . Each output dimension depends on each input dimension. Let  $x \in R^n$   $x = (x_1, x_2, \dots, x_n)$  be the input vector. The connections between the input layer and the hidden layer are represented by a weight matrix  $W$  and bias vector  $b$ . If the hidden layer has  $m$  neurons, the weight matrix  $W$  will have dimensions  $m \times n$ , where  $n$  is the number of input features and  $m$  is the number of neurons in the hidden layer. The bias vector  $b$  has dimension  $m$ .

The output  $h = (h_1, h_2, \dots, h_m)$  of the hidden layer is computed as:

$$z = Wx + b \tag{2.4}$$

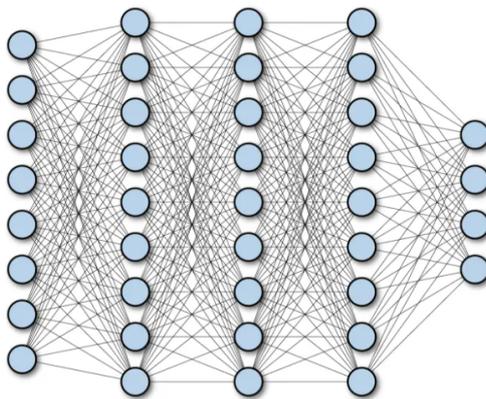


Figure 2.8: Example of a fully connected network (adapted from Ramsundar 2019)

where  $z$  is the pre-activation vector of the hidden layer. Each element of  $z$  is the weighted sum of the inputs, plus the bias term. The pre-activation vector is then passed through an activation function  $f$ , such as the Rectified Linear Unit (ReLU), sigmoid and Tanh.

If the network has a single output, we can define the output layer similarly. Let  $W_o$  be the weight matrix connecting the hidden layer to the output layer, and  $b_o$  be the bias for the output layer. The output  $y$  of the network is computed as:

$$y = W_o h + b_o \quad (2.5)$$

The output  $y$  can then be passed through an activation function depending on the task, softmax for classification, linear activation for regression.

## 2.5.2 Convolutional neural networks

Convolutional Neural Networks (CNNs) are a specialized type of neural network primarily used for processing structured grid data, like images. Unlike fully connected networks, CNNs exploit the spatial hierarchies of features by using convolutional layers, which apply local receptive fields and weight sharing, significantly reducing the number of parameters and computational complexity.

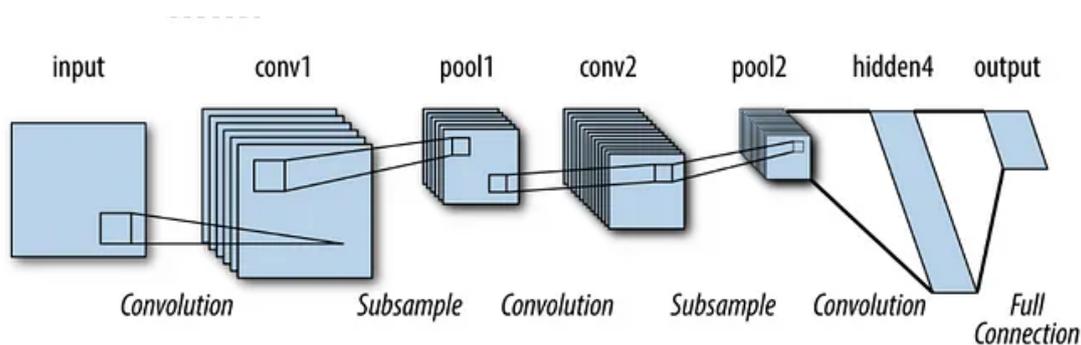


Figure 2.9: Example of a convolutional neural network (adapted from Ramsundar 2019)

A convolutional layer applies a set of learnable filters (or kernels) to the input to extract features such as edges, textures, and patterns. The convolution operation is defined as:

$$z_{ij}^{(l)} = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} W_{mn}^{(l)} x(i+m)(j+n)^{(l-1)} + b^{(l)} \quad (2.6)$$

where  $z_{ij}^{(l)}$  is the output feature map at position  $(i, j)$  in layer  $l$ ;  $x(i+m)(j+n)^{(l-1)}$  represents the input from the previous layer at position  $(i+m, j+n)$ ;  $W_{mn}^{(l)}$  is the filter or kernel weights of size  $k \times k$ ; and  $b^{(l)}$  represents the bias term.

Each filter slides across the input, performing an element-wise multiplication followed by a summation, producing a feature map. The number of filters determines the depth of the output feature map.

After convolution, an activation function is applied to introduce non-linearity, allowing the network to learn complex features. The most common activation function is ReLU (Rectified Linear Unit) as  $f(z) = \max(0, z)$ ; this helps in preventing vanishing gradients and improves the training speed.

Pooling layers downsample feature maps to reduce computational complexity while retaining important spatial information. The two main types of pooling are max pooling (takes the maximum value in a given window, e.g.  $k \times k$ ) and average pooling (computes the average value in a window). For a pooling window of size  $k \times k$  with stride  $s$  that controls how much the filter moves at each step, the output is computed as:

$$O = \frac{I - k}{s} + 1 \quad (2.7)$$

where  $O$  is the feature map size output,  $k$  is the kernel size,  $s$  is the stride.

After multiple convolutional and pooling layers, the feature maps are flattened into a vector and passed through fully connected layers for classification or regression tasks. The output layer applies an activation function based on the task, for example softmax for classification:

$$P(y_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (2.8)$$

A key properties of the convolutional layer are the local receptive field where each neuron connects only to a small region of the input instead of the whole input. Other things is that the same filter is applied across the whole input, reducing the number of parameters compared to fully connected layers. CNNs can be extended to operate on higher-dimensional data, adapting their convolutional filters to match the input structure. The two most common extensions are 2D CNNs and 3D CNNs.

## 2D Convolutional Neural Networks

2D Convolutional Neural Networks (2D CNNs) are the standard CNNs architecture used for processing grayscale and RGB images. In the context of 2D CNN, we consider a two-dimensional signal represented by  $x_{ij}$ . The convolution operation employs a  $n \times n$  kernel, denoted as  $K \in \mathbb{R}^{n \times n}$ . Kernels are generally odd-sized to allow for proper centering on the current pixel. The output feature map  $Z$  at position  $(i, j)$  for a single convolutional layer is computed as Equation 2.5.2.

This operation is applied to all channels, and in the case of multiple input channels (e.g., RGB images), the convolution extends as:

$$z_{ij}^{(l)} = \sum_{c=0}^{C-1} \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} W_{mnc}^{(l)} x(i+m)(j+n)c^{(l-1)} + b^{(l)} \quad (2.9)$$

where  $C$  represents the number of input channels.

## 3D CNN

3D CNN extend the concept of 2D CNN to process three-dimensional data. They are particularly useful for analyzing volumetric data or sequences of 2D images. For 3D CNNs, the main difference is that the convolution kernel extends across the depth (D) dimension, making the operation:

$$z_{ij}^{(l)} = \sum_{c=0}^{C-1} \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} \sum_{p=0}^{k-1} W_{mncp}^{(l)} x(i+m)(j+n)(k+p)c^{(l-1)} + b^{(l)} \quad (2.10)$$

where the additional summation over  $p$  accounts for depth in volumetric data. Similar to 2D CNN, 3D CNN typically employ multiple kernels at each layer. The outputs produced by these kernels are referred to as "feature maps," emphasizing their volume-like nature in the 3D context. This approach allows 3D CNN to effectively process and analyze three-dimensional data, making them valuable for tasks involving volumetric image analysis, video processing, and medical imaging applications.

## 2.6 Autoencoders

Autoencoders (AEs) are a class of neural networks primarily used for unsupervised learning. Their objective is to learn a compressed, meaningful representation of input data while preserving essential information. They are particularly useful for nonlinear dimensionality reduction, feature learning (capturing underlying patterns), data denoising (removing noise from images/sounds), anomaly detection (learning normal patterns and flagging deviations) and missing data imputation (reconstructing incomplete datasets) (Vincent et al. 2008). The fundamental idea behind autoencoders is to reconstruct their input after passing it through a low-dimensional bottleneck, ensuring that only the most relevant information is retained.

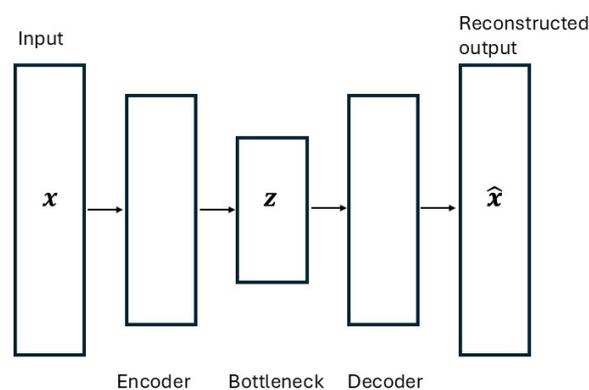


Figure 2.10: Example of a simple autoencoder

The loss function typically minimizes the reconstruction error between the input  $x$  and the reconstructed output  $\hat{x}$ , often using Mean Squared Error (MSE):

$$L = \|x - \hat{x}\| \quad (2.11)$$

where  $\hat{x} = D_\phi(E_\theta(x))$ , with  $E_\theta$  represents the encoder that mapping input  $x$  to a latent representation  $z$ ; and the  $D_\phi$  is the decoder that reconstruct  $x$  from  $z$ .

The encoder-decoder framework is a pair of encoding and decoding functions,  $E_\theta : \mathbb{R}^K \rightarrow \mathbb{R}^{n_z}$  and  $D_\phi : \mathbb{R}^{n_z} \rightarrow \mathbb{R}^K$  that are parameterised by a set of learnable coefficients  $\theta$  and  $\phi$ , respectively. Since the latent space dimension  $n_z \in \mathbb{N}$  is typically much smaller than the input dimension  $K$  ( $n_z \ll K$ ), the network is forced to learn a compact representation (Rumelhart et al. 1986).

### 2.6.1 Variational autoencoders

Variational Autoencoders (VAE) are a class of generative models in the field of neural networks. They consist of two interconnected components: the encoder and decoder networks, which are trained simultaneously. Given some input data  $\mathbf{x}$ , the encoder is designed to approximate the posterior distribution of the latent variables  $q_\phi(\mathbf{z} | \mathbf{x})$ , where  $z$  is the latent space, under some assumed prior distribution  $p(\mathbf{z})$  over the latent variables (typically, a multivariate Gaussian prior is used, that is,  $p(\mathbf{z}) \sim \mathcal{N}(\cdot)$ ), while the decoder is trained to reconstruct the input data by sampling from the approximated posterior distribution  $p_\theta(\hat{\mathbf{x}} | \mathbf{z})$ , where  $\hat{\mathbf{x}}$  is the reconstructed input. In other words, the encoder network maps inputs to low-dimensional latent representations, and the decoder network acts as the generative model. The approximation of the true but intractable posterior distribution is obtained by maximising the lower bound of the evidence (ELBO), which can be expressed as follows:

$$\mathbf{ELBO} = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [\log p_\theta(\mathbf{x} | \mathbf{z})] - D_{KL}(q_\phi(\mathbf{z} | \mathbf{x}) || p(\mathbf{z})) \quad (2.12)$$

where  $\theta$  represents the parameters of the generative model,  $\phi$  represents the parameters of the inference model, and KL indicates Kullback-Leibler divergence. The Kullback-Leibler divergence  $\mathcal{D}_{KL}$  quantifies the dissimilarity between the learnt latent distribution and a previously specified distribution  $p(\mathbf{z})$ , which normally is a multivariate Gaussian distribution. By minimising the KL divergence, the model is incentivised to shape a latent space that adheres to the target Gaussian distribution.

During the training phase of a VAE, the encoder maps input data to a probability distribution rather than a fixed vector. For each latent variable  $z_j$  where  $z = (z_1, z_2, \dots, z_{n_z})$  is the latent vector in and  $j$  indexes its components, the encoder learns two key parameters: mean:  $\mu_j(\mathbf{x})$ , and standard deviation:  $\sigma_j(\mathbf{x})$ . A realization  $z_j$  is then sampled from the distribution  $Z_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$  and passed to the decoder for output generation.

To encourage the variational approximate posterior to resemble a multivariate Gaussian with diagonal covariance, a regularization term is employed.

This approach allows VAE to learn a structured latent space that balances between accurate reconstruction and adherence to a predefined distribution, enabling both effective data compression and generative capabilities.

## 2.6.2 Multi-Channel variational autoencoder.

VAE have proven to be powerful generative models for learning representations of complex data. However, many real-world scenarios involve multiple data sources or modalities, which standard VAE are not designed to handle efficiently. This limitation led to the development of Multi-Channel Variational Autoencoders (MCVAE), an extension of the VAE framework that can jointly model multiple data sources or channels. This advancement allows for more comprehensive modeling of complex, multi-modal data structures.

Consider a MCVAE with  $M$  channels. For each channel  $m$ , let  $x_m$  be the input data,  $z$  the shared latent variable, and  $\theta_m$  the parameters of the generative model. The joint probability distribution of the generative model is defined as  $[p(x_1, \dots, x_M, z) = p(z) \prod_{m=1}^M p(x_m|z, \theta_m)]$  where  $p(z)$  is typically a standard normal distribution  $\mathcal{N}(0, I)$ . The Evidence Lower Bound (ELBO) for the multi-channel VAE is:

$$\text{ELBO} = \mathbb{E}_{q(z|\mathbf{x}, \phi)} \left[ \sum_{m=1}^M \log p(x_m|z, \theta_m) \right] - \text{KL}(q(z|\mathbf{x}, \phi) \| p(z)) \quad (2.13)$$

where  $x = x_1, \dots, x_M$  and the first term represents the reconstruction term and encourages the model to accurately reconstruct the input data for each channel, and the second term, the KL divergence term regularizes the approximate posterior to be close to the prior distribution of  $z$ . Since the true posterior distribution  $p(z|x_1, \dots, x_M)$  is intractable, we approximate it with  $q(z|x_1, \dots, x_M, \phi)$ , typically chosen as a multivariate Gaussian. The MCVAE is trained to minimize the ELBO, promoting accurate reconstruction of all channels while maintaining a regularized latent space. The shared latent representation  $z$

captures common information across all channels, facilitating the joint modeling of heterogeneous data types (Antelmi et al. 2019).

## 2.7 Segmentation

### 2.7.1 UNet

U-Net is a deep learning architecture primarily used for semantic segmentation tasks in image analysis. U-Net's architecture consists of two main parts: encoder and decoder. The encoder is designed to capture context and features from the input image. It consists of several convolutional layers followed by max pooling operations, which progressively reduce the spatial dimensions of the input while increasing the depth of feature maps. This allows the network to learn high-level features at various levels of abstraction. The decoder aims to upsample the feature maps back to the original image size. This section in the model includes upsampling layers that increase the spatial resolution of the feature maps, followed by convolutional layers. A key feature of U-Net is the use of skip connections, which concatenate feature maps from the encoder to the decoder. This helps retain spatial information that may be lost during the downsampling process, allowing for more precise localization of features in the final segmentation map (Ronneberger et al. 2015).

U-Net architectures perform well with small dataset as some found in the field of medical imaging where annotated data is normally few. In order to produce a precise segmentation, a network needs to extract both, what objects are in the image (semantic information) as well as exactly which pixels belong to them (spatial information).

A common loss function for segmentation in medical imaging is the categorical cross-entropy loss, which measures the difference between the predicted and true class probabilities for each sample:

$$L = - \sum_{i=1}^N \sum_{c=1}^K y_{i,c} \log(\hat{y}_{i,c}), \quad (2.14)$$

where  $k$  is the number of samples,  $K$  is the number of classes,  $y_{i,c}$  is a binary indicator (0 or 1) if class label  $c$  is the correct classification for sample  $i$ ,  $\hat{y}_{i,c}$  is the predicted probability that sample  $i$  belongs to class  $c$ , typically obtained via the softmax function (Azad et al. 2022; Siddique et al. 2020).

### 2.7.2 nnUNet

nnUNet is a self-adapting framework based on the U-Net architecture commonly used for medical image segmentation that automatically adapts to different datasets and tasks (Isensee et al. 2020). It automatically configures the network architecture, hyperparameters, and other settings to optimize performance for a given dataset and empirical optimization of other parameters through trial-and-error. nnUNet can handle diverse medical image segmentation tasks, including 2D, 3D, and multi-modal data.

nnU-Net is a deep learning-based segmentation method that automatically configures itself, including preprocessing, network architecture, training and post-processing for any new task in the biomedical domain (Isensee et al. 2020).

## 2.8 Genome-wide association studies

Genome-wide association studies (GWAS) are used to identify genetic variants associated with specific phenotypes by analyzing the genomes of large groups of individuals. A phenotype can be a binary trait (such as disease presence or absence) or a continuous trait (such as height or body mass index). GWAS has become a widely adopted method for uncovering genetic factors that contribute to the risk of common and complex diseases. By testing millions of SNPs (see Section 2.3) across the genome, GWAS offers an unbiased approach to identifying associations between genetic markers and traits. By scanning the genomes of many individuals, GWAS seeks to pinpoint specific genetic variations that may increase susceptibility to particular conditions, providing valuable insights into the genetic basis of complex diseases (Uffelmann et al. 2021; Abdellaoui et al. 2023).

The most common approach in GWAS are the single-variant models because they allow for an exhaustive, genome-wide search for genetic associations with traits.

### 2.8.1 Single-variant models.

A single-variant model in the context of GWAS refers to a statistical approach where each genetic variant, SNP, is tested individually for its association with a phenotype.

The most common statistical approach used is a linear or logistic regression model, depending on whether the trait is continuous or binary.

For an individual  $i$ , with genotype  $G_i$  at a given SNP, the association model is typically written as:

$$Y_i = \beta_0 + \beta_1 G_i + \epsilon_i \quad (2.15)$$

where  $Y_i$  is the phenotype (continuous trait) of individual  $i$ ,  $G_i$  is the genotype at the SNP (coded as 0, 1, or 2, depending on the number of minor alleles),  $\beta_0$  is the intercept (representing the expected value of the phenotype when the SNP genotype  $G_i$  is zero),  $\beta_1$  is the effect size (or regression coefficient) that estimates how much the phenotype changes per one-unit increase in the SNP genotype,  $\epsilon_i$  is the error term, which accounts for unexplained variation in the phenotype.

For binary traits like the presence or absence of a disease (e.g., diabetes or heart disease), the association is modeled using a logistic regression equation:

$$\log \frac{P(Y_i = 1)}{P(Y_i = 0)} = \beta_0 + \beta_1 G_i \quad (2.16)$$

where  $P(Y_i = 1)$  is the probability that individual  $i$  has the trait (i.e., the disease), and  $P(Y_i = 0)$  for controls (i.e., disease absent) (Kim et al. 2013).

In GWAS, different inheritance models can be used to encode the genotype variable  $G_i$  in the regression model, affecting how SNP effects are estimated (Alliance et al. 2009):

- **Recessive model:**  $\mathbf{AA} \mapsto 0$ ,  $\mathbf{Aa} \mapsto 0$ ,  $\mathbf{aa} \mapsto 1$ . It is necessary to carry two copies of the alternative allele (A) to observe a difference in the phenotype.
- **Dominant model:**  $\mathbf{AA} \mapsto 0$ ,  $\mathbf{Aa} \mapsto 1$ ,  $\mathbf{aa} \mapsto 1$ . A single copy of the alternative allele is enough to affect the phenotype. In other words, as long as an individual carries at least one alternative allele, they will have the same effect on the phenotype.
- **Additive model:**  $\mathbf{AA} \mapsto 0$ ,  $\mathbf{Aa} \mapsto 1$ ,  $\mathbf{aa} \mapsto 2$ , i.e. the effect of the SNP is assumed to increase linearly with the number of alternative alleles. Most Common in GWAS.

Throughout this work, we will always assume an additive model.

**Hypothesis testing and statistical significance** Each SNP is tested for association using a hypothesis test:

$$H_0 : \beta_1 = 0, \quad (\text{no association})$$

$$H_A : \beta_1 \neq 0, \quad (\text{significant association})$$

The p-value from the regression analysis determines whether the SNP is significantly associated with the trait. Due to the large number of SNPs tested ( millions), a Bonferroni correction or False Discovery Rate (FDR) correction is applied to control for multiple testing. A SNP is considered genome-wide significant if  $p < 5 \times 10^{-8}$ .

## 2.9 Mendelian randomization

Mendelian randomization (MR) is a method in genetic epidemiology that uses genetic variants as instrumental variables to investigate causal relationships between modifiable risk factors (exposures) and health outcomes. It leverages the random assortment of genes from parents to offspring to mimic a randomized controlled trial (the information of this section was based on Julian et al. 2021; Julian et al. 2022).

The core principle of MR relies on three key assumptions:

- The genetic variant (Z) is associated with the exposure of interest (X).

$$\text{Cov}(Z, X) \neq 0$$

- The genetic variant is not associated with confounders (U).

$$Z \perp\!\!\!\perp U$$

- The genetic variant affects the outcome (Y) only through its effect on the exposure.

$$Z \perp\!\!\!\perp Y \mid X, U$$

**Instrument selection.** In practical MR analyses, the selection of appropriate SNVs as instruments begins with setting an arbitrary p-value threshold for identifying candidate genetic variants. Typically, a p-value threshold of  $p < 5 \times 10^{-8}$  is used to ensure that the selected variants are strongly associated with the exposure. This threshold is derived from GWAS to minimize the risk of false positives.

**Proxies.** When a direct exposure instrument is not available in the outcome dataset, a suitable proxy is identified. Tools like the Ensembl server are commonly used for this purpose (Cunningham et al.

2022). A proxy is a genetic variant that is highly correlated with the missing instrument and serves as a substitute in the analysis.

**Clumping.** In MR, SNVs are typically clumped using a linkage disequilibrium  $R^2$  threshold (often set at 0.001) and a genetic distance cut-off of 10,000 kilobases. A European reference panel is commonly employed for clumping to ensure that the selected SNVs are independent of each other.

**Harmonisation.** Harmonization is the process of aligning the effects of genetic instruments on both the exposure and the outcome so that the effect sizes (beta values) are expressed per additional copy of the same allele. To reduce errors, palindromic alleles with a minor allele frequency ( $MAF$ )  $> 0.42$  are often excluded from the analysis (Hartwig et al. 2016).

**Removal of pleiotropic genetic variants and outliers.** Pleiotropy occurs when a genetic instrument influences multiple traits, which can bias MR estimates. Instruments that demonstrate stronger associations with the outcome than with the exposure are considered pleiotropic and are removed from the analysis using statistical methods (Hemani et al. 2018).

### 2.9.1 Inverse variance weighted

The inverse variance weighted (IVW) method combines causal estimates from individual genetic instruments to provide an overall estimate of the causal effect. It assumes a linear relationship between the genetic instruments and the exposure, and it provides an unbiased causal estimate when the IVs are valid.

The IVW estimate can be represented as:

$$\hat{\beta}_{IVW} = \frac{\sum_i w_i \cdot \beta_{G_i \rightarrow Y}}{\sum_i w_i} \quad (2.17)$$

where  $w_i = \frac{1}{SE(\beta_{G_i \rightarrow Y})^2}$  is the weight for each instrument based on the inverse of the variance of the effect estimate, and  $\beta_{G_i \rightarrow Y}$  is the causal estimate for each instrument.

The IVW method relies on several key assumptions for valid causal inference:

- No horizontal pleiotropy: The genetic variants affect the outcome only through the exposure and not through other independent pathways.

- Linear relationship: There is a linear relationship between the genetic variant's effect on the exposure and its effect on the outcome.
- No measurement error: The genetic associations with the exposure are measured without error (although this assumption is relaxed in some cases).

The IVW estimate  $\hat{\beta}_{IVW}$  provides a weighted average of the causal effects estimated by each genetic variant, with more precise variants (those with smaller variances) contributing more to the overall estimate. This approach is similar to a meta-analysis of individual genetic variants' effects, weighted by their inverse variances.

## Chapter 3

# Literature review

This chapter begins with an introduction to CVDs. It then reviews published articles on the use of deep learning and retinal imaging, organized into two main sections: (1) the use of retinal imaging for assessing CVD risk factors; and (2) the use of retinal imaging for assessing CVD events. Each of these sections is further divided based on the imaging modality used, fundus photography or OCT. The final section focuses on studies that integrate genetic data, retinal imaging, and CVDs, including those that apply causal analysis. In Sections 3.2.3 and 3.3.1, we also discuss the limitations and gaps identified in the current literature.

### 3.1 Cardiovascular diseases

CVDs continue to represent a significant global health challenge, affecting over 500 million individuals worldwide. Specifically, in 2021, CVD accounted for 20.5 million deaths (World-Heart et al., 2023). It is particularly concerning that up to 80% of premature myocardial infarction (MI) and stroke cases could potentially be prevented through early detection (World-Heart et al., 2023). Moreover, the burden of CVD disproportionately impacts low- and middle-income countries, where nearly four out of every five CVD-related deaths worldwide occur (Lindstrom et al. 2023).

Currently, tools such as QRISK (quantitative risk assessment tool for cardiovascular disease) are utilised in primary care settings by healthcare professionals to identify patients at higher risk of various CVDs. These tools are commonly employed during health assessments to evaluate a patient's risk based on factors such as demographic details (e.g., ethnicity, age, gender), clinical indicators (e.g., cardiac volume measurements, blood markers, indicators of obesity, etc.), and socioeconomic data (Li et al. 2019).

Notably, early identification of individuals at risk is crucial, as premature CVD is highly preventable. Effective primary prevention strategies have been shown to reduce CVD mortality and morbidity, as demonstrated in several previous clinical studies (Lindstrom et al. 2023).

Recent research has highlighted the potential of retinal imaging as an innovative, non-invasive tool for systemic vascular assessment (Guo et al. 2020; Yeung et al. 2020). The retinal and choroidal microvasculature, which share embryological and physiological similarities with cerebral and coronary vasculature, serve as sensitive indicators of systemic vascular pathology (Farrah et al. 2020). Retinal examinations may thus provide early insights into microvascular dysfunction associated with CVD, including stroke and MI (Anderson et al. 1995; Farrah et al. 2020).

The clinical utility of retinal imaging is further supported by its non-invasive nature, widespread availability in optometric and ophthalmologic practice, and established role in diagnosing systemic conditions such as diabetes mellitus, hypertension, and atherosclerosis (Wagner et al. 2020; Flammer et al. 2013). Emerging evidence suggests that retinal vascular phenotypes, including vessel tortuosity, caliber changes, and fractal dimension, correlate with cardiovascular risk profiles, offering a novel avenue for early CVD detection (Hanssen et al. 2022). Furthermore, advances in AI-based image analysis have enhanced the precision and scalability of retinal biomarkers for large-scale risk prediction (Wong et al. 2022).

### **3.2 Deep learning and retinal imaging**

Numerous previous studies have investigated the application of AI in predicting both CVDs risk factors and CVDs events through retinal imaging (Wong et al. 2022). This approach has garnered significant interest due to the ease with which retinal images can be obtained during routine visits to ophthalmologists or optometrists, as well as the cost-effectiveness of employing such modalities. These factors present a valuable opportunity for the early detection of cardiovascular disease.

In this section, we will categorise the studies into two groups. First, we examine research that utilises retinal imaging to assess CVD risk factors such as age, sex, and blood pressure. Next, we review papers that investigate the relationship between retinal phenotypes and CVD events. In each of these categories, we will focus on studies that employ only fundus photography and OCT data, as these are the data utilised in this thesis.

### **3.2.1 Retinal imaging for assessing CVD risk factors**

Recent advancements in AI have been notably applied in the assessment of CVD risk through retinal imaging. Various studies have investigated different AI techniques and retinal imaging modalities to identify and evaluate CVD risk factors (Table 3.2.1).

#### **Fundus Photography**

Poplin et al. 2018 were the first to predict cardiovascular risk factors from retinal images using DL methods. They employed a convolutional neural network (CNN) to analyze fundus photography (FP) images, aiming to predict multiple CVDs risk factors, including age, sex, smoking status, blood pressure, body mass index (BMI), and HbA1c level. Their model was validated on two datasets: UK Biobank and EyePACS-2K. The study reported mean absolute errors (MAE) of 3.26 and 3.42 for age prediction, respectively. For gender classification, they achieved an area under the curve (AUC) of 0.97 for both validation sets. The model highlighted blood vessels when predicting risk factors such as age, sex, and systolic blood pressure (SBP).

Similarly, Vaghefi et al. 2019 employed a CNN to predict smoking status using FP images, achieving an accuracy of 88%. Their model highlighted the perivascular region and the fovea as key features.

Other studies have also utilized DL to predict various risk factors, including non-modifiable risk factors (age, sex), modifiable lifestyle factors (drinking, smoking, BMI, waist-to-hip ratio), modifiable medical and biological factors (blood pressure, body composition, hemoglobin A1c, relative fat mass, glucose, insulin, lipid profile, diabetes-related measures, C-reactive protein, hematocrit, mean corpuscular hemoglobin concentration, bilirubin, hematological parameters, kidney and liver function, and thyroid function markers), and hormonal factors (sex hormone-binding globulin, estradiol, testosterone) (Gerrits et al. 2020; Zhang et al. 2020; Rim et al. 2020).

Sabanayagam et al. 2020 explored the use of CondenseNet for detecting chronic kidney disease (CKD) using FP images, achieving an accuracy of 90%, highlighting the applicability of AI in detecting systemic conditions from retinal images. Similarly, Zhang et al. 2021 utilized the ResNet-50 architecture, incorporating metadata alongside FP images to predict CKD. Their study reported an AUC of 0.91 when combining image data with additional patient information, demonstrating superior performance compared to using individual modalities alone.

Diaz-Pinto et al. 2022 employed a combined model of mcVAE and ResNet-50 to analyse FP, in conjunc-

tion with cardiac magnetic resonance (CMR) and demographic data. This study focused on predicting left ventricular mass (LVM) and left ventricular end-diastolic volume (LVEDV), achieving correlation coefficients of  $R=0.65$  and  $R=0.45$ , respectively. This demonstrates a stronger association between retinal features and LVM compared to LVEDV.

Rim et al. 2020 utilised EfficientNet on FP to predict coronary artery calcium (CAC) scores, a crucial indicator of atherosclerotic plaque burden. Their model achieved AUC of 0.73, indicating moderate predictive capability. Son et al. 2020 applied the Inception-v3 model to FP to classify individuals into low and high CAC score groups, with their model achieving an AUC of 0.83.

Lau et al. 2018 employed a pattern recognition neural network (PRNN) method to detect various retinal vessel measurements, including arteriovenous nicking, arteriolar occlusion, and tortuosity. The model aimed to predict the grade of age-related white matter changes, achieving a sensitivity of 0.93 and specificity of 0.82.

Zee et al. 2021 implemented a ML framework combining a ResNet50 CNN for retinal feature extraction with a classification tree model to analyze retinal vascular parameters, including the central retinal artery equivalent (CRAE), central retinal vein equivalent (CRVE), and fractal dimensions of retinal vessels to predict age-related white matter changes (ARWMC), achieving high diagnostic performance with a sensitivity of 0.90 and specificity of 0.97.

Khan et al. 2022 used a support vector machine (SVM) to evaluate a broad range of retinal biomarkers, including arteriovenous ratio, tortuosity index, and multifractal parameters, to predict pial collateral blood flow in acute ischaemic stroke. Their model demonstrated a sensitivity of 0.74 and specificity of 0.71, reflecting balanced performance in identifying critical blood flow alterations during stroke events.

Nusinovici et al. 2022 applied the Visual Geometry Group (VGG) neural network to FP images, developing a model called 'RetiAGE' to predict the probability of an individual being  $\geq 65$  years old. The model demonstrated high performance, achieving an AUC of 0.968 and an area under the precision-recall curve (AUPRC) of 0.83 in the Korean Health Screening study cohort. When validated on the UK Biobank dataset, it showed lower performance, with an AUROC of 0.756 and AUPRC of 0.399. Recently, Zhu et al. 2022 employed the Xception model to predict retinal age, a proxy for biological age, achieving a MAE of 2.9 years, further underscoring the potential of advanced AI models for age estimation from retinal images.

### **Optical coherence tomography**

In the realm of OCT imaging, Hassan et al. 2021 employed BagNet33 to predict chronological age and sex, achieving a MAE of 4 years with a coefficient of determination ( $R^2$ ) of 0.77 for age prediction, and an AUC of 86% for sex classification. Similarly, Chueh et al. 2021 applied ResNet18 to OCT images for sex prediction, reporting an AUC of 0.91. The MAE for age prediction using DL on macular OCT was  $5.78 \pm 0.29$  years. Lastly, Munk et al. 2021 employed a CNN on OCT images to predict age and sex, achieving an accuracy of 84% and an MAE of 5.625 years.

Menten et al. 2023 investigated the use of counterfactual generative adversarial networks (GANs) for predicting retinal ageing through OCT images, with their model achieving a MAE of 3.1 years.

Table 3.1: Overview of studies using AI for predicting CVD risk factors from retinal images

Paper	AI Method	Retina Modality	CVD risk factors	Database (n)	Performance Metrics
Poplin et al. 2017	CNN	FP	Age, gender, smoking status, systolic blood pressure, BMI, HbA1c levels	UKbiobank (120090); Eye-PACS (1684896)	MAE: 3.26%, 3.42%; AUC: 0.97, 0.97; AUC: 0.71, n/a; MAE: 11.35, n/a; MAE: 3.29, n/a; MAE: n/a, 1.39
Vaghefi et al. 2019	CNN	FP	Smoking status	Auckland (165103)	Accuracy: 88%
Gerrits et al. 2020	MobileNet-V2	FP	Age, sex, systolic blood pressure, diastolic blood pressure, Haemoglobin A1c, relative fat mass, testosterone	Qatar Biobank (12000)	MAE: 2.78 years; AUC:0.97; MAE: 8.96 mmHg; MAE: 6.84 mmHg; MAE: 0.61%; MAE: 5.68 units; MAE: 3.76 nmol/L
Rim et al. 2020	VGG16	FP	Age, sex, height, BMI, DBP, SBP	Korea, China and Singapore (236257)	AUC: 0.96; MAE: 2.43 years; MAE: 5.20 years; MAE: 2.15; MAE: 7.20; MAE: 9.29
Zhu et al. 2022	Xception	FP	Retinal age prediction	UKbiobank (19200)	MAE: 2.9 years
Sabanayagam et al. 2020	CondenseNet	FP	Chronic kidney disease detection	SEED (11673)	Accuracy: 90%

Paper	AI Method	Retina Modality	CVD risk factors	Database (n)	Performance Metrics
Zhang et al. 2021	ResNet-50	FP (metadata)	Chronic kidney disease detection, type 2 diabetes	Henan, China (1222)	AUC: 0.91
Hassan et al. 2021	BagNet33	OCT	Age, sex	UKbiobank (66767)	Accuracy: 92%
Chueh et al. 2021	ResNet18	OCT	Age, sex	Taiwan (6147)	MAE: 5.78; AUC: 0.91
Menten et al. 2023	Counterfactual GANs	OCT	Retinal aging	UKbiobank (85709)	MAE: 3.1 years
Diaz-Pinto et al. 2022	mcVAE+ResNet-50	Fundus Photographs (CMR and demographics)	Left ventricular mass, left ventricular end diastolic volume	UKbiobank (5663)	R:0.65 (LVM); R:0.45 (LVEDV) 0.45
Nusinovici et al. 2022	VGG	FP	Probability of an individual being $\geq 65$ years old	Korea (161554)	AUC: 0.968; AUPRC: 0.83
Rim et al. 2021	EfficientNet	Fundus Photographs	CAC prediction	Korea (5590)	AUC: 0.73
Son et al. 2020	Inception-v3	Fundus Photographs	Low and high CAC scores	Korea (44184)	AUC:0.83
Lau et al. 2018	Pattern recognition neural network	Retinal vessel measurements, arteriole-venous nicking and arteriole occlusion, hemorrhages, tortuosity, bifurcation coefficients, asymmetry of branches and bifurcation angles	Age-related white matter changes grade	CU-RISK (180)	Sensitivity: 0.93; specificity:0.82

Paper	AI Method	Retina Modality	CVD risk factors	Database (n)	Performance Metrics
Zee et al. 2021	ResNet50 CNN + classification tree model	Central retinal artery equivalent, central retinal vein equivalent, arteriole-venule ratio calculate, arteriole-venous nipping, arteriole occlusions, presence of hemorrhage and exudates, tortuosity, arterioles and venules bifurcation coefficients, arterioles and venules bifurcation angles, arterioles, venules asymmetry and fractal dimensions	Age-related white matter changes grade	CU-RISK (240)	Sensitivity: 0.90; specificity:0.97
Khan et al. 2022	SVM	Arteriovenous ratio, Tortuosity index, Monofractal Df, Vessel width, Multifractal, Curve asymmetry, Singularity length, $f(\alpha)_{max}$	Pial collaterals blood flow in acute ischemic stroke	Qatar (35)	Sensitivity: 0.74; specificity: 0.71
Munk et al. 2021	CNN	OCT	Age, sex	Bern, Switzerland (85536)	MAE: 5.625 years; AUC: 84%

### 3.2.2 Retinal imaging for assessing CVD events

The use of AI in predicting CVD events through retinal imaging has garnered significant attention in recent years. Table 3.2 summarises various studies employing diverse AI methodologies to analyse retinal images for assessing CVD risk, with notable performance metrics.

#### Fundus photography

Diaz-Pinto et al. 2022 implemented a logistic regression model using demographic data alongside LVM and LVEDV measurements derived from FP. Their model focused on predicting myocardial infarction, achieving an AUC of 0.80. Poplin et al. 2017 employed the Inception-v3 model to analyse fundus photographs for predicting major adverse cardiac events (MACE), reporting an AUC of 0.70.

Nusinovici et al. 2022 showed that the RetiAGE had a moderate predictive accuracy for cardiovascular mortality, achieving an AUROC of 0.70. The model demonstrated 76% sensitivity in identifying individuals who died from CVD within a 10-year follow-up period.

Al-Absi et al. 2022 used the DenseNet-121 model on FP to differentiate between individuals with and without CVDs, reporting an accuracy of 0.76, which demonstrates the model's potential in distinguishing between individuals with and without CVD.

Ma et al. 2021 applied a hybrid model combining Inception and ResNet-v2 to FP for predicting ischaemic CVD. The study reported a  $R^2$  of 0.88, indicating a strong correlation between predicted and actual outcomes, and thus high predictive accuracy for ischaemic events.

Rudnicka et al. 2022 employed a Cox regression model to develop a retinal risk score based on age and sex, using fundus photographs. The study aimed to predict circulatory mortality, stroke, and MI, with  $R^2$  ranging from 0.33 to 0.44 for circulatory mortality, 0.21 to 0.35 for stroke, and 0.15 to 0.23 for MI.

Dai et al. 2020 utilised a CNN to analyse FP for predicting hypertension. The model reported an AUC of 0.65, indicating moderate predictive accuracy for hypertension based on retinal images.

Mueller et al. 2022 applied a deep attention-based multiple instance learning approach to FP to distinguish between individuals with and without peripheral arterial disease. The model achieved an AUC of 0.89, reflecting high accuracy in differentiating peripheral arterial disease from other conditions.

Qu et al. 2022 employed the ResNet50 model to analyse various retinal features, including arteriovenous nicking, occlusion, haemorrhages, exudates, vessel tortuosity, and bifurcation angles, to predict

ischaemic and haemorrhagic stroke. The model achieved AUCs of 0.93 for ischaemic stroke and 0.95 for haemorrhagic stroke, indicating excellent performance in stroke prediction based on retinal vascular characteristics.

Gupta 2019 used a SVM on FP to predict stroke, achieving an accuracy of 0.95. This result underscores the high precision of SVM in identifying stroke from retinal images.

### **Optical coherence tomography**

Zekavat et al. 2023 employed logistic regression and Cox proportional hazards models to analyse retinal layer thinning using OCT images. This study predicted a broad spectrum of cardiovascular conditions, including coronary atherosclerosis, hypertensive heart and renal diseases, and peripheral vascular disease. The model's performance was assessed using z-scores, demonstrating its comprehensive capacity to predict a range of CVD outcomes based on retinal layer measurements.

Zhong et al. 2021 utilised multivariable logistic regression, incorporating clinical ECG and OCTA data, to analyse the densities of the superficial and deep capillary plexuses in the perifovea. Their model aimed to predict coronary artery disease (CAD) in patients with suspected angina, achieving a sensitivity of 0.813 and a specificity of 0.850, indicating strong diagnostic performance.

Zhou et al. 2023 introduced RETFound, a deep learning model trained on FP and OCT to predict MI, heart failure (HF), and ischaemic stroke (IS), achieving AUCs of 0.737, 0.794, and 0.754, respectively.

Table 3.2: Overview of studies using AI for predicting CVD events from retinal images

Paper	AI Method	Retina Modality	CVD Events	Database (n)	Performance Metrics
Diaz-Pinto et al. 2022	Logistic regression model	LVM/LVEDV (demographics)	Myocardial Infarction	UKbiobank (5663)	AUC:0.80
Poplin et al. 2017	Inception-v3	FP	Major adverse cardiac events	UKbiobank (120090); Eye-PACS (1684896)	MACE: 0.70
Nusinovici et al. 2022	PhenoAGE + RetiAGE system	FP	Cardiovascular mortality	Korea (161554)	AUROC of 0.70
Zekavat et al. 2023	Logistic regression, Cox Proportional hazard	Layer Thinning of OCT layers	Coronary atherosclerosis, hypertensive heart/renal disease, hypertensive chronic kidney disease, peripheral vascular disease, angina pectoris, nonspecific chest pain, cardiac dysrhythmias, ischemic heart disease, atrial fibrillation/flutter, essential hypertension	UKbiobank (44823)	z-score
Al-Absi et al. 2022	DenseNet-121	FP	CVD vs. no CVD	Qatar Biobank (1839)	Accuracy:0.76
Zhong et al. 2021	Multivariable logistic regression (Model clinical ECG OCTA)	Superficial and deep capillary plexus densities in the perifovea	CAD in suspected angina patients	Guangdong, China (795)	sensitivity: 0.813; specificity: 0.850

Paper	AI Method	Retina Modality	CVD Events	Database (n)	Performance Metrics
Ma et al. 2021	Inception-Resnet-v2	FP	Ischaemic CVD	China (841916)	$R^2$ : 0.88
Rudnicka et al. 2022	Cox regression model	Retinal risk score (Age and sex)	Circulatory mortality, stroke, MI prediction	UKbiobank (66326)	$R^2$ : 0.33-0.44 (Circulatory mortality); $R^2$ : 0.21-0.35 (Stroke); $R^2$ : 0.15-0.23 (MI);
Dai et al. 2020	CNN	FP	Hypertension	Shenyang, China (2012)	AUC: 0.65
Mueller et al. 2022	Deep attention-based Multiple Instance Learning	FP	Peripheral arterial disease vs No peripheral arterial disease	Bonn, Germany (135)	AUC:0.89
Qu et al. 2022	ResNet50	Arteriole-venous nipping and occlusion, hemorrhages and exudates, Vessel tortuosity, bifurcation coefficients, Branch asymmetry and bifurcation angles	Ischaemic and haemorrhagic stroke	Shenzhen, China (711)	AUC: 0.93 (Ischaemic stroke); AUC: 0.95 (haemorrhagic stroke)
Gupta 2019	SVM	FP	Stroke	NIH-eyeGENE (200)	Accuracy:0.95

<b>Paper</b>	<b>AI Method</b>	<b>Retina Modality</b>	<b>CVD Events</b>	<b>Database (n)</b>	<b>Performance Metrics</b>
Zhou et al. 2023	RETFound	FP and OCT	MI, HF, IS	MEH-MIDAS, Kaggle EyePACS (904,170 FP, 736,442 OCT)	AUC MI: 0.737 ; AUC HF:0.794 ;AUC IS: 0.754

### 3.2.3 Limitations in the current literature review

Despite its promising outlook, challenges remain in translating AI based CVD prediction into clinical practice. These challenges include issues related to retinal image quality, data accessibility, and the interpretability and generalizability of DL algorithms.

Most of the studies using retinal imaging has fundus photographs as modalities for the clinical settings. While retinal fundus photographs offer a two-dimensional depiction of retinal vasculature, OCT with its high-resolution 3D imaging capabilities allows for quantitative assessment of the thickness and structure of different retinal layers and microvasculature, providing insights not possible with fundus photography alone. The potential of 3D OCT imaging lies in its ability to detect subtle abnormalities in retinal microstructure and microvasculature that may go unnoticed in 2D images, making it a valuable tool for identifying early disease indicators (Farrah et al. 2020).

A major limitation of deep learning systems is their inherent “black-box” nature, which makes it difficult to interpret how features are extracted and how predictions are made. To address this, some studies have employed explainability techniques, such as heatmaps, to highlight the most discriminative regions in the decision-making process. Many heatmap-based methods, such as Grad-CAM, tend to produce patchy and non-continuous regions, which makes it difficult to correlate highlighted areas with meaningful anatomical structures. For instance, Zhou et al. 2023 identified the optic nerve head and retinal vessels as key discriminative regions in fundus photography for CVDs prediction. However, their qualitative interpretability results lacked clear spatial coherence, as the heatmaps highlighted only small, scattered regions rather than consistently delineating these structures. Similarly, in their OCT analysis, they suggested that the nerve fiber layer and ganglion cell layer were the most critical features for prediction. Yet, their heatmaps failed to provide clear, layer-specific activation, making it challenging to determine which retinal layers significantly influenced the model’s predictions.

Furthermore, several studies (Diaz-Pinto et al. 2022; Zekavat et al. 2023) have not incorporated any interpretability analysis, further limiting the clinical applicability of their deep learning models. The absence of robust explainability methods hinders trust and adoption in medical practice, emphasizing the need for more advanced and reliable interpretability techniques beyond conventional heatmaps.

In studies utilizing OCT imaging from the UK Biobank, a common approach has been to use only the 64th B-scan out of the available 128 B-scan volumes (Menten et al. 2023; Zhou et al. 2023). While this selection may be a practical compromise, it risks excluding valuable structural information contained in

the full volume, potentially affecting model performance and clinical relevance.

There is a notable gap in research exploring the integration of multiple retinal imaging modalities for CVDs prediction. Many CVD-related retinal biomarkers manifest differently across imaging techniques, yet only a limited number of studies have incorporated more than one modality in their analyses. For instance, Zhou et al. 2023 successfully trained the RETFound model separately on FP and OCT images, achieving strong predictive performance for MI, HF, and IS, with AUC values of 0.737, 0.794, and 0.754, respectively. However, training a model that integrates both modalities could further enhance predictive accuracy and robustness by leveraging complementary structural and vascular information from FP and OCT. A multimodal approach could enable the detection of a broader range of retinal biomarkers, some of which may not be discernible in a single imaging modality, thereby improving diagnostic precision. Moreover, in real-world clinical practice, ophthalmologists and cardiologists rely on multiple imaging techniques to inform their diagnoses. Therefore, a DL model trained on both FP and OCT could provide AI-driven insights that more closely align with clinical decision-making, ultimately facilitating a more comprehensive and reliable assessment of CVD risk.

### **3.3 Genetic and causal factors associated to cardiovascular disease risk via retinal imaging**

In this section, we focus in relevant literature that has explored intricate relationships between retinal phenotypes, CVD events, and CVD risk factors, using genetic insights to deepen our understanding of these associations. The synthesis of findings from several key studies highlights the retinal phenotypes, associated CVDs, related genes, and, in some cases, the identified causal relationships.

Tomasoni et al. 2023 examined the tortuosity of arteries, veins, and a combination of both vessel types in relation to type 2 diabetes, coronary heart disease, myocardial infarction, and other cardiovascular conditions. The study identified significant associations with genes such as HMG20A, WDR12, COL4A2, MYOZ2, LHFPL2, LRCH1, CSK, and CPLX3. Their findings suggest that increased arterial and venous tortuosity is linked to lower BMI and other cardiovascular risk factors, highlighting the genetic underpinnings of these retinal changes.

Jiang et al. 2023 focused on arteriolar and venular tortuosity and width, linking these retinal phenotypes to waist-hip ratio (as an indicator of obesity) and cardiovascular function via the COL4A2 and PDE3A genes. The study found that increased arteriolar tortuosity causally influences blood pressure, offering

insights into how retinal vascular changes reflect systemic cardiovascular conditions.

Zekavat et al. 2023 explored fractal dimension (FD) and vascular density in relation to elevated blood pressure and type 2 diabetes. Although specific genes were not mentioned, the study reported that individuals with higher blood pressure and diabetes exhibited lower retinal vascular density and FD, illustrating the impact of systemic conditions on retinal vasculature.

Vela et al. 2023 investigated a range of retinal vascular parameters, including artery and vein temporal angle, tortuosity, vascular density, and diameters, in relation to hypertension and heart disease. The study identified the FLT1 gene as a key marker, noting that elevated blood pressure correlates with reduced retinal vascular density, and that BMI influences venous and arterial diameter variability.

Sergouniotis et al. 2024 analysed the inner and outer retinal limits using OCT, associating these with the AFAP1L1 and ABLIM3 genes related to waist-hip ratio. While the study did not establish causal relationships, it provided valuable genetic links to retinal structural changes.

Veluchamy et al. 2017 explored arteriolar and venular tortuosity, CRAE, CRVE, and other retinal measurements in relation to heart rate and waist-hip ratio, identifying the ACTN4 and COL4A2 genes. Although causal relationships were not established, the study demonstrated significant genetic associations with retinal vascular phenotypes.

Currant et al. 2020 focused on the retinal nerve fibre layer and ganglion cell-inner plexiform layer, linking these retinal layers to BMI via the INVS/FOXO3 and IGFBP3 genes. While causal relationships were not discussed, the study highlighted important genetic correlations between retinal structure and body mass.

These studies emphasise the intricate interplay between retinal phenotypes, cardiovascular and metabolic diseases, and genetic factors. The repeated identification of genes such as COL4A2 and FLT1 across multiple studies underscores their central role in retinal vascular health and systemic disease, pointing to potential targets for future research and therapeutic interventions. Performance metrics, including association strength and causal influence, provide quantitative evidence of these relationships, reinforcing the utility of retinal imaging in understanding and predicting cardiovascular and metabolic health.

<b>Paper</b>	<b>Retina Phenotypes</b>	<b>CMD and Gene</b>	<b>Causal Relationships</b>
Tomasoni et al. 2023	Tortuosity measures of arteries, veins, and combining both vessel types	type 2 diabetes HMG20A, coronary heart disease WDR12, myocardial infarction WDR12, waist-hip ratio (obesity) COL4A2 and MYOZ2, pulse pressure LHFPL2 and LRCH1, diastolic blood pressure CSK and CPLX3, mean arterial pressure CSK	arteries, and increased venous tortuosity reduced BMI
Jiang et al. 2023	Arteriolar tortuosity, arteriolar width, venular tortuosity and venular width	waist-hip ratio (obesity) COL4A2, cardiovascular function PDE3A	increased AT causally influence blood pressure
Zekavat et al. 2021	FD and vascular density	N/A	elevated blood pressure have lower retinal vascular density and FD, type 2 diabetes also had lower retinal vascular density
Vela et al. 2023	Artery and vein temporal angle, ratio and vein and artery tortuosity, ratio vascular density, artery and ratio central retinal, artery and vein std diameter, vein central retinal, bifucartions, vein and artery vascular density, artery and vein and ratio median diameter	hypertension and heart disease FLT1	elevated BP tend to have lower retinal vascular density, BMI may cause higher variability in venous and arterial diameter
Sergouniotis et al. 2024	The inner- and outer-most limits of the retina phenotypes of OCT	AFAP1L1/ABLIM3 waste-hip ratio	N/A
Veluchamy et al. 2017	Arteriolar tortuosity, maximum arteriolar tortuosity, venular tortuosity, maximum venular tortuosity, CRAE, CRVE, arteriole-to-venule ratio, nonvascular optic disc radius	heart rate ACTN4, waist-hip ratio (obesity) COL4A2	N/A
Currant et al. 2020	Retinal nerve fibre layer and ganglion cell inner plexiform layer	BMI INVS/FOXO3, Body Mass IGFBP3	N/A

### 3.3.1 Limitations in the Current Literature Review

Retinal segmentation has been widely studied using fundus imaging, with most studies focusing on arterial and venous metrics such as vessel width, arteriolar and venular tortuosity, and other morphological features (Tomasoni et al. 2023; Jiang et al. 2023). However, research utilizing OCT imaging remains limited, and when OCT is used, it often focuses on only a few retinal layers rather than providing a comprehensive analysis (Sergouniotis et al. 2024). This gap restricts a deeper understanding of retinal microstructure and its potential links to systemic diseases.

Additionally, research exploring the causal relationships between retinal biomarkers and broader phenotypic traits, diseases, or cardiovascular conditions using OCT imaging remains limited (Zekavat et al. 2023; Sergouniotis et al. 2024). Existing studies primarily focus on association analyses, which do not establish causality. In contrast, causal investigations have been more extensively conducted using fundus photographs or phenotypes derived from them, highlighting a gap in OCT-based research.

Current segmentation methods in large-scale studies, such as Topcon Advanced Boundary Segmentation (TABS), have been applied in datasets like the UK Biobank, achieving an average processing time of 58 seconds per volume, including data fetching and analysis making it computationally intensive and may not be optimal for large-scale applications (Keane et al. 2016). In contrast, deep learning-based approaches, particularly CNN-based models, offer the potential to accelerate segmentation while requiring fewer manually labeled ground truth images.

Among deep learning models, U-Net has gained popularity in medical image segmentation due to its efficiency and accuracy in handling large datasets (Zekavat et al. 2023). More advanced architectures, such as nnU-Net, have emerged, offering further improvements in segmentation performance. Despite these advancements, the application of U-Net to OCT B-scan volumes remains largely unexplored, presenting a promising avenue for future research. Integrating U-Net-based approaches into OCT analysis could enhance automation, improve segmentation accuracy, and facilitate large-scale retinal imaging studies. Future studies should explore these methodologies to unlock their full potential in clinical and epidemiological applications.

## Chapter 4

# Predicting cardiovascular diseases using optical coherence tomography imaging and demographic and clinical data

### 4.1 Introduction

This chapter is based on our paper Maldonado-Garcia et al. 2024, currently available on arXiv and under review at Frontiers in Artificial Intelligence. This study introduces a predictive model that integrates various types of data, including features derived from 3D OCT imaging through a self-supervised deep neural network, along with patient demographic and clinical details. The aim is to detect individuals at risk of Myocardial Infarction (MI) or stroke within five years after image capture. To the best of our knowledge, this is a pioneering investigation into the application of 3D OCT imaging and artificial intelligence for automatically identifying patients vulnerable to adverse CVD incidents. The main contributions of this research are: (i) the development of a self-supervised feature selection VAE integrated with a multimodal RF classification model, which effectively combines diverse patient data, including OCT imaging, and clinical variables; (ii) the introduction of an innovative method for enhancing model interpretability through vector field, enabling detailed localization of retinal features that significantly contribute to the accurate identification of patients at risk of adverse cardiovascular events; and (iii) the identification of the choroidal layer as the key feature influencing the model's predictive accuracy for cardiovascular disease risk.

We described the most relevant studies in the Chapter 3 Section 3.2 employing fundus photographs, OCT imaging as CVD risk factors, CVD biomarkers, CVD events predictors. In this study, we propose the use of OCT due to its high-resolution 3D imaging capabilities, which allow for a quantitative assessment of the thickness and structure of different retinal layers and microvasculature. This provides insights that are not possible with fundus photography alone, and the potential of OCT has not been as extensively studied compared to fundus photography. The strength of 3D OCT imaging lies in its ability to detect subtle abnormalities in retinal microstructure and microvasculature that may go unnoticed in 2D images, making it a valuable tool for identifying early disease indicators (Farrah et al. 2020). This advancement in OCT technology has transformed retinal imaging by enabling the visualization of the chorioretinal microcirculation, which can serve as an early sign of microvascular disease.

## 4.2 Database

In this research, we utilised retinal OCT imaging data sourced from the UK Biobank, captured using the Topcon 3D OCT 1000 Mark 2 system (details regarding OCT are explained in Section 2.2.1). A total of 68,109 and 67,681 participants underwent retinal imaging for their right and left eyes, respectively at the Initial assessment visit (2006-2010). To ensure only high-quality images were included, we automatically evaluated image quality using a quality index (QI) detailed in a prior study (Stein et al. 2006). This QI is a globally accurate quality assessment algorithm derived from the intensity ratio, which is based on a histogram covering the entire image, and the tissue signal ratio, indicating the ratio of highly reflective pixels to less reflective ones. The QI is calculated as the product of two terms referred to as Intensity Ratio (IR) and Tissue Signal Ratio (TSR). The IR is akin to the signal-to-noise ratio (SNR), but rather than considering the maximum SNR value among all A-scans, it encompasses the entire image. Meanwhile, TSR represents the ratio of highly reflective pixels to those with lower reflectivity. Figure 4.1 shows the distribution of the QI from both eyes.

We excluded images applying a quality score threshold, resulting in the exclusion of 14,573 images for the left eye and 20,873 images for the right eye, leaving 53,108 and 47,236 remaining images, respectively. Among these, we identified 2,448 (left eye) and 2,228 (right eye) images from participants who had experienced a stroke or MI event, referred to as CVD+ participants. However, only images from the left eye of 875 participants and the right eye of 791 participants were taken before the CVD event. Furthermore, we omitted 131 patients with diabetes and/or cardiomyopathy for the left eye and 121 patients for the right eye. We also removed cases to address data imbalance, resulting in a final

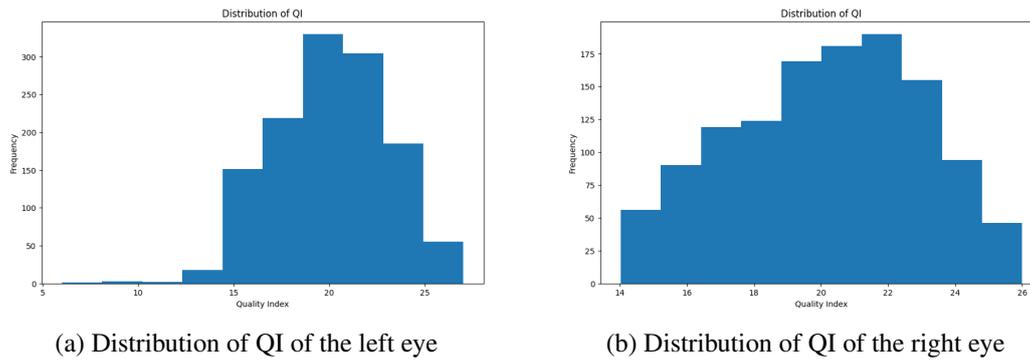


Figure 4.1: Comparison of the distribution of the Quality Index from both eyes.

cohort of 612 subjects for both eyes. A visual representation of the participant selection and exclusion criteria used to establish the cohort for this study is depicted as a STROBE diagram in Figure 4.2.

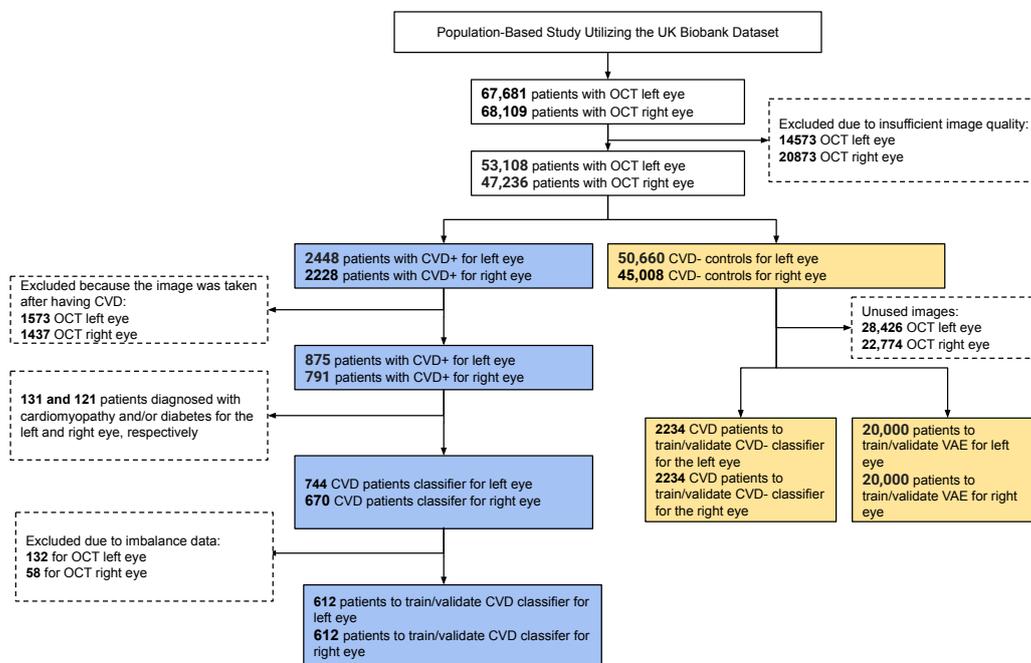


Figure 4.2: STROBE flow chart describing participant inclusion and exclusion criteria applied to define the study cohort.

This study exclusively encompasses patients who experienced MI or stroke within a five-year period after OCT image acquisition. Figure 4.3 presents a bar plot illustrating the methods used to determine the onset date of stroke or MI for the 612 CVD+ subjects involved in our classification task. The majority of cases, approximately 350, were attributed to hospital primary records, followed by hospital secondary records. The fewest cases were associated with death contributory records.

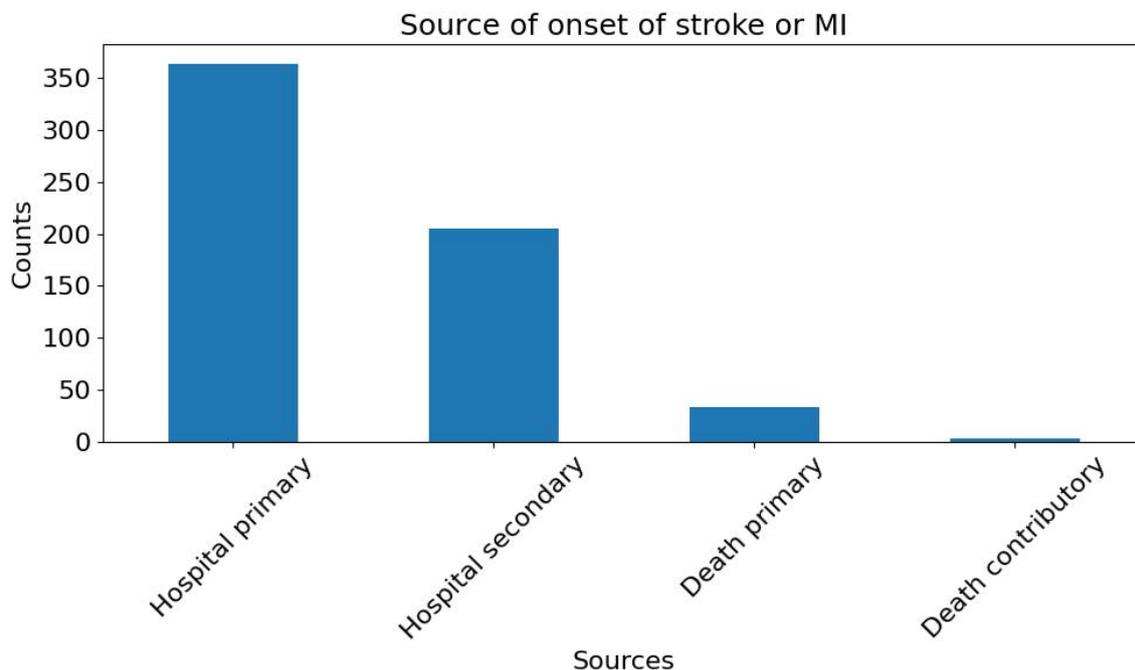


Figure 4.3: Bar plots showing the methods used to determine the onset date of stroke or MI for the 612 CVD+ subjects involved in our classification task

The size of the CVD+ group was determined based on the application of specific inclusion and exclusion criteria, as outlined in the STROBE diagram in Figure 4.2, resulting in a final count of 612 participants with OCT images of both eyes. For the non-CVD group, 2,234 participants were propensity score-matched based on sex and age or CVD- for OCT images of both eyes. The essential patient characteristics used to match the CVD+ and CVD- groups included demographic factors and clinical measurements, which are detailed in Table 4.1. The average age of individuals with and without CVD was  $60.78 \pm 6.47$  years, showing no significant difference between the two groups. The majority of participants in the UK Biobank cohort were of white ethnicity, with similar proportions in both groups. The average body mass index (BMI) was  $28.31 \pm 4.45$  kg/m<sup>2</sup> for those with CVD and  $27.43 \pm 4.33$  kg/m<sup>2</sup> for those without. In terms of blood pressure readings, individuals with CVD had a systolic blood pressure (SBP) of  $147.26 \pm 19.57$  mm Hg, while those without CVD had an SBP of  $145.1 \pm 18.75$  mm Hg. The diastolic blood pressure (DBP) was  $84.75 \pm 10.23$  mm Hg for individuals with CVD and  $83.22 \pm 9.73$  mm Hg for those without. The mean level of haemoglobin A1c (HbA1c) was  $36.52 \pm 4.32$  mmol/L for individuals with CVD and  $36.59 \pm 6.61$  for those without. A notable percentage of participants reported

being current alcohol consumers, accounting for 90.69% of individuals with CVD and 91.83% of those without. These participant characteristics for both groups are summarized in Table 4.1.

Characteristics	CVD+	CVD-
Number of subjects	612	2234
Age: Mean (s.d), years	60.78 (6.47)	60.78 (6.47)
Gender: F, M %	29.74, 70.26	29.74, 70.26
Ethnicity, %	90.18 White, 4.26 Mixed, 3.93 Asian or Asian British, 0.33 Black or Black British, 0.16 Chinese, 1.15 Other ethnic group	89.22 White, 4.25 Mixed, 4.41 Asian or Asian British, 0.82 Black or Black British, 0.49 Chinese, 0.82 Other ethnic group
BMI: Mean, kg/m <sup>2</sup>	28.31 (4.45)	27.43 (4.33)
SBP: Mean, mm Hg	147.26 (19.57)	145.1 (18.75)
DBP: Mean, mm Hg	84.75 (10.23)	83.22 (9.73)
HbA1c: Mean, mmol/mol	36.52 (4.32)	36.59 (6.61)
Alcohol consumption: N, P, C, NA %	3.59, 5.72, 90.69, 0	3.92, 3.92, 91.83, 0.33

Table 4.1: Characteristics of patients in the CVD+ and CVD- sets. N, Never. P, Previous. C, Current. NA, Not answer.

We used an age-sex-matched cohort for controls in CVD and non-CVD scenarios with a ratio of 1:3, respectively. One significant benefit of employing an age-sex-matched study cohort is that it helps mitigate the effects of confounders that might heavily influence the predictive model. Machine learning models may capture spurious correlations (i.e. to learn shortcuts) between predictors and targets, such as, for example, linking age or gender to the presence of pathology, unless care is taken when defining the predictors and study cohort (Brown et al. 2022). Figure 4.4 illustrates the age and sex in the CVD+ and CVD- patient groups, which were used to train and evaluate the predictive model proposed in this study. The construction of the metadata incorporated eight clinical variables, namely sex, age, HbA1c, systolic and diastolic blood pressure, alcohol consumption, and body mass index. The decision was made to omit the smoking variable from the study analysis because a significant number of participants did not provide responses to the relevant questionnaire item. From now on, the term metadata will be used to describe patient details, including demographic and clinical history information.

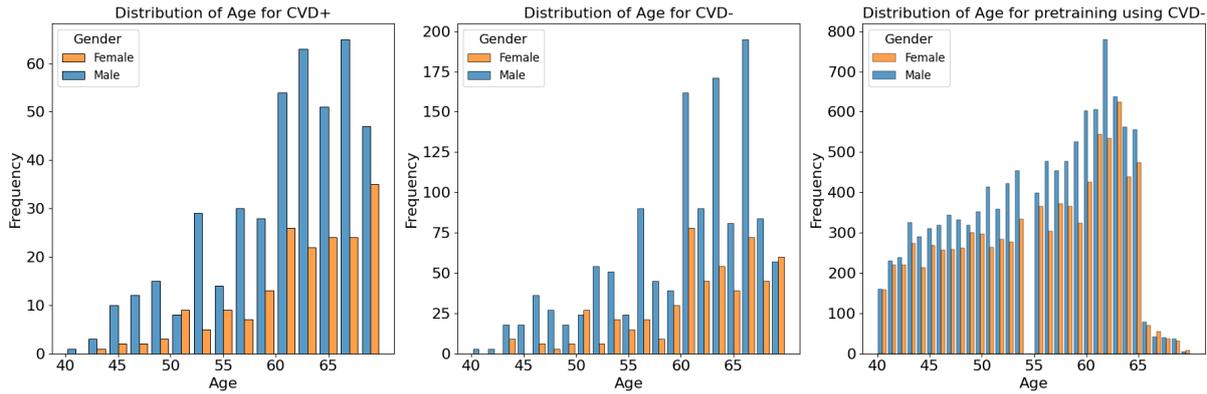


Figure 4.4: Distribution of age-sex cohort match. M, Male. F, Female. The left histogram illustrates the total of CVD+ labeled data used solely for the classification task. The middle histogram shows the total of CVD- for the classification task, while the right histogram illustrates CVD- used for the pretraining task.

## 4.3 Methodology

### 4.3.1 Framework of self-supervised feature selection Variational Autoencoder and multimodal random forest classification

This study proposes a predictive model for classifying patients into CVD+ and CVD- categories, comprising a Variational Autoencoder (VAE) (Kingma et al. 2014) to extract features in a self-supervised manner from retinal OCT images, and a Random Forest (RF) classifier which combines the former with patient metadata and uses the resulting set of multimodal features as predictors. The proposed model consists of two stages, a self-supervised feature extraction stage, and a subsequent classification stage. A schematic diagram of the overall predictive framework is shown in Figure 4.5.

#### Self-supervised VAE

In the first stage of the proposed model, a VAE is used to learn latent representations for B-scan OCT images. We describe VAEs with more detail in Chapter 2 Section 2.6.1.

The loss function utilised for training the proposed self-supervised VAE comprises two key elements: (1) the loss of mean square error (MSE)  $\mathcal{L}_{MSE}$ , detailed in Equation (4.2), which evaluates the discrepancy in reconstruction between the original data ( $\mathbf{x}_i$ ) and the reconstructed data ( $\hat{\mathbf{x}}_i$ ), and (2) the loss of Kullback-Leibler divergence  $\mathcal{L}_{KL}$ , illustrated in Equation(4.3). KL divergence quantifies the dissimilarity between the learnt latent distribution and a previously specified distribution  $p(\mathbf{z})$ , which, in this scenario, is a multivariate Gaussian distribution. The parameters of the learned distribution  $q(\mathbf{z})$  are its mean ( $\mu_i$ ) and variance ( $\sigma_i^2$ ) (the derivation of the KL divergence between the  $q(\mathbf{z}|x)$  and the  $p(\mathbf{z})$

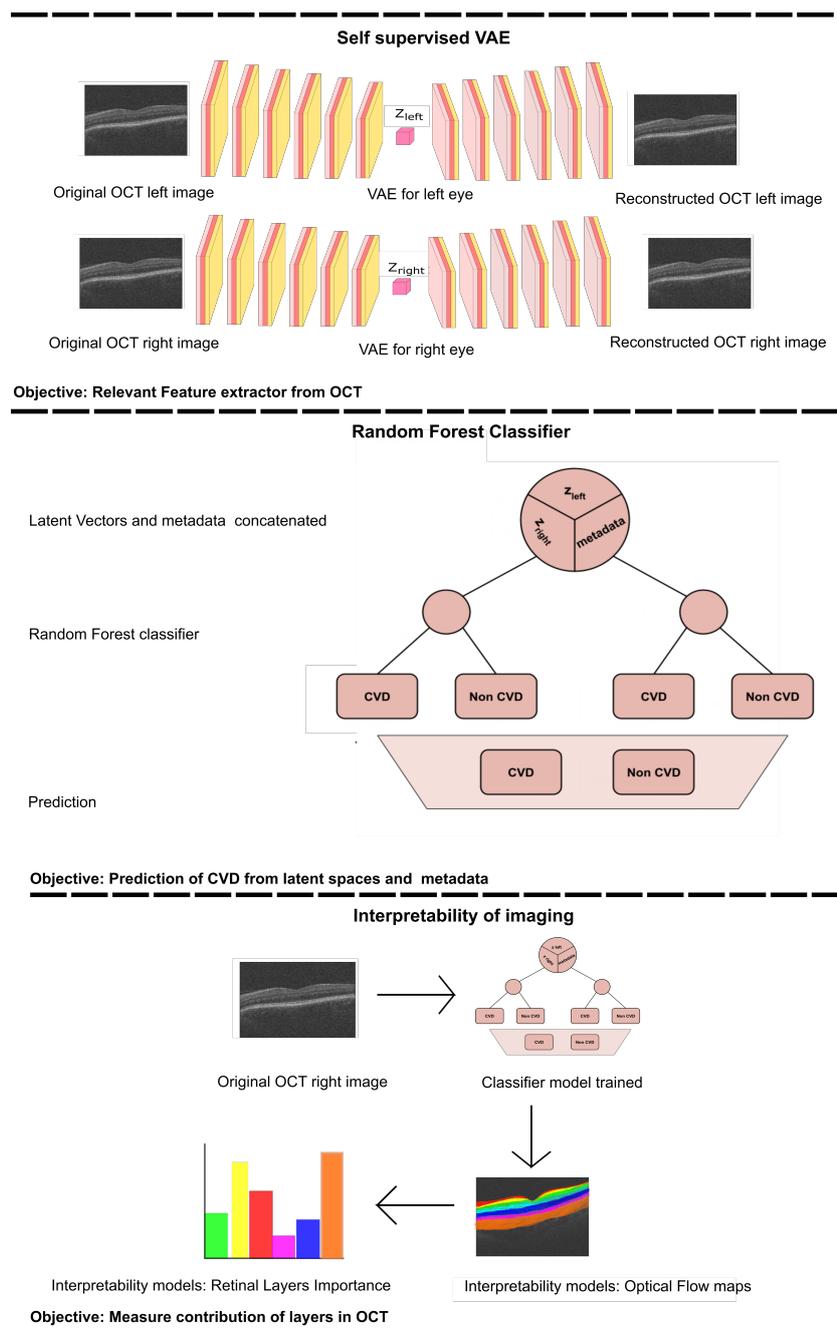


Figure 4.5: The workflow diagram illustrates the comprehensive process of training the Variational Autoencoder (VAE) and subsequently using it to acquire the latent vectors (upper section). These latent vectors are then combined with metadata and serve as inputs to the Random Forest (RF) classifier (middle section). Finally, we perform an interpretability analysis by perturbing the most relevant features, reconstructing the corresponding image and computing the vector field between the perturbed reconstructions (lower section).  $z_{left}$  represents the latent vector obtained from the training of the VAE for the left eye.  $z_{right}$  corresponds to the latent vector acquired from training the VAE for the right eye.

is explained in Kingma et al. 2014). By minimising the KL divergence, the model is incentivised to shape a latent space that adheres to the target Gaussian distribution. The integration of these two loss components steers the VAE towards the dual objective of reducing reconstruction errors and aligning the

learnt latent distribution with the intended prior distribution.

$$\mathcal{L}_{VAE} = \mathcal{L}_{MSE} + \beta \mathcal{L}_{KL} \quad (4.1)$$

$$\mathcal{L}_{MSE} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \hat{\mathbf{x}}_i)^2 \quad (4.2)$$

$$\mathcal{L}_{KL} = \frac{1}{2} [1 + \log(\sigma_i^2) - \sigma_i^2 - \mu_i^2] \quad (4.3)$$

We trained independent VAEs for the left and right eye, to learn unique latent features from the OCT images. Subsequently, a classifier used these learnt features to predict the probability of an individual's prospective CVD incidence.

### Classification

Using the features acquired from the VAE in the previous stage, we trained a Random Forest (RF) classifier to distinguish between individuals in the CVD+ and CVD- categories, as illustrated in Figure 4.5. The input for this process consists of the latent vector representation of each OCT image generated by the VAE for each eye, which is merged with a vector containing the relevant patient information. RF are a type of ensemble machine learning method that involves multiple decision trees, each of which is trained on a randomly selected subset of training data (Breiman 2001)(see Section 2.4.1). Using the power of numerous decision trees and incorporating random feature selection, this ensembling technique enhances the generalisability of the predictive model to new data by reducing model variance by averaging predictions from the trees in the ensemble. RF have been widely applied in medical settings for both classification and regression tasks, including in previous studies related to CVD diagnosis (Khozeimeh et al. 2022; Yang et al. 2020). One notable advantage of RF compared to other classification algorithms is their ability to easily handle multimodal data that include various data types (such as categorical, ordinal, and continuous). Decision trees within the ensemble operate independently, and their combined predictions are aggregated to produce the final RF prediction for a given input using majority voting for classification tasks. This structure also provides feature importance, which enhances the explainability of the model's decisions. Additionally, RF is computationally efficient when compared to more complex models, such as neural networks, as it does not require GPU resources for training.

### 4.3.2 Model explainability

Predictive models based on machine/deep learning algorithms, proposed for identifying risk of disease from medical imaging often fail to report both local and global explanations for the model's predictions. This is especially prevalent in the case of deep learning-based approaches that are often treated as black boxes, with little information provided on the mechanism by which models arrive at specific predictions. Local explanations provide insights to individual decisions/predictions of the model. For example, this may involve identifying specific input variables/regions of an image that had the most influence on the model's prediction for that instance. On the other hand, global explanations describe the model's behaviour across predictions for all instances in all classes of interest. Specifically, global explanations provide information on the most common discriminative features identified by the model for all instances in each class of interest. Providing both local and global explanations of model behaviour is essential for developing responsible AI in healthcare applications, as it can help identify systematic biases in data and mitigate for the same (e.g., learning of 'short-cuts' is a common issue encountered in the application of deep learning-based methods for predictive tasks using medical images), build trust in AI systems by improving transparency in model decision making, and may even provide new insights to previously known associations between image-derived phenotypes and the presence or progression of diseases. Therefore, in this study, we employ distinct techniques to provide both local and global explanations for the proposed predictive model.

#### Local explanations

To provide local explanations of the behaviour of the model and elucidate how the model uses OCT imaging-derived features to classify instances in the CVD+ group, we first selected the best performing RF classifier according to AUC value. Subsequently, based on the selected RF classifier, we identified the latent variable derived from the OCT image with the highest importance assigned by the RF, which we denoted  $\mathbf{z}_{max}$ . To visually assess the regions of the retina in OCT images that contribute significantly to the prediction of CVD, we propose a novel vector field-based latent traversal approach that evaluates the impact of perturbing the most important latent feature  $\mathbf{z}_{max}$  on subsequent reconstructions of OCT images. Specifically, given an image  $\mathbf{x}$ , we compute the corresponding latent vector  $\mathbf{z}$  using the trained encoder network. Next, we perturb only the dimension  $\mathbf{z}_{max}$  of the calculated latent vector and reconstruct the corresponding image. This perturbation is performed by multiplying the latent dimension  $\mathbf{z}_{max}$  by a scalar value, which in this case is the standard deviation of this latent component, calculated

across the training population and defined as  $\sigma_{max}$ . The remaining latent variables in the computed latent vector are left unchanged, resulting in a perturbed latent vector  $\hat{\mathbf{z}}$ . Finally, we reconstruct the input OCT image  $\hat{\mathbf{x}}$  using the perturbed latent vector  $\hat{\mathbf{z}}$ .

To visualise the regions in the reconstructed OCT images affected by the altered latent vector  $\hat{\mathbf{z}}$ , we examine the differences between the initial image  $\mathbf{x}$  and the reconstructed image  $\hat{\mathbf{x}}$  derived from  $\hat{\mathbf{z}}$ . In this context, we calculate the vector field between these images using the Lucas-Kanade algorithm (Lucas et al. 1981). The resulting vector field, showing the magnitude of the vector field for the moving pixels, was then superimposed on the original image, as shown in Figure 4.9. The vector field indicates how the pixels between the images (that is,  $\hat{\mathbf{x}}$  and  $\hat{\mathbf{x}}$ ) change due to the latent traversal from  $\mathbf{z}$  to  $\hat{\mathbf{z}}$ . The estimated vector field between  $\mathbf{x}$  and  $\hat{\mathbf{x}}$  helps to visually illustrate the regions in the OCT image that were altered by modifying the latent component  $\mathbf{z}_{max}$ . This aids in pinpointing the areas of the image influenced by alterations to the crucial latent variable for accurately classifying a patient’s CVD risk based on their OCT image(s), and consequently, helps to understand which retinal areas are informative for distinguishing between the CVD+ and CVD- patient groups.

### Global explanations

To provide global explanations of the behaviour of the model, we calculate the importance of the characteristics assigned to each characteristic by the RF in each predictive model investigated. As mentioned previously, the RF in each predictive model were trained using a reduced set of characteristics identified by RFE. Feature importance is calculated in RF as the average Gini information gain for any given feature, calculated across all decision trees in the forest. The feature importance values for all classifiers studied in this work, across all test set instances, are summarized as bar plots later. Additionally, we calculate the relative importance of the type / channel of data used as inputs/predictors in this study, namely, OCT images of the left and right eye and patient metadata, for the task of distinguishing between the CVD+ and CVD- groups.

## 4.4 Experiments

All experiments were carried out with an NVIDIA Tesla M60 GPU. The model was trained using PyTorch (v1.10.2) and a grid search strategy was used with five-fold cross-validation to determine the best hyperparameters (Table A.1). The data set was divided into training, validation, and test sets in a ratio of 6: 2: 2. The encoder and decoder networks were constructed with six 2D convolution layers each; the

encoder used Rectified Linear Unit (ReLU) activations, while Leaky Rectified Linear Unit (LeakyReLU) activations were used in the decoder (see Table A.2).

During the classification phase, we conducted a thorough investigation into the impact of latent representations derived from the OCT images of the left and right eyes, along with patient metadata, on the predictive task. This was accomplished by generating seven distinct datasets from the same group of 2846 patients, each comprising different combinations of data sources: (i) latent representations from the left eye only (LE); (ii) latent representations from the right eye only (RE); (iii) latent representations from both eyes (BE); (iv) metadata only (MTDT); (v) left eye with metadata (LE-MTDT); (vi) right eye with metadata (RE-MTDT); and (vii) both eyes with metadata (BE-MTDT). Random Forest classifiers were trained separately on each of these seven datasets, as depicted in Figure 4.5 for the BE-MTDT dataset. Finally, the optimal hyperparameter values of RF classifiers were determined through a combination of grid search and empirical experimentation, to identify the best performing RF model for each specific dataset (see Table A.3). We divided the dataset into training, validation, and test sets, following a distribution ratio of  $\sim 5 : 2 : 3$ , respectively (resulting 1882 patients in the training set and 964 patients in the held-out unseen test set). Grid search was performed using five-fold cross-validation, while an independent, unseen test set remained fixed throughout all experiments to evaluate all trained classifiers fairly. Furthermore, a feature selection method using Recursive Feature Elimination (RFE) (Guyon et al. 2002) was employed to mitigate overfitting and train the model with the most relevant variables for classification. RFE is to select features by recursively considering smaller and smaller sets of features. In the majority of cases, the model was trained with the top 10 most significant features, with the exception of the RE case, where, we used only 5 variables to avoid overfitting.

To evaluate the effectiveness of our model, we compared predictive performance against the QRISK3 algorithm (Li et al. 2019), the current gold standard used by healthcare professionals / cardiologists to assess the patient’s risk of stroke or heart attack (acute myocardial infarction), in a 10-year period. The QRISK score was calculated within the specified test dataset, following the methodological guidelines outlined in (Li et al. 2019). The evaluation of the QRISK3 score involved entering essential variables such as the cholesterol-to-HDL ratio, age, SBP, standard deviation of SBP, smoking status, BMI, Townsend score, sex, weight, height, and ethnicity based on available data. For our classification task, we evaluated the model performance using a range of metrics. Accuracy, precision, sensitivity, and specificity were determined by calculating true positives, true negatives, false positives, and false negatives (using a classification probability threshold of  $t = 0.5$ , i.e. if the predicted probability is

$\geq 0.5$ , the patient is classified as CVD+, else as CVD−). The receiver operating characteristic (ROC) curve was constructed by computing the true positive rate and false positive rate. The area under the ROC curve (AUROC) was then employed as a performance measure to assess both our model and the QRISK3 algorithm.

## 4.5 Results

### 4.5.1 Classification performance

As discussed previously, we trained and evaluated the performance of several RF classifiers, where each classifier was trained and evaluated independently using seven different combinations of data types obtained from the same set of patients (comprising CVD+ and CVD− groups). Specifically, the datasets used were LE, RE, BE, LE-MTDT, RE-MTDT, BE-MTDT and MTDT. Henceforth, for brevity, we refer to classifiers trained and evaluated on these datasets as LE-RF, RE-RF, BE-RF, LE-MTDT-RF, RE-MTDT-RF, BE-MTDT-RF, and MTDT-RF. The performance of all seven classifiers was evaluated and compared using the same unseen test set (which contains 964 patient data, 834 CVD− and 130 CVD+), and using the same set of evaluation metrics outlined in Section 4.4.

The rationale for comparing all seven classifiers against each other was to:

- Assess whether combining latent features learnt from OCT images of both eyes (BE) provided greater discriminative power than using those from either left (LE) or right eye (RE) alone
- Compare the discriminative power of OCT image-derived latent features against patient metadata
- Evaluate the discriminative power gained by enriching OCT image-derived latent features with patient metadata

The performance of all seven classifiers on the unseen test set is summarised in Table 4.2. These results show that the BE-MTDT-RF classifier consistently outperformed all six other classifiers, with statistically significant differences in p-values (refer to Table 4.3). This suggests that combining information from OCT images of both eyes (ie, learnt latent representations) with patient metadata was more informative in distinguishing between the CVD+ and CVD− groups. In terms of evaluating the effectiveness of OCT image-derived characteristics and metadata information for classifying CVD+ and CVD− patients, results for the BE-MTDT-RF, LE-MTDT-RF, RE-MTDT-RF and MTDT-RF classifiers indicate that combining latent features learnt from OCT images with patient metadata (i.e. BE-MTDT-RF, LE-

MTDT-RF, RE-MTDT-RF), consistently improves classification performance relative to using patient metadata alone (MTDT-RF).

Classification Criteria				
Classifiers	Accuracy	Sensitivity	Specificity	AUC
LE-MTDT-RF	0.68	0.69	0.68	0.72
RE-MTDT-RF	0.67	0.69	0.67	0.70
<b>BE-MTDT-RF</b>	<b>0.70</b>	<b>0.70</b>	<b>0.70</b>	<b>0.75</b>
MTDT-RF	0.52	0.52	0.52	0.57
LE-RF	0.61	0.56	0.62	0.64
RE-RF	0.58	0.53	0.59	0.62
BE-RF	0.64	0.57	0.65	0.67

Table 4.2: Predictive analysis of cardiovascular disease (CVD) metrics utilizing UK Biobank data across seven distinct cases.

Specifically, the BE-MTDT-RF classifier demonstrated the highest performance, achieving an accuracy of 0.70, sensitivity of 0.70, specificity of 0.70, and an AUC score of 0.75. Furthermore, combining latent features from RE or LE OCT images with metadata (i.e., LE-MTDT-RF and RE-MTDT-RF) resulted in improvements of 13 – 17% across all classification metrics relative to the MTDT-RF classifier. Notably, the MTDT-RF classifier exhibited the lowest values across all classification metrics, with an accuracy, sensitivity, and specificity of 0.52, and an AUC score of 0.57. The BE-RF, RE-RF, and LE-RF classifiers also consistently outperformed the MTDT-RF classifier in all classification metrics. However, they were outperformed by classifiers that incorporated OCT image-derived features with patient metadata (i.e., BE-MTDT-RF, LE-MTDT-RF, RE-MTDT-RF).

Figure 4.6 presents four histograms depicting true positives, true negatives, false positives, and false negatives for the seven classifiers investigated. A consistent observation across all our results is that the BE-MTDT-RF classifier misclassified fewer instances in the CVD+ group, than all other classifiers, which is consistent with the classification metrics summarised in Table 4.2. Similarly, the LE-MTDT-RF and RE-MTDT-RF classifiers exhibited good sensitivity by incurring few false negative errors percentages (11%) in the CVD+ group were incorrectly classified. The MTDT-RF classifier yielded fewer true positives and true negatives (12% in both) compared to the other classifiers that utilized only OCT features, which aligns with the results presented in Table 4.2.

A significant observation in the results is that including both eyes was advantageous for both cases, BE-MTDT-RF and BE-RF, compared to their counterparts, LE-MTDT-RF and RE-MTDT-RF, RE-RF, and

Classifier 1	Classifier 2	$\chi^2$	P value
BE-MTDT-RF	LE-MTDT-RF	15.21	$9.261 \times 10^{-05}$
BE-MTDT-RF	RE-MTDT-RF	21.16	$4.22 \times 10^{-06}$
BE-MTDT-RF	MTDT-RF	93.23	$4.65 \times 10^{-22}$
BE-MTDT-RF	BE-RF	10.59	0.0011
BE-MTDT-RF	LE-RF	27.43	$1.63 \times 10^{-07}$
BE-MTDT-RF	RE-RF	43.83	$3.57 \times 10^{-11}$
BE-RF	LE-RF	21.16	$4.22 \times 10^{-06}$
BE-RF	RE-RF	37.78	$7.89 \times 10^{-10}$
LE-MTDT-RF	RE-MTDT-RF	6	$9.61 \times 10^{-05}$
LE-RF	RE-RF	17.06	$4.22 \times 10^{-06}$

Table 4.3: Comparison between various classifiers based on different configurations of eye data (both eyes, left eye, right eye) and metadata inclusion. The chi-squared ( $\chi^2$ ) values and associated p-values indicate the statistical significance of differences between these classifiers.

LE-RF, respectively. Furthermore, in both scenarios, with/without metadata, the left eye consistently provided improved classification performance compared to the right eye. This finding was statistically significant, as indicated by the p-values reported in Table 4.3. This observation is in concordance with the global explanations of models' predictions summarised in Figure 4.7, wherein features attributed to the left eye were found to be more discriminative (i.e., had higher feature importance) than those of the right eye. We posit that the impact on the latent vectors associated with the left eye is interconnected with the superior image quality of the images of the left eye within our cohort (see Figure 4.1). The UKBB standard operating procedure stipulated that the second eye imaged was consistent with the left eye. This protocol was not randomised. As a result, there may be potential systematic disparities between left and right eye OCTs (for example, left eye scans might consistently exhibit better quality because they are the second scan performed and patients are potentially more adept at following instructions). These collective findings underscore the improved performance achieved by integrating retina OCT imaging and metadata in the classification task.

Subsequently, we conducted a comparative analysis of the best classifier identified from the previous experiments, namely BE-MTDT-RF, and the QRISK3 algorithm, the current clinical standard for assessing patients at risk of stroke or MI. In our age-sex matched UK Biobank cohort, BE-MTDT-RF achieved modestly higher discriminative performance (AUC: 0.75 vs. 0.60; Table 4.4). However, this difference requires careful interpretation. QRISK3's relatively lower performance here may reflect the cohort's age-sex matching, which attenuates the algorithm's reliance on its two strongest predictors (age

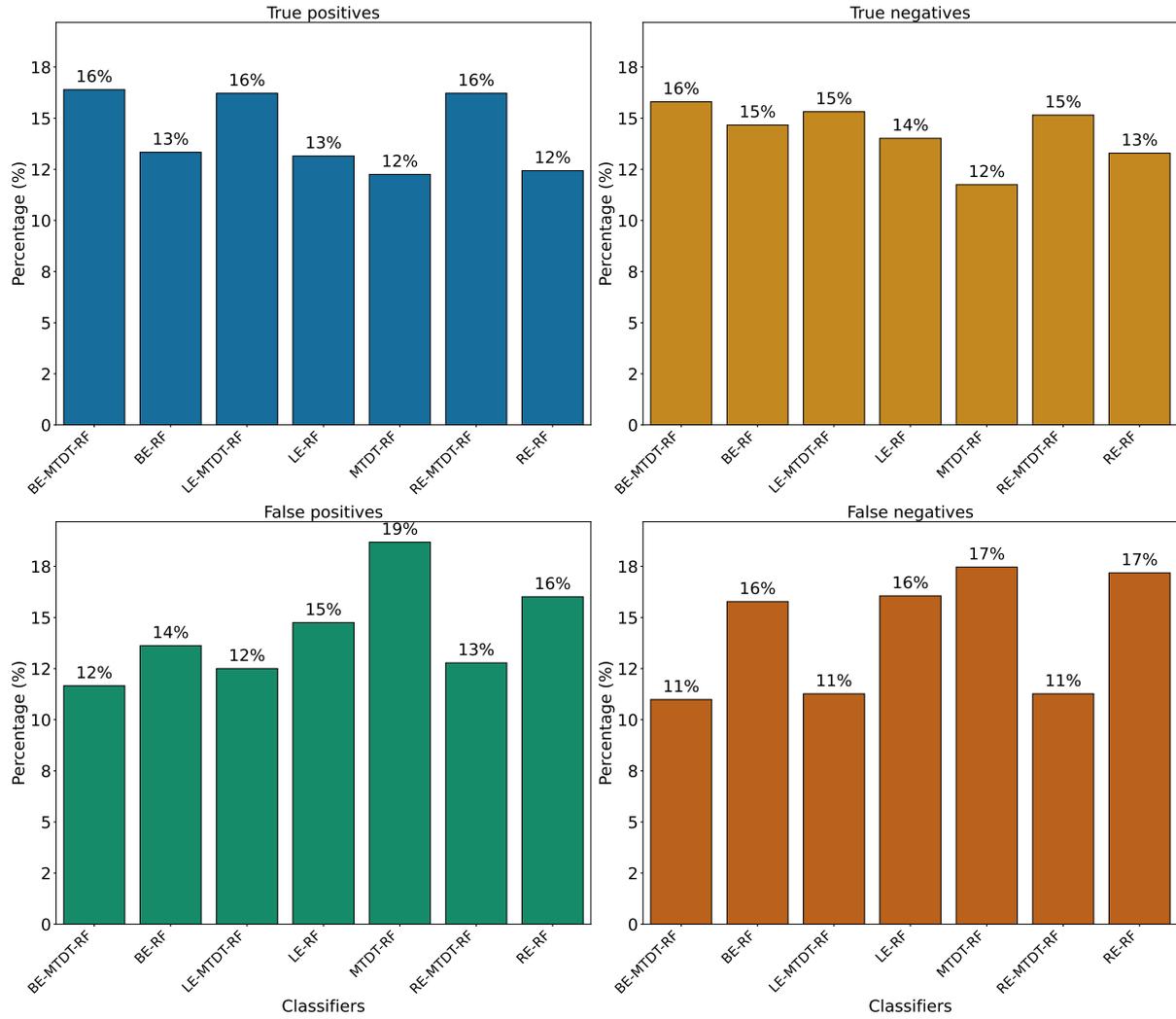


Figure 4.6: Comparison of classifier performance in terms of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Values above the bars represent the corresponding percentages for each classifier.

and sex) (Ghosa et al. 2022). In contrast, our model leverages retinal microvascular features, which may capture risk heterogeneity independent of demographics, potentially identifying high-risk individuals overlooked by QRISK3’s macrovascular/metabolic focus. While promising, this advantage must be validated in prospective, unmatched cohorts to assess real-world utility, as QRISK3 remains robust in general populations.

Modality	Accuracy	Sensitivity	Specificity	AUC-Value
<b>BE-MTDT-RF</b>	<b>0.70</b>	<b>0.70</b>	<b>0.70</b>	<b>0.75</b>
QRISK	0.55	0.60	0.545	0.60

$$\chi^2 = 95.72, df = 1, p = 1.31 \times 10^{-22}$$

Table 4.4: Comparison of classification metric results between our model employing both ocular data and metadata (BE-MTDT) and the QRISK algorithm. McNemar’s Test (\* $p < 0.005$ )

### 4.5.2 Ablation study

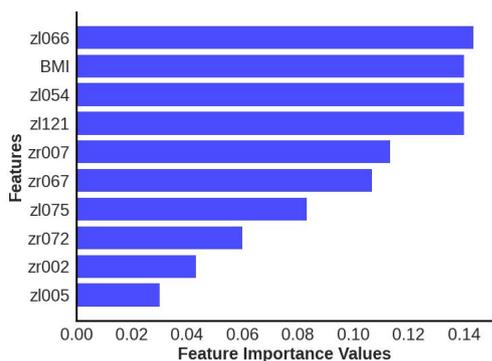
We conducted an ablation study comparing the BE-MTDT-RF model, which achieved superior results, against a Multilayer Perceptron (MLP) algorithm. Table 4.5 provides the classification performance metrics evaluated on the same holdout set used for the experiments reported in Table 4.2. For the MLP model, we conducted a comprehensive grid search to identify its optimal configuration, ensuring a fair comparison (Table A.4). Both models were evaluated using ocular data and metadata as input features. The BE-MTDT-RF model demonstrated superior performance across all evaluated metrics. Beyond its better performance, the BE-MTDT-RF model offers additional advantages: (i) RF feature importance analysis provides insights into the contribution of individual features, aligning with our exploratory goals and enhancing domain-specific understanding; (ii) Unlike MLPs, RF are less computationally demanding and do not require GPU acceleration, making them more practical for our analysis.

Classification criteria				
Modality	Accuracy	Sensitivity	Specificity	AUC-Value
<b>BE-MTDT-RF</b>	<b>0.7</b>	<b>0.7</b>	<b>0.7</b>	<b>0.75</b>
MLP	0.65	0.635	0.65	0.719

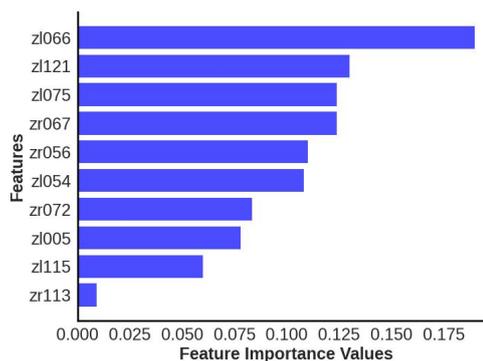
Table 4.5: Comparison of classification metric results between our model employing both ocular data and metadata (BE-MTDT-RF) and the MLP algorithm.

### 4.5.3 Model explainability

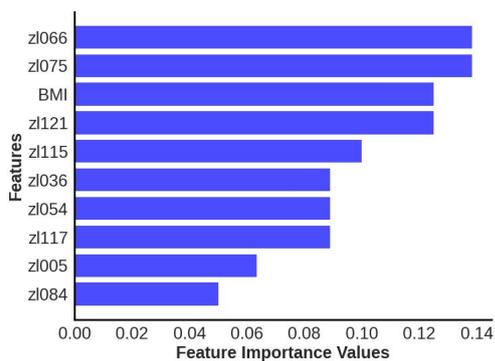
In order to provide global explanations for the behavior of all classifiers investigated in this study, we analysed the most important features identified by each model (refer to Figure 4.7) for distinguishing between the  $CVD+$  and  $CVD-$  groups. Important features identified for the best performing classifier, namely, BE-MTDT-RF in particular, provided some noteworthy insights. As highlighted in Figure 4.7(a), we found that a latent variable learned from the left-eye OCT image, denoted  $z_{l066}$ , had the most influence on the classifier’s ability to separate  $CVD+$  and  $CVD-$  patient groups. Additionally, among the top 10 most important features identified for the BE-MTDT-RF classifier, 9 of the features pertained to latent variables learned from the left-eye OCT image (see Figure 4.8). BMI was the only feature from the basic set of patient metadata used to train the classifier, that was identified to have a significant influence on the classifier’s predictions. Furthermore, looking at the global explanations summarised in Figure 4.7(b), (e) and (f), we observe that latent variable  $z_{l066}$  consistently ranks among the top two most important features for the LE-MTDT-RF, BE-RF and LE-RF classifiers, respectively. This indicates that the retinal features encoded by  $z_{l066}$  are consistently considered to be relevant across



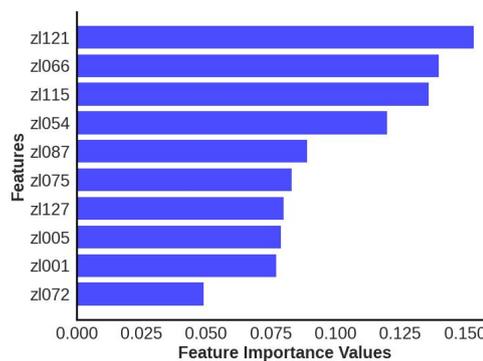
(a) BE-MTDT-RF



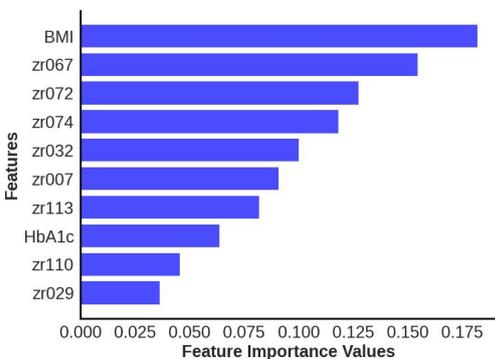
(b) BE-RF



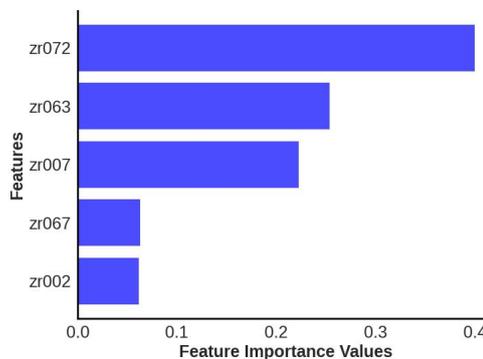
(c) LE-MTDT-RF



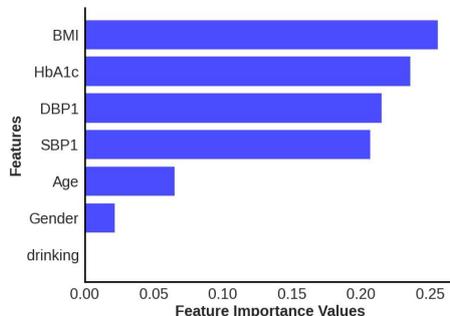
(d) LE-RF



(e) RE-MTDT-RF



(f) RE-RF



(g) MTDT-RF

Figure 4.7: Calculation of feature importance magnitudes for the seven different classifiers investigated, where each classifier uses different combinations of data channels/modalities. Latent variables starting with 'zr' refer to the right eye, while those starting with 'zl' refer to the left eye.

all four classifiers presented in Figure 4.7(a), (b), (e) and (f). Among the classifiers which combined retinal OCT-image derived features with patient metadata, namely, BE-MTDT-RF, LE-MTDT-RF and RE-MTDT-RF, we observed that only two features, namely, BMI and HbA1c, ranked among the top 10 most important features for the classification task. Both features are known and established cardiovascular risk factors, and importantly, we infer from these results that the latent variables learned from the retinal OCT images, had a greater influence on the classifiers' predictions than the patient metadata variables. As previously highlighted, we hypothesize that the significant importance of the latent variables corresponding to the left eye can be attributed to the superior image quality of left-eye OCT images within our cohort (as illustrated in Figure 4.1). As a result, corresponding latent vectors  $\mathbf{z}$  effectively capturing image features that potentially enhance predictive capabilities.

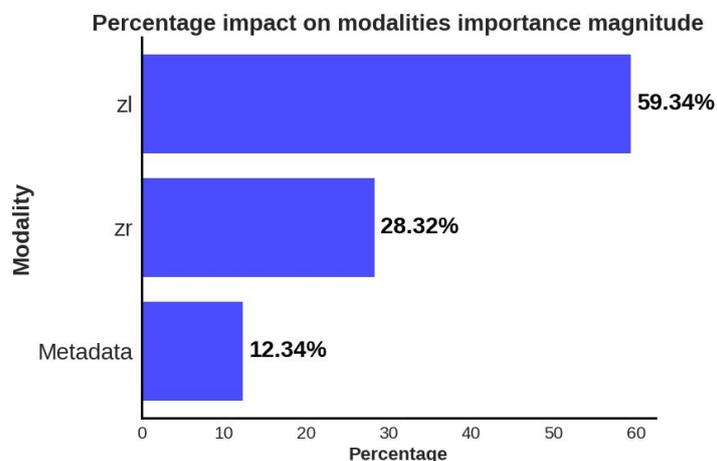


Figure 4.8: Global explanations of features from different data modalities/channels which were considered important by the predictive model for separating the  $CVD+$  and  $CVD-$  groups. Bar plot summarises the relative importance of latent variables from left ( $zl$ ) and right ( $zr$ ) eye OCT images and patient metadata, as percentages.

Using the insights gained from analysing the global explanations of classifier behavior summarised in Figure 4.7, we propose a novel approach based on latent space traversals to translate the former into local explanations that provide insights to regions of the OCT image that contain relevant information for correctly identifying patients at risk of cardiovascular disease. Specifically, having identified latent variable  $zl066$ , derived from left-eye OCT images as being the most important feature for classification, our local explainability approach (refer to Section 4.3.2) begins by perturbing the values of the latent variable for any image in the  $CVD+$  group, reconstructs the OCT image using the perturbed latent representation (using the pretrained VAE) and then estimates vector field maps between the original and perturbed OCT image reconstructions, seeking to pinpoint the specific image regions that change as a

result of the perturbation.

We conducted a qualitative analysis by estimating the vector field maps between the original reconstructed B-scan OCT images and their perturbed counterparts for all CVD patients in the test set. By overlaying the estimated vector field map onto the original OCT images, we visualized the retinal image features encoded by the latent variable of interest. Although this process was applied to all high-risk CVD patients, we present results for five representative cases here (Figure 4.9), with each row representing one patient. For each patient, we analyzed three B-scans: the 1st (top row), 64th (middle row), and 128th (bottom row).

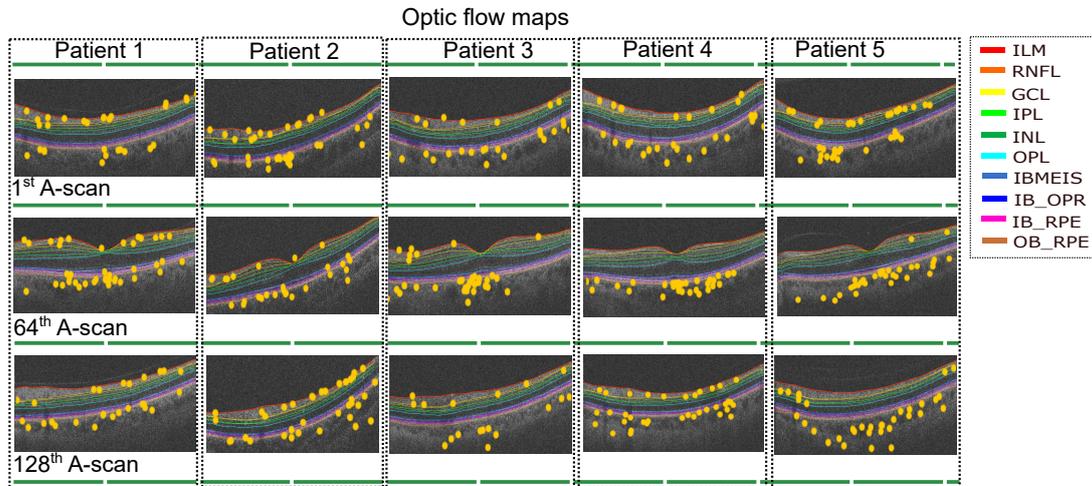


Figure 4.9: Vector field maps presented for three different B-scans for the left eye. The top row corresponds to the 1st B-scan, the middle image for the 64th B-scan while the bottom row depicts the final B-scan. The yellow circles represent the regions that the vector field maps highlight when modifying the latent variable  $z_{l066}$ .

The vector field maps generated by our model highlighted the choroidal layer in the majority of B-scans, with additional identification of layers adjacent to the choroid, including the retinal pigment epithelium (RPE). Additionally, the vector field maps highlighted the regions in the inner retinal layers, likely corresponding to the retinal nerve fiber layer (RNFL) and ganglion cell layer (GCL) (Zhou et al. 2023). Although some other layers received some, if less, emphasis, the main focus remained on the choroidal layer and the innermost layers. The vector field maps provide precise localisation of the image regions modelled by latent variable  $z_{l066}$  (visualised as landmarks, as shown in Figure 4.9), thereby providing local explanations for the most discriminative regions within the OCT, and providing insights to which retinal layers may contain relevant information for predicting risk of CVD in patients. In particular, these local explanations highlighted the relevance of information contained within the choroidal layer of the retina, for distinguishing between CVD+ and CVD- patient groups. Furthermore, vector field

maps demonstrated the most consistent choroidal representation across all B-scans compared to other established explainability methods (see Figure A.1).

## 4.6 Discussion

Our findings indicate that the use of retinal OCT images in conjunction with VAE and multimodal RF classification has potential to identify patients at risk of CVD (within a five-year interval). Our investigation included the deployment of a self-supervised VAE coupled with an RF classifier framework, which incorporates B-scan OCT images and metadata as distinct modalities. This integration allowed our model to discern the specific attributes within the OCT images that contribute significantly to the prediction of CVD. Importantly, our study distinguishes itself by interpreting the particular OCT image features (at both the global i.e. class/category, and local i.e. instance, levels), which are relevant to the classification task and thereby provide insights to the key regions of the retinal image that are most discriminative. To the best of our knowledge, some studies have ventured into the application of OCT within a primary care framework for CVD. However, these studies were limited in their explanatory capacity regarding the effects of including images from both eyes and different types of patient data, and used a small portion of the OCT B scans, limiting the information from the entire volume. Nevertheless, the performance results show promising outcomes for OCT as a modality in the primary care of CVDs (Zhou et al. 2023). Additionally, a key benefit of the proposed approach is that it lends itself to explaining model predictions in both a global (across all instances) and local (instance-specific) sense, and thereby, provides insight into which retinal layers contain the most relevant information to identify risk of CVD.

Interestingly, our results suggest that choroidal morphology is a predictor of identifying patients at risk of CVD, which is consistent with previous studies (Yeung et al. 2020) that have reported significant associations between choroidal characteristics and the risk of stroke and acute myocardial infarction. Given that the choroid has the highest flow per perfused volume of any human tissue and that there is growing evidence that changes in the choroid microvasculature can be indicative of systemic vascular pathology (Ferrara et al. 2021), our findings offer clinical interpretability to the predictions of our classifier. Currently, UK Biobank images are captured using a spectral domain (SD) OCT (Keane et al. 2016). SD-OCT images suffer significant light scattering at the choroid, which limits the resolution of this layer. Despite this limited resolution, it is encouraging to observe that the proposed approach focused on features within the choroidal layer to identify patients at risk of stroke or myocardial infarction. To provide

a greater context to the key findings reported in this study, it appears probable that once image modalities with deeper tissue penetration, such as swept source OCT, become available at scale in population imaging initiatives (such as UK Biobank), the predictive performance of learning-based systems such as ours will improve (Copete et al. 2013). In addition to the choroidal characteristics, our results indicated that the inner retinal layers contributed to the classification, including the retinal nerve fibre layer (RNFL), the ganglion cell layer (GCL), and the inner plexiform layer (IPL). These aspects of the neurosensory retina consist of retinal ganglion cells, their synapses with bipolar cells, and their axons. Regarding the choroid, thinning and defects in these layers have received extensive study in relation to established CVD, but their role as predictors of future disease has received limited attention (Chen et al. 2023; Matuleviūtė et al. 2022). Mechanisms that may underpin the role of neurosensory retina morphology as a predictor of CVD are yet to be elucidated, although it could be hypothesised that subclinical ocular circulatory pathology could explain morphological changes through local ischaemic damage (Chen et al. 2023), or neuronal degeneration could even occur through silent subclinical cerebral ischaemic vascular changes manifesting in the inner retina through transneuronal retrograde degeneration (Langner et al. 2022; Park et al. 2013).

While our DL model demonstrates superior predictive performance compared to QRISK3, a direct comparison between the two is not entirely fair due to fundamental differences in their design and target population. QRISK3 has been extensively validated across large, diverse population datasets, such as the QResearch (Hippisley-Cox et al. 2017) and Clinical Practice Research Datalink (CPRD) cohorts (Livingstone et al. 2021), and is widely used in clinical practice. Besides, QRISK3 incorporates well-established clinical risk factors, including age, cholesterol, diabetes status, and smoking history, which have strong, well-documented associations with CVD (Parsons et al. 2023). In contrast, our model relies solely on retinal imaging, an emerging modality for CVD risk assessment that, while promising, has not yet been as extensively validated as traditional risk models. Furthermore, our cohort consists of age- and sex-matched participants, which may attenuate the algorithm's reliance on these two strong predictors (age and sex; Ghosa et al. 2022) and limit generalizability to real-world, unmatched populations. Given that QRISK3 is already clinically validated and widely adopted, further research is needed to determine how retinal imaging-based models can complement or enhance traditional risk assessments in a clinically meaningful way.

A key limitation of this study stems from the varying class imbalance ratios across different stages of the analysis. Initially, the CVD+ to CVD- ratio in the OCT imaging dataset was 1:30, reflecting the natural

prevalence in the UK Biobank. After exclusions for data quality, this ratio improved to 1:4, which was used for model training. While this balanced approach enhanced the classifier's ability to learn CVD+ patterns, it may have introduced optimism bias in performance metrics. The test set used a 1:8 ratio to better approximate real-world conditions, but this intermediate prevalence still differs from both the natural distribution and the training set. Consequently, while the model shows strong discriminative power (AUC 0.75), its sensitivity and specificity may not directly translate to clinical settings with different prevalence rates. Future work should validate these findings in entirely unmodified cohorts and explore techniques like cost-sensitive learning to account for natural class imbalances (Fernández et al. 2018).

Additionally, our dataset is derived from the UK Biobank, which is known to be less representative of the general UK population. Participants in UK Biobank tend to be older, more likely to be female, live in less socioeconomically deprived areas, have lower obesity rates, consume less alcohol, smoke less, and generally have fewer health conditions (Fry et al. 2017). This healthy volunteer bias could influence our model's performance and limit its generalizability to broader clinical settings.

Although the focus of this study was on spectral domain OCT imaging, future improvements to the presented work could include the use of more informative retinal imaging modalities such as swept source OCT imaging or wide-field OCT angiography (OCTA) imaging. We hypothesise that learning representations from multi-modal retinal imaging (e.g. fundus photographs, OCT, OCTA) may improve the classification performance of the proposed approach.

In general, this investigation has supported the utility and prospective predictive value of OCT imaging to identify people at risk of stroke or acute myocardial infarction, suggesting it may be useful for future clinical risk prediction.

## Chapter 5

# Integrating deep Learning with fundus and optical coherence tomography for cardiovascular disease prediction

### 5.1 Introduction

This chapter builds upon a conference paper originally presented at the PRIME workshop of the MIC-CAI conference (Maldonado-Garcia et al. 2025). In this chapter we employ both retina images available from the UKBiobank, fundus photographs and OCT, for CVD prediction. This study presents a novel predictive model that integrates various types of retinal data, including features derived from OCT imaging and fundus photographs, by employing a multichannel variational autoencoder and a transformer network classifier. The key contribution of this research is the development of an advanced predictive model that synergistically combines multi-modal retinal patient data to identify individuals at risk of developing CVD within five years of image capture. This approach aims to enhance our understanding of the systemic changes that contribute to an increased risk of cardiovascular diseases in patients.

A literature review on the use of retinal imaging is discussed in Chapter 3, Section 3.2. In this study, we propose the use of two retinal imaging modalities (fundus and optical coherence tomography). The intricate nature of biological tissues, organs, and disease processes requires the integration of complementary information provided by multimodal imaging techniques. This approach offers the medical community a more accurate characterization of disease phenotypes and enables the extraction of clinically relevant

quantitative data (Rajiah et al. 2019; Govindarajan et al. 2005).

Retinal imaging has shown significant promise in predicting CVDs, particularly due to its accessibility during routine ophthalmological or optometric exams. This accessibility suggests that cardiovascular health screening could become an ancillary benefit of standard eye assessments. However, the literature reveals a lack of studies that utilize both imaging modalities concurrently for CVD prediction. Most existing research fails to fully leverage the potential of combined data or conduct comprehensive comparative analyses of dual-modality approaches versus single-modality methods (Zhou et al. 2023; Huang et al. 2023).

Therefore, the development of advanced statistical and machine learning techniques for the effective integration and analysis of these diverse data sources is crucial for improving patient care.

## 5.2 Dataset

For the present study, we utilized retinal OCT and fundus imaging data obtained from the UK Biobank. We focused on retinal imaging of the left eye, as we assumed that prior familiarity with the testing procedure appeared to yield scans with systematically higher overall quality (Sergouniotis et al. 2024; Patel et al. 2016). During the baseline visit (2006-2010), a total of 67,656 participants underwent retinal imaging for both modalities. Quality control exclusion criteria for OCT and fundus images were applied in accordance with established protocols (Zekavat et al. 2023; Fu et al. 2019). Following rigorous quality assessment, 43,097 participants with high-quality OCT images and 41,271 participants with high-quality fundus images were identified as suitable for further analysis. Among these, 5,125 patients with OCT data and 3,911 patients with fundus data experienced a CVD event. After excluding patients whose images were acquired post-CVD events and those diagnosed with ocular diseases or diabetes, the final cohort comprised 2,142 patients for OCT images and 1,652 patients for fundus images who suffered CVD (CVD+). Notably, 977 CVD+ patients had both retinal modalities available. For the patients without CVD or control group (CVD-), 38,886 participants had OCT data, and 38,553 participants had fundus data. Of these, 21,758 participants had both retinal modalities available.

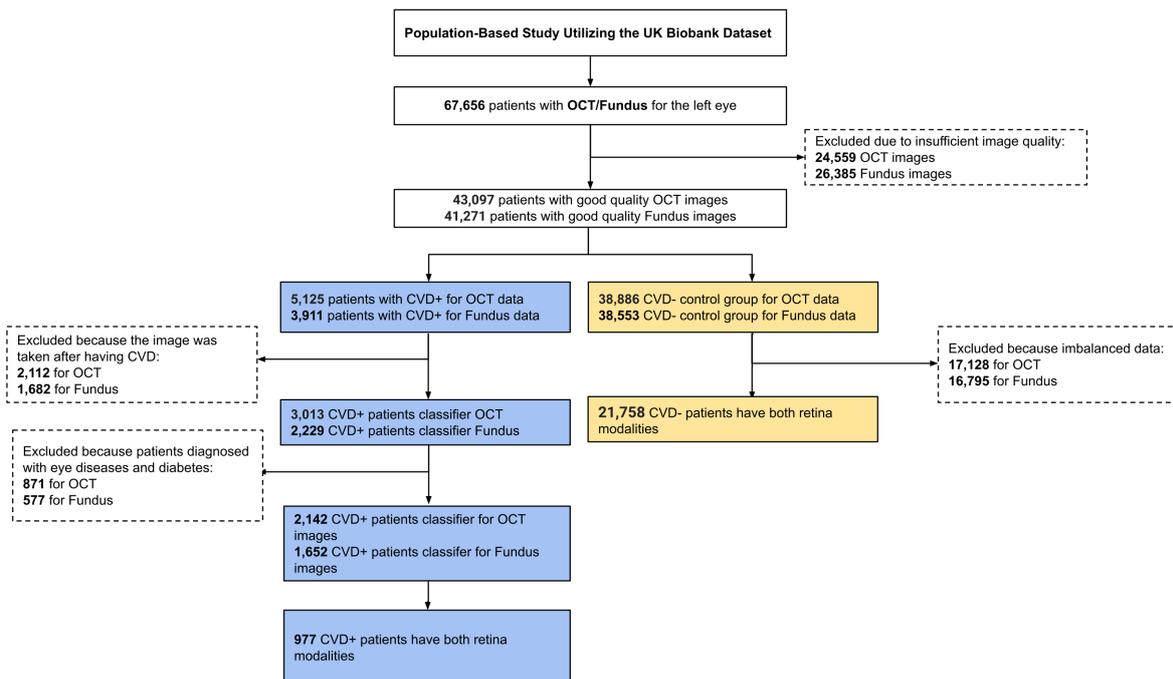


Figure 5.1: STROBE flow chart outlines the patient selection process for a study using the UK Biobank dataset with OCT and Fundus images of the left eye.

Figure 5.1 illustrates the inclusion and exclusion criteria for our study. The focus of the study is on the left eye’s OCT and Fundus images. The study balances the cohorts by excluding cases with image quality issues, post-diagnosis images, other eye diseases, and imbalanced data, resulting in a robust dataset for assessing cardiovascular disease using retinal imaging.

## 5.3 Methodology

### 5.3.1 Framework of Task-Aware Multi-Channel Variational Autoencoder (task-aware MCVAE)

This study proposes a predictive model for classifying patients into  $CVD+$  and  $CVD-$ . The methodology comprises two stages: (1) training a Multi-Channel Variational Autoencoder (MCVAE) (Antelmi et al. 2019) using both retinal modalities as inputs from approximately 18,000  $CVD-$  patients, with the objective of extracting features from retinal OCT and fundus images, and (2) implementing a transformer-based binary classifier. The inputs consist of volumetric OCT B-scans and fundus photographs, which are utilized to learn a compressed latent representation of the high-dimensional image data. This latent representation is subsequently employed by the transformer classifier to differentiate between  $CVD-$

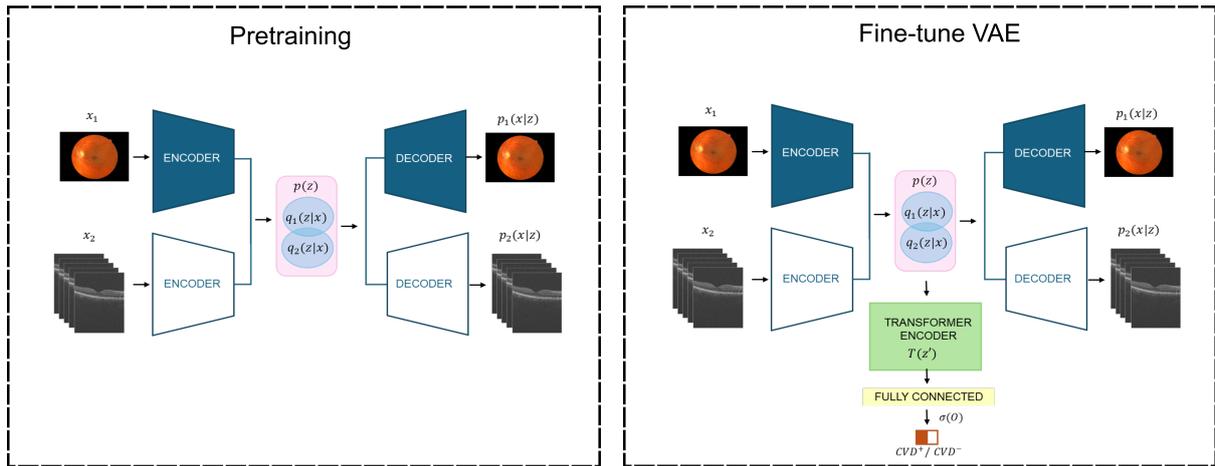


Figure 5.2: The illustration shows the architecture of our model. During the pre-training phase (left), fundus photographs and OCT images are processed through a 2D CNN and a 3D CNN encoder-decoder network (in a multi-channel VAE configuration), respectively. In the fine-tuning phase (right), the images are processed through the pre-trained encoder-decoder networks. The resulting latent vectors are aggregated and input into a transformer architecture, followed by a fully connected layer. The model undergoes end-to-end training in both phases.

subjects and those at risk of future CVD (CVD+) (see Fig. 5.2).

The detail of Multi-channel Variational Autoencoder (MCVAE) is given in the Section 2.6.2.

### Transformer encoder classifier

To further utilize the latent representation  $z$  obtained from the MCVAE, we introduce a classifier built using a transformer architecture (Vaswani et al. 2017). The input to the classifier is the latent space representation  $z$ , which is fed into a series of transformer encoder layers. We propose using a transformer as the Multi-Head Self-Attention mechanism has demonstrated the ability to focus on different parts of the input representation, capturing dependencies and interactions between different latent dimensions (Benarabet et al. 2022). The final output of the transformer encoder layers is passed through a fully connected layer followed by a softmax activation function to produce classification probabilities. The transformer encoder classifier is trained to minimize a binary cross-entropy loss function, enabling it to learn from the latent representations and make accurate predictions.

### Losses

The loss function used to train the proposed model is presented in Equation (5.1), and is composed of: (1) the mean square error (MSE)  $\mathcal{L}_{MSE}$  loss (see Equation (5.2)) for the reconstruction error between the original data ( $\mathbf{x}_i$ ) and reconstructed data ( $\hat{\mathbf{x}}_i$ ); (2) the binary cross-entropy loss  $\mathcal{L}_{BCE}$  (see Equation

(5.3)) for the classification task, where,  $y_i^g$  and  $y_i^p$  denote the ground truth class label and predicted class probability for the  $i^{\text{th}}$  sample; and (3) the derivation of KL divergence loss  $\mathcal{L}_{KL}$  (see Equation (5.2; details in Kingma et al. 2014)).

$$\mathcal{L}_{task-aware-MCVAE} = \mathcal{L}_{recon} + \mathcal{L}_{class} \quad (5.1)$$

$$\mathcal{L}_{recon} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \hat{\mathbf{x}}_i)^2 + \frac{1}{2} \sum_{i=1}^N [1 + \log(\sigma_i^2) - \sigma_i^2 - \mu_i^2] \quad (5.2)$$

$$\mathcal{L}_{class} = \mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^N [y_i^g \log(y_i^p) + (1 - y_i^g) \log(1 - y_i^p)] \quad (5.3)$$

### Model interpretation

To elucidate the model’s behavior and provide local explanations for its classification decisions, particularly in relation to the CVD+ group, we employed a two-step approach leveraging both feature importance analysis and visualization techniques. Initially, we utilized the SHAP (SHapley Additive exPlanations) library to compute SHAP values (Lundberg et al. 2017), enabling the identification of latent variables derived from both retinal imaging modalities that contribute most significantly to the CVD classifications. This approach allows for a quantitative assessment of feature importance within the learned latent space. Subsequently, we applied the methodology of vector field maps, as developed in Chapter 4 Section 4.3.2. This process focused on the latent vectors exhibiting the highest contribution to the classification outcomes. We systematically altered these critical latent vectors, visualized the resulting reconstructions of the modified latent representations, and subsequently calculated the vector field between the original and perturbed reconstructions.

## 5.4 Experiments

The experimental procedures were conducted utilizing an NVIDIA Tesla M60 GPU. The model was implemented using PyTorch (v1.10.2), and optimal hyperparameters were determined through a grid search strategy employing five-fold cross-validation (see Table B.1). The dataset was partitioned into training, validation, and test sets in a 5:2:3 ratio. For the fundus network, both encoder and decoder architectures comprised six 2D convolutional layers. The OCT image network utilized four 3D con-

volutional layers for both encoder and decoder components. In both instances, the encoder employed Rectified Linear Unit (ReLU) activations, while the decoder implemented Leaky Rectified Linear Unit (LeakyReLU) activations (refer to Tables B.2 and B.3).

Our end-to-end model architecture consists of two primary stages: the Multi-Channel Variational Autoencoder (MCVAE) training stage and the task-aware MCVAE transformer classifier training stage. The model accepts two types of retinal images as input: fundus photographs and B-scan OCT images. Each modality is processed by its respective encoder. The MCVAE is optimized to minimize the reconstruction loss (the discrepancy between original and reconstructed images) and the KL divergence between the latent variable distributions and the prior  $p(z)$ . In the fine-tuning phase, the latent representations  $z$  serve as input to the multimodal transformer classifier. The transformer encoder processes these latent representations, utilizing its capacity to capture complex dependencies and interactions within the latent space. It enhances the latent features through multi-head self-attention mechanisms and feed-forward neural networks, thereby learning richer representations. The transformer encoder’s output is subsequently processed through a fully connected layer. During this stage, the model is trained to minimize the classification loss, employing a binary cross-entropy loss function that quantifies the difference between predicted and actual CVD classifications. All subjects’ images used for pre-training the MCVAE were excluded from subsequent experiments, where the task-aware MCVAE was trained and evaluated to identify patients at risk of future CVD based on their retinal images. We employed an age-sex-matched cohort for the  $CVD+$  and  $CVD-$  groups. This approach has been demonstrated to offer a significant advantage in mitigating the influence of confounding variables on the predictive model (Zhou et al. 2023; Brown et al. 2022). During the classification phase, we conducted a comprehensive investigation into the impact of latent representations derived from OCT and fundus imaging on the predictive task. This was accomplished by generating three datasets from the same cohort of around 3,000 patients, each comprising different combinations of data sources: (i) latent representations from fundus photographs; (ii) latent representations from OCT imaging; and (iii) latent representations from both fundus and OCT modalities.

## 5.5 Results

### 5.5.1 Classification performance

In this study, we aim to evaluate the efficacy of two retinal imaging modalities, fundus photography and OCT, for predicting patients at risk of having a CVD event. To accomplish this objective, we have

constructed datasets comprising three distinct cohorts: one utilizing solely fundus images, another employing only OCT images, and a third integrating both retinal modalities (Fundus-OCT). To evaluate the model's efficacy, we have employed the three classifiers on test set not used for learning, comprising 1,000 CVD- subjects and 100 CVD+ subjects. Our primary aim is to investigate whether the amalgamation of latent features derived from OCT images and fundus photographs provides superior discriminative power compared to the utilization of either modality in isolation.

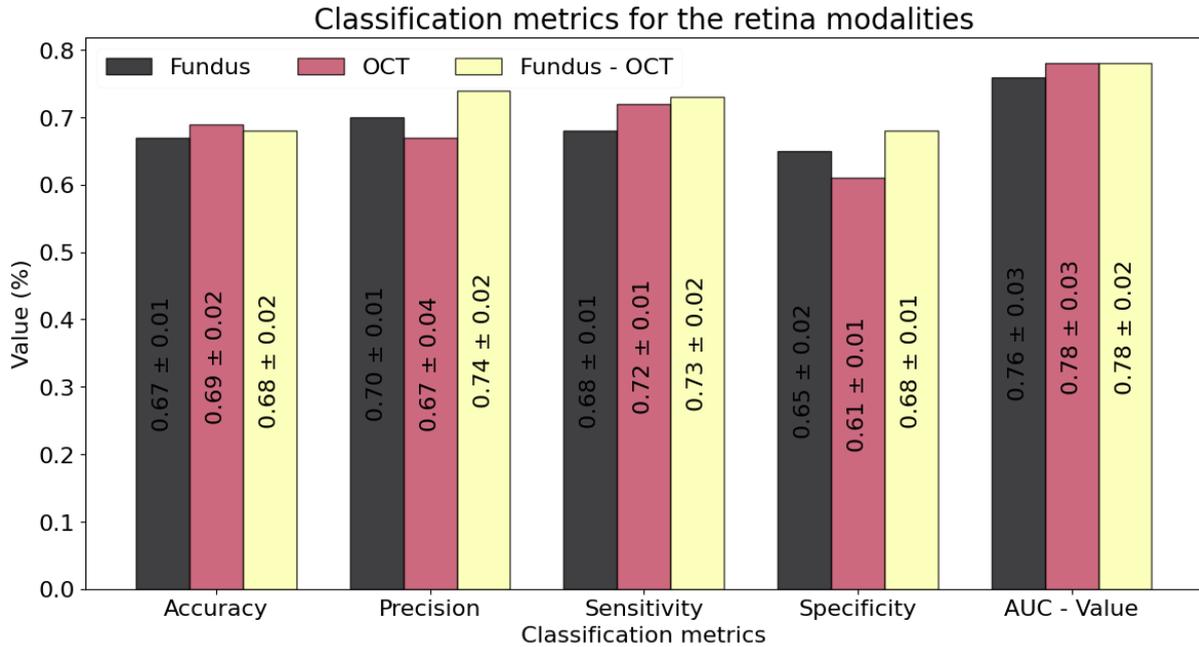


Figure 5.3: The bar chart displays the classification metrics for Fundus, OCT, and combined Fundus-OCT modalities. The metrics shown include Accuracy, Precision, Sensitivity, Specificity, and AUC (Area Under the Curve) values, with their respective standard deviations. The colors represent different modalities: Fundus (black), OCT (red), and Fundus-OCT (yellow). Each bar is labeled with its corresponding value.

The results indicate that the combined Fundus-OCT classifier generally demonstrates superior or equivalent performance to the individual Fundus and OCT classifiers across the majority of metrics, with particular emphasis on precision and sensitivity. The observed minor variations in accuracy and AUC among the modalities fall within the margins of error, suggesting that while the integration of modalities tends to enhance performance, these improvements may not always achieve statistical significance. Nevertheless, the combined Fundus-OCT modality consistently exhibits higher or equal performance across various metrics, lending support to the hypothesis that the integration of multiple retinal imaging modalities can provide more comprehensive and discriminative features for CVD prediction (see Figure 5.3). The overlapping confidence intervals (CI) suggest that the differences in performance metrics among the modalities are not highly significant (refer to Table 5.1). However, the consistent perfor-

mance improvement observed with the combined modality (Fundus-OCT) indicates practical benefits in the simultaneous utilization of both imaging modalities.

<b>Metric</b>	<b>Fundus CI</b>	<b>OCT CI</b>	<b>Fundus-OCT CI</b>
Accuracy	(0.662, 0.677)	(0.6827, 0.697)	(0.675, 0.687)
Precision	(0.693, 0.707)	(0.662, 0.678)	(0.733, 0.746)
Sensitivity	(0.674, 0.688)	(0.713, 0.727)	(0.723, 0.738)
Specificity	(0.641, 0.658)	(0.606, 0.619)	(0.672, 0.687)
AUC	(0.753, 0.767)	(0.772, 0.787)	(0.773, 0.788)

Table 5.1: 95% CI for classification metrics of the Fundus, OCT, and Fundus-OCT classifiers. Each interval represents the range within which the true metric value is expected to lie with 95% confidence, providing insight into the precision and reliability of the classifiers' performance.

The superior performance of combined fundus photography and OCT in cardiovascular risk stratification stems from their complementary capabilities in assessing different aspects of microvascular pathology. Fundus photography provides comprehensive two-dimensional visualization of the retinal vasculature, enabling evaluation of structural changes such as arteriolar narrowing, venular widening, and altered vascular branching patterns, all established markers of systemic microvascular dysfunction. Meanwhile, OCT offers three-dimensional, depth-resolved quantification of both neural retina and choroidal layers, detecting early subclinical alterations like choroidal thinning (indicative of impaired ocular blood flow) and retinal nerve fiber layer loss (reflecting end-organ damage from hypertension). The synergistic value arises because fundus images primarily assess the superficial vascular plexus, while OCT specifically evaluates the deeper choroidal circulation and neural tissue integrity - together providing a more complete assessment of ocular microvascular health. This combined approach captures both the structural vascular changes (via fundus) and the resulting tissue-level consequences (via OCT), mirroring the panvascular nature of cardiovascular disease pathophysiology. Additionally, while fundus photography excels at detecting established microvascular damage through visible vascular alterations, OCT's ability to identify preclinical changes (e.g., choroidal thinning before visible retinopathy develops) provides earlier detection capability. The integration of these modalities thus offers a more comprehensive evaluation of microvascular health across different disease stages, explaining their enhanced predictive performance compared to either modality alone in cardiovascular risk assessment.

Furthermore, we conducted a comparative analysis between the Fundus-OCT classifier, which demonstrated the highest performance across classification metrics, and the QRISK3 algorithm, the current clinical standard for assessing patients at risk of stroke or myocardial infarction. The results of this comparison are presented in Table 5.2, where our model outperformed the QRISK3 algorithm in terms of

accuracy, sensitivity, specificity, and AUC. The observed superior performance of our model compared to QRISK3 in age- and sex-matched cohorts may be partially attributed to the diminished contribution of QRISK3’s strongest predictors (age and sex) (Ghosa et al. 2022) within this study design. As QRISK3 was developed and validated in general population cohorts where age and sex contribute substantially to risk prediction, its performance may be attenuated in matched cohorts. This methodological approach, while controlling for demographic confounders, may introduce selection biases that artificially enhance our model’s apparent performance, as it relies more heavily on retinal biomarkers that are less dependent on these demographic factors.

Classification metrics				
Modality	Accuracy	Sensitivity	Specificity	AUC-Value
<b>Fundus-OCT</b>	<b>0.68 ± 0.02</b>	<b>0.73 ± 0.02</b>	<b>0.68 ± 0.01</b>	<b>0.78 ± 0.02</b>
QRISK	0.55	0.615	0.545	0.59

Table 5.2: Comparative analysis of our model (Fundus-OCT) and QRISK algorithm in predictive classification metric.

### 5.5.2 Ablation study

Confusion matrices for the classification of retinal images using Fundus, OCT, and Fundus-OCT modalities (see Figure 5.4). Each matrix displays the true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) for each modality, along with the totals for prediction positives, prediction negatives, and overall totals. To evaluate the efficacy of our end-to-end training approach, we conducted an ablation study in which only the classifier component was trained, omitting the training of the MC-VAE. In this experimental scenario, our loss function was restricted to the classification loss exclusively (refer to Equation (5.3)).

Consistent with our previous findings, the utilization of both retinal modalities yielded superior outcomes across nearly all classification metrics. The results demonstrate that, when compared to our comprehensive end-to-end model, the classification metrics for all three classifiers (Fundus, OCT, and Fundus-OCT) were significantly diminished in the ablation study. This observed discrepancy suggests that the end-to-end training approach enhances the model’s capacity to learn and focus on the salient features that distinguish  $CVD+$  and  $CVD-$  patients, thereby improving overall classification performance. The classification metrics for the retinal modalities in the ablation study are presented in Table 5.3. It is evident that while the Fundus-OCT classifier maintains its superior performance relative to the individual Fundus and OCT classifiers, the overall performance is attenuated compared to the end-to-end trained

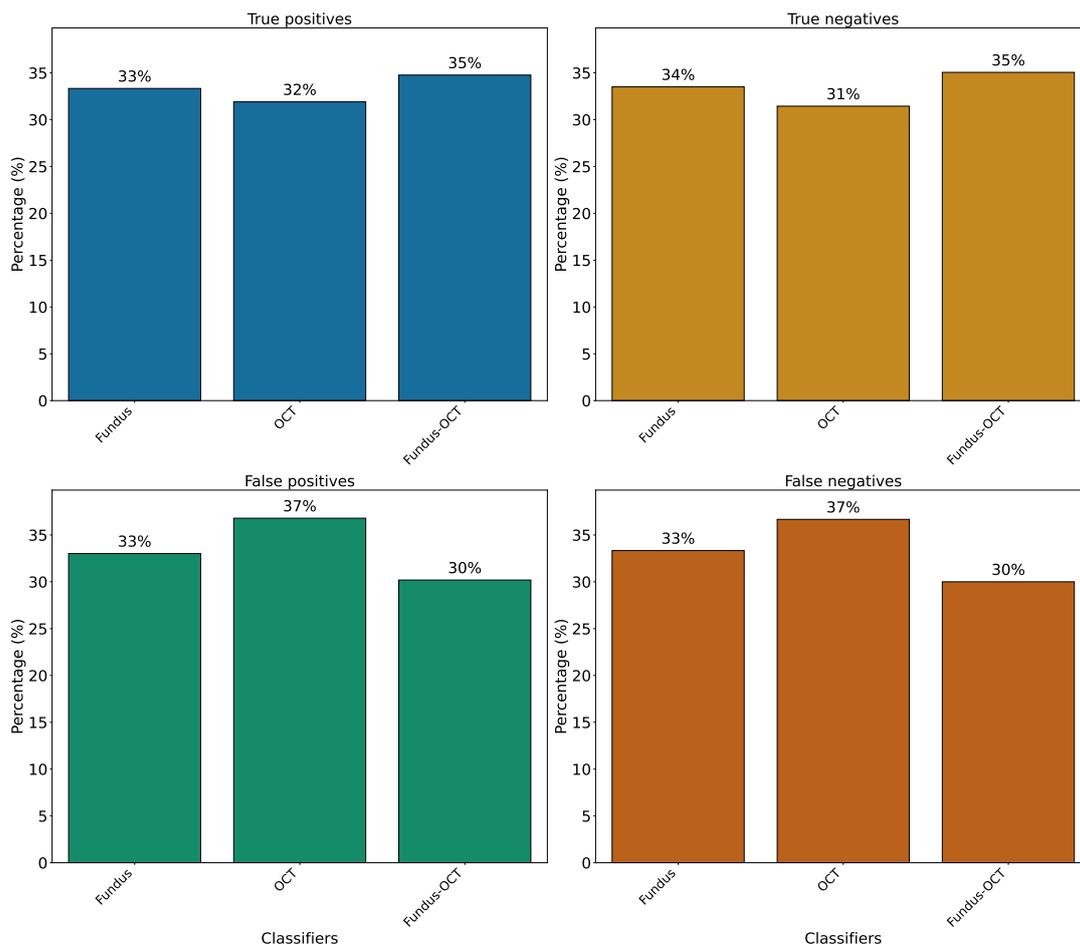


Figure 5.4: Comparison of classifier performance in terms of TP, TN, FP, and FN. Values above the bars represent the corresponding percentages for each classifier.

model. Specifically, we observed reductions in all classification metrics. These findings underscore the critical importance of comprehensive end-to-end training for optimizing classifier performance.

Metrics	Fundus	Fundus CI	OCT	OCT CI	Fundus - OCT	Fundus - OCT CI
Accuracy	0.58	0.56 - 0.60	0.62	0.60 - 0.64	0.60	0.58 - 0.62
Precision	0.60	0.58 - 0.62	0.57	0.54 - 0.60	0.62	0.59 - 0.65
Sensitivity	0.58	0.56 - 0.60	0.56	0.52 - 0.60	0.62	0.59 - 0.65
Specificity	0.56	0.53 - 0.59	0.58	0.55 - 0.61	0.55	0.52 - 0.58
AUC - Value	0.60	0.57 - 0.63	0.63	0.60 - 0.66	0.64	0.63 - 0.65

Table 5.3: Classification metrics for the retina modalities (Ablation Study). This table compares the performance of classifiers trained without the end-to-end approach, highlighting the reduced efficacy across all metrics when the MCVAE is not trained alongside the classifier.

### 5.5.3 Interpretation.

Following the identification of latent vectors exhibiting the highest contribution to the prediction of patients at elevated risk of future cardiac events using the SHAP values (see Figure 5.5), we systematically

altered the  $z$  values of both OCT and fundus images that demonstrated the highest values. For each case, we reconstructed the retinal modality utilizing the modified  $z$  values and subsequently estimated the vector field maps between the original reconstructed image and the perturbed reconstructed image. In Figure 5.6, we present examples from five randomly selected patients purely for visualization purposes, all identified as being at high risk of developing CVD. For simplicity, in the case of OCT, we only show the 64th B-scan example, although the entire volume was processed. Figure 5.6(a) displays the qualitative OCT results, while Figure 5.6(b) shows the fundus results, where the yellow points highlight the most pronounced differences.

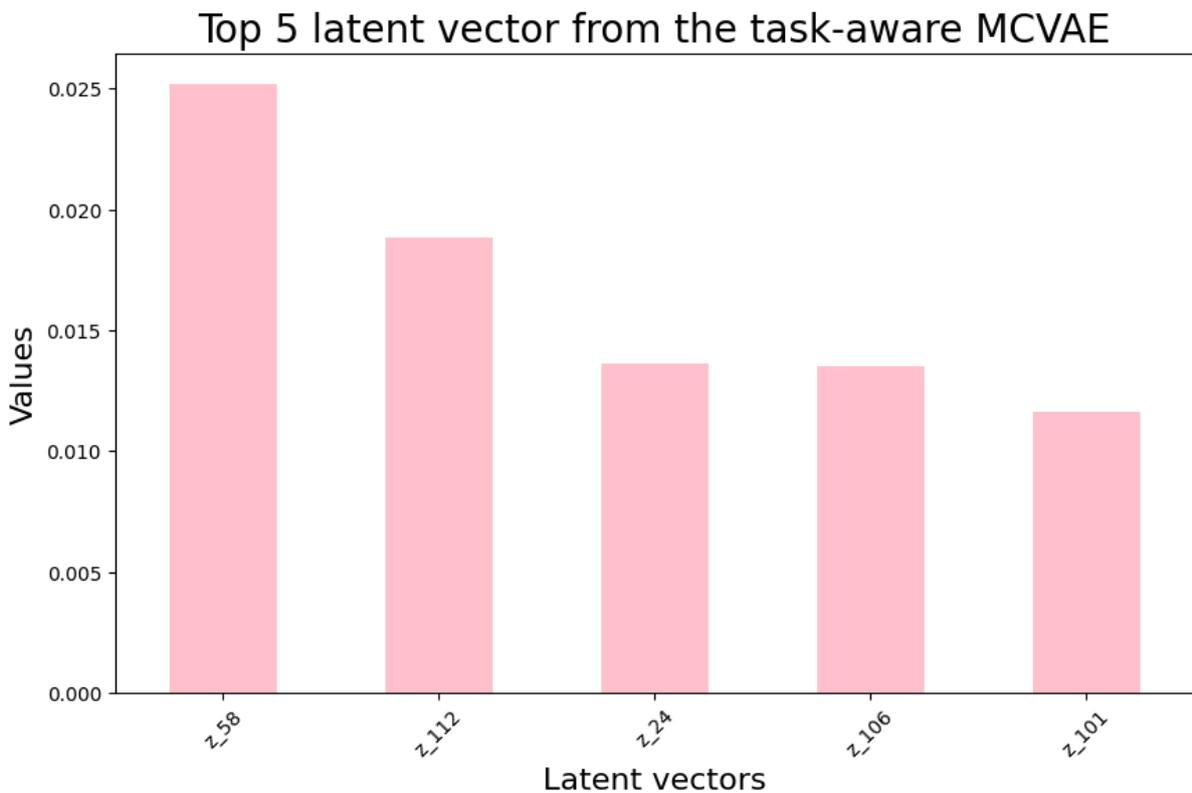


Figure 5.5: The top 5 latent vectors  $z_{58}$ ,  $z_{112}$ ,  $z_{24}$ ,  $z_{106}$ , and  $z_{101}$  are identified based on their SHAP values, which measure the contribution of each feature to the model’s prediction of cardiovascular disease risk.

Figure 5.5 illustrates the top 5 latent vectors that have the greatest influence on the prediction of CVD risk, as identified using the SHAP methodology. These vectors represent the features extracted by the task-aware MCVAE model that contribute most significantly to the predictive performance. Specifically, the vectors  $z_{58}$ ,  $z_{112}$ ,  $z_{24}$ ,  $z_{106}$ , and  $z_{101}$  are shown to have the highest SHAP values, indicating their importance in the model’s decision-making process.

The vector field maps generated by our model predominantly emphasized the choroidal layer in the

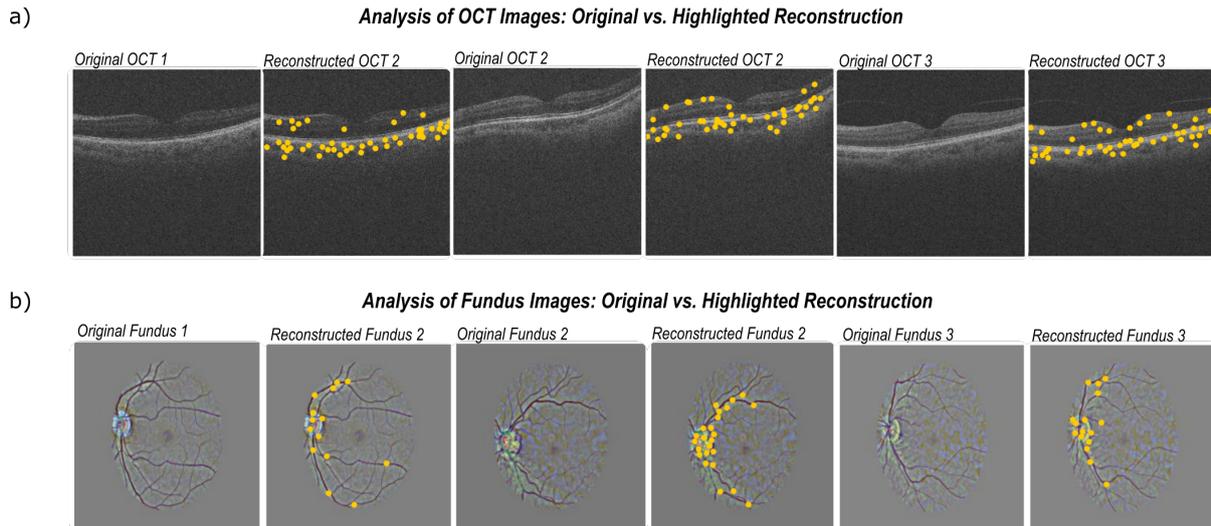


Figure 5.6: a) The top row displays a sequence of Optical Coherence Tomography (OCT) images, highlighting the different layers of the retina with yellow markers indicating regions of interest. b) The bottom row shows corresponding fundus images with vascular structures and regions of interest also marked in yellow. These images are utilized for analyzing and predicting cardiovascular disease risks based on retinal biomarkers.

OCT volume, as well as the retinal pigment epithelium (RPE). These results suggest that the choroidal layer may play a pivotal role in distinguishing between CVD+ and CVD- patients. In the case of fundus images, blood vessels were prominently highlighted in the majority of patients in the unseen test set, along with the optic disc. This observation aligns with established clinical indicators of retinal vascular health and its association with cardiovascular risk. These findings provide quantitative evidence of the specific retinal features that our model prioritizes in its classification process, offering insights into the potential biomarkers for cardiovascular risk assessment using multi-modal retinal imaging.

## 5.6 Discussion

Our findings demonstrate the efficacy of latent representations derived from retinal OCT images and fundus photographs, learned through a MCVAE, in predicting CVD when employed within a multimodal transformer classifier. This outcome corroborates previous studies that have identified relevant vascular biomarkers and cardiac health indicators within retinal imaging (Maldonado-Garcia et al. 2024; Zhou et al. 2023). Our study utilized a pre-trained MCVAE that integrates B-scan OCT images and fundus photographs as distinct modalities, enabling our model to identify specific attributes within both types of retinal images that significantly contribute to CVD prediction. To our knowledge, only one other study has employed both fundus and OCT modalities for predicting certain CVDs; however, that study was limited to a single B-scan of the OCT, did not train the modalities jointly, and provided limited

detail on the advantages and contributions of each modality (Zhou et al. 2023). A key contribution of our study is the interpretation of specific features in OCT images and fundus photographs that are relevant to the classification task, providing insights into the most discriminative regions of the retinal image. Our results suggest that choroidal morphology is a significant predictor of CVD risk, aligning with previous studies reporting associations between choroidal characteristics and the risk of stroke and acute myocardial infarction (Yeung et al. 2020). Similarly, our model highlighted the optic disc and prominent veins in fundus images, which have been extensively studied and recognized as indicators of cardiovascular health. Abnormalities in retinal vessel caliber, branching patterns, and overall vascular geometry have been consistently correlated with CVD risk in previous studies. Furthermore, alterations in the morphology of the optic disc, including variations in its size, shape, and coloration, have been associated with various cardiovascular risk factors (Guo et al. 2020). These features are instrumental in elucidating the systemic impacts of cardiovascular health on the retinal microvasculature.

Several important limitations of this study warrant consideration when interpreting our results. First, our exclusive reliance on UK Biobank data means our findings inherit the known limitations of this resource, particularly its reduced representativeness of the general UK population. The cohort's well-documented "healthy volunteer" bias, with participants being predominantly white, more likely female, and exhibiting healthier behaviors (lower smoking and alcohol consumption rates) than population averages, may affect the generalizability of our model to more diverse clinical settings (Fry et al. 2017).

The comparison between our retinal imaging-based model and QRISK3 requires nuanced interpretation. While we observed superior discriminative performance in our age-sex matched cohorts, this advantage must be contextualized. QRISK3's predictive power derives substantially from age and sex factors (accounting for approximately half of its risk calculation (Ghosa et al. 2022)), which were deliberately controlled in our study design. In clinical practice, QRISK3 benefits from extensive validation across diverse populations through the QResearch and CPRD databases (Hippisley-Cox et al. 2017; Livingstone et al. 2021), and incorporates well-established risk factors including cholesterol levels, diabetes status, and smoking history. In unmatched general populations, QRISK3's AUC typically reaches 0.75-0.80, suggesting our results may overstate the relative advantage of OCT in clinical practice (Hippisley-Cox et al. 2017; Li et al. 2019). Our retinal imaging approach, while demonstrating promise as a potential complementary tool, requires further validation to establish its clinical utility alongside these established risk assessment methods.

Methodologically, our use of varying class balance ratios across study phases (from the initial 1:20 CVD+

to CVD– ratio in the full dataset to 1:20 in model training and 1:10 in testing) introduced important considerations. While this progressive balancing strategy improved model stability during development, it may have created optimistic performance estimates compared to real-world prevalence rates. Future validation efforts should prioritize evaluation in unmodified cohorts reflecting natural disease prevalence and explore alternative approaches like cost-sensitive learning to better handle inherent class imbalances (Fernández et al. 2018).

Looking ahead, two key research directions emerge to advance this work. First, the integration of multi-organ imaging data could potentially enhance the model’s predictive accuracy and provide a more holistic view of cardiovascular health. Additionally, the inclusion of an external validation dataset will be crucial for assessing the generalizability of our model across diverse populations and clinical settings. This step is essential for establishing the robustness and clinical applicability of our approach in real-world scenarios.

The concordance between our model’s emphasis on these specific retinal features and established clinical indicators of cardiovascular risk provides validation for our approach. It suggests that our deep learning model has learned to identify and prioritize clinically relevant features in both OCT and fundus images for cardiovascular risk prediction. This alignment between machine learning-derived features and known clinical markers not only enhances the interpretability of our model but also reinforces its potential clinical utility. However, substantial work remains to establish whether and how retinal imaging can meaningfully contribute to clinical decision-making pathways for cardiovascular disease prevention and management.

## Chapter 6

# Mendelian randomisation studies of OCT-derived retinal phenotypes and cardiovascular disease

### 6.1 Introduction

In this chapter, we explore an alternative approach to understanding the use of retinal imaging in relation to CVDs. Specifically, we introduce a framework that employ Mendelian randomisation to explore the causal relationships between CVDs and OCT phenotypes, utilising the large-scale genetic data from the UK Biobank alongside retinal images to investigate genetic variants that influence the thickness of retinal layers. The aim is to investigate the biological mechanisms underlying the association between identified retinal biomarkers and cardiovascular diseases, thereby establishing potential causal links. This study makes three key contributions: i) the application of the state-of-the-art nnU-Net segmentation model for 10 OCT layers using 50 ground truth images, ii) the identification of novel genetic loci associated with OCT-derived phenotypes, and iii) the establishment of a causal relationship between OCT features and cardiovascular events through Mendelian randomisation.

Recent genetic investigations into retinal microvasculature and vasculature have uncovered numerous previously unidentified loci, as well as genes and pathways implicated in CVD (Tomasoni et al. 2023; Jiang et al. 2023; Zekavat et al. 2021; Vela et al. 2023; Currant et al. 2020), a review of the literature on retinal imaging and genetic data is discussed in Chapter 3, Section 3.3. In this study, we propose

integrating OCT imaging with genetic data to gain a more comprehensive understanding of the complex biological mechanisms underlying systemic diseases such as CVD. OCT was specifically chosen as it has been less extensively studied compared to fundus imaging, and offers valuable insights into the thickness of retinal layers, which these images provide. Furthermore, we demonstrate through Mendelian randomisation that CVD, along with some well-known risk factor for the disease, is strongly implicated in the causal pathway affecting OCT layer thickness. This approach has the potential to uncover new insights that could contribute to the development of enhanced diagnostic tools and therapeutic strategies.

## **6.2 Dataset**

### **6.2.1 OCT imaging data from the UK Biobank**

For this study, we utilized retinal OCT imaging data from the UK Biobank (see Chapter 2 section 2.2.1 for more references). We focused on retinal imaging of the left eye, which we found that is generally of higher quality than the right eye (see Figure 4.1). During the baseline visit (2006-2010), retinal imaging was performed on 67,339 participants for both modalities. We applied quality control exclusion criteria for OCT data according to established protocols (Zekavat et al. 2023). First, we excluded participants with an image quality score (a measure of scan signal strength) below 45, as defined by UK Biobank (Data-Field-28552), retaining 61,708 subjects. Next, we excluded the bottom 20% of scans based on the ILM indicator quality, which quantifies boundary sharpness and detects artifacts (e.g., blinks, signal fading, or segmentation errors) (Data-Field-28542; Ko et al. 2017). This step removed an additional 18,611 participants, ensuring robust data for downstream analyses. Finally, we focused on individuals within a genetically homogenous white European cohort (Data-Field-22006). The final dataset for the primary analysis comprised 42,350 subjects (see Figure 6.1).

### **6.2.2 SNP microarrays used in the UK Biobank.**

The genotypic data available from the UK Biobank, was employed two closely related SNP arrays to generate its extensive genotypic dataset (Bycroft et al. 2017). This ambitious genotyping effort encompassed over 480,000 participants, utilizing technology from Affymetrix to capture a wide spectrum of genetic markers.

The initial phase of genotyping involved 49,550 participants, who were assessed using the UK BiLEVE Axiom array. This array provided coverage of 807,411 distinct genetic markers. Following this, the majority of participants, 438,427 individuals, were genotyped using the UK Biobank Axiom array, which

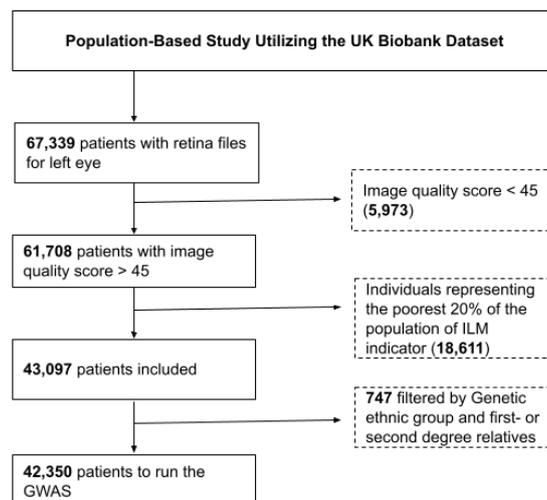


Figure 6.1: STROBE flow chart describing participant inclusion and exclusion criteria applied to define the study cohort.

expanded the coverage to 825,927 markers. Despite being separate arrays, these tools were designed with significant overlap, sharing 95% of their marker content to ensure consistency across the study population.

The arrays were custom-designed for the UK Biobank project, with marker selection guided by several strategic considerations:

- **Phenotype-specific markers:** Approximately 45,000 markers were chosen based on their relevance to particular traits or conditions of interest.
- **Genomic hotspots:** Around 47,000 markers were selected from regions of the genome known to be of particular scientific or clinical importance.
- **Coding variants:** A substantial portion of the arrays, about 125,000 markers, focused on variations within protein-coding regions of genes.
- **Population genetics:** The arrays were designed to provide comprehensive genome-wide coverage, facilitating accurate imputation in European populations. This coverage spanned both common variants (those with a minor allele frequency greater than 5%) and low-frequency variants (those with a minor allele frequency between 1% and 5%).

This carefully curated approach to marker selection ensured that the UK Biobank's genetic data would be a valuable resource for a wide range of genetic studies, from common disease risk to population structure analysis.

### 6.2.3 Genotype imputation in the UK Biobank.

The UK Biobank's genetic dataset was significantly enhanced through the application of genotype imputation, a sophisticated statistical technique that expands the scope of genetic information beyond directly assayed markers. This process leverages densely genotyped reference panels to infer genotypes at loci not directly measured in the study sample, thereby increasing the resolution and power of genetic analyses.

Imputation serves multiple crucial functions in genetic research:

- It amplifies the statistical power of association studies.
- It enables more precise identification of causal variants through fine-mapping.
- It facilitates the integration of findings across studies using different genotyping platforms.

The UK Biobank imputation effort utilized two distinct reference panels: i) A composite panel combining data from The UK10K project (Ciampi et al. 2015) and The 1000 Genomes Phase 3 (Oleksyk et al. 2015) (This panel encompassed 12,570 haplotypes and over 87 million bi-allelic markers); ii) The Haplotype Reference Consortium (HRC) panel (McCarthy et al. 2016) (Included nearly 65,000 haplotypes and covered approximately 40 million markers).

The imputation process employed a modified version of the IMPUTE2 algorithm, which uses a hidden Markov model to estimate missing genotypes. When a SNP was present in both panels, the HRC imputation was preferentially selected. This comprehensive approach resulted in a vastly expanded dataset comprising: 92,693,895 autosomal genetic variants, including SNPs, short insertions and deletions (indels), and large structural variants across 487,442 individuals.

To ensure standardization and facilitate integration with other genetic databases, dbSNP Reference SNP (rs) IDs were assigned to markers wherever possible. These identifiers were sourced from rs ID lists in the UCSC genome annotation database, based on the GRCh37 assembly of the human genome. This imputation strategy significantly broadened the scope of genetic variation captured in the UK Biobank, creating a rich resource for diverse genetic investigations, from common disease risk to complex trait analysis.

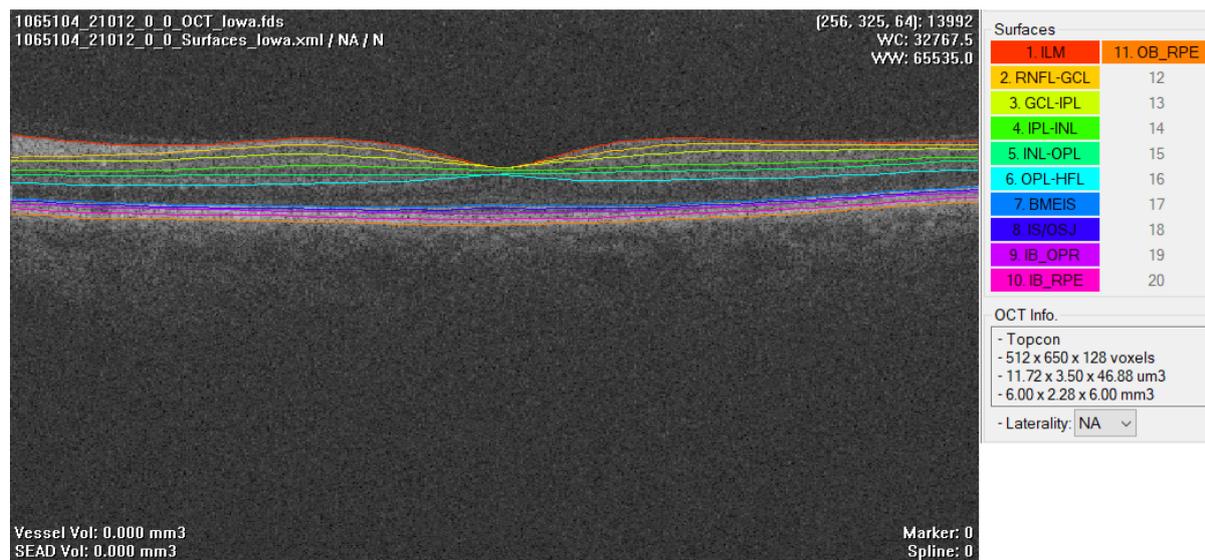


Figure 6.2: Example of a 64th B-scan OCT, displaying the boundaries of 10 layers. The corresponding labels for these layers are provided in the figure on the right.

## 6.3 Methodology

### 6.3.1 Optical coherence tomography imaging segmentation

The delineation of retinal layers was established through The Iowa Reference Algorithms (Retinal Image Analysis Lab, Iowa Institute for Biomedical Imaging, Iowa City, IA) (Abràmoff et al. 2010), identifying boundaries for 10 distinct layers across 50 OCT volumetric scans which will be the ground truth masks for the segmentation model. The boundaries of the OCT volumes are: **ILM** (Internal limiting membrane), **RNFL** Retinal nerve fiber layer, **GCL** Ganglion cell layer, **IPL** Inner plexiform layer, **INL** Inner nuclear layer, **OPL** Outer plexiform layer, **BMEIS** Boundary of myoid and ellipsoid of inner segments, **IS/OSJ** Inner/outer segment junction, **RPE** Retinal pigment epithelium, **IB\_OPR** Inner boundary of Outer segment PR/RPE complex, **IB\_RPE** Inner boundary of RPE, **OB\_RPE** Outer boundary of RPE (see Figure 6.2).

Subsequently, layer boundaries are translated into a format suitable for semantic segmentation where we assign labels to individual pixels in the image to mark the specific boundaries of different regions in the image. The 50 original OCT volume images and their corresponding 50 ground truth segmentation masks were used to train the state-of-the-art nn-UNet model (Isensee et al. 2020), while the remaining OCT volumes were reserved for testing. This demonstrates the advantage of using deep learning algorithms, where a small amount of ground truth data is sufficient to train the model efficiently, requiring minimal computational resources and completing the training process in a short time (less than 2 hours for training and under 10 hours overall). The segmented output obtained by the nn-UNet were then used

to calculate the thickness maps of each layer (totaling 10 layers).

### 6.3.2 Phenotypes of thickness maps of variational autoencoder

To calculate the thickness of each retinal layer, we analysed OCT images where each layer had been delineated using segmentation masks (binary images) highlighting specific retinal structures. For each of the 512 points along the horizontal axis, we measured the vertical distance between the upper and lower boundaries of the segmentation mask corresponding to the layer of interest. The final result is a volume consisting of 10 slices, each representing one of the layers where each slice has dimensions of 128 x 512, where 128 corresponds to the projection of each A-scan of the volume, and 512 represents the width of each scan. We did this process for all the UK Biobank patients that met the inclusion criteria in this study (see Figure 6.1).

### 6.3.3 Mendelian randomization

We performed Mendelian randomization (MR) analyses using published GWAS summary statistics from the EBI GWAS Catalog, CARDIoGRAMplusC4D, and FinnGen databases. To avoid sample overlap bias, we excluded all UK Biobank-derived GWAS from our exposure datasets. While our restriction to European-ancestry cohorts limits the generalizability of findings to other populations, it minimizes confounding due to population stratification and ensures robust instrument strength. Full details of the GWAS sources, including accession links and sample sizes, are provided in Appendix C Section C.2.1.

We used the 64 latent vectors derived from the VAE of retinal layer thicknesses, as described in Section 6.3.2. The MR methodology, including instrument selection, and causality assumptions, is detailed in Chapter 2 Section 2.9.

## 6.4 Experiments

All experiments, both for the segmentation task and the autoencoder, were conducted using an NVIDIA Tesla M60 GPU. For the segmentation of the OCT B-scan volumes, the nnUNet was trained over 100 epochs. The optimal model was determined to be the 3D variant, based on the best hyperparameters identified. For the VAE model, we utilised PyTorch (v1.10.2) and implemented a grid search strategy with five-fold cross-validation to select the best hyperparameters (see Table C.1). The dataset was split into training, validation, and test sets in a ratio of 6:2:2. The final encoder and decoder networks were each constructed using three 2D convolutional layers; the encoder employed Rectified Linear Unit (ReLU)

activations, while Leaky Rectified Linear Unit (LeakyReLU) activations were used in the decoder. The latent dimension was set to 64 (see Table C.2).

The UK Biobank used SNP microarrays to genotype all participants. Quality control of the genetic data involved filtering variants according to several criteria: a minor allele frequency (MAF) threshold of 1%, a Hardy–Weinberg equilibrium test with a significance threshold of  $p < 10^{-5}$ , and an imputation information score greater than 0.3. After applying these filters, 8,000,193 genotyped variants were retained. Retinal phenotypes were adjusted for various covariates, including sex, age, age<sup>2</sup>, height, weight, refractive error (calculated as spherical error + 0.5 × cylindrical error), and 10 genetic principal components.

For the GWAS, we employed an additive linear model implemented in BGENIE (BGENIE 2018). During preprocessing, all latent vectors were inverse rank-normalised before being modelled with BGENIE.

Mendelian randomisation analysis was conducted using the TwoSampleMR package in R (Hemani et al. 2018). In this analysis, summary statistics from genetic association studies related to various CVD and their associated risks were utilised as the exposure variable (see details in C.2.1). These studies were selected based on genome-wide significance ( $p < 5 \times 10^{-8}$ ). The latent vectors obtained from the VAE of the OCT-derived phenotypes served as the outcome variable, with the same set of technical covariates applied as in the aforementioned GWAS. The inverse variance weighting (IVW) method was employed for this analysis.

## 6.5 Results

### 6.5.1 Results of qualitative segmentation

An example of the qualitative results is shown in Figure 6.3. Visually, the segmentation masks appear to be of high quality, with clear separation of the retinal layers. This indicates that the model has successfully learned the corresponding pixels for each layer, demonstrating its effectiveness for the subsequent task of determining the thickness of each layer.

### 6.5.2 Results from qualitative and quantitative analysis of thickness maps

An example of a reconstructed subject is shown in Figure 6.4. The resulting images display promising outcomes, suggesting that the model is effectively learning the most representative features of the thickness maps. This allows the model to capture the characteristics of the variation in layer thickness across

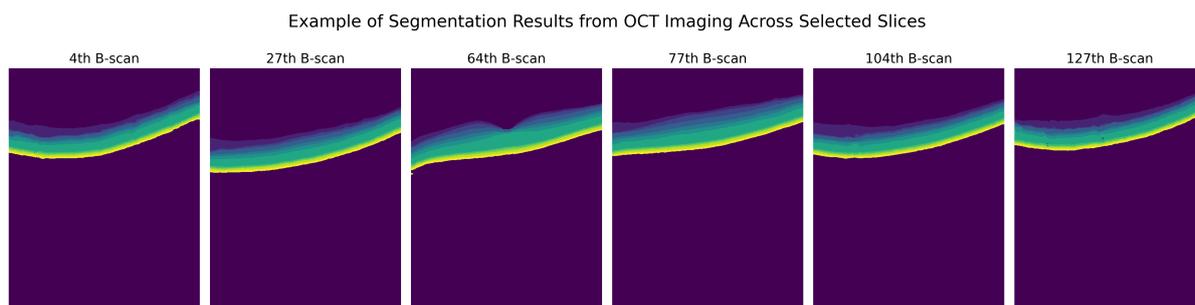


Figure 6.3: An example of the nnUNet segmentation results for the ten layers of OCT.

patients, yielding high-quality results in OCT scans.

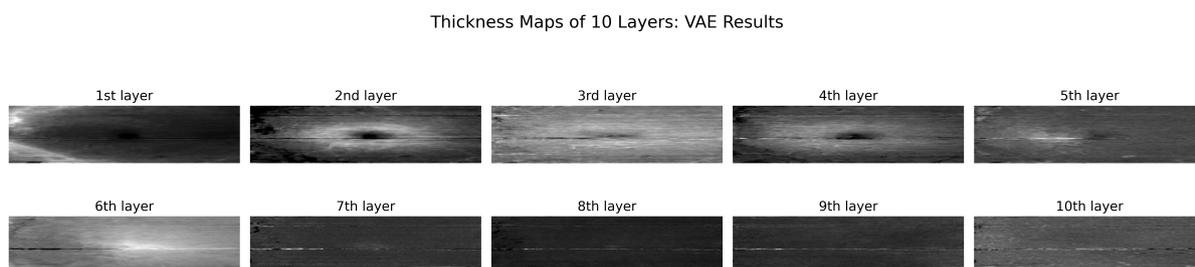


Figure 6.4: Thickness maps generated from Optical Coherence Tomography (OCT) data, illustrating the Variational Autoencoder (VAE) model's performance in analyzing 10 layers of tissue.

The table 6.1 presents the Root Mean Square Error (RMSE) values for the 10 different retinal layers, indicating the accuracy of a predictive model or measurement method in quantifying the thickness or properties of these layers.

Layer	RMSE
Layer 1 (ILM-RNFL)	$0.22 \pm 0.08$
Layer 2 (RNFL-GCL)	$0.27 \pm 0.11$
Layer 3 (GCL-IPL)	$0.14 \pm 0.07$
Layer 4 (IPL-INL)	$0.21 \pm 0.08$
Layer 5 (INL-OPL)	$0.16 \pm 0.07$
Layer 6 (OPL-BMEIS)	$0.10 \pm 0.05$
Layer 7 (BMEIS-IS/OSJ)	$0.53 \pm 0.13$
Layer 8 (BMEIS-IB_OPR)	$0.64 \pm 0.14$
Layer 9 (IB_OPR-IB_RPE)	$0.23 \pm 0.10$
Layer 10 (IB_RPE-OB_RPE)	$0.17 \pm 0.06$

Table 6.1: Root Mean Square Error (RMSE) for each retinal layer.

Layer 6 (OPLBMEIS) has the lowest RMSE of  $0.10 \pm 0.05$ , suggesting that the model or method used to predict or measure the thickness of this layer is the most accurate among all layers. Otherwise, layer 8 (BMEIS-IB\_OPR) has the highest RMSE of  $0.64 \pm 0.14$ , indicating that this layer's thickness is the most challenging to predict or measure accurately, with the largest discrepancies between the predicted

and actual values; also, this layer has a relatively high variability ( $0.64 \pm 0.14$ ), which suggests that the errors are not consistent and can vary widely. Layers with RMSE values between these extremes, such as Layer 7 (BMEIS-IS/OSJ) with an RMSE of  $0.53 \pm 0.13$ , also suggest areas where prediction accuracy could be improved.

### **6.5.3 Results of the GWAS**

We demonstrate that this approach effectively discovers additional biologically relevant genetic associations. It identifies 90 loci with study-wide significance, expanding on previous knowledge. All discovered genes are provided in Table 6.2.

Our GWAS revealed 17 novel retinal associations (20% of significant loci), highlighting previously unrecognized genetic influences on retinal structure. Several of these loci are implicated in ocular pathologies. For example, ERCC6 is linked to age-related macular degeneration (AMD), with studies showing altered mRNA levels in RPE cells of healthy and early AMD donor eyes (Baas et al. 2010). Another gene, NGLY1, is associated with congenital disorder of deglycosylation, leading to reduced tear production and other ocular abnormalities in affected children (Adam et al. 2019). Additionally, HSP90B1 plays a critical role in retinal nervous system function (Nasaré et al. 2015).

We also replicated 54 known retinal loci (64%), many with well-established phenotypic roles. The *Sine oculis homeobox* gene SIX6 regulates glaucoma development (Carnes et al. 2014), while PITX3 contributes to congenital posterior subcapsular cataract (Wu et al. 2019). RDH5 is associated with familial fleck retina and night blindness (Qian et al. 2022), and mutations in LAMC1 disrupt ocular basement membranes, including the inner limiting membrane, as demonstrated in zebrafish models (Richardson et al. 2017). Finally, GUCY2D, essential for photoreceptor function, is critical for normal vision (Neubauer et al. 2022).

Table 6.2: This table presents the results of our GWAS, including: significant SNPs, their closest genes, prior known associations with retinal phenotypes from existing GWAS literature, and classification as either novel or previously reported in retinal GWAS.

SNP	Closest Gene	Prior Retinal GWAS Associations	Retinal Association Status
1:200383761_GTTCCTG	PDE6B	None reported	Novel
AACAAATAAAAAT_G			
rs56029302	CDC42BPA	Cataracts	Established
rs12408663	NBL1	Retinal thickness	Established
rs3737306	ALOX5	None reported	Novel
rs4838417	ERCC6	None reported	Novel
rs56238729	LINC02640	Optic nerve measure- ment	Established
rs10998176	COL13A1	Optic disc size	Established
rs4601657	PITRM1	None reported	Novel
rs2817706	CNNM1	None reported	Novel
rs4919505	PITX3	Lens, Ocular stroma, and its derivatives	Established
rs2071754	PAX6	Optic cup	Established
rs12408663	NBL1	Retinal thickness	Established
rs72930261	TPCN2	Retinal vascular frac- tial dimension	Established

SNP	Closest Gene	Prior Retinal GWAS Associations	Retinal Association Status
rs1126809	TYR	Melanin for pigmentation in the eye, including the iris and RPE	Established
rs10753550	E2F2	Overexpression relate to increase retinal RPE cell density in vivo	Established
rs12791393	OAF	Retinal thickness	Established
rs12319755	PLEKHA5	None reported	Novel
rs3138142	RDH5	Refractive error, High myopia, Spherical equivalent	Established
rs117821803	LINC02410	Refractive error	Established
rs2373980	HSP90B1	None reported	Novel
rs566259299	TRPV4	retinal ganglion cells and bipolar cells	Established
rs5442	CDCA3	Glaucoma, Refractive error, Myopia	Established
rs72680588	GCH1	None reported	Established
rs2350892	SIX6	Glaucoma	Novel

SNP	Closest Gene	Prior Retinal GWAS Associations	Retinal Association Status
rs56073585	VSX2	retinal progenitor cells	Established
rs7152160	FLRT2	Development of the zebrafish eyes	Established
rs35337422	RD3L	Refractive error, GCIPL thickness, myopia	Established
rs2614205	UNC13C	None reported	Novel
rs796924936	ALDH1A3	Retinoic acid gradient	Established
rs1800407	OCA2	Ocular albinism, eye color	Established
rs759019844	CHRNA7	None reported	Novel
rs139966754	MIR3178	Retinal thickness	Established
rs4635359	DYNLRB2-AS1	Refractive error, Macular thickness, Retinal thickness	Established
rs726799	HS3ST3A1	Retinal thickness	Established
rs538508139	BCAS3	Optic cup area	Established
rs7503221	NPLOC4	Benign neoplasm of eye, Cataracts	Established

SNP	Closest Gene	Prior Retinal GWAS Associations	Retinal Association Status
rs201723937	GUCY2D	Retinal guanylyl cyclase	Established
rs17452020	COLEC12	Retinal thickness	Established
rs7229033	FECH	None reported	Novel
rs34085461	LAMC1	lens, retina, and RPE of the developing zebrafish eye	Established
1:183342754_CT_C	-	None reported	Novel
rs2386994	TULP2	ocular diseases	Established
rs199806394	FLRT3	Retinal Ganglion Cells	Established
rs6433036	LRP2	Optic disc size	Established
rs17581284	PDK1	None reported	Established
rs11129195	NGLY1	None reported	Novel
rs2371070	SLC4A7	Color vision disorder	Established
rs73058466	LIMD1	None reported	Novel
rs115456027	MIR9-2HG	Vertical cup-disc ratio, Retinal venular width	Established
rs35437406	TFAP2D	Retinal thickness	Established
rs13057533	LIF-AS1	age-related macular degeneration	Established

SNP	Closest Gene	Prior Retinal GWAS Associations	Retinal Association Status
rs3770778	STRN	Retinal thickness	Established
rs11681837	RNU6-1168P	AMD	Established
rs112174475	NGLY1	None reported	Novel
rs77877421	FOXP1	Optic disc size	Established
rs200857348	MREG	Macular thickness	Established
rs5752638	Photoreceptor cell layer thickness phenotypes	MN1	Established
2:48792184_CTT_C	None reported	-	
rs115483925	VWA5B2	Retinal thickness	Established
rs377713857	MITF	Retinal vascular fractal density	Established
rs7614016	ADCY5	age-related macular degeneration	Established
rs13080798	VEPH1	pigmented layer of retina	Established
rs2326451	MIR588	Optic disc size	Established
rs34790261	GRB10	Retinal thickness	Established
rs55634298	EBAG9	None reported	Novel
rs60862542	EBAG9	None reported	Novel
rs10643278	STRIP2	pigmented layer of retina	Established

<b>SNP</b>	<b>Closest Gene</b>	<b>Prior Retinal GWAS Associations</b>	<b>Retinal Association Status</b>
rs10124312	CDKN2A	Vertical cup-disc ratio	Established
rs11143996	VPS13A	Aone reported	Established
rs7834615	BPNT2	None reported	Novel
rs3857971	CLVS1	Optic disc size	Established
rs13261390	GDAP1	None reported	Novel
rs759784176	PKD1L1	Cataract	Established
rs10781177	RORB-AS1	Photoreceptor cell layer thickness phenotypes	Established

#### 6.5.4 Results of mendelian randomisation

Using MR, we assessed causal relationships between cardiovascular traits and latent representations derived from a 64-dimensional retinal autoencoder space. We analysed three CVDs (CAD, stroke, MI) and two risk factors (body fat, triglycerides).

To comprehensively assess the direction of causality, we performed bidirectional MR analyses:

- **Forward MR:** Cardiovascular traits (exposure) → Retinal latent vectors (outcome)
- **Reverse MR:** Retinal latent vectors (exposure) → Cardiovascular traits (outcome)

Table 6.3 presents only the significant results of the forward MR (Cardiovascular Traits → Retinal Latent Vectors) analysis, which we will explain in detail below:

**Coronary Artery Disease:** We found seven latent variables that exhibited significant  $p$ -values using the inverse variance weighting (IVW) method. Notably, both positive and negative effects were observed depending on the dataset (e.g., positive in z006 and negative in z007), indicating heterogeneity in the results across studies.

**Body Fat:** Five latent vectors displayed significant  $p$ -values. Most IVW results were negative, suggesting that increased body fat is likely to decrease the measured outcome, with notable significance (e.g.,  $p = 0.0028$  for z008).

**Stroke:** This cardiac event showed the fewest associations, with only two latent vectors exhibiting significant  $p$ -values in the exposure analysis. The IVW method yielded mixed results (e.g., causal in z020 and protective in z034), albeit with some significant  $p$ -values, suggesting potential causal relationships.

**Triglycerides:** Predominantly negative effects were observed with the IVW method, indicating a protective effect, with some significant results (e.g.,  $p = 0.0081$  for z057). When considering triglycerides as an exposure, we identified eight latent vectors associated with causal/protective effects.

**Myocardial Infarction:** Four latent vectors showed significant  $p$ -values in the IVW analysis. The IVW method generally indicated positive effects with very significant  $p$ -values (e.g.,  $p = 0.0011$  for z017), suggesting a potential causal effect of the studied exposures.

The significant  $p$ -values observed across various results from different datasets and the IVW method underscore that these exposures (CVD risk/event GWAS) may causally influence outcomes (changes in retinal layer thickness) through MR analysis.

Exposure	Outcome	Method	Effect size	SE	<i>p</i> -value
coronary artery disease	z006	IVW	0.021842734	0.046360357	0.033799024
coronary artery disease	z007	IVW	-0.047641087	0.01497833	0.001469362
coronary artery disease	z011	IVW	0.032862319	0.016667996	0.048657362
coronary artery disease	z022	IVW	0.043992163	0.013013543	0.000723565
coronary artery disease	z027	IVW	-0.043314214	0.016455424	0.008483018
coronary artery disease	z028	IVW	-0.033791813	0.011483565	0.003254435
coronary artery disease	z048	IVW	0.040541422	0.01504963	0.007063227
Body Fat	z008	IVW	-0.221130867	0.073915693	0.002774621
Body Fat	z021	IVW	0.139399257	0.066586534	0.036304078
Body Fat	z023	IVW	-0.178709589	0.074242137	0.01607897
Body Fat	z045	IVW	-0.274720864	0.096375735	0.004364794
Body Fat	z050	IVW	-0.194423637	0.076286391	0.01081557
Stroke	z020	IVW	0.064005961	0.025608992	0.012441962
Stroke	z034	IVW	-0.086440108	0.032131075	0.007140194
Triglycerides	z008	IVW	0.05960648	0.023847513	0.012437427
Triglycerides	z010	IVW	0.051006721	0.023062696	0.02699053
Triglycerides	z020	IVW	-0.060597755	0.026446227	0.021942757
Triglycerides	z030	IVW	-0.054716527	0.024062892	0.02297221
Triglycerides	z046	IVW	-0.063050367	0.02615641	0.015930102
Triglycerides	z049	IVW	0.060201327	0.029662129	0.042400407
Triglycerides	z057	IVW	-0.060881105	0.02298727	0.008085692
Triglycerides	z060	IVW	-0.06260239	0.025140471	0.012770569
Myocardial Infarction	z017	IVW	0.054949462	0.016777536	0.001055944
Myocardial Infarction	z020	IVW	-0.083064999	0.020502557	5.09E-05
Myocardial Infarction	z031	IVW	0.051770067	0.01343862	0.000149076
Myocardial Infarction	z033	IVW	0.055573735	0.015306058	0.00036604

Table 6.3: Summary of statistically significant results from forward MR analyses. The table presents effect size estimates, standard errors, and *p*-values for associations between exposures (e.g., cardiovascular traits) and outcomes (retinal latent vectors). IVW, inverse-variance weighted.

Table 6.4 shows the most suggestive (though non-significant) associations from reverse MR analyses (Retinal Latent Vectors → Cardiovascular Traits), reporting the lowest observed *p*-values alongside effect estimates. This suggests retinal changes are more likely a consequence of CVD than a causative factor, aligning with known mechanisms of microvascular damage secondary to hypertension (Chen et al. 2012). While the retina shares vascular systemic circulation (Zhong et al. 2021), our findings do not support a primary role of retinal structure in driving CVD pathogenesis.

To our knowledge, prior studies investigating the causal relationship between retinal traits and CVD remain limited. As summarized in Chapter 3, Section 3.3 (Table 3.3), only a subset of existing work has explored this direction of causality, and, based on current literature, no studies have reported causal associations using OCT-derived phenotypes. For example, elevated blood pressure has been linked to reduced retinal vascular density in observational and genetic studies (Zekavat et al. 2021; Vela et al.

Exposure	Outcome	Method	Effect size	SE	<i>p</i> -value
z000	coronary artery disease	IVW	0.033451859	0.021344829	0.117065290
z008	coronary artery disease	IVW	0.040025622	0.027487363	0.145351891
z010	coronary artery disease	IVW	0.010044497	0.06013645	0.1387896
z002	Body Fat	IVW	0.01667006	0.012199793	0.171806659
z007	Body Fat	IVW	0.0134027927	0.022744478	0.155567563
z013	Body Fat	IVW	-0.004805159	0.047811931	0.192214959
z015	Stroke	IVW	-0.0145653114	0.016850169	0.138736739
z054	Stroke	IVW	-0.013502069	0.024147223	0.157605511
z058	Stroke	IVW	-0.090462195	0.047771911	0.087537266
z003	Triglycerides	IVW	-0.052877788	0.0291568659	0.097091343
z032	Triglycerides	IVW	-0.011882036	0.017266189	0.149134731
z053	Triglycerides	IVW	0.003489408	0.023697221	0.188293494
Z012	Myocardial Infarction	IVW	-0.007949401	0.016774837	0.163555656
Z024	Myocardial Infarction	IVW	-0.015039617	0.049185166	0.176758156
Z031	Myocardial Infarction	IVW	0.017659624	0.023860922	0.145923536

Table 6.4: This table summaries the top associations (lowest *p*-values) from analyses testing retinal latent vectors (exposures) against cardiovascular outcomes, including effect estimates ( $\beta$ ), standard errors (SE), and statistical significance (*p*values).

2023).

Our bidirectional MR supports a unidirectional causal pathway (cardiovascular traits  $\rightarrow$  retinal structure), with no evidence for reverse causation. This reinforces the retina’s role as a “window” to systemic vascular health rather than an active contributor to CVD risk.

## 6.6 Discussion

In this study, we employed MR to enhance our understanding of the relationship between CVDs and retinal phenotypes. We present the findings from a genome-wide assessment using OCT images of the retina from 42,350 individuals. The genome-wide association analysis of ten retinal layer thicknesses identified 90 genetic loci. Comparative genetic associations revealed multiple ocular and systemic traits linked to OCT-derived phenotypic thickness. To the best of our knowledge, this is the first study to conduct an in-depth investigation into the causal relationships between OCT-derived features and CVD.

Our GWAS identified several genes associated with ocular diseases, retinal characteristics, or components. While some of these genes had been previously reported, others represent novel findings in the context of retinal phenotypes. For instance, genes such as SIX6, frequently highlighted in other studies, are linked to retinal measurements and ocular diseases such as glaucoma (Carnes et al. 2014). A particularly notable gene is LAMC1, mutations in which lead to defects in the inner limiting membrane of the

zebrafish eye (Lee et al. 2007). The identification of genes previously associated with ocular traits lends credibility to the methodology employed in this study.

In relation to genes linked to cardiac events, biomarkers, or risk factors, several, including E2F2, were identified in our study. E2F2 has been associated with vascular contractility and blood pressure regulation. Mice deficient in the E2F2 gene exhibit significantly higher systolic and diastolic blood pressure levels compared to wild-type mice (Zhou et al. 2009). Another gene, NR5A2, shows reduced expression in metabolic syndrome. Exercise has been shown to increase NR5A2 expression in patients with metabolic syndrome, suggesting it as a potential therapeutic target for improving this condition through physical activity (Meng et al. 2023).

Our bidirectional MR analysis supports a unidirectional causal pathway, with cardiovascular traits influencing retinal structure but no evidence for reverse causation. We observed both causal and protective effects of CVD risks and events on retinal layer thickness. Genetically elevated body fat percentage showed protective associations with OCT-autoencoder phenotypes, aligning with prior work linking adiposity to retinal layer alterations (Zekavat et al. 2023). Myocardial infarction, stroke, and coronary artery disease exhibited significant, likely causal relationships with OCT-derived phenotypes.

These findings reinforce the retina's role as a "window" to systemic vascular health rather than an active driver of CVD risk. Supporting this, our group previously identified specific OCT features, including choroid layer thickness, GCL, and RNFL, as potential early predictors of MI and stroke (Maldonado-Garcia et al. 2024). In contrast, fundus imaging (e.g., macular and venous metrics) has been more extensively studied for CVD prediction (Diaz-Pinto et al. 2022), highlighting the need to expand OCT-based causal frameworks.

Nevertheless, we acknowledge the limitations of our study, particularly its reliance on UK Biobank data, which suffers from underrepresentation of non-White ethnicities (< 5% South Asian, < 2% Black African/Caribbean) (Sudlow et al. 2015), overrepresentation of urban and affluent populations, and a "healthy volunteer" bias (participants are healthier and wealthier than the UK general population) (Fry et al. 2017). These biases may limit the generalizability of our findings, especially to younger, lower-income, or non-European ancestry groups.

While our use of latent representations may partially capture universal biological pathways, replication in more diverse cohorts is essential to validate these results. Future studies should prioritize inclusive recruitment, ancestry-aware GWAS methods, and cross-population MR frameworks to assess causal

robustness (Cai et al. 2021).

# Chapter 7

## Conclusions

In this chapter, I bring this thesis to an end by summarising the key findings of our work. Then, I highlight limitations of this work and, finally, some avenues for future research.

### 7.1 Summaries and key findings

#### 7.1.1 Using optical coherence tomography as a predictive tool for cardiovascular diseases.

We developed a predictive model that integrates imaging data with non-imaging data to predict cardiovascular issues within a five-year period post-acquisition. This work is the first to provide detailed evidence of how OCT images can be utilised in clinical settings to estimate risk of CVD. Our framework employs a self-supervised representation learning approach based on a VAE with convolutional layers, pre-trained on OCT images from 20,000 healthy subjects. After feature extraction, the learned latent representations were used as input features for a RF classifier, which integrated these with non-imaging data (e.g., patient demographics and clinical variables) to identify patients at risk of stroke or myocardial infarction.

Our proposed method demonstrated that the latent variables learned from the retinal OCT images, had a greater influence on the classifiers' estimation of risk of myocardial infarction or stroke than conventional metadata. Specifically, in age- and sex-matched cohorts free from confounding these biomarkers, our model achieved balanced accuracy, sensitivity, and specificity of 0.70 and an AUC value of 0.75, using OCT imaging data from both eyes and associated metadata. In contrast, the model using only metadata produced lower results, with accuracy, sensitivity, and specificity values of 0.52 and and AUC

value of 0.57.

A significant contribution of this work is the development of a novel explainability AI method that aids in interpreting the OCT image features influencing the model's predictions. This method, based on vector field, provides more visually interpretable results compared to commonly used techniques, such as occlusion-based methods, thereby offering a deeper understanding of the model's decision-making process. Specifically, our findings highlight the choroid layer in OCT images as the most important feature for predictions.

### **7.1.2 A multimodal deep learning method using fundus photographs and optical coherence tomography for predicting cardiovascular diseases.**

We developed an innovative end-to-end DL method that utilises retinal phenotypes extracted from both fundus photographs and OCT images to classify CVDs, including myocardial infarction, stroke, angina, heart failure, and rheumatic heart disease. The two-stage architecture first pretrains a MCVAE on 18,000 healthy retinas, then fine-tunes with a transformer classifier on 3,000 matched CVD+/CVD- cases, where the input data consist of both retinal imaging modalities.

Our proposed framework demonstrated that the combination of both retinal modalities resulted in significantly better performance compared to using a single retinal modality, achieving accuracy, precision, sensitivity, specificity, and AUC values of 0.68, 0.74, 0.73, 0.68, and 0.78, respectively, for the optimal model (Fundus-OCT). These findings highlight the complementary value of multimodal retinal imaging: fundus photography captures the superficial vascular plexus, while OCT specifically assesses the choroidal circulation, together enabling a holistic evaluation of microvascular health.

Furthermore, we demonstrated that training a multimodal model to simultaneously extract relevant features from the input data and utilise those features for classification yields better results than training each task separately, with CI of 0.773 and 0.788 for AUC values for the first approach, compared to a CI of 0.63 - 0.65 for the AUC values of the second approach. This dual-task optimisation enhances the overall performance of the model.

Additionally, we showed that our explainability model, originally developed using vector field maps for OCT images, is also effective with other medical imaging modalities, such as fundus photographs, suggesting its potential applicability to other medical imaging for different organs.

### **7.1.3 Genetic basis and causal links influencing OCT-derived phenotypes and cardiovascular diseases.**

We conducted GWAS on OCT-derived phenotypic data from the UK Biobank, which were extracted in an unsupervised manner. To achieve this, we implemented the state-of-the-art nn-UNet model to train segmentation of ten OCT layers using ground truth data from only 50 patients, obtained via a non-deep learning algorithm. We highlight the advantages of using a deep learning–based segmentation model, which required only a small amount of annotated data to scale to the full dataset. Besides, this approach proved to be more efficient compared to other studies that require more extensive segmentations with fewer segmented layers. We then converted the ten segmented layers into thickness maps, which served as input for our 3D convolutional VAE architecture. This study is noteworthy for using a latent vector that represents novel features of retinal structure across ten OCT layers, making it the study with the most segmented OCT layers using DL techniques. This approach facilitated the discovery of previously unidentified 17 genetic loci associated with retinal thickness, and replicated 54 known retinal loci.

Furthermore, we established, for the first time, causal links between CVD and associated risks and OCT layer thickness through MR studies of OCT phenotypes. Using the IVW method, we obtained significant p-values of  $< 0.05$ , indicating that CVD risk factors such as body fat, and triglycerides, as well as CVD events like coronary artery disease, myocardial infarction, and stroke, have causal and/or protective relationships with certain OCT-derived features. Reverse MR analyses suggested that retinal thickness alterations are more likely a consequence of CVD rather than a causative factor.

This GWAS of 3D OCT retinal images identifies genetic variants linked to retinal layer thickness and suggests potential causal relationships between cardiovascular diseases and retinal layer thickness.

## **7.2 Implications and significance**

**Clinical implications:** Our research suggests that OCT and fundus photography can reveal microvascular alterations that may enable earlier identification and targeting of individuals at risk for cardiovascular disease in a noninvasive manner. While the models developed in this thesis are not yet ready for clinical deployment, the findings highlight the potential for future studies to further explore and validate their clinical utility.

**Biological implications:** We identified previously unreported genetic loci associated with retinal layer thickness and established causal relationships between cardiovascular risk factors and OCT-derived mor-

phological phenotypes. These findings enhance our understanding of the biological link between retinal structure and cardiovascular health, suggesting that retinal changes may function as indicators to systemic vascular disease.

**Methodological implications:** We highlight the effectiveness of using predictive DL techniques alongside medical imaging to obtain more detailed imaging-derived phenotypes, which enhance the accuracy of CVDs prediction. The frameworks we proposed could be applied to other medical imaging challenges, thereby broadening the impact of our research beyond just CVDs prediction. Additionally, the development of a novel explainability method for AI models significantly enhances the interpretability of DL models, making them more transparent and trustworthy for clinical use. This is particularly important in medical AI, where understanding the rationale behind predictions is crucial for gaining the trust of clinicians and patients.

### 7.3 Limitations, challenges, and opportunities for future research

In this section, we outline some of the limitations of this work and, in certain cases, propose ideas for addressing them in future studies.

**External validation dataset.** The primary limitation of this work is the lack of accessible datasets that include both OCT images and comprehensive information on whether patients have experienced cardiovascular events, along with corresponding clinical and demographic data. While more datasets are available for fundus imaging that include cardiovascular event information, they often do not include the other retinal modalities like OCT. To the best of our knowledge, no existing datasets meet all these criteria and are accessible for validation purposes.

**Limitations of the UK Biobank and impact on generalizability.** The UK Biobank is an invaluable resource for large-scale biomedical research, but its design introduces several well-documented limitations that must be considered when interpreting our findings. Participants are predominantly of white European ancestry and exhibit a “healthy volunteer” bias, with lower rates of smoking, drinking alcohol, obesity, and socioeconomic deprivation compared to the general UK population. Besides, UK Biobank’s CVD outcomes rely on hospital admissions and death registers, potentially missing milder or undiagnosed cases. Nevertheless, replication in more diverse cohorts, is essential before clinical deployment.

**Class imbalance across study phases.** A key limitation stems from inconsistent class imbalance ratios across experimental stages (Chapters 4–5). While the original UK Biobank retina dataset preserved a clinically realistic CVD+-to-control ratio, pre-processing exclusions created mismatched training distributions. Although balancing improved feature learning for CVD+ cases, it may have inflated performance metrics (optimism bias). The test set adopted an intermediate ratio approximating clinical prevalence but remained discordant with both natural distributions and the training environment. Future work should validate findings in unmodified cohorts and explore cost-sensitive learning to address inherent imbalances.

**Comparative limitations with QRISK3.** While our model demonstrated superior discriminative performance compared to QRISK3 in age-sex matched cohorts, this comparison has important considerations. QRISK3’s lower performance may partly reflect our intentional age-sex matching, which neutralizes its strongest predictors. In unmatched general populations, QRISK3’s AUC typically reaches highest performance, suggesting our results may overstate the relative advantage of OCT in clinical practice. Additionally, QRISK3 has been externally validated in > 10 million patients across diverse settings, whereas our OCT model requires validation in comparable real-world cohorts to assess generalizability beyond the UK Biobank’s selected population.

**Incorporating additional retinal imaging modalities.** Recently, OCT-Angiography (OCTA) has shown promise as a retinal imaging technique that could be more effective for predicting cardiovascular diseases. The use of more informative retinal imaging modalities with deeper tissue penetration, such as swept source OCT or wide-field OCTA imaging, could potentially enhance prediction models. We hypothesize that incorporating representations from multi-modal retinal imaging (e.g., fundus photographs, OCT, OCTA) may further improve the classification performance of the proposed approach.

**Utilising multi-organ data.** A promising direction for future research involves the integration of data from multiple organs. Combining information from various sources could significantly improve the performance of predictive models. Resources such as the UK Biobank, which offer extensive datasets, present the opportunity to incorporate medical images from different organs, such as liver and pancreas imaging, or cardiac magnetic resonance imaging (MRI), all of which have been utilised to assess cardiovascular risk factors and events. This approach would enable a more comprehensive investigation of cardiovascular and cardiometabolic conditions.

## 7.4 Final remarks

In this thesis, we have demonstrated the potential of combining deep learning methodologies with retinal imaging to make meaningful contributions to cardiovascular disease research, supporting the idea that deeply multidisciplinary approaches can yield valuable insights into the mechanisms of health and disease.

We hope that the novel scientific findings on the clinical relevance of retinal phenotypes in relation to cardiovascular disease, together with the methodological advancements presented in this work, will serve as a foundation for improving patient care and inspiring further research in related domains.

# References

- Abdellaoui, Abdel, Yengo, Loic, Verweij, Karin J. H., and Visscher, Peter M. (2023). “15 years of GWAS discovery: Realizing the promise.” In: *American journal of human genetics*.
- Abràmoff, Michael David, Garvin, Mona Kathryn, and Sonka, Milan (2010). “Retinal Imaging and Image Analysis”. In: *IEEE Reviews in Biomedical Engineering* 3, pp. 169–208.
- Al-Absi, Hamada R. H., Islam, Mohammad Tariqul, Refaee, Mahmoud Ahmed, Chowdhury, Muhammad Enamul Hoque, and Alam, Tanvir (2022). “Cardiovascular Disease Diagnosis from DXA Scan and Retinal Images Using Deep Learning”. In: *Sensors (Basel, Switzerland)* 22.
- Adam, Margaret P, Hh, Ardinger, Pagon, Roberta A., Se, Wallace, Ljh, Bean, Stephens, Karen G., and Amemiya, Anne R (2019). “NGLY1-Related Congenital Disorder of Deglycosylation – GeneReviews®”. In.
- Alliance, Genetic, Genetic, The New York-Mid-Atlantic Consortium for, and Services, Newborn Screening (2009). *Understanding Genetics: A New York, Mid-Atlantic Guide for Patients and Health Professionals*. Washington (DC): Genetic Alliance.
- Amal, Saeed, Safarnejad, Lida, Omiye, Jesutofunmi A., Ghanzouri, Ilies, Cabot, John H., and Ross, Elsie Gyang (2022). “Use of Multi-Modal Data and Machine Learning to Improve Cardiovascular Disease Care”. In: *Frontiers in Cardiovascular Medicine* 9.
- Anderson, Todd J., Uehata, Akimi, Gerhard, Marie D., Meredith, Ian T., Knab, S, Delagrang, Danielle, Lieberman, Eric H., Ganz, Peter, Creager, Mark A., and Yeung, Alan C (1995). “Close relation of endothelial function in the human coronary and peripheral circulations.” In: *Journal of the American College of Cardiology* 26 5, pp. 1235–41.

- Antelmi, Luigi, Ayache, Nicholas, Robert, Philippe, and Lorenzi, Marco (2019). “Sparse Multi-Channel Variational Autoencoder for the Joint Analysis of Heterogeneous Data”. In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. PMLR, pp. 302–311.
- Aumann, Stephan, Donner, Sebastian, Fischer, Julian, et al. (2019). “Optical Coherence Tomography (OCT): Principle and Technical Realization”. In: *High Resolution Imaging in Microscopy and Ophthalmology: New Frontiers in Biomedical Optics*. Ed. by Josef F Bille. Springer, Cham. Chap. 3, pp. 59–85.
- Azad, Reza, Aghdam, Ehsan Khodapanah, Rauland, Amelie, Jia, Yiwei, Avval, Atlas Haddadi, Bozorgpour, Afshin, Karimijafarbigloo, Sanaz, Cohen, Joseph Paul, Adeli, Ehsan, and Merhof, Dorit (2022). “Medical Image Segmentation Review: The Success of U-Net”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, pp. 10076–10095.
- Baas, Dominique C., Despriet, Dominiek D. G., Gorgels, Theo G. M. F., Bergeron-Sawitzke, Julie, Uitterlinden, André G., Hofman, Albert, Duijn, Cornelia M. van, Merriam, Joanna E., Smith, R. Theodore, Barile, Gaetano, Brink, Jacoline B. ten, Vingerling, Johannes R., Klaver, Caroline C. W., Allikmets, Rando, Dean, Michael, Bergen, Arthur A.B., and Bergen, Arthur A.B. (2010). “The ERCC6 Gene and Age-Related Macular Degeneration”. In: *PLoS ONE* 5.
- Band, Gavin and Marchini, Jonathan (2018). *BGENIE: BGEN File Reader and Genetic Association Tool*. <https://github.com/oxfordmm/bgenie>. Version: 1.4.
- Batty, G. David, Gale, Catharine R., Kivimäki, Mika, Deary, Ian J., and Bell, Steven (2020). “Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis”. In: *The BMJ* 368.
- Benarab, Charaf Eddine and Gui, Shenglin (2022). “CNN-Trans-Enc: A CNN-Enhanced Transformer-Encoder On Top Of Static BERT representations for Document Classification”. In: *ArXiv* abs/2209.06344.
- Bharadwaj, Arpita, Appukuttan, Binoy, Wilmarth, Phillip A., Pan, Yuzhen, Stempel, Andrew J., Chipps, Timothy J., Benedetti, Eric, Zamora, David O., Choi, Dongseok, David, Larry, and Smith, Justine R.

- (2013). “Role of the retinal vascular endothelial cell in ocular disease”. In: *Progress in Retinal and Eye Research* 32, pp. 102–180.
- Bishop, Christopher M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*.
- Breiman, L. (2001). “Random Forests”. In: *Machine Learning* 45, pp. 5–32.
- Brown, Alex, Tomaev, Nenad, Freyberg, Jana von, Liu, Yuan, Karthikesalingam, Alan, and Schrouff, Jessica (2022). “Detecting shortcut learning for fair medical AI using shortcut testing.” In: *Nature communications* 14 1, p. 4314.
- Bycroft, Clare and al., et (2017). “Genome-wide genetic data on 500,000 UK Biobank participants”. In: *Nature* 551.7681, pp. 207–212.
- Bycroft, Clare, Freeman, Colin, Petkova, Desislava, Band, Gavin, Elliott, Lloyd T., Sharp, Kevin, Motyer, Allan, Vukcevic, Damjan, Delaneau, Olivier, O’Connell, Jared, Cortes, Adrian, Welsh, Samantha, Young, Alan, Effingham, Mark, McVean, Gil, Leslie, Stephen, Allen, Naomi E., Donnelly, Peter, and Marchini, Jonathan (2018). “The UK Biobank resource with deep phenotyping and genomic data”. In: *Nature* 562, pp. 203–209.
- Cai, Mingxuan, Xiao, Jiashun, Zhang, Shunkang, Wan, Xiang, Zhao, Hongyu, Chen, Gang, and Yang, Can (2021). “A unified framework for cross-population trait prediction by leveraging the genetic correlation of polygenic traits.” In: *American journal of human genetics*.
- Carnes, Megan Ulmer, Liu, Yangfan P., Allingham, Rand R., Whigham, Benjamin T., Havens, Shane, Garrett, Melanie E., Qiao, Chun-yan, Qiao, Chun-yan, Katsanis, Nicholas, Wiggs, Janey L., Pasquale, Louis R., Ashley-Koch, Allison E., Oh, Edwin C. T., Hauser, Michael A., Brilliant, Murray H., Budenz, Donald L., Chin, Hemin R., Bailey, Jessica N. Cooke, Fingert, John H., Friedman, David S., Gaasterland, Douglas E., Gaasterland, Terry, Haines, Jonathan L., Kang, Jae Hee, Lee, Richard K., Lichter, Paul R., Liu, Yutao, Loomis, Stephanie J, McCarty, Cathy Essentia, Moroi, Sayoko E., Pericak-Vance, Margaret A., Realini, Anthony, Richards, Julia E, Schuman, Joel S., Scott, William K., Singh, Kuldev, Sit, Arthur J., Vollrath, Douglas, Weinreb, Robert N., Wollstein, Gadi, Yaspan, Brian L., Zack, Donald J., and Zhang, Kang (2014). “Discovery and Functional Annotation of SIX6 Variants in Primary Open-Angle Glaucoma”. In: *PLoS Genetics* 10.

- Chen, Yanping, Yuan, Yixiong, Zhang, Shiran, Yang, Shaopeng, Zhang, Junyao, Guo, Xiao-lei, Huang, Wenyong, Zhu, Zhuoting, He, Mingguang, and Wang, Wen (2023). “Retinal nerve fiber layer thinning as a novel fingerprint for cardiovascular events: results from the prospective cohorts in UK and China”. In: *BMC Medicine* 21.
- Chen, Yuh-Lien, Wong, Tien Yin, Cheung, Carol Yim-Lui, Ikram, M. Kamran, Sabanayagam, Charumathi, and Wong, Tien Yin (2012). “Retinal Microvasculature as a Model to Study the Manifestations of Hypertension”. In: *Hypertension* 60, pp. 1094–1103.
- Chueh, Kuan-Ming, Hsieh, Yi-Ting, Chen, Homer H., Ma, Islam, and Huang, Sheng-Lung (2021). “Identification of Sex and Age from Macular Optical Coherence Tomography and Feature Analysis Using Deep Learning.” In: *American journal of ophthalmology*.
- Ciampi, Antonio, Greenwood, Celia M. T., Hendricks, Audrey E., Li, Rui, Metrustry, Sarah, Oualkacha, Karim, Tachmazidou, Ioanna, Xu, Chang-Jiang, Zeggini, Eleftheria, et al. (2015). “The UK10K project identifies rare variants in health and disease”. In: *Nature* 526.7571, pp. 82–90.
- Copete, Sergio, Flores-Moreno, Ignacio, Montero, Javier Antonio, Duker, Jay S., and Ruiz-Moreno, José M. (2013). “Direct comparison of spectral-domain and swept-source OCT in the measurement of choroidal thickness in normal eyes”. In: *British Journal of Ophthalmology* 98, pp. 334–338.
- Cunningham, Fiona, Allen, James E, Allen, Jamie, Alvarez-Jarreta, Jorge, Amode, M Ridwan, Armean, Irina M, Austine-Orimoloye, Olanrewaju, Azov, Andrey G, Barnes, If, Bennett, Ruth, et al. (2022). “Ensembl 2022”. In: *Nucleic acids research* 50.D1, pp. D988–D995.
- Currant, Hannah, Hysi, Pirro G., Fitzgerald, Tesca, Gharahkhani, Puya, Bonnemaier, Pieter W. M., Atan, Denize, Aung, Tin, Charng, Jason, Choquet, Heene, Craig, Jamie E., Hewitt, Alex W., Khaw, Peng Tee, Klaver, Caroline C. W., Kubo, Michiaki, Ong, Jue-Sheng, Pasquale, Louis R., Reisman, Charles A., Simcoe, Mark J., Thiadens, Alberta A.H.J., Duijn, Cornelia M. van, Yazar, Seyhan, Jorgenson, Eric, Macgregor, Stuart, Hammond, Christopher J., Mackey, David A., Wiggs, Janey L., Foster, Paul J., Patel, Praveen J., Birney, Ewan, and Khawaja, Anthony P. (2020). “Genetic variation affects morphological retinal phenotypes extracted from UK Biobank optical coherence tomography images”. In: *PLoS Genetics* 17.

- D'Agostino, Ralph B., Pencina, Michael J., Massaro, Joseph M., and Coady, Sean A. (2013). "Cardiovascular Disease Risk Assessment: Insights from Framingham." In: *Global heart* 8 1, pp. 11–23.
- Dai, Guangzheng, He, Wei, Xu, Ling, Pazo, Eric E, Lin, Tiezhu, Liu, Shasha, and Zhang, Chenguang (2020). "Exploring the effect of hypertension on retinal microvasculature using deep learning on East Asian population". In: *PLoS ONE* 15.
- Diaz-Pinto, Andrés, Ravikumar, Nishant, Attar, Rahman, Suinesiaputra, Avan, Zhao, Yitian, Levelt, Eylem, Dall'armellina, E, Lorenzi, Marco, Chen, Qingyu, Keenan, Tiarnán D. L., Agrón, Elvira, Chew, Emily Y., Lu, Zhi-gang, Gale, Chris P., Gale, Richard, Plein, Sven, and Frangi, Alejandro F. (2022). "Predicting myocardial infarction through retinal scans and minimal personal information". In: *Nature Machine Intelligence* 4, pp. 55–61.
- Farrah, Tariq E., Dhillon, Baljean, Keane, Pearse A., Webb, David John, and Dhaun, Neeraj (2020). "The eye, the kidney, and cardiovascular disease: old concepts, better tools, and new horizons". In: *Kidney International* 98, pp. 323–342.
- Farrah, Tariq E., Webb, David John, and Dhaun, Neeraj (2019). "Retinal fingerprints for precision profiling of cardiovascular risk". In: *Nature Reviews Cardiology* 16, pp. 379–381.
- Federation, World Heart (2023). "World Heart Report 2023: Confronting the World's Number One Killer. Geneva, Switzerland". In: *World Heart Federation*.
- Fernández, Alberto, García, Salvador, Galar, Mikel, Prati, Ronaldo Cristiano, Krawczyk, B., and Herrera, Francisco (2018). "Learning from Imbalanced Data Sets". In: *Cambridge International Law Journal*.
- Ferrara, Mariantonia, Lugano, Gaia, Sandinha, Maria Teresa, Kearns, Victoria R., Geraghty, Brendan, and Steel, David H. W. (2021). "Biomechanical properties of retina and choroid: a comprehensive review of techniques and translational relevance". In: *Eye* 35, pp. 1818–1832.
- Flammer, Josef, Konieczka, Katarzyna, Bruno, Rosa Maria, Virdis, Agostino, Flammer, Andreas J., and Taddei, Stefano (2013). "The eye and the heart". In: *European Heart Journal* 34, pp. 1270–1278.
- Fry, Anna, Littlejohns, Thomas J., Sudlow, Cathie L. M., Doherty, Nicola, Adamska, Ligia, Sprosen, Tim, Collins, Rory, and Allen, Naomi E. (2017). "Comparison of Sociodemographic and Health-

- Related Characteristics of UK Biobank Participants With Those of the General Population”. In: *American Journal of Epidemiology* 186, pp. 1026–1034.
- Fu, H., Wang, Boyang, Shen, Jianbing, Cui, Shanshan, Xu, Yanwu, Liu, Jiang, and Shao, Ling (2019). “Evaluation of Retinal Image Quality Assessment Networks in Different Color-spaces”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*.
- Gerrits, Nele, Elen, Bart, Craenendonck, Toon van, Triantafyllidou, Danai, Petropoulos, Ioannis N., Malik, Rayaz Ahmed, and Boever, Patrick De (2020). “Age and sex affect deep learning prediction of cardiometabolic risk factors from retinal images”. In: *Scientific Reports* 10.
- Ghosal, Samit and Sinha, Binayak (2022). “Using machine learning to assess cardiovascular risk in type 2 diabetes mellitus patients without established cardiovascular disease: The MARK-2 analysis”. In: *International Journal of Diabetes and Technology* 1, pp. 114–118.
- Girach, Zain, Sarian, Arni, Maldonado-García, Cynthia, Ravikumar, Nishant, Sergouniotis, Panagiotis I., Rothwell, Peter M., Frangi, Alejandro F., and Julian, Thomas H. (2024). “Retinal imaging for the assessment of stroke risk: a systematic review.” In: *Journal of neurology*.
- Govindarajan, Gurushankar, Whaley-Connell, Adam, Mugo, Maryann N., Stump, Craig S., and Sowers, James R. (2005). “The Cardiometabolic Syndrome as a Cardiovascular Risk Factor”. In: *The American Journal of the Medical Sciences* 330, pp. 311–318.
- Guo, Shao-Hua, Yin, Songtao, Tse, Gary, Li, Guangping, Su, Long, and Liu, Tong (2020). “Association Between Caliber of Retinal Vessels and Cardiovascular Disease: a Systematic Review and Meta-Analysis”. In: *Current Atherosclerosis Reports* 22, pp. 1–13.
- Gupta, Ankit (2019). “StrokeSave: A Novel, High-Performance Mobile Application for Stroke Diagnosis using Deep Learning and Computer Vision”. In: *ArXiv abs/1907.05358*.
- Guyon, Isabelle M, Weston, Jason, Barnhill, Stephen D., and Vapnik, Vladimir Naumovich (2002). “Gene Selection for Cancer Classification using Support Vector Machines”. In: *Machine Learning* 46, pp. 389–422.
- Hanssen, Henner, Streese, Lukas, and Vilser, Walthard (2022). “Retinal vessel diameters and function in cardiovascular risk and disease”. In: *Progress in Retinal and Eye Research* 91.

- Hartwig, Fernando Pires, Davies, Neil Martin, Hemani, Gibran, and Davey Smith, George (2016). *Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique*.
- Hassan, Osama N., Menten, Martin J., Bogunović, Hrvoje, Schmidt-Erfurth, Ursula Margarethe, Lotery, Andrew John, and Rueckert, Daniel (2021). “Deep Learning Prediction Of Age And Sex From Optical Coherence Tomography”. In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pp. 238–242.
- Hemani, Gibran, Zheng, Jie, Elsworth, Benjamin L., Wade, Kaitlin H., Haberland, Valeriia, Baird, Denis A, Laurin, Charles, Burgess, Stephen, Bowden, Jack, Langdon, Ryan J., Tan, Vanessa Y., Yarmolinsky, James, Shihab, Hashem A., Timpson, Nicholas John, Evans, David M., Relton, Caroline L., Martin, Richard M., Smith, George Davey, Gaunt, Tom R., and Haycock, Philip C. (2018). “The MR-Base platform supports systematic causal inference across the human phenome”. In: *eLife* 7.
- Hippisley-Cox, Julia, Coupland, Carol A. C., and Brindle, Peter M (2017). “Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study”. In: *The BMJ* 357.
- Huang, David, Swanson, Eric A., Lin, Charles P., Schuman, Joel S., Stenson, William G., Chang, Warren, Hee, Michael Richard, Flotte, Thomas, Gregory, Kenton, Puliafito, Carmen A., and Fujimoto, James G. (1991). “Optical Coherence Tomography”. In: *LEOS '92 Conference Proceedings*, pp. 656–657.
- Huang, Yu, Cheung, Carol Y, Li, Dawei, Tham, Yih-Chung, Sheng, Bin, Cheng, Ching-Yu, Wang, Ya Xing, and Wong, Tien Yin (2023). “AI-integrated ocular imaging for predicting cardiovascular disease: advancements and future outlook.” In: *Eye*.
- Isensee, Fabian, Jaeger, Paul F., Kohl, Simon A. A., Petersen, Jens, and Maier-Hein, Klaus (2020). “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation”. In: *Nature Methods* 18, pp. 203–211.
- Jiang, Xiaofan, Hysi, Pirro G., Khawaja, Anthony P., Mahroo, Omar A., Xu, Zihe, Hammond, Christopher J., Foster, Paul J., Welikala, Roshan Alex, Barman, Sarah Ann, Whincup, Peter H., Rudnicka, Alicja Regina, Owen, Christopher G., and Strachan, David P. (2023). “GWAS on retinal vasculometry phenotypes”. In: *PLOS Genetics* 19.

- Jordan, Michael I. and Mitchell, Thomas (2015). “Machine learning: Trends, perspectives, and prospects”. In: *Science* 349, pp. 255–260.
- Julian, Thomas H., Boddy, Sarah L., Islam, Mahjabin, Kurz, Julian M, Whittaker, Katherine J, Moll, Tobias, Harvey, Calum, Zhang, Sai, Snyder, Michael P, McDermott, Christopher J., Cooper-Knock, Johnathan, and Shaw, Pamela J (2021). “A review of Mendelian randomization in amyotrophic lateral sclerosis.” In: *Brain : a journal of neurology*.
- Julian, Thomas H., Cooper-Knock, Johnathan, Macgregor, Stuart, Guo, Hui, Aslam, Tariq Mehmood, Sanderson, Eleanor, Black, Graeme C.M., and Sergouniotis, Panagiotis I. (2022). “Phenome-wide Mendelian randomisation analysis identifies causal factors for age-related macular degeneration”. In: *eLife* 12.
- Keane, Pearse A., Grossi, Carlota M., Foster, Paul J., Yang, Qi, Reisman, Charles A, Chan, Kinpui, Peto, Tunde, Thomas, Dhanes, and Patel, Praveen J. (2016). “Optical Coherence Tomography in the UK Biobank Study – Rapid Automated Analysis of Retinal Thickness for Large Population-Based Studies”. In: *PLoS ONE* 11.
- Khan, Adnan, Boever, Patrick De, Gerrits, Nele, Akhtar, Naveed, Saqqur, Maher, Ponirakis, Georgios, Gad, Hoda, Petropoulos, Ioannis N., Shuaib, Ashfaq, Faber, James E., Kamran, Saadat, and Malik, Rayaz Ahmed (2022). “Retinal vessel multifractals predict pial collateral status in patients with acute ischemic stroke”. In: *PLoS ONE* 17.
- Khozeimeh, Fahime, Sharifrazi, Danial, Izadi, Navid Hoseini, Joloudari, Javad Hassannataj, Shoeibi, Afshin, Alizadehsani, Roohallah, Tartibi, Mehrzad, Hussain, Sadiq, Sani, Zahra Alizadeh, Khodatars, Marjane, et al. (2022). “RF-CNN-F: random forest with convolutional neural network features for coronary artery disease diagnosis based on cardiac magnetic resonance”. In: *Scientific Reports* 12.1, p. 11178.
- Kim, Wonkuk, Londono, Douglas, Zhou, Lisheng, Xing, Jinchuan, Nato, Alejandro Q., Musolf, Anthony M., Matise, Tara Cox, Finch, Stephen J., and Gordon, Derek (2013). “Single-Variant and Multi-Variant Trend Tests for Genetic Association with Next-Generation Sequencing That Are Robust to Sequencing Error”. In: *Human Heredity* 74, pp. 172–183.

- Kingma, Diederik P. and Ba, Jimmy (2015). “Adam: A Method for Stochastic Optimization”. In: *CoRR* abs/1412.6980.
- Kingma, Diederik P. and Welling, Max (2014). “Auto-Encoding Variational Bayes”. In: *CoRR* abs/1312.6114.
- Ko, Fang, Foster, Paul J., Strouthidis, Nicholas G., shweikh, yusrah, Yang, Qi, Reisman, Charles A, Muthy, Zaynah A., Chakravarthy, Usha, Lotery, Andrew John, Keane, Pearse A., Tufail, Adnan, Grossi, Carlota M., and Patel, Praveen J. (2017). “Associations with Retinal Pigment Epithelium Thickness Measures in a Large Cohort: Results from the UK Biobank.” In: *Ophthalmology* 124 1, pp. 105–117.
- Langner, Sönke, Terheyden, Jan Henrik, Geerling, Clara F, Kindler, Christine, Keil, Vera C W, Turski, Christopher A, Turski, Gabrielle N., Behning, Charlotte, Wintergerst, Maximilian W. M., Petzold, Gabor C., and Finger, Robert P. (2022). “Structural retinal changes in cerebral small vessel disease”. In: *Scientific Reports* 12.
- Lau, Alexander Yuk-lun, Mok, Vincent Chung-Tong, Lee, Jack, Fan, Yuhua, Zeng, Jinsheng, Lam, Bonnie Yin Ka, Wong, Adrian, Kwok, Chloe, Lai, Maria, and Zee, Benny Chung Ying (2018). “Retinal image analytics detects white matter hyperintensities in healthy adults”. In: *Annals of Clinical and Translational Neurology* 6, pp. 98–105.
- Lee, Jiwoon and Gross, Jeffrey M. (2007). “Laminin beta1 and gamma1 containing laminins are essential for basement membrane integrity in the zebrafish eye.” In: *Investigative ophthalmology & visual science* 48 6, pp. 2483–90.
- Leica Microsystems (2025). *A Guide to OCT*. URL: <https://www.leica-microsystems.com/science-lab/medical/a-guide-to-oct/>.
- Li, Yan, Sperrin, Matthew, and Staa, Tjeerd Pieter van (2019). “R package “QRISK3”: an unofficial research purposed implementation of ClinRisk’s QRISK3 algorithm into R”. In: *F1000Research*.
- Lindstrom, Megan et al. (2023). “DEATHS FROM CARDIOVASCULAR DISEASE SURGED 60 GLOBALLY OVER THE LAST 30 YEARS: REPORT”. In: *World Heart Federation*.
- Littlejohns, Thomas J., Sudlow, Cathie L. M., Allen, Naomi E., and Collins, Rory (2019). “UK Biobank: opportunities for cardiovascular research”. In: *European Heart Journal* 40, pp. 1158–1166.

- Livingstone, Shona, Morales, Daniel R., Donnan, Peter T., Payne, Katherine, Thompson, Alexander James, Youn, Ji-Hee, and Guthrie, Bruce (2021). “Effect of competing mortality risks on predictive performance of the QRISK3 cardiovascular risk prediction tool in older people and those with co-morbidity: external validation population cohort study”. In: *The Lancet. Healthy Longevity* 2, e352–e361.
- Lucas, Bruce D. and Kanade, Takeo (1981). “An Iterative Image Registration Technique with an Application to Stereo Vision”. In: *International Joint Conference on Artificial Intelligence*.
- Lundberg, Scott M. and Lee, Su-In (2017). “A Unified Approach to Interpreting Model Predictions”. In: *Neural Information Processing Systems*.
- Ma, Yanjun, Xiong, Jianhao, Zhu, Yidan, Ge, Zongyuan, Hua, Rong, Fu, Meng, Li, Chenglong, Wang, Bin, Dong, Li, Zhao, Xin, Chen, Jili, Rong, Ce, He, Chao, Chen, Yuzhong, Wang, Zhaohui, Wei, Wen Bin, Xie, Wuxiang, and Wu, Yangfeng (2021). “Deep learning algorithm using fundus photographs for 10-year risk assessment of ischemic cardiovascular diseases in China.” In: *Science bulletin* 67 1, pp. 17–20.
- Maldonado-Garcia, Cynthia, Bonazzola, Rodrigo, Ferrante, Enzo, Julian, Thomas H, Sergouniotis, Panagiotis I., Ravikumar, Nishant, and Frangi, Alejandro F. (2024). “Predicting risk of cardiovascular disease using retinal OCT imaging”. In: *ArXiv abs/2403.18873*.
- Maldonado-Garcia, Cynthia, Zakeri, Arezoo, Frangi, Alejandro F., and Ravikumar, Nishant (2025). “Integrating Deep Learning with Fundus and Optical Coherence Tomography for Cardiovascular Disease Prediction”. In: ed. by Islem Rekik, Ehsan Adeli, Sang Hyun Park, and Celia Cintas, pp. 161–172.
- Matuleviūtė, Indrė, Sidaraitė, Agnė, Tatarunas, Vacis, Veikutienė, Audronė, Dobilienė, Olivija, and Zaliuniene, Dalia (2022). “Retinal and Choroidal Thinning—A Predictor of Coronary Artery Occlusion?” In: *Diagnostics* 12.
- Mccarroll, Steven, Kuruvilla, Finny G., Korn, Joshua M., Cawley, Simon E., Nemes, James, Wysoker, Alec, Shapero, Michael H., Bakker, Paul I. W. de, Maller, Julian B., Kirby, Andrew W., Elliott, Amanda, Parkin, Melissa, Hubbell, Earl, Webster, Teresa A., Mei, Rui, Veitch, Jim, Collins, Patrick J., Handsaker, Robert E., Lincoln, Stephen E., Nizzari, Marcia M., Blume, John E., Jones, Keith W., Rava, Richard P., Daly, Mark J., Gabriel, S., and Altshuler, David M (2008). “Integrated detection and

- population-genetic analysis of SNPs and copy number variation”. In: *Nature Genetics* 40, pp. 1166–1174.
- McCarthy, Shane and al., et (2016). “A reference panel of 64,976 haplotypes for genotype imputation”. In: *Nature Genetics* 48.10, pp. 1279–1283.
- Meng, Lingxiu, Dong, Fusheng, and Deng, Junguo (2023). “NR5A2 as a potential target for exercise to improve metabolic syndrome”. In: *Aging (Albany NY)* 15, pp. 2485–2502.
- Menten, Martin J., Holland, Robbie, Leingang, Oliver, Bogunovic ´, Hrvoje, Hagag, Ahmed M., Kaye, Rebecca, Riedl, Sophie, Traber, Ghislaine Lieselotte, Hassan, Osama N., Pawlowski, Nick, Glocker, Ben, Fritsche, Lars G., Scholl, Hendrik P. N., Sivaprasad, Sobha, Schmidt-Erfurth, Ursula, Rueckert, Daniel, and Lotery, Andrew John (2023). “Exploring healthy retinal aging with deep learning”. In: *Ophthalmology Science*.
- Milosevic, Marko, Jin, Qingchu, Singh, Akarsh, and Amal, Saeed (2024). “Applications of AI in multi-modal imaging for cardiovascular disease”. In: *Frontiers in Radiology* 3.
- Mishra, Chandra and Tripathy, Koushik (2024). *Fundus Camera*. Updated 2023 Aug 25. Treasure Island (FL): StatPearls Publishing.
- Mueller, Simon, Wintergerst, Maximilian W. M., Falahat, Peyman, Holz, Frank G., Schaefer, Christian Alexander, Schahab, Nadjib, Finger, Robert P., and Schultz, Thomas (2022). “Multiple instance learning detects peripheral arterial disease from high-resolution color fundus photography”. In: *Scientific Reports* 12.
- Munk, Marion R., Kurmann, Thomas, Márquez-Neila, Pablo, Zinkernagel, Martin Sebastian, Wolf, Sebastian, and Sznitman, Raphael (2021). “Assessment of patient specific information in the wild on fundus photography and optical coherence tomography”. In: *Scientific Reports* 11.
- Nasaré, Aléx Martins, Tedesco, Roberto Carlos, Cristovam, Priscila Cardoso, Cenedese, Marcos A, Galisteo, Andrés Jimenez, Andrade, Heitor Franco de, Gomes, José Álvaro Pereira, Guimarães, Érik V, Barbosa, Helene Santos, and Alonso, Luís Garcia (2015). “Toxoplasma Gondii Infection of Chicken Embryos Causes Retinal Changes and Modulates HSP90B1 Gene Expression: A Promising Ocular Toxoplasmosis Model”. In: *European Journal of Microbiology & Immunology* 5, pp. 316–320.

- Neubauer, Jonas, Hahn, Leo, Birtel, Johannes, Boon, Camiel J. F., Issa, Peter Charbel, and Fischer, M. Dominik (2022). “GUCY2D-Related Retinal Dystrophy with Autosomal Dominant Inheritance—A Multicenter Case Series and Review of Reported Data”. In: *Genes* 13.
- Nusinovici, Simon, Rim, Tyler Hyungtaek, Yu, Marco, Lee, Geunyoung, Tham, Yih-Chung, Cheung, Ning, Chong, Crystal, Soh, Zhi Da, Thakur, Sahil, Lee, Chan Joo, Sabanayagam, Charumathi, Lee, Byoung Kwon, Park, Sungha, Kim, Sung Soo, Kim, Hyeon Chang, Wong, Tien Yin, and Cheng, Ching-Yu (2022). “Retinal photograph-based deep learning predicts biological age, and stratifies morbidity and mortality risk”. In: *Age and Ageing* 51.
- Oleksyk, Taras K., C, Adam Gonçalo R. David M. Richard M. Gonçalo R. David R. Auton Abecasis Altshuler Durbin Abecasis Bentley, Auton, Adam, Abecasis, Gonçalo R., Altshuler, David M, Durbin, Richard, Bentley, D. R., Chakravarti, Aravinda, Clark, Andrew G., Donnelly, Peter, Eichler, Evan E., Flicek, Paul, Gabriel, S., Gibbs, Richard A., Green, Eric D., Hurles, Matthew E., Knoppers, Bartha Maria, Korbel, Jan O., Lander, Eric S., Lee, Charles, Lehrach, Hans, Mardis, Elaine R., Marth, Gabor T., McVean, Gil, Nickerson, Deborah A., Schmidt, Jeanette, Sherry, Stephen T., Wang, Jun, Wilson, Richard K., Barnes, Kathleen C., Beiswanger, Christine M, Burchard, Esteban González, Bustamante, Carlos D., Cai, Hongyu, Cao, Hongzhi, Gerry, Norman P., Gharani, Neda, Gignoux, Christopher R., Gravel, Simon, Henn, Brenna M., Jones, Danielle A., Jorde, Lynn B., Kaye, Jane, Keinan, Alon, Kent, Alastair, Kerasidou, Angeliki, Li, Yingrui, Mathias, Rasika A., Moreno-Estrada, Andrés, Ossorio, Pilar N., Parker, Michael, Resch, Alissa M., Rotimi, Charles N., Royal, Charmaine D.M., Sandoval, Karla, Su, Yeyang, Sudbrak, Ralf, Tian, Zhongming, Tishkoff, Sarah A., Toji, Lorraine H., Tyler-Smith, Chris, Via, Marc, Wang, Yuhong, Yang, Huanming, Yang, Ling, Zhu, Jia, Brooks, Lisa D., Felsenfeld, Adam L., Mcewen, J. E., Vaydylevich, Yekaterina, Duncanson, Audrey, Dunn, Michael, Schloss, Jeffery A., Garrison, Erik P., Kang, Hyun Min, Marchini, Jonathan, and McCarthy, Shane A. (2015). “A global reference for human genetic variation”. In: *Nature* 526, pp. 68–74.
- Park, Hae-Young Lopilly, Park, Young Gun, Cho, A Hyun, and Park, Chan Kee (2013). “Transneuronal retrograde degeneration of the retinal ganglion cells in patients with cerebral infarction.” In: *Ophthalmology* 120 6, pp. 1292–9.
- Parsons, Ruth E., Liu, Xiaonan, Collister, Jennifer A, Clifton, David A., Cairns, Benjamin J., and Clifton, Lei A. (2023). “Independent external validation of the QRISK3 cardiovascular disease risk prediction model using UK Biobank”. In: *Heart* 109, pp. 1690–1697.

- Patel, Praveen J., Foster, Paul J., Grossi, Carlota M., Keane, Pearse Andrew, Ko, Fang, Lotery, Andrew John, Peto, Tunde, Reisman, Charles A., Strouthidis, Nicholas G., and Yang, Qi (2016). “Spectral-Domain Optical Coherence Tomography Imaging in 67 321 Adults: Associations with Macular Thickness in the UK Biobank Study.” In: *Ophthalmology* 123 4, pp. 829–40.
- Poplin, Ryan, Varadarajan, Avinash V., Blumer, Katy, Liu, Yun, McConnell, Michael V., Corrado, Greg S, Peng, Lily H., and Webster, Dale R. (2017). “Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning”. In: *Nature Biomedical Engineering* 2, pp. 158–164.
- (2018). “Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning”. In: *Nature Biomedical Engineering* 2, pp. 158–164.
- Qian, Tianwei, Gong, Qiaoyun, Shen, Hangqi, Li, Caihua, Wang, Gao, Xu, Xun, Schrauwen, Isabelle, and Wang, Weijun (2022). “Novel variants in the RDH5 Gene in a Chinese Han family with fundus albipunctatus”. In: *BMC Ophthalmology* 22.
- Qu, Yiming, Zhuo, Yuanyuan, Lee, Jack, Huang, Xingxian, Yang, Zhuoxin, Yu, Haibo, Zhang, Jinwen, Yuan, Weiqu, Wu, Jiaman, Owens, David, and Zee, Benny Chung-ying (2022). “Ischemic and haemorrhagic stroke risk estimation using a machine-learning-based retinal image analysis”. In: *Frontiers in Neurology* 13.
- Rajiah, Prabhakar Shantha, MacNamara, James P., Chaturvedi, Abhishek, Ashwath, Ravi, Fulton, Nicholas, and Goerne, Harold (2019). “Bands in the Heart: Multimodality Imaging Review.” In: *Radiographics : a review publication of the Radiological Society of North America, Inc*, p. 180176.
- Ramsundar, Bharath and Zadeh, Reza Bosagh (2019). “TensorFlow for Deep Learning”. In.
- Richardson, Rose, Tracey-White, Dhani C., Webster, Andrew R., Webster, Andrew R., Moosajee, Mariya, and Moosajee, Mariya (2017). “The zebrafish eye—a paradigm for investigating human ocular genetics”. In: *Eye* 31, pp. 68–86.
- Rim, Tyler Hyungtaek, Lee, Chan Joo, Tham, Yih-Chung, Cheung, Ning, Yu, Marco, Lee, Geunyoung, Kim, Youngnam, Ting, Daniel Shu Wei, Chong, Crystal, Choi, Yoon Seong, Yoo, Tae Keun, Ryu, Ik Hee, Baik, Su Jung, Kim, Young Ah, Kim, Sung Kyu, Lee, Sang-Hak, Lee, Byoung-Kwon, Kang, Seok-Min, Wong, Edmund Yick Mun, Kim, Hyeon Chang, Kim, Sung Soo, Park, Sungha, Cheng, Ching-Yu, and Wong, Tien Yin (2021). “Deep-learning-based cardiovascular risk stratification using

- coronary artery calcium scores predicted from retinal photographs.” In: *The Lancet. Digital health* 3 5, e306–e316.
- Rim, Tyler Hyungtaek, Lee, Geunyoung, Kim, Youngnam, Tham, Yih-Chung, Lee, Chan Joo, Baik, Su Jung, Kim, Young Ah, Yu, Marco, Deshmukh, Mihir, Lee, Byoung-Kwon, Park, Sungha, Kim, Hyeon Chang, Sabayanagam, Charumathi, Ting, Daniel Shu Wei, Wang, Ya Xing, Jonas, Jost Bruno, Kim, Sung Soo, Wong, Tien Yin, and Cheng, Ching-Yu (2020). “Prediction of systemic biomarkers from retinal photographs: development and validation of deep-learning algorithms.” In: *The Lancet. Digital health* 2 10, e526–e536.
- Ronneberger, Olaf, Fischer, Philipp, and Brox, Thomas (2015). “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *ArXiv abs/1505.04597*.
- Ruder, Sebastian (2016). “An overview of gradient descent optimization algorithms”. In: *ArXiv abs/1609.04747*.
- Rudnicka, Alicja Regina, Welikala, Roshan Alex, Barman, Sarah Ann, Foster, Paul J., Luben, Robert N., Hayat, Shabina A, Khaw, Kay-Tee, Whincup, Peter H., Strachan, David P., and Owen, Chris G. (2022). “Artificial intelligence-enabled retinal vasculometry for prediction of circulatory mortality, myocardial infarction and stroke”. In: *The British Journal of Ophthalmology* 106, pp. 1722–1729.
- Rumelhart, David E, Hinton, Geoffrey E, and Williams, Ronald J (1986). “Learning representations by back-propagating errors”. In: *nature* 323.6088, pp. 533–536.
- Sabanayagam, Charumathi, Xu, Dejiang, Ting, Daniel Shu Wei, Nusinovici, Simon, Banu, Riswana, Hamzah, Haslina, Lim, Cynthia Ciwei, Tham, Yih-Chung, Cheung, Carol Yim-lui, Tai, E. Shyong, Wang, Ya Xing, Jonas, Jost B., Jonas, Jost B., Cheng, Ching-Yu, Lee, Mong Li, Hsu, Wynne, and Wong, Tien Yin (2020). “A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations.” In: *The Lancet. Digital health* 2 6, e295–e302.
- Saine, Patrick J. and Tyler, Marshall E. (2024). *Fundus Photography Overview*. 2nd. Butterworth-Heinemann Medical. ISBN: 0750673729.
- Schmidhuber, Jürgen (2014). “Deep learning in neural networks: An overview”. In: *Neural networks : the official journal of the International Neural Network Society* 61, pp. 85–117.

- Sergouniotis, Panagiotis, Diakite, Adam, Gaurav, Kumar, Eye, UK Biobank, Consortium, Vision, Birney, Ewan, and Fitzgerald, Tomas (2024). “Autoencoder-based phenotyping of ophthalmic images highlights genetic loci influencing retinal morphology and provides informative biomarkers”. In: *Bioinformatics* 41.
- Siddique, Nahian Alam, Paheding, Sidike, Elkin, Colin P., and Devabhaktuni, Vijay Kumar (2020). “U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications”. In: *IEEE Access* 9, pp. 82031–82057.
- Son, Jaemin, Shin, Joo Young, Chun, Eun Ju, Jung, Kyu-Hwan, Park, Kyu Hyung, and Park, Sang Jun (2020). “Predicting High Coronary Artery Calcium Score From Retinal Fundus Images With Deep Learning Algorithms”. In: *Translational Vision Science & Technology* 9.
- Stein, Daniel M, Ishikawa, Hiroshi, Hariprasad, Roopa, Wollstein, Gadi, Noecker, Robert J., Fujimoto, James G., and Schuman, Joel S. (2006). “A new quality assessment parameter for optical coherence tomography”. In: *British Journal of Ophthalmology* 90, pp. 186–190.
- Stranger, Barbara E., Nica, Alexandra C., Forrest, Matthew S., Dimas, Antigone S., Bird, Christine P., Beazley, Claude, Ingle, Catherine E., Dunning, Mark J., Flicek, Paul, Koller, Daphne, Montgomery, Stephen B., Tavaré, Simon, Deloukas, Panos, and Dermitzakis, Emmanouil T. (2007). “Population genomics of human gene expression”. In: *Nature Genetics* 39, pp. 1217–1224.
- Sudlow, Cathie L. M., Gallacher, John E.J., Allen, Naomi E., Beral, Valerie, Burton, Paul, Danesh, John, Downey, Paul, Elliott, Paul, Green, Jane, Landray, Martin J, Liu, Bette, Matthews, Paul M., Ong, Giok J., Pell, Jill P., Silman, Alan J, Young, Alan, Sprosen, Tim, Peakman, Tim C, and Collins, Rory (2015). “UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age”. In: *PLoS Medicine* 12.
- Tomasoni, Mattia, Beyeler, Michael Johannes, Vela, Sofia Ortin, Mounier, Ninon, Porcu, Eleonora, Corre, Tanguy, Krefl, Daniel, Button, Alexander Luke, Abouzeid, Hana, Lazaros, K, Bochud, Murielle, Schlingemann, Reinier O., Bergin, Ciara, and Bergmann, S. (2023). “Genome-wide Association Studies of Retinal Vessel Tortuosity Identify Numerous Novel Loci Revealing Genes and Pathways Associated With Ocular and Cardiometabolic Diseases”. In: *Ophthalmology Science* 3.

- Uffelmann, Emil, Huang, Qin Qin, Munung, Nchangwi Syntia, Vries, Jantina de, Okada, Yukinori, Martin, Alicia R., Martin, Hilary C., Lappalainen, Tuuli, and Posthuma, Danielle (2021). “Genome-wide association studies”. In: *Nature Reviews Methods Primers* 1.
- UK Biobank (2024). *UK Biobank - UK Biobank*. <https://www.ukbiobank.ac.uk>. Accessed on September 24, 2024. URL: <https://www.ukbiobank.ac.uk>.
- Vaghefi, Ehsan, Yang, Song, Hill, Sophie, Humphrey, Gayl, Walker, Natalie, and Squirrell, David M. (2019). “Detection of smoking status from retinal images; a Convolutional Neural Network study”. In: *Scientific Reports* 9.
- Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N, Kaiser, Łukasz, and Polosukhin, Illia (2017). “Attention is all you need”. In: *Advances in neural information processing systems*, pp. 5998–6008.
- Vela, S. Ortin, Beyeler, Morin, Trofimova, Olga, Tomasoni, Mattia, Iuliani, Ilaria, Presby, D., Hoogewoud, Florence, and Bergmann, S. (2023). “Phenotypic and Genetic Characteristics of Retinal Vascular Parameters and their Association with Diseases”. In: *medRxiv*.
- Veluchamy, Abirami, Ballerini, Lucia, Vitart, Veronique, Schraut, Katharina E., Kirin, Mirna Abaffy, Campbell, Harry, Joshi, Peter K., Relan, Devanjali, Harris, Sarah E., Brown, Ellie, Vaidya, Suraj S., Dhillon, Baljean, Zhou, Kaixin, Pearson, Ewan R., Hayward, Caroline, Polaek, Ozren, Deary, Ian J., MacGillivray, Thomas J., Wilson, James F., Trucco, Emanuele, Palmer, Colin N. A., and Doney, Alex S. F. (2017). “Novel Genetic Locus Influencing Retinal Venular Tortuosity Is Also Associated With Risk of Coronary Artery Disease”. In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 39, pp. 2542–2552.
- Vincent, Pascal, Larochelle, Hugo, Bengio, Yoshua, and Manzagol, Pierre-Antoine (2008). “Extracting and composing robust features with denoising autoencoders”. In: *Proceedings of the 25th international conference on Machine learning*, pp. 1096–1103.
- Visscher, Peter M., Wray, Naomi R., Zhang, Qian, Sklar, Pamela, McCarthy, Mark I., Brown, Matthew A., and Yang, Jian (2017). “10 Years of GWAS Discovery: Biology, Function, and Translation.” In: *American journal of human genetics* 101 1, pp. 5–22.

- Voulodimos, Athanasios, Doulamis, Nikolaos D., Doulamis, Anastasios D., and Protopapadakis, Eftychios E. (2018). “Deep Learning for Computer Vision: A Brief Review”. In: *Computational Intelligence and Neuroscience* 2018.
- Wagner, Siegfried Karl, Fu, Dun Jack, Faes, Livia, Liu, Xiaoxuan, Huemer, Josef, Khalid, Hagar, Ferraz, Daniel Araújo, Korot, Edward, Kelly, Christopher J., Balaskas, Konstantinos, Denniston, Alastair K. O., and Keane, Pearse A. (2020). “Insights into Systemic Disease through Retinal Imaging-Based Oculomics”. In: *Translational Vision Science & Technology* 9.
- Wang, Jun, Lin, Min, Crenshaw, Andrew, Hutchinson, Amy, Hicks, Belynda D., Yeager, Meredith, Berndt, Sonja, Huang, Wen-Yi, Hayes, Richard B., Chanock, Stephen J, Jones, Robert C, and Ramakrishnan, Ramesh (2009). “High-throughput single nucleotide polymorphism genotyping using nanofluidic Dynamic Arrays”. In: *BMC Genomics* 10, pp. 561–561.
- Waring, Jonathan, Lindvall, Charlotta, and Umeton, Renato (2020). “Automated machine learning: Review of the state-of-the-art and opportunities for healthcare”. In: *Artificial intelligence in medicine* 104, p. 101822.
- Warwick, Alasdair N., Curran, Katie, Hamill, Barbra, Stuart, Kelsey Vernon, Khawaja, Anthony P., Foster, Paul J., Lotery, Andrew J., Quinn, Michael, Madhusudhan, Savita, Balaskas, Konstantinos, Peto, Tunde, Brait, N. T. D. S. J. P. G. T. R. U. M. S. A. P. B. A. A. C. Allen Aslam Atan Barman Barrett Bishop Black, Allen, Naomi E., Aslam, T., Atan, Denize, Barman, Sarah Ann, Barrett, Jenny H., Bishop, Paul, Black, Gregory, Braithwaite, Thomas, Carare, Roxana O., Chakravarthy, Usha, Chan, Michelle P Y, Chua, Sharon Y. L., Day, Alexander, Desai, Parul, Dhillon, Bal, Dick, Andrew D., Doney, Alex, Egan, Catherine A., Ennis, S., Foster, Paul J., Fruttiger, Marcus, Gallacher, James B., Garway-Heath, David, Gibson, J M, Guggenheim, Jeremy A., Hammond, Chris J., Hardcastle, Alison, Harding, Sp, Hogg, Ruth, Hysi, P G, Keane, P. A., Khaw, P. T., Khawaja, Anthony P., Lascaratos, Gerassimos, Littlejohns, Thomas J., Lotery, Andrew J., Luthert, Philip J., MacGillivray, Thomas J., Mackie, Sarah, Mcguinness, B., Mckay, G., McKibbin, Mary Adeline, Moore, Tony, Morgan, Jonathan H., Oram, Risa Grace, O’sullivan, E., Owen, Chris G., Patel, Praveen J., Paterson, Euan Neil, Peto, Tunde, Petzold, Andreas, Pontikos, Nikolas, Rahi, Jugnoo S., Rudnicka, A. R., Sattar, Nikhat, Self, J. E., Sergouniotis, Panagiotis, Sivaprasad, S, Steel, David, Stratton, Ira W., Strouthidis, Nicholas G., Sudlow, C. L., Sun, Z., Tapp, Roland W., Thomas, Dhanes, Trucco, Emmanuel, Tuffail, A., Viswanathan, A., Vitart, Veronique, Weedon, Mj, Williams, K., Williams, Citra, Woodside,

- John M., Yates, Max M., Yip, Jennifer Lai Yee, and Zheng, Y. (2022). “UK Biobank retinal imaging grading: methodology, baseline characteristics and findings for common ocular diseases”. In: *Eye* 37, pp. 2109–2116.
- Wong, Dragon Y.L., Lam, Mary C., Ran, An Ran, and Cheung, Carol Yim-lui (2022). “Artificial intelligence in retinal imaging for cardiovascular disease prediction: current trends and future directions”. In: *Current Opinion in Ophthalmology* 33, pp. 440–446.
- Wu, Zehua, Meng, Delong, Fang, Chengbo, Li, Jian, Zheng, Xiujie, Lin, Jiansuo, Zeng, Haijiang, Lv, Sihan, Zhang, Zhen-Ning, Luan, Bing, Zhong, Zi-lin, and Chen, Jianjun (2019). “PITX3 mutations associated with autosomal dominant congenital cataract in the Chinese population”. In: *Molecular Medicine Reports* 19, pp. 3123–3131.
- Yang, Li, Wu, Haibin, Jin, Xiaoqing, Zheng, Pinpin, Hu, Shiyun, Xu, Xiaoling, Yu, Wei, and Yan, Jing (2020). “Study of cardiovascular disease prediction model based on random forest in eastern China”. In: *Scientific reports* 10.1, p. 5245.
- Yeung, Shanna C., You, Yuyi, Howe, Kathryn L., and Yan, Peng (2020). “CHOROIDAL THICKNESS IN PATIENTS WITH CARDIOVASCULAR DISEASE: A REVIEW.” In: *Survey of ophthalmology*.
- Zee, Benny C Y, Wong, Yanny Wing Yan, Lee, Jack, Fan, Yuhua, Zeng, Jinsheng, Lam, Bonnie Yin Ka, Wong, Adrian, Shi, Lin, Lee, Allen Ting Chun, Kwok, Chloe, Lai, Maria, Mok, Vincent Chung-Tong, and Lau, Alexander Yuk-lun (2021). “Machine-learning method for localization of cerebral white matter hyperintensities in healthy adults based on retinal images”. In: *Brain Communications* 3.
- Zekavat, Seyedeh Maryam and et.al (2023). “Insights into human health from phenome- and genome-wide analyses of UK Biobank retinal optical coherence tomography phenotypes”. In: *medRxiv*.
- Zekavat, Seyedeh Maryam, Raghu, Vineet K., Trinder, Mark, Ye, Yixuan, Koyama, Satoshi, Honigberg, Michael C., Yu, Zhi, Pampana, Akhil, Urbut, Sarah, Haidermota, Sara, O’Regan, Declan P., Zhao, Hongyu, Ellinor, Patrick T., Segrè, Ayellet V., Elze, Tobias, Wiggs, Janey L., Martone, James, Adelman, Ron A., Zebardast, Nazlee, Priore, Lucian Del, Wang, Jay C., and Natarajan, Pradeep (2021). “Deep Learning of the Retina Enables Phenome- and Genome-Wide Analyses of the Microvasculature”. In: *Circulation* 145, pp. 134–150.

- Zhang, Kang, Liu, Xiaohong, Xu, Jie, Yuan, Jin, Cai, Wenjia, Chen, Ting, Wang, Kai, Gao, Yuanxu, Nie, Sheng, Xu, Xiaodong, Qin, Xiaoqi, Su, Yuandong, Xu, Wenqin, Olvera, Andrea, Xue, Kanmin, Li, Zhihuan, Zhang, Meixia, Zeng, Xiaoxi, Zhang, Charlotte L., Li, Oulan, Zhang, Edward, Zhu, Jie, Xu, Yiming, Kermany, Daniel S., Zhou, Kaixin, Pan, Ying, Li, Shaoyun, Lai, Iat Fan, Chi, Ying, Wang, Changuang, Pei, Michelle, Zang, Guangxi, Zhang, Qi, Lau, Johnson Yiu-Nam, Lam, Dennis S. C., Zou, Xiaoguang, Wumaier, Aizezi, Wang, Jianquan, Shen, Yin, Hou, Fan Fan, Zhang, Ping, Xu, Tao, Zhou, Yong, and Wang, Guangyu (2021). “Deep-learning models for the detection and incidence prediction of chronic kidney disease and type 2 diabetes from retinal fundus images”. In: *Nature Biomedical Engineering* 5, pp. 533–545.
- Zhang, Li, Yuan, Meng, An, Zhen, Zhao, Xiangmei, Wu, Hui, Li, Haibin, Wang, Ya, Sun, Beibei, Li, Huijun, Ding, Shibin, Zeng, Xiang, Chao, Ling, Li, Pan, and Wu, Weidong (2020). “Prediction of hypertension, hyperglycemia and dyslipidemia from retinal fundus photographs via deep learning: A cross-sectional study of chronic diseases in central China”. In: *PLoS ONE* 15.
- Zhang, Shuhe, Webers, Carroll A B, and Berendschot, Tos T.J.M. (2024). “Computational single fundus image restoration techniques: a review”. In: *Frontiers in Ophthalmology* 4.
- Zhong, Pingting, Qin, Jie, Li, Zhixi, Jiang, Lei, Peng, Qingsheng, Huang, Manqing, Lin, Ying-wen, Liu, Baoyi, Li, Cong, Wu, Qiaowei, Kuang, Yu, Cui, Shirong, Yu, Honghua, Liu, Zaiyi, and Yang, Xiaohong (2021). “Development and Validation of Retinal Vasculature Nomogram in Suspected Angina Due to Coronary Artery Disease”. In: *Journal of Atherosclerosis and Thrombosis* 29, pp. 579–596.
- Zhou, Junlan, Zhu, Yan, Cheng, Min, Dinesh, Deepika, Thorne, Tina M., Poh, Kian Keong, Liu, Dongxu, Botros, Chantal, Tang, Yao Liang, Reisdorph, Nichole A., Kishore, Raj, Losordo, Douglas, and Qin, Gangjian (2009). “Regulation of Vascular Contractility and Blood Pressure by the E2F2 Transcription Factor”. In: *Circulation* 120, pp. 1213–1221.
- Zhou, Yukun, Chia, Mark A, Wagner, Siegfried K., Ayhan, Murat S., Williamson, Dominic J, Struyven, Robbert R., Liu, Timing, Xu, Moucheng, Lozano, Mateo G., Woodward-Court, Peter, Kihara, Yuka, Ser, Naomi John E. J. Thomas Tariq Paul Graeme Panagiotis Den Allen Gallacher Littlejohns Aslam Bishop Black, Allen, Naomi, Gallacher, John E. J., Littlejohns, Thomas, Aslam, Tariq, Bishop, Paul, Black, Graeme, Sergouniotis, Panagiotis I., Atan, Denize, Dick, Andrew D., Williams, Cathy, Barman, Sarah Ann, Barrett, Jenny H., Mackie, Sarah, Braithwaite, Tasanee, Carare, Roxana O., En-

- nis, Sarah, Gibson, Jane, Lotery, Andrew J., Self, Jay, Chakravarthy, Usha, Hogg, Ruth E., Paterson, Euan N., Woodside, Jayne, Peto, Tunde, McKay, Gareth J., McGuinness, Bernadette, Foster, Paul J., Balaskas, Konstantinos, Khawaja, Anthony P., Pontikos, Nikolas, Rahi, Jugnoo S, Lascaratos, Gerassimos, Patel, Praveen J., Chan, Michelle, Chua, Sharon Y. L., Day, Alexander, Desai, Parul, Egan, Catherine, Fruttiger, Marcus, Garway-Heath, David F., Hardcastle, Alison, Khaw, Sir Peng Tee, Moore, Tony, Sivaprasad, Sobha, Strouthidis, Nicholas G., Thomas, Dhanes, Tufail, Adnan, Viswanathan, Ananth C., Dhillon, Bal, Macgillivray, Tom, Sudlow, Cathie, Vitart, Veronique, Doney, Alexander, Trucco, Emanuele, Guggeenheim, Jeremy A., Morgan, James, Hammond, Chris J., Williams, Katie M., Hysi, Pirro G., Harding, Simon, Zheng, Yalin, Luben, Robert, Luthert, Phil, Sun, Zihan, Mckibbin, Martin, O'Sullivan, Eoin, Oram, Richard, Weedon, Mike, Owen, Chris G., Rudnicka, Alicja, Sattar, Naveed, Steel, David, Stratton, Irene, Tapp, Robyn, Yates, Max M., Petzold, Axel, Madhusudhan, Savita, Altmann, André, Lee, Aaron Y., Topol, Eric J., Denniston, Alastair Keith, Alexander, Daniel C., and Keane, Pearse Andrew (2023). "A foundation model for generalizable disease detection from retinal images". In: *Nature* 622, pp. 156–163.
- Zhu, Zhuoting, Chen, Yifan, Wang, Wei, Wang, Yueye, Hu, Wenyi, Shang, Xianwen, Liao, Huan, Shi, Danli, Huang, Yu, Ha, Jason, Tan, Zachary, Kiburg, Katerina V., Zhang, Xueli, Tang, Shulin, Yu, Honghua, Yang, Xiaohong, and He, Mingguang (2022). "Association of Retinal Age Gap With Arterial Stiffness and Incident Cardiovascular Disease". In: *Stroke* 53, pp. 3320–3328.
- Zhu, Zhuoting, Shi, Danli, Guankai, Peng, Tan, Zachary, Shang, Xuedong, Hu, Wenyi, Liao, Huan, Zhang, Xueli, Huang, Yu, Yu, Honghua, Meng, Wei, Wang, Wei, Ge, Zongyuan, Yang, Xiaohong, and He, Mingguang (2022). "Retinal age gap as a predictive biomarker for mortality risk". In: *British Journal of Ophthalmology* 107, pp. 547–554.

# Appendix A

## Supplementary information to Chapter 4

### A.1 Supplementary information to 4.3

Hyperparameter	Tested Values	Optimal Value
CNN channels	[128, 256, 128] [128, 256, 128, 64] [128, 256, 128, 128] [128, 256, 128, 128, 64] [128, 256, 128, 128, 64, 64] [128, 256, 128, 128, 64, 32]	[128, 256, 128, 128, 64, 64]
Batch size	4, 8, 16, 32	8
Learning rate	Log-scale: $10^{-1}$ to $10^{-6}$ ; Multiples: 0.3, 0.5, 0.7, 0.9	0.0001
Weight decay	Log-scale: $10^{-1}$ to $10^{-6}$ ; Multiples: 0.3, 0.5, 0.7, 0.9	0.0001
Latent space size	128, 256, 512, 1024	128
Train/Val/Test split ratio	4:3:3, 5:3:2, 5:2:3, 6:2:2	6:2:2

Table A.1: Hyperparameter grid search for the VAE model. Tested values include architectural choices (CNN channels, latent space size), optimization parameters (learning rate, weight decay), and data splits. The optimal configuration was selected via 5 cross-validation.

Table A.2 shows the architectural design of the VAE used in our model, detailing the encoder and decoder components. It presents the structure of each network, including the layer types, configurations, and dimensional transformations that occur between layers. The encoder extracts features from the input RGB image, while the decoder reconstructs the image from the latent variables. The training of our model occurred in two phases, initially engaging in self-supervised learning for each eye independently

<b>Encoder</b>	<b>Decoder</b>
Input: $128 \times 224 \times 224$ (RGB image)	Latent variables, Fully Connected, Reshape
Conv2D(128, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(128) ReLU	Conv2D(64, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(64) ReLU
Conv2D(256, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(256) ReLU	Conv2D(64, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(64) ReLU
Conv2D(128, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(128) ReLU	Conv2D(128, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(128) ReLU
Conv2D(128, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(128) ReLU	Conv2D(128, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(128) ReLU
Conv2D(64, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(64) ReLU	Conv2D(256, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(256) ReLU
Conv2D(64, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(64) ReLU	Conv2D(128, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(128) ReLU
Latent variables, Fully Connected	Output: $128 \times 224 \times 224$ (Reconstructed RGB image)

Table A.2: Architectural design of the VAE, showing the encoder and decoder structures with their respective layers, configurations, and dimensional transformations.

to acquire latent representations, which were then utilised to initialise the subsequent fine-tuning phase.

Hyperparameter	Tested Values	Optimal Value
Random state	0,1,2	BE-MTDT-RF: 1 BE-RF: 1 LE-MTDT-RF:2 LE-RF: 1 RE-MTDT-RF:0 RE-RF: 2 MTDT-RF: 1
Estimators	100–1000 (step 50)	BE-MTDT-RF: 300 BE-RF: 100 LE-MTDT-RF: 500 LE-RF: 500 RE-MTDT-RF: 700 RE-RF: 100 MTDT-RF: 400
Min samples split	10–500 (step 10)	BE-MTDT-RF: 210 BE-RF: 410 LE-MTDT-RF: 160 LE-RF: 360 RE-MTDT-RF: 260 RE-RF: 460 MTDT-RF: 360
Min samples leaf	10–500 (step 10)	BE-MTDT-RF: 60 BE-RF: 60 LE-MTDT-RF: 260 LE-RF: 10 RE-MTDT-RF: 10 RE-RF: 10 MTDT-RF: 260
Max features	0.25, 0.5, 0.75	BE-MTDT-RF: 0.25 BE-RF: 0.25 LE-MTDT-RF: 0.25 LE-RF: 0.25 RE-MTDT-RF: 0.25 RE-RF: 0.75 MTDT-RF: 0.5
Max depth	1, 2	BE-MTDT-RF: 1 BE-RF: 2 LE-MTDT-RF: 1 LE-RF: 1 RE-MTDT-RF: 1 RE-RF: 2 MTDT-RF: 1
Criterion	gini, entropy	BE-MTDT-RF: gini BE-RF: gini LE-MTDT-RF: entropy LE-RF: entropy RE-MTDT-RF: entropy RE-RF: entropy MTDT-RF: gini

Table A.3: Hyperparameter grid search for RF models. Tested ranges and optimal values are shown for each variant (BE/LE/RE-MTDT-RF, BE/LE/RE-RF, and MTDT-RF). Step sizes are indicated in parentheses for ranged parameters.

Hyperparameter	Tested Values	Optimal Value
number channels of CNN	[1] [8, 1] [16, 1] [32, 1] [64, 1] [32, 16, 1]	[16, 1]
batch size	4,8,16,32	8
learning rate	Log-scale: $10^{-1}$ to $10^{-6}$ ; Multiples: 0.3, 0.5, 0.7, 0.9	0.00001
Weight decay	Log-scale: $10^{-1}$ to $10^{-6}$ ; Multiples: 0.3, 0.5, 0.7, 0.9	0.000001
Ratio splits	4:3:3, 5:3:2, 5:2:3, 6:2:2	5:3:2

Table A.4: Hyperparameter grid search for the MLP model. Tested values include architectural choices (CNN channels), optimization parameters (learning rate, weight decay), and data splits. The optimal configuration was selected via 5 cross-validation.

<b>Variables used in QRISK3</b>
Gender
Age
Atrial fibrillation
Atypical antipsy
Regular steroid tablets
Erectile disfunction
Migraine
Rheumatoid arthritis
Chronic kidney disease
Severe mental illness
Systemic lupus erythematosis
Blood pressure treatment
Diabetes1
Diabetes2
Weight
Height
Ethnicity
Heart attack relative
Cholesterol HDL ratio
Systolic blood pressure
Std systolic blood pressure
Smoke
Townsend

Table A.5: List of Variables used to calculate QRISK3

## A.2 Supplementary information to 4.5

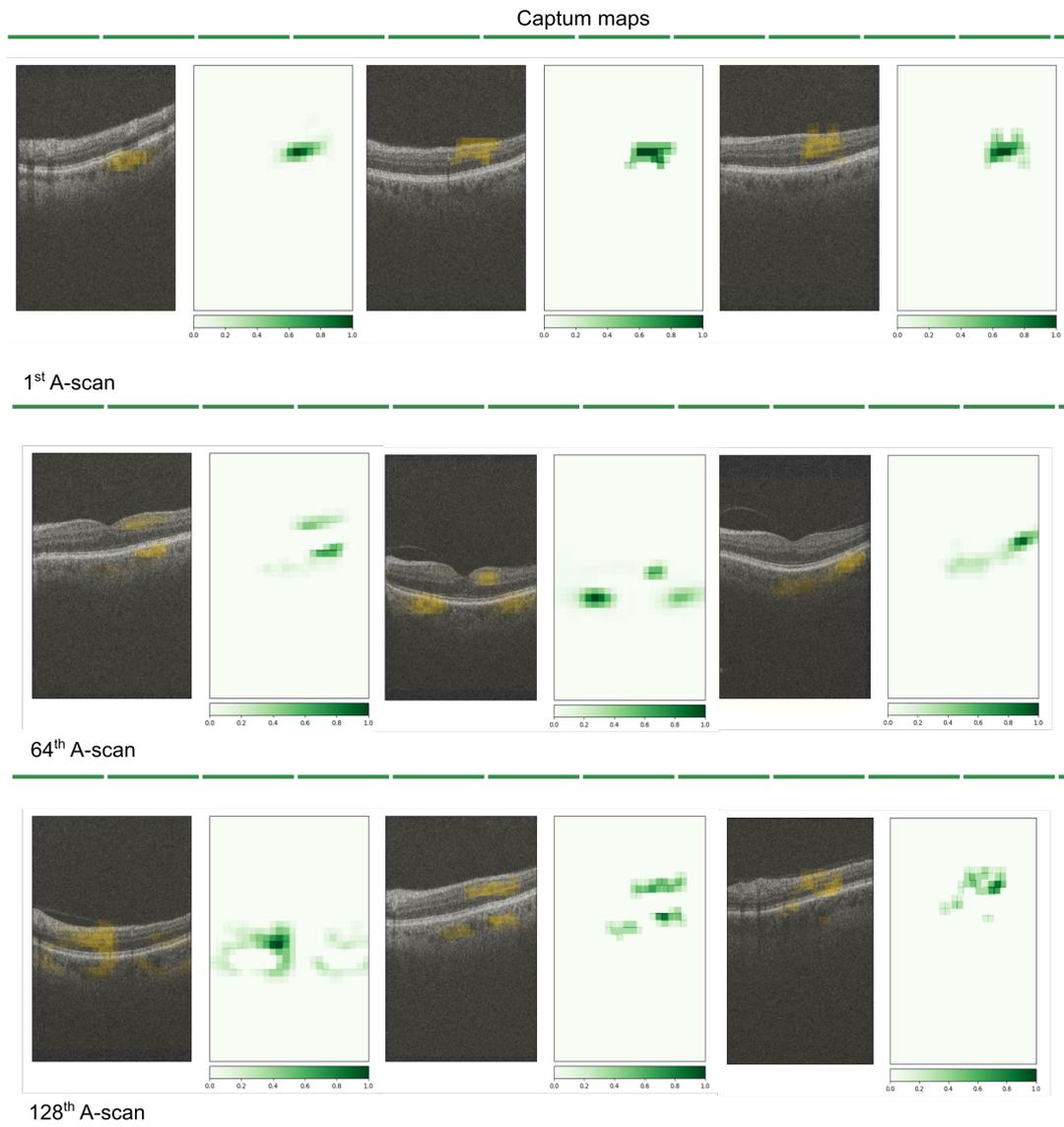


Figure A.1: Captum images present three different B-scans for the left eye. The top row corresponds to the 1st B-scan, the middle image to the 64th B-scan, and the bottom row depicts the final B-scan. The yellow zones highlight regions of the oblation method when modifying the latent variable  $z_{l066}$ .

## A.3 Links to Docker image and Github

The Docker image containing the computational environment and dependencies used for the experiments in Chapter 4 is available at: [Docker/scclmg/OCT-VAE-RF](https://hub.docker.com/r/scclmg/OCT-VAE-RF)

The code used in this study in Chapter 4 is available on GitHub at the following address: [Github/Predicting-CVD-using-OCT-images](https://github.com/Scclmg/Predicting-CVD-using-OCT-images)

## Appendix B

# Supplementary information to Chapter 5

### B.1 Supplementary information to 5.2

The provided heatmaps (refer to Figures B.1 and B.2) illustrate the distribution of patients with various CVDs across two different diagnostic imaging methods: Fundus and OCT. Each heatmap visually represents the number of patients diagnosed with different CVDs, where the elements along the diagonal indicate the number of patients with a given diagnosis and the off-diagonal elements represent the number of patients with co-occurrence of pairs of diagnoses.

**Fundus (Figure B.1)** : The highest concentration of stroke patients is observed, with a count of 690. MI has the highest number of patients, totaling 1000. It also shows some overlap with Stroke (64), Angina (150), and Heart Failure (78). There are 530 patients with Angina, with notable overlaps with MI (150) and Heart Failure (39). A significant number of patients (320) are reported with Heart Failure (HF), with minor overlaps with other CVDs. Very few patients fall into rheumatic heart, with minimal overlap with other diseases.

**OCT (Figure B.2)** : The highest concentration is 930 for Stroke, with minor overlaps with MI (94) and other CVDs. MI also shows the highest number of patients (1400), with overlaps in Angina (230) and HF (110). There are 740 patients with Angina, with significant overlaps with MI (230) and Heart Failure (66). A considerable number of patients (460) are recorded with HF, with overlaps in Angina (66) and Rheumatic Heart (32). Similar to the Fundus heatmap, RH shows a small number of patients, with minor overlaps with other diseases.

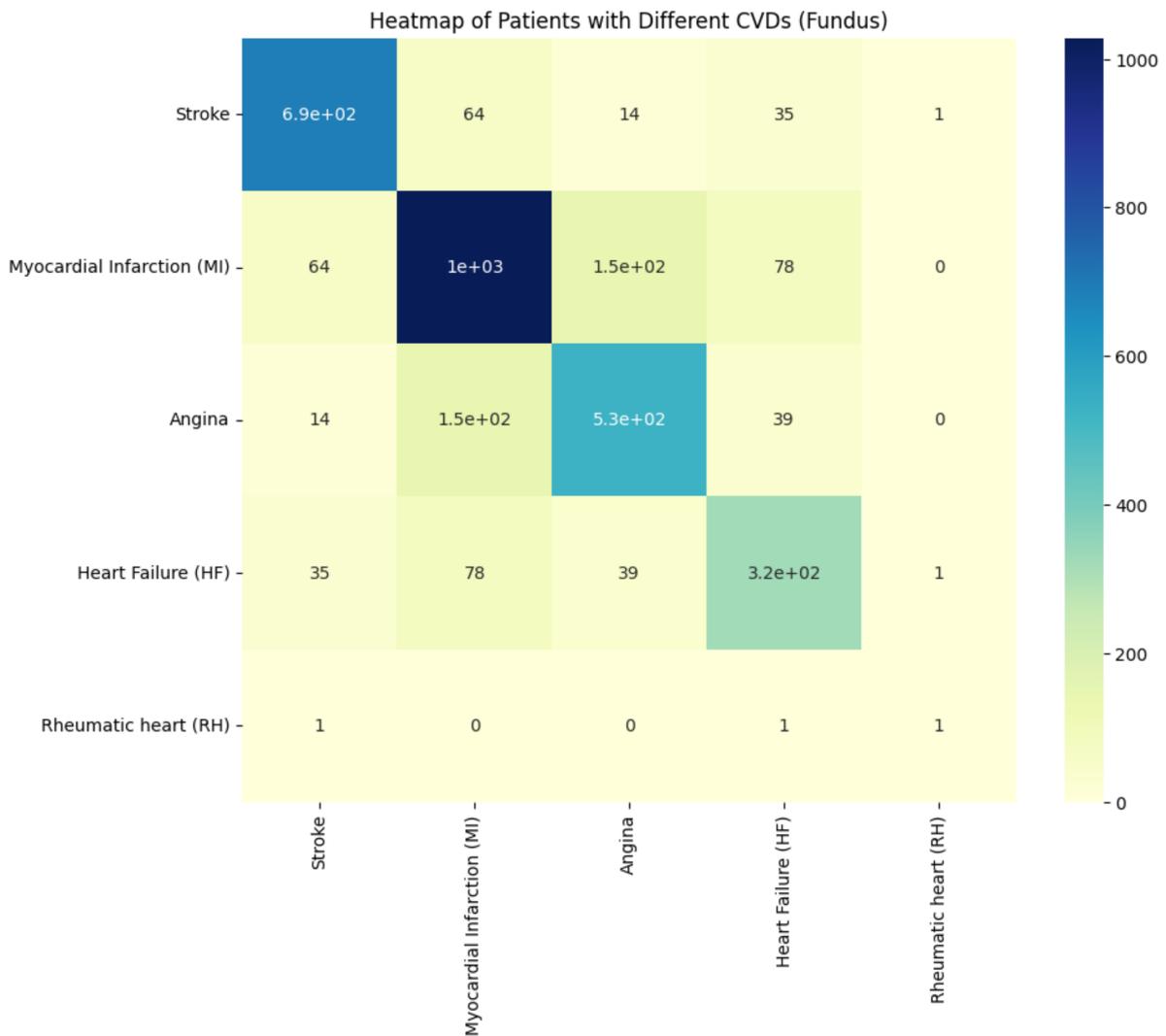


Figure B.1: Heatmaps indicating the number of patients with different CVDs and fundus photographs, where the diagonal represents the number of patients with a given diagnosis and the off-diagonal elements indicate co-occurrence of pairs of diagnoses.



Figure B.2: Heatmaps indicating the number of patients with different CVDs and OCT, where the diagonal represents the number of patients with a given diagnosis and the off-diagonal elements indicate co-occurrence of pairs of diagnoses.

## B.2 Supplementary information to 5.3

Hyperparameter	Tested Values	Optimal Value
CNN channels	Fundus: [3, 128, 256, 512] [3, 128, 256, 512, 1024] [3, 128, 256, 512, 1024, 2048] OCT: [1, 32, 64] [1, 32, 64, 128] [1, 32, 64, 128, 256] [1, 32, 64, 128, 256, 512]	Fundus: [3, 128, 256, 512, 1024, 2048] OCT: [1, 32, 64, 128, 256, 512]
Batch size	4, 8, 16, 32	Fundus: 2 OCT: 4 Fundus-OCT: 4
Learning rate	Log-scale: $10^{-1}$ to $10^{-6}$ Multiples: 0.3, 0.5, 0.7, 0.9	Fundus: 0.001 OCT: 0.0001 Fundus-OCT: 0.0001
Weight decay	Log-scale: $10^{-1}$ to $10^{-6}$ Multiples: 0.3, 0.5, 0.7, 0.9	Fundus: 0.0000001 OCT: 0.000001 Fundus-OCT: 0.000001
Latent space size	128, 256, 512, 1024	Fundus: 2048 OCT: 512
Train/Val/Test split ratio	4:3:3, 5:3:2, 5:2:3, 6:2:2	Fundus: 5:2:3 OCT: 5:2:3 Fundus-OCT: 5:2:3

Table B.1: Hyperparameter grid search for the MCVAE model. Tested values include architectural choices (CNN channels, latent space size), optimization parameters (learning rate, weight decay), and data splits. The optimal configuration was selected via 5-fold cross-validation.

<b>Encoder</b>	<b>Decoder</b>
Input: $3 \times 224 \times 224$ (RGB image)	Latent variables, Fully Connected, Reshape
Conv2D(3, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(3) ReLU	Conv2D(2048, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(2048) ReLU
Conv2D(128, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(128) ReLU	Conv2D(1024, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(1024) ReLU
Conv2D(256, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(256) ReLU	Conv2D(512, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(512) ReLU
Conv2D(512, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(512) ReLU	Conv2D(256, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(256) ReLU
Conv2D(1024, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(1024) ReLU	Conv2D(128, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(128) ReLU
Conv2D(2048, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(2048) ReLU	Conv2D(3, $3 \times 3$ , stride= $2 \times 2$ , padding= $1 \times 1$ ) BatchNorm2d(3) ReLU
Latent variables, Fully Connected	Output: $3 \times 224 \times 224$ (Reconstructed RGB image)

Table B.2: Architectural design of the VAE for fundus images, showing the encoder and decoder structures with their respective layers, configurations, and dimensional transformations.

Encoder	Decoder
Input: $1 \times 128 \times 224 \times 224$ (RGB image)	Latent variables, Fully Connected, Reshape
Conv3D(1, $3 \times 3 \times 3$ , stride= $2 \times 2 \times 2$ , padding= $1 \times 1 \times 1$ ) BatchNorm3d(1) ReLU	Conv3D(128, $3 \times 3 \times 3$ , stride= $2 \times 2 \times 2$ , padding= $1 \times 1 \times 1$ ) BatchNorm3d(128) ReLU
Conv3D(32, $3 \times 3 \times 3$ , stride= $2 \times 2 \times 2$ , padding= $1 \times 1 \times 1$ ) BatchNorm3d(32) ReLU	Conv3D(64, $3 \times 3 \times 3$ , stride= $2 \times 2 \times 2$ , padding= $1 \times 1 \times 1$ ) BatchNorm3d(64) ReLU
Conv3D(64, $3 \times 3 \times 3$ , stride= $2 \times 2 \times 2$ , padding= $1 \times 1 \times 1$ ) BatchNorm3d(64) ReLU	Conv3D(32, $3 \times 3 \times 3$ , stride= $2 \times 2 \times 2$ , padding= $1 \times 1 \times 1$ ) BatchNorm3d(32) ReLU
Conv3D(128, $3 \times 3 \times 3$ , stride= $2 \times 2 \times 2$ , padding= $1 \times 1 \times 1$ ) BatchNorm3d(128) ReLU	Conv2D(1, $3 \times 3 \times 3$ , stride= $2 \times 2 \times 2$ , padding= $1 \times 1 \times 1$ ) BatchNorm3d(1) ReLU
Latent variables, Fully Connected	Output: $1 \times 128 \times 224 \times 224$ (Reconstructed RGB image)

Table B.3: Architectural design of the VAE for OCT images, showing the encoder and decoder structures with their respective layers, configurations, and dimensional transformations.

### B.3 Links to Docker image and Github

The Docker image containing the computational environment and dependencies used for the experiments in Chapter 5 is available at: [Docker/scclmg/OCT-MCVAE](https://github.com/scclmg/OCT-MCVAE)

The code used in this study in Chapter 5 is available on GitHub at the following address: [Github/Multimodal-Retina](https://github.com/Multimodal-Retina)

# Appendix C

## Supplementary information to Chapter 6

### C.1 Supplementary information to 6.2

Hyperparameter	Tested Values	Optimal Value
CNN channels	[10,32] [10, 32, 64] [10, 32, 64, 128]	[10, 32, 64]
Batch size	4, 8, 16, 32	8
Learning rate	Log-scale: $10^{-1}$ to $10^{-6}$ Multiples: 0.3, 0.5, 0.7, 0.9	0.0001
Weight decay	Log-scale: $10^{-1}$ to $10^{-6}$ Multiples: 0.3, 0.5, 0.7, 0.9	0.00001
Latent space size	32,64,128	64
Train/Val/Test split ratio	4:3:3, 5:3:2, 6:2:2, 7:2:1	7:2:1

Table C.1: Hyperparameter grid search for the VAE model. Tested values include architectural choices (CNN channels, latent space size), optimization parameters (learning rate, weight decay), and data splits. The optimal configuration was selected via 5-fold cross-validation.

### C.2 Supplementary information to 6.3

#### C.2.1 GWAS databases links used as exposure data in the Mendelian randomization Analysis

Coronary Artery Disease: Cardiogramplus

Body fat percentage: GCST003435

Encoder	Decoder
Input: $10 \times 128 \times 512$	Latent variables, Fully Connected, Reshape
Conv3D(10, $3 \times 3 \times 3$ , stride= $2 \times 2 \times 2$ , padding= $1 \times 1 \times 1$ ) BatchNorm3d(10) ReLU	Conv3D(64, $3 \times 3 \times 3$ , stride= $2 \times 2 \times 2$ , padding= $1 \times 1 \times 1$ ) BatchNorm3d(64) ReLU
Conv3D(32, $3 \times 3 \times 3$ , stride= $2 \times 2 \times 2$ , padding= $1 \times 1 \times 1$ ) BatchNorm3d(32) ReLU	Conv3D(32, $3 \times 3 \times 3$ , stride= $2 \times 2 \times 2$ , padding= $1 \times 1 \times 1$ ) BatchNorm3d(32) ReLU
Conv3D(64, $3 \times 3 \times 3$ , stride= $2 \times 2 \times 2$ , padding= $1 \times 1 \times 1$ ) BatchNorm3d(64) ReLU	Conv3D(10, $3 \times 3 \times 3$ , stride= $2 \times 2 \times 2$ , padding= $1 \times 1 \times 1$ ) BatchNorm3d(10) ReLU
Latent variables, Fully Connected	Output: $1 \times 128 \times 512$ (Reconstructed RGB image)

Table C.2: Architectural design of the VAE for thickness maps of OCT retinal layers, showing the encoder and decoder structures with their respective layers, configurations, and dimensional transformations.

Stroke: GCST90104534

Triglycerides: GCST002216

Myocardial Infarction: FinnMI

Diabetes: FinnDiabetes

### C.3 Links to Docker image and Github

The Docker image containing the computational environment and dependencies used for the nn-UNet experiments in Chapter 6 is available at: [Docker/scclmg/OCT-nn-UNet](#)

The code used in this study in Chapter 6 is available on GitHub at the following address: [Github/GWAS-pipeline](#)