



OPEN ACCESS

EDITED BY

Thomas Hartung,
Johns Hopkins University, United States

REVIEWED BY

Chi Ma,
Capital Medical University, China
Adnan Mohsin Abdulzeez,
University of Duhok, Iraq

*CORRESPONDENCE

Cynthia Maldonado-Garcia
✉ cynthiamalgar@gmail.com

†These authors share senior authorship

RECEIVED 07 May 2025

ACCEPTED 24 October 2025

PUBLISHED 18 November 2025

CITATION

Maldonado-Garcia C, Bonazzola R, Ferrante E,
Julian TH, Sergouniotis PI, Ravikumar N and
Frangi AF (2025) Predicting cardiovascular
disease risk using retinal optical coherence
tomography imaging.
Front. Artif. Intell. 8:1624550.
doi: 10.3389/frai.2025.1624550

COPYRIGHT

© 2025 Maldonado-Garcia, Bonazzola,
Ferrante, Julian, Sergouniotis, Ravikumar and
Frangi. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Predicting cardiovascular disease risk using retinal optical coherence tomography imaging

Cynthia Maldonado-Garcia^{1*}, Rodrigo Bonazzola¹,
Enzo Ferrante², Thomas H. Julian^{3,4},
Panagiotis I. Sergouniotis^{3,4,5,6}, Nishant Ravikumar^{1†} and
Alejandro F. Frangi^{7,8,9,10†}

¹Centre for Computational Imaging and Simulation Technologies in Biomedicine, School of Computing, University of Leeds, Leeds, United Kingdom, ²Research Institute for Signals, Systems and Computational Intelligence, sinc(i), Consejo Nacional de Investigaciones Científicas y Técnicas – Universidad Nacional del Litoral (CONICET-UNL), Santa Fe, Argentina, ³Division of Evolution, Infection and Genomics, Faculty of Biology, Medicine, and Health, School of Biological Sciences, University of Manchester, Manchester, United Kingdom, ⁴Manchester Royal Eye Hospital, Manchester University NHS Foundation Trust, Manchester, United Kingdom, ⁵Manchester Centre for Genomic Medicine, Saint Mary's Hospital, Manchester University NHS Foundation Trust, Manchester, United Kingdom, ⁶European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom, ⁷Division of Informatics, Imaging, and Data Sciences, Faculty of Biology, Medicine, and Health, School of Health Sciences, University of Manchester, Manchester, United Kingdom, ⁸Faculty of Science and Engineering, School of Computer Science, University of Manchester, Manchester, United Kingdom, ⁹Christabel Pankhurst Institute, University of Manchester, Manchester, United Kingdom, ¹⁰NIHR Manchester Biomedical Research Centre, Manchester Academic Health Science Centre, Manchester, United Kingdom

Introduction: Cardiovascular Diseases (CVD) are the leading cause of death globally. Non-invasive, cost-effective imaging techniques play a crucial role in early detection and prevention of CVD. Optical Coherence Tomography (OCT) has gained recognition as a noninvasive method of detecting microvascular alterations that might enable earlier identification and targeting of at-risk patients. In this study, we investigated the potential of OCT as an additional imaging technique to predict future CVD events.

Methods: We analyzed retinal OCT data from the UK Biobank. The dataset included 612 patients who suffered a Myocardial Infarction (MI) or stroke within five years of imaging and 2,234 controls without CVD (total: 2,846 participants). A self-supervised deep learning approach based on Variational Autoencoders (VAE) was used to extract low-dimensional latent representations from high-dimensional 3D OCT images, capturing structural and morphological features of retinal and choroidal layers. These latent features, along with clinical data, were used to train a Random Forest (RF) classifier to differentiate between patients at risk of future CVD events (MI or stroke) and healthy controls.

Results: Our model achieved an AUC of 0.75, sensitivity of 0.70, specificity of 0.70, and accuracy of 0.70. The choroidal layer in OCT images was identified as a key predictor of future CVD events, revealed through a novel model explainability approach.

Discussion: Our findings demonstrate the potential of retinal OCT imaging, when combined with advanced deep learning methods, as a predictive tool for identifying individuals at increased risk of CVD events.

KEYWORDS

optical coherence tomography, cardiovascular diseases, multimodal, deep learning, variational autoencoder

1 Introduction

Cardiovascular Diseases (CVDs) continue to pose a significant global health challenge, affecting more than 500 million individuals worldwide. Specifically, in 2021, CVDs led to 20.5 million deaths. It is concerning to observe that up to 80% of premature Myocardial Infarction (MI) and stroke cases could potentially be prevented if detected early. Moreover, the burden of CVDs disproportionately impacts low- and middle-income countries, where nearly four out of every five CVDs-related deaths worldwide occur (Lindstrom et al., 2023). Currently, tools like the third version of the QRISK cardiovascular disease risk prediction algorithm (QRISK3) are utilized in primary care settings by healthcare professionals to pinpoint patients at higher risk of various CVDs. These tools are commonly employed during health assessments to assess a patient's risk based on factors such as demographic details (e.g., ethnicity, age, sex), clinical indicators (e.g., cardiac volume measurements, blood markers, indicators of obesity, etc.), and socioeconomic data (Li et al., 2019), but they require clinical infrastructure and laboratory testing, which may be limited in resource-constrained settings. This underscores the need for scalable, low-cost screening methods. Retinal optical coherence tomography (OCT) offers a promising solution: it is non-invasive, relatively inexpensive, and increasingly available in optometry and primary care settings. Retinal and choroidal microvasculature are sensitive indicators of systemic vascular health and have been linked to cerebral and coronary vascular disease. Thus, retinal imaging provides a unique window into microvascular health and enables the early identification of individuals at risk of CVD (Wong et al., 2022; Farrah et al., 2020; Chew et al., 2025; Arefin, 2025). Leveraging retinal imaging as a point-of-care screening tool could support overburdened health systems by enabling opportunistic risk assessment during routine eye exams and triaging high-risk patients for further cardiovascular evaluation.

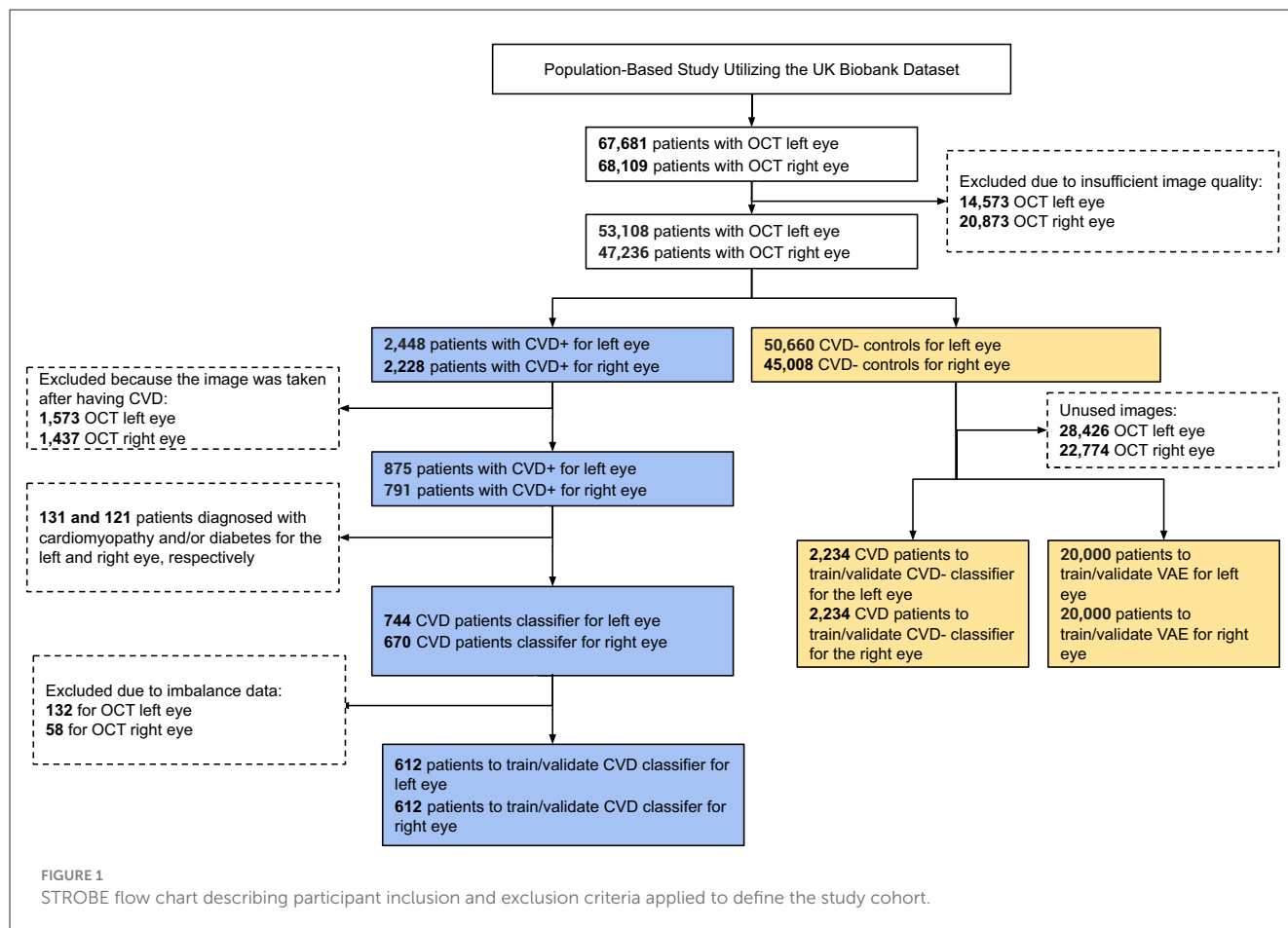
Numerous prior studies have investigated the application of Artificial Intelligence (AI) in predicting CVDs risk factors using fundus photographs (Poplin et al., 2018; Nusinovič et al., 2022), as well as CVDs events (Diaz-Pinto et al., 2022; Cheung et al., 2020; Abdollahi et al., 2024). While retinal fundus photographs provide a two-dimensional surface view of the retinal vasculature, they are limited in their ability to capture information about the deeper retinal and choroidal layers. In contrast, OCT offers high-resolution three-dimensional cross-sectional imaging, enabling precise quantification of retinal layer thicknesses, choroidal architecture, and microstructural integrity. This depth-resolved information allows the detection of subtle microvascular changes that may precede clinically visible alterations on fundus photography, facilitating earlier detection of disease indicators and more precise monitoring of progression (Farrah et al., 2020). Furthermore, the volumetric nature of OCT data supports robust extraction of quantitative biomarkers, providing richer input for machine learning models and potentially improving risk stratification performance. This technological advancement in OCT has revolutionized retinal imaging by facilitating visualization of the chorioretinal microcirculation, which can serve as an early sign of microvascular disease. However, there remains limited research utilizing OCT as a predictor of CVDs (Maldonado-Garcia et al., 2022; Zhou et al., 2023).

This study introduces a predictive model that integrates 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 MI or stroke within five years after image capture. To the best of our knowledge, this is among the first investigations applying 3D OCT imaging and artificial intelligence for automatically forecasting patients vulnerable to adverse CVD incidents. The model advances the field in several ways employing a self-supervised Variational Autoencoder (VAE) combined with a multimodal Random Forest (RF) classifier to effectively integrate diverse patient data, and it incorporates an innovative vector-field-based explainability method to localize retinal features that contribute most to risk prediction, and it identifies the choroidal layer as a key predictive feature for cardiovascular risk.

2 Materials and methods

2.1 Database

In this research, we utilized retinal OCT imaging data sourced from the UK Biobank, captured using the Topcon 3D OCT 1000 Mark 2 system. OCT, a non-invasive imaging technique, employs light waves to generate intricate images of ocular structures like the retina, choroid, and optic nerve. The imaging procedure occurred in a dimly lit environment without pupil dilation, utilizing the 3D $6 \times 6 \text{ mm}^2$ macular volume scan mode, which consists of 128 horizontal B-scans arranged in a $6 \times 6 \text{ mm}$ raster pattern. The UK Biobank contains a vast array of health-related information from more than 500,000 participants in the UK, encompassing genetics, demographics, clinical measurements, lifestyle aspects, and medical imaging. Specifically, during their baseline visit [Instance 0, "Initial assessment visit (2006–2010)"], a total of 68,109 and 67,681 participants underwent retinal imaging for their right and left eyes, respectively. 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. We excluded images representing the lowest 20% as indicated by the QI indicator, 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 to minimize potential confounding, as both conditions are associated with retinal microvascular and choroidal alterations that could obscure associations specific to future myocardial infarction or stroke risk (Diaz-Pinto et al., 2022; Rakusiewicz et al., 2021; Yu et al., 2023). We also removed cases to address data imbalance, resulting in a final 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 1](#). This study exclusively encompasses patients who experienced MI or stroke within a five-year period after OCT image acquisition. The acquisition details of reported CVD incidents' sources are elaborated in [Supplementary Figure S1](#).

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 1](#), resulting in a final count of 612 participants with OCT images of both eyes. For the non-CVD group or CVD-, 2,234 participants were propensity score-matched based on sex and age 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 1](#). The average age of individuals with and without CVDs was 60.78 ± 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 ± 4.45 kg/m² for those with CVDs and 27.43 ± 4.33 kg/m² for those without. In terms of blood pressure readings, individuals with CVDs had a Systolic Blood Pressure (SBP) of 147.26 ± 19.57 mm Hg, while those without CVDs had an SBP of 145.1 ± 18.75 mm Hg. The Diastolic Blood Pressure (DBP) was 84.75 ± 10.23 mm Hg for individuals with CVDs and 83.22 ± 9.73 mm Hg for those without. The mean level of Hemoglobin

A1c (HbA1c) was 36.52 ± 4.32 mmol/L for individuals with CVDs and 36.59 ± 6.61 for those without. A notable percentage of participants reported being current alcohol consumers, accounting for 90.69% of individuals with CVDs and 91.83% of those without. These participant characteristics for both groups are summarized in [Table 1](#).

We used an age-sex-matched cohort for the two groups. The relative proportion of patients included in the study in the two groups were 1:3 in the CVD and non-CVD group respectively. We used a 1:3 case:control matching ratio to maximize statistical power while keeping computational requirements tractable. Utilizing an age-sex-matched study cohort offers a significant advantage by helping to mitigate the influence of confounding variables that could bias the predictive model ([Zhou et al., 2023](#)). For instance, since age and sex are known to impact the risk of developing CVDs, matching groups based on these characteristics ensures that any observed differences in CVDs incidence are more directly associated with the retinal features, which are the variables of interest in this study. Previous research has indicated that matching on age and sex in the UK Biobank reduces variability within groups and enhances homogeneity among participants, thereby potentially improving the statistical power of the study ([Batty et al., 2019, 2020](#)). Furthermore, machine learning models may capture spurious correlations (i.e. to learn shortcuts) between predictors and targets, such as, for example, linking age or sex to the presence of pathology, unless care is taken when defining the predictors and

TABLE 1 Characteristics of patients in the CVD+ and CVD- sets.

Characteristics	CVD+	CVD-
Number of subjects	612	2,234
Age: Mean (s.d.), years	60.78 (6.47)	60.78 (6.47)
Sex: 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 ²	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)
HbA1c: %	2.43	2.44
Alcohol consumption: N, P, C, NA %	3.59, 5.72, 90.69, 0	3.92, 3.92, 91.83, 0.33

N, Never; P, Previous; C, Current; NA, Not answer.

study cohort (Brown et al., 2022). Figure 2 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.

2.2 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 (Figure 3). The framework integrates a VAE (Kingma and Welling, 2013) for self-supervised feature extraction from retinal OCT images, and a RF classifier that combines the learned representations with patient metadata. The proposed model consists of two stages, a self-supervised feature extraction stage, and a subsequent classification stage.

2.2.1 Self-supervised VAE

In the first stage of the proposed model, a VAE is used to learn latent representations for B-scan OCT images. VAE are a widely used type of generative model in neural networks, comprising a pair of networks or network branches that are trained together, called the encoder and decoder networks/branches. Given some input data (\mathbf{x}), the encoder is designed to approximate the posterior

distribution of the latent variables ($q_{\phi}(\mathbf{z} | \mathbf{x})$), 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 ($p_{\theta}(\hat{\mathbf{x}} | \mathbf{z})$). 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 maximizing the Evidence Lower Bound (ELBO), which can be expressed as follows:

$$\text{ELBO} = \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} [\log p_{\theta}(\mathbf{x} | \mathbf{z})] - D_{KL}(q_{\phi}(\mathbf{z} | \mathbf{x}) || p(\mathbf{z})) \quad (1)$$

The loss function utilized 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 3, 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. KL divergence quantifies the dissimilarity between the learnt latent distribution $q(\mathbf{z})$ and a previously specified distribution $p(\mathbf{z})$, which, in this scenario, the prior $p(\mathbf{z})$ is a multivariate Gaussian distribution. The parameters of the learned distribution $q(\mathbf{z})$ are its mean (μ_i) and variance (σ_i^2). By minimizing 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 toward the dual objective of reducing reconstruction errors and aligning the learnt latent distribution with the intended prior distribution.

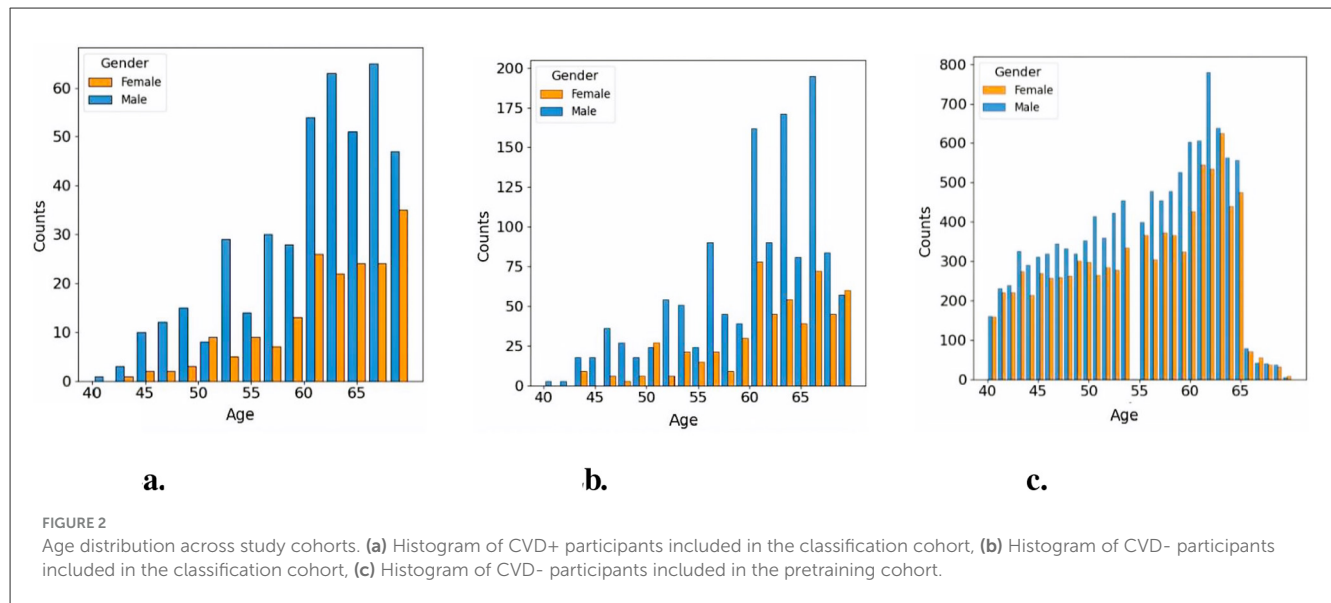
$$\mathcal{L}_{VAE} = \mathcal{L}_{MSE} + \beta \mathcal{L}_{KL} \quad (2)$$

$$\mathcal{L}_{MSE} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \hat{\mathbf{x}}_i)^2 \quad (3)$$

$$\mathcal{L}_{KL} = \frac{1}{2} [1 + \log(\sigma_i^2) - \sigma_i^2 - \mu_i^2] \quad (4)$$

Given the relatively small number of CVD events, training a fully supervised, well-known Convolutional Neural Network (CNN) could lead to overfitting. A self-supervised VAE mitigates this risk by learning compact latent representations from the full OCT dataset without outcome labels, capturing meaningful microstructural variability. By combining reconstruction and KL divergence losses, the VAE creates a structured latent space that captures informative patterns in retinal layers relevant to cardiovascular risk. These latent features can then be used by downstream classifiers to improve prediction performance compared to using clinical variables alone. Prior studies in medical imaging show that VAEs or autoencoder-based pretraining improve classification performance when labeled data are limited (Ehrhardt and Wilms, 2022; Rais et al., 2024).

For this study, we trained independent VAEs for each 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.



2.2.2 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 3. 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). 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.

2.3 Experiment details

All experiments were carried out on an NVIDIA Tesla M60 GPU using PyTorch (v1.10.2). A grid search strategy with five-fold cross-validation was employed to determine the optimal hyperparameters (Supplementary Table S1). The dataset

was divided into training, validation, and test sets in a 6:2:2 ratio. The encoder and decoder networks each consisted of six 2D convolutional layers; the encoder used Rectified Linear Unit (ReLU) activations, while the decoder used LeakyReLU activations (see Supplementary Table S2). Model training was performed in two phases: an initial self-supervised learning phase to acquire latent representations independently for each eye, followed by a fine-tuning phase initialized with the learned weights. Specifically, the left and right eye VAEs were trained independently from scratch. When training a model using both eyes simultaneously, we initialized the networks with the previously trained left/right VAE weights to reduce computational cost and accelerate convergence.

The VAE was trained with a latent space size of 128 using KL annealing to improve training stability. The KL divergence weight (β) was gradually increased from 0.001 at the start of training to a maximum of 0.01 over the first 20 epochs, after which it was held constant for the remaining training epochs, allowing the model to prioritize reconstruction early on before encouraging disentangled latent representations (Ichikawa and Hukushima, 2024). The VAE was trained for 100 epochs in total. Other hyperparameters, including CNN channels, batch size, learning rate, and weight decay, were optimized via grid search, with the tested and optimal values reported in Supplementary Table S1. Hyperparameter details for downstream Random Forest models are summarized in Supplementary Table S3.

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 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

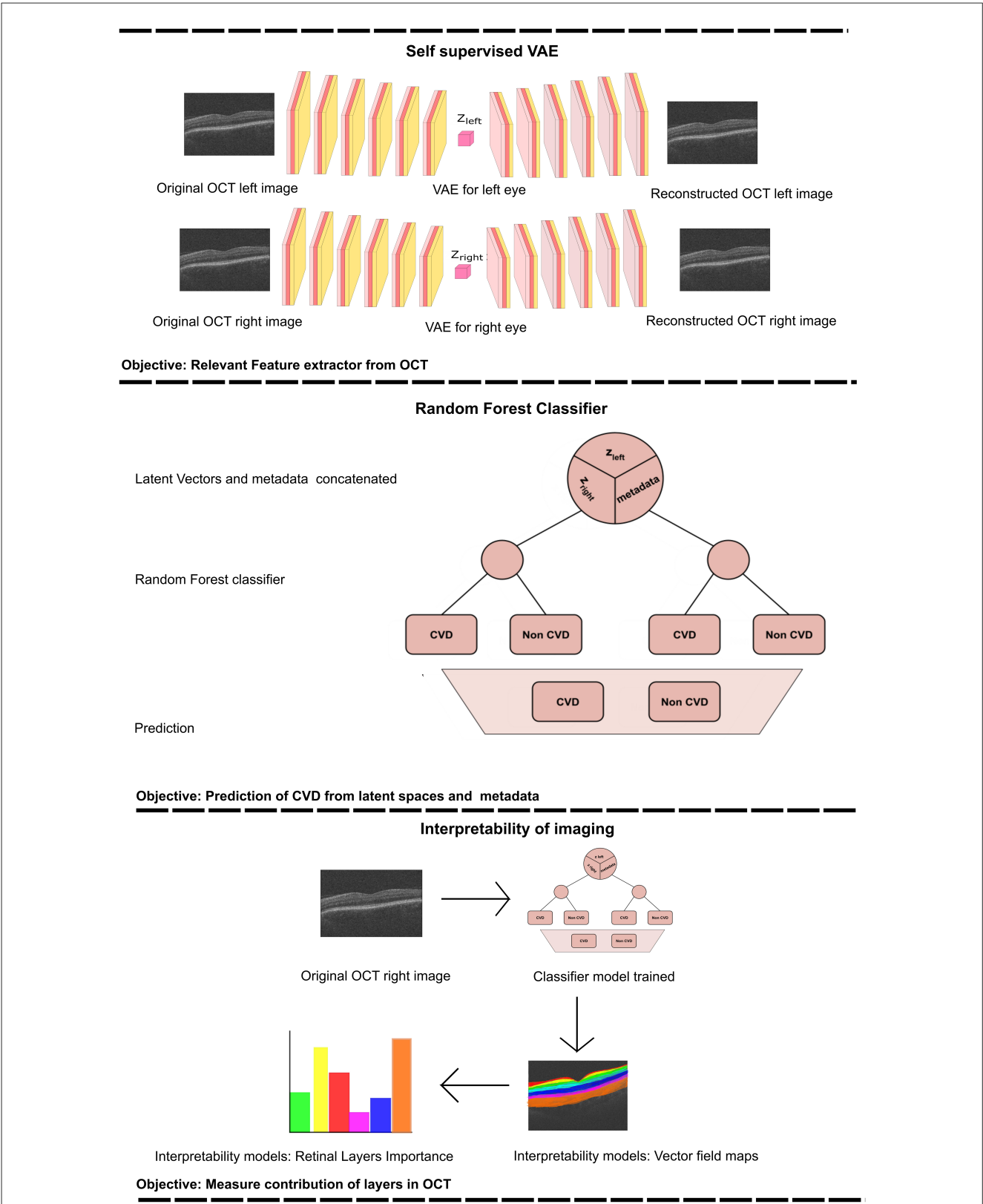


FIGURE 3
The provided workflow diagram illustrates the comprehensive process of training the Variational Autoencoder (VAE) and subsequently using it to acquire the latent vectors (upper section). These latent vectors are then combined with metadata and serve as inputs to the Random Forest (RF) classifier (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.

seven datasets, as depicted in Figure 3 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 Supplementary Table S3). We divided the dataset into training, validation, and test sets, following a distribution ratio of $\sim 5:2:3$, respectively (resulting, 1423 patients in the training set, 459 patients in the validation 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. During the optimal hyperparameter search, we employed Recursive Feature Elimination (RFE) as a feature selection technique to address overfitting and train the model with the most pertinent variables for classification. In the majority of cases, the model was trained with the top 10 most significant features, with the exception of the RE case, where the optimal conditions involved using only 5 variables.

To evaluate the effectiveness of our model, we compared predictive performance against the QRISK3 algorithm, the current gold standard used by healthcare professionals / cardiologists to assess the patient's risk of stroke or heart attack, in a 10-year period. The QRISK3 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 which are described in Supplementary Table S4. In cases where certain information was missing, we consistently represented these gaps as "0" when evaluating the QRISK3 scores on the test dataset. For our classification task, we evaluated the model performance using a range of metrics. Accuracy, 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 ≥ 0.5 , the patient is classified as CVD+, else as CVD-). The Area Under the Receiver Operating curve (AUROC) was then employed as a performance measure to assess both our model and the QRISK3 algorithm.

2.4 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 behavior 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 behavior 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.

To provide local explanations of the behavior 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 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 z_{max} on subsequent reconstructions of OCT images. Specifically, given an image x , we compute the corresponding latent vector z using the trained encoder network. Next, we perturb only the dimension z_{max} of the calculated latent vector and reconstruct the corresponding image. This perturbation is performed by multiplying the latent dimension z_{max} by a scalar value, which in this case is the standard deviation of this latent component, calculated throughout the training population, defined as σ_{max} . The remaining latent variables in the computed latent vector are left unchanged, resulting in a perturbed latent vector (\hat{z}). Finally, we reconstruct the input OCT image (\hat{x}) using the perturbed latent vector \hat{z} .

To visualize the regions in the reconstructed OCT images affected by the altered latent vector \hat{z} , we examine the differences between the initial image x and the reconstructed image \hat{x} derived from \hat{z} . In this context, we calculate the vector field between these images using the Lucas-Kanade algorithm (Lucas and Kanade, 1981). The resulting vector field, showing the magnitude of the displacement vector for the moving pixels, was then superimposed on the original image to create a visual representation. The vector field indicates how the pixels between the images (that is, x and \hat{x}) change due to the latent traversal from z to \hat{z} . The estimated vector field between x and \hat{x} helps to visually illustrate the regions in the OCT image that were altered by modifying the latent component 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 CVDs 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.

To provide global explanations of the behavior 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.

3 Results

3.1 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 2.3. The rationale for comparing all seven classifiers against each other was to - (i) 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; (ii) compare the discriminative power of OCT image-derived latent features against patient metadata; and (iii) 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 summarized in Table 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 Supplementary Table S5). 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). 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 presents four histograms depicting true positives (Figure 4a), true negatives (Figure 4b), false positives (Figure 4c),

and false negatives (Figure 4d) 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 summarized in Table 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 (Figure 4d). The MTDT-RF classifier yielded fewer true positives and true negatives (12% in both) compared to the other classifiers that utilized only OCT features (Figures 4a, b, respectively), which aligns with the results presented in Table 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 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 Supplementary Table S6. This observation is in concordance with the global explanations of models' predictions summarized in Figure 5, 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. Confusion matrices and precision-recall curves for all classifiers are provided in Supplementary Figures S2, S3.

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 3).

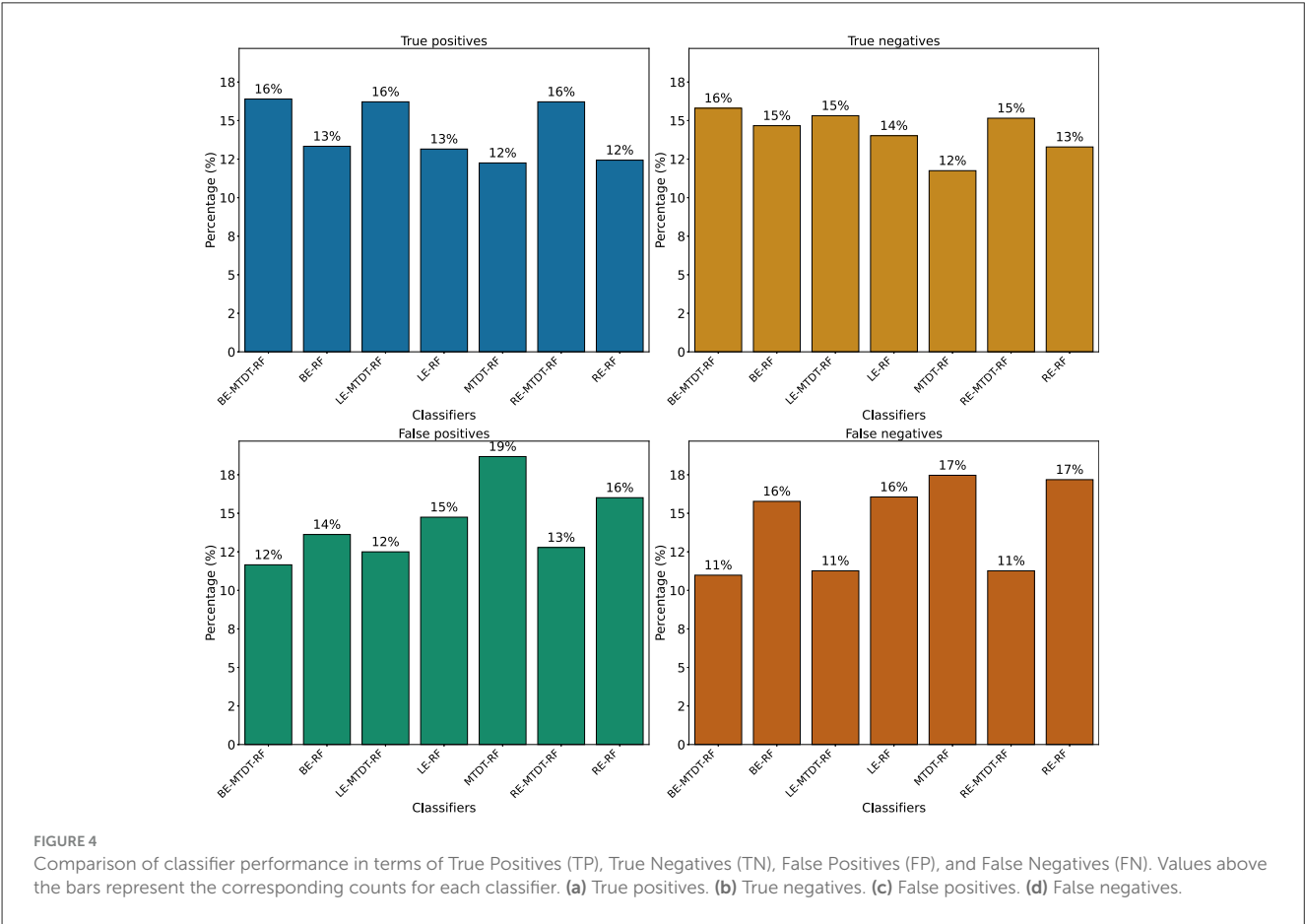
3.2 Model explainability

In order to provide local explanations for the behavior of all classifiers investigated in this study, we analyzed the most important features identified by each model (refer to Figure 5) 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 5a, we found that a latent variable learned from the left-eye OCT image, denoted z1066, 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. 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 local explanations summarized in Figures 5b–d, we observe that latent variable z1066 consistently ranks among the top two most important features for the BE-RF, LE-MTDT-RF, and LE-RF classifiers, respectively. This indicates that the retinal features encoded by z1066 are consistently considered to be relevant across all four classifiers presented in Figures 5a–d. Among the classifiers which combined retinal OCT-image derived features

TABLE 2 Predictive analysis of Cardiovascular Disease (CVD) metrics utilizing UK Biobank data across seven distinct cases.

Classification metrics				
Classifier	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)
LE-MTDT-RF	0.68 (0.65–0.71)	0.69 (0.66–0.73)	0.68 (0.65–0.72)	0.72 (0.69–0.75)
RE-MTDT-RF	0.67 (0.65–0.69)	0.69 (0.67–0.71)	0.67 (0.65–0.69)	0.70 (0.68–0.72)
BE-MTDT-RF	0.70 (0.67–0.73)	0.70 (0.67–0.73)	0.70 (0.67–0.72)	0.75 (0.72–0.78)
MTDT-RF	0.52 (0.50–0.54)	0.52 (0.50–0.54)	0.52 (0.50–0.54)	0.57 (0.55–0.59)
LE-RF	0.61 (0.58–0.64)	0.56 (0.53–0.59)	0.62 (0.59–0.65)	0.64 (0.61–0.67)
RE-RF	0.58 (0.56–0.60)	0.53 (0.51–0.55)	0.59 (0.57–0.61)	0.62 (0.60–0.64)
BE-RF	0.64 (0.62–0.67)	0.57 (0.55–0.59)	0.65 (0.63–0.68)	0.67 (0.64–0.70)

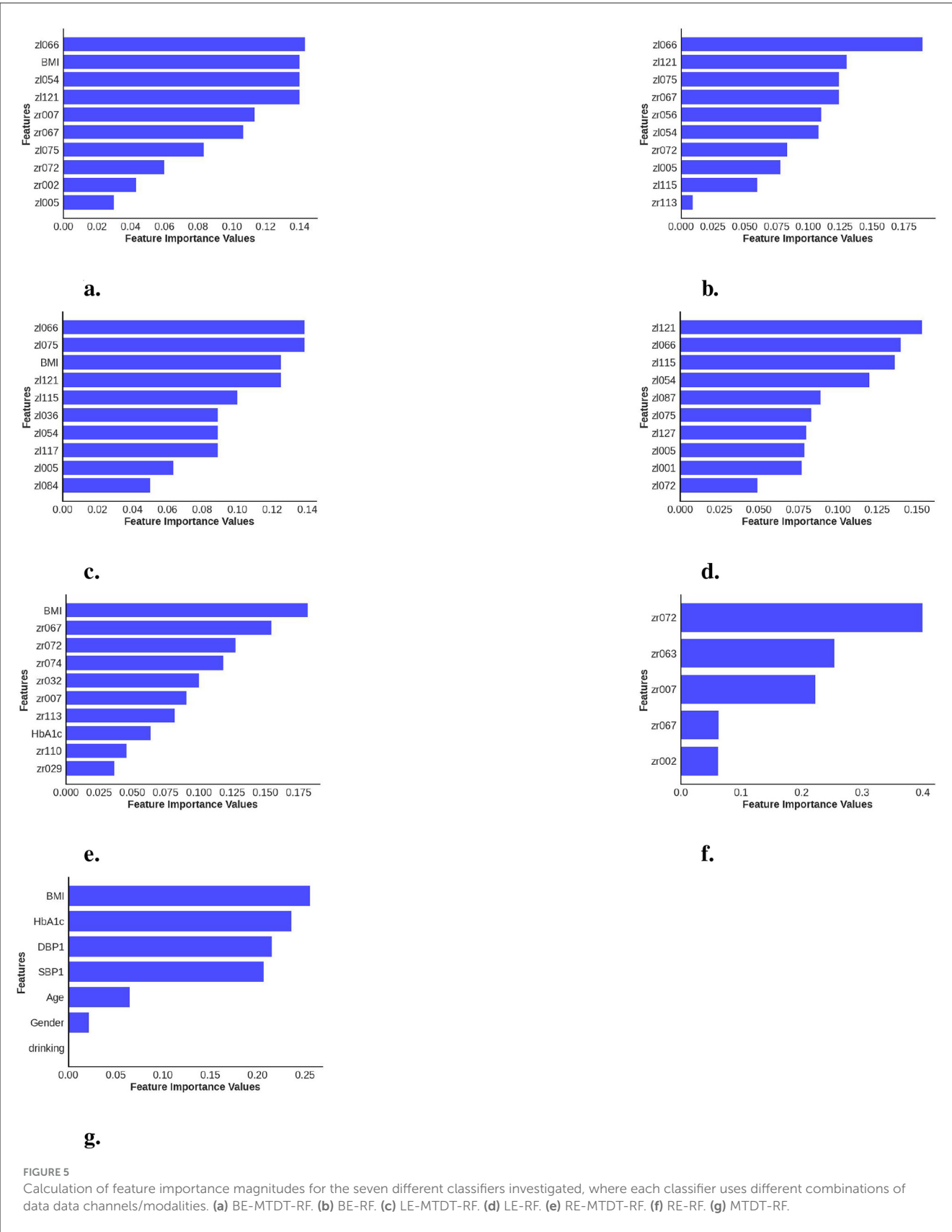
All metrics are reported as mean values with 95% confidence intervals (CI). Bold values indicate the highest performance for each classification metric.



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.

Using the insights gained from analyzing the global explanations of classifier behavior summarized in Figure 6, 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 z1066, derived from left-eye OCT images, as being the most important feature for classification, our local explainability approach (refer to Section 2.4) 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

TABLE 3 Comparison of classification metric results between our model employing both ocular data and metadata (BE-MTDT) and the QRISK algorithm.

Classification metrics				
Modality	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)
BE-MTDT-RF	0.70 (0.67–0.73)	0.70 (0.67–0.73)	0.70 (0.67–0.72)	0.75 (0.72–0.78)
QRISK	0.55 (0.53–0.57)	0.60 (0.58–0.62)	0.55 (0.52–0.58)	0.60 (0.57–0.63)
$\chi^2 = 95.72, df = 1, p = 1.31 \times 10^{-22}$				

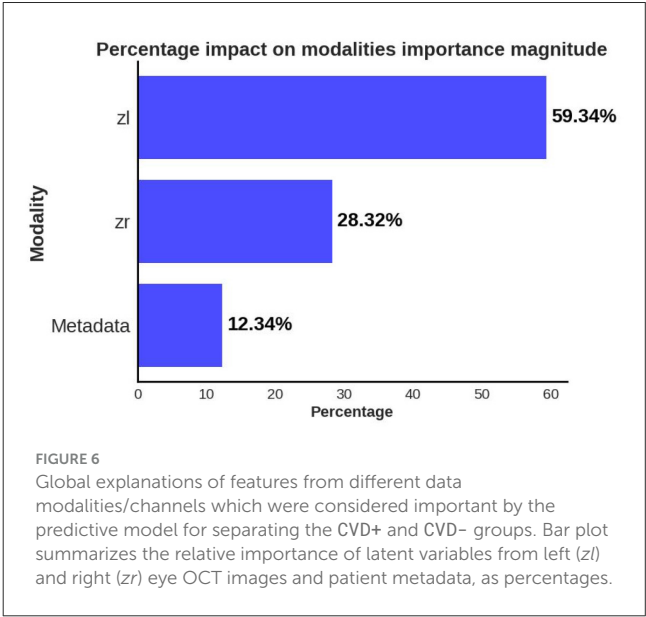
All metrics are reported as mean values with 95% confidence intervals (CI). McNemar's Test (* $p < 0.005$). Bold values indicate the highest performance for each classification metric.

latent variable of interest. Although this process was applied to all high-risk CVD patients, we present results for five representative cases here (Figure 7), with each row representing one patient. We considered 3 B-scans per patient, including the 1st one (upper row), 64th (middle row) and 128th (bottom row). Finally, by overlaying the estimated vector field maps onto the original OCT images, we obtained visual interpretations of the retinal image features encoded by the latent variable of interest.

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 localization of the image regions modeled by latent variable z1066 (visualized as landmarks, as shown in Figure 7), 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. To further validate the robustness of our findings, we applied an occlusion-based approach (Supplementary Figure S4), which produced qualitatively similar results. Notably, our vector field perturbation method highlights regions more consistently across the entire image and all B-scans. For instance, the choroid is captured as a whole rather than partially, as observed with the comparison method.

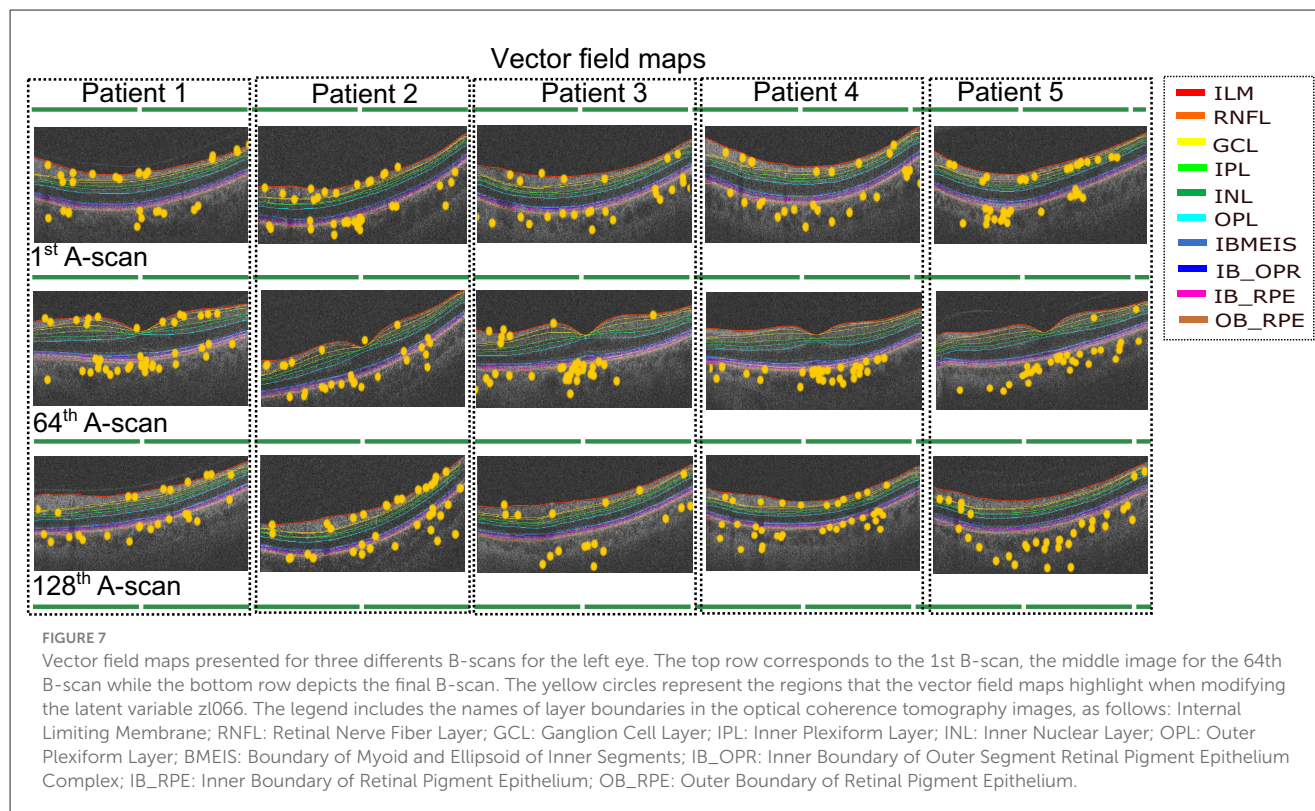
4 Discussion

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 CVDs (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 CVDs. 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 CVDs. 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 (Maldonado-Garcia et al., 2022; 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.

Our results show that combining OCT-derived latent representations from both eyes with patient metadata yields the highest discriminative power for distinguishing between the CVD+ and CVD- groups (AUC: 0.75). While classifiers trained on left-eye data slightly outperformed those trained on right-eye data (AUC: 0.72 vs. 0.70 for LE-MTDT-RF and RE-MTDT-RF, respectively), the combined model achieved superior performance, suggesting



that information from both eyes provides complementary features that improve predictive accuracy. Notably, left-eye features ranked highest in both global and local feature importance analyses. We hypothesize that this finding may be partly explained by systematic differences in image quality between eyes. In our cohort, left-eye scans exhibited higher average quality (see [Supplementary Figure S5](#)), likely because the UK Biobank protocol consistently acquired the left eye second, giving participants time to adapt to the procedure and potentially resulting in fewer motion artifacts. Although this acquisition order may introduce subtle biases favoring left-eye images, the improved performance of the combined model indicates that our findings are robust to these effects. Future studies using datasets with randomized or balanced acquisition sequences will help confirm that this performance difference is not primarily driven by protocol-related artifacts.

Additionally, in our age- and sex-matched cohort, the model correctly assigns a higher risk score to individuals who will experience a CVD event approximately 75% of the time, compared to only 60% for QRISK3. Importantly, this improved performance is achieved using a minimal set of demographic and clinical variables combined with retinal OCT-derived features, highlighting the potential of non-invasive retinal imaging to identify high-risk individuals more accurately while reducing false positives and unnecessary follow-up interventions.

Interestingly, our results suggest that choroidal morphology is a predictor of identifying patients at risk of CVDs, 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. This suggests potential clinical applications, such as adapting standardized OCT acquisition protocols to enhance choroidal visibility ([Chew et al., 2025](#)). Techniques like enhanced depth imaging, swept-source OCT (which provides deeper penetration and better delineation of the choroid), or wide-field OCT angiography (OCTA; which offers complementary functional information by visualizing blood flow and quantifying vascular biomarkers) could be integrated into screening workflows. While swept-source OCT currently remains less widely available and more costly than spectral-domain OCT, ongoing technological advances and decreasing costs are likely to improve accessibility in both clinical and research settings. We hypothesize that combining multimodal retinal imaging (e.g. OCT, OCTA) with AI-based risk prediction may improve the classification performance of the proposed approach by capturing both anatomical and functional signatures of systemic vascular health, potentially improving the sensitivity and robustness of cardiovascular risk stratification ([Copete et al., 2013](#); [Runsewe et al., 2024](#); [Germanese et al., 2024](#)). In addition to the choroidal characteristics, our results indicated that the inner retinal layers contributed to the classification, including

the Retinal Nerve Fiber 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 CVDs, but their role as predictors of future disease has received limited attention (Chen et al., 2023; Matuleviciūtė et al., 2022). Mechanisms that may underpin the role of neurosensory retina morphology as a predictor of CVDs are yet to be elucidated, although it could be hypothesized 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).

Our study has several limitations. First, the analyses were conducted using data from the UK Biobank, a volunteer-based cohort known to have a “healthy volunteer” bias, with participants generally exhibiting lower smoking rates, healthier lifestyles, and lower prevalence of chronic conditions compared with the general population. This may have resulted in an underestimation of absolute CVD risk and may limit the generalizability of our results to more diverse or higher-risk populations (Fry et al., 2017). Additionally, although we performed internal validation, we were unable to perform external validation, as no publicly available datasets currently combine 3D OCT volumes with longitudinal cardiovascular outcomes. Future research using clinical populations with broader demographic and health profiles, as well as independent external validation, is essential to confirm the robustness and generalizability of our findings. Such efforts could also explore the inclusion of additional clinical variables in predictive models, provided these variables are readily obtainable in typical eye care settings, thereby enhancing decision-making while maintaining practical applicability. Additionally, 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 populations [e.g., QResearch (Hippisley-Cox et al., 2017), CPRD (Livingstone et al., 2021)] and is widely adopted in clinical practice. Our matched cohort may also reduce the predictive impact of key variables like age and sex (Ghosal and Sinha, 2022), further limiting direct comparability.

5 Conclusion

This study presents a predictive framework comprising a self-supervised representation learning approach based on a VAE and a Random Forest classifier, which effectively integrates multi-modal imaging (OCT imaging) and non-imaging (e.g. patient demographic and clinical variables) data, to identify patients at risk of stroke or myocardial infarction. Future work should explore multi-modal retinal imaging integration to enhance CVD prediction.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.ukbiobank.ac.uk/>.

Author contributions

CM-G: Writing – original draft, Validation, Investigation, Writing – review & editing, Methodology, Formal analysis, Visualization, Data curation. RB: Writing – review & editing, Writing – original draft. EF: Writing – original draft, Writing – review & editing. TJ: Writing – original draft, Writing – review & editing. PS: Writing – original draft, Writing – review & editing. NR: Supervision, Conceptualization, Writing – review & editing, Writing – original draft. AF: Conceptualization, Supervision, Writing – review & editing, Writing – original draft.

Funding

The authors declare that financial support was received for the research and/or publication of this article. This project was funded by the following: The Royal Academy of Engineering INSILEX (grant no. CiET1819\19) (AF); Consejo Nacional de Humanidades Ciencias y Tecnologías (CONAHCYT) (scholarship no. 766588) (CM-G); UKRI Frontier Research Guarantee INSILICO (grant no. EP/Y030494/1) (RB, NR, and AF); the MRC Clinical Research Training Fellowship (grant no. MR/Z504105/1) (TJ); the Wellcome Trust (grant no. 224643/Z/21/Z) (PS); and the NIHR Manchester Biomedical Research Centre (grant no. NIHR 203308) (TJ, PS, and AF).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of

their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abdollahi, M., Jafarizadeh, A., Ghafouri-Asbagh, A., Sobhi, N., Pourmoghtader, K., Pedrammehr, S., et al. (2024). Artificial intelligence in assessing cardiovascular diseases and risk factors via retinal fundus images: A review of the last decade. *Wiley Interdiscipl. Rev. Data Mining Knowl. Discov.* 14:e1560. doi: 10.1002/widm.1560
- Arefin, M. A. O. S. (2025). Advancements in AI-enhanced OCT imaging for early disease detection and prevention in aging populations. *Int. J. Multidiscipl. Res. Sci. Eng. Technol.* 8, 1430–1444.
- Batty, G. D., Gale, C. R., Kivimäki, M., Deary, I. J., and Bell, S. (2019). Generalisability of results from uk biobank: comparison with a pooling of 18 cohort studies. *medRxiv* 368:19004705. doi: 10.2139/ssrn.3437793
- Batty, G. D., Gale, C. R., Kivimäki, M., Deary, I. J., and Bell, S. (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. *BMJ* 368:m131. doi: 10.1136/bmj.m131
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Brown, A., Tomasev, N., von Freyberg, J., Liu, Y., Karthikesalingam, A., and Schrouff, J. (2022). Detecting shortcut learning for fair medical AI using shortcut testing. *Nat. Commun.* 14:4314. doi: 10.1038/s41467-023-39902-7
- Chen, Y., Yuan, Y., Zhang, S., Yang, S., Zhang, J., lei Guo, X., et al. (2023). Retinal nerve fiber layer thinning as a novel fingerprint for cardiovascular events: results from the prospective cohorts in UK and China. *BMC Med.* 21:7. doi: 10.1186/s12916-023-02728-7
- Cheung, C., Xu, D., Cheng, C.-Y., Sabanayagam, C., Tham, Y.-C., Yu, M., et al. (2020). A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. *Nat. Biomed. Eng.* 5, 498–508. doi: 10.1038/s41551-020-00626-4
- Chew, E. Y., Burns, S. A., Abraham, A. G., Bakhoun, M. F., Beckman, J. A., Chui, T. Y., et al. (2025). Standardization and clinical applications of retinal imaging biomarkers for cardiovascular disease: a roadmap from an NHLBI workshop. *Nat. Rev. Cardiol.* 22, 47–63. doi: 10.1038/s41569-024-01060-8
- Copete, S., Flores-Moreno, I., Montero, J. A., Duker, J. S., and Ruiz-Moreno, J. M. (2013). Direct comparison of spectral-domain and swept-source oct in the measurement of choroidal thickness in normal eyes. *Br. J. Ophthalmol.* 98, 334–338. doi: 10.1136/bjophthalmol-2013-303904
- Diaz-Pinto, A., Ravikumar, N., Attar, R., Suinesiaputra, A., Zhao, Y., Levelt, E., et al. (2022). Predicting myocardial infarction through retinal scans and minimal personal information. *Nat. Mach. Intellig.* 4, 55–61. doi: 10.1038/s42256-021-00427-7
- Ehrhardt, J., and Wilms, M. (2022). “Autoencoders and variational autoencoders in medical image analysis,” in *Biomedical Image Synthesis and Simulation* (London: Elsevier), 129–162.
- Farrah, T. E., Dhillon, B., Keane, P. A., Webb, D. J., and Dhaun, N. (2020). The eye, the kidney, and cardiovascular disease: old concepts, better tools, and new horizons. *Kidney Int.* 98, 323–342. doi: 10.1016/j.kint.2020.01.039
- Ferrara, M., Lugano, G., Sandinha, M. T., Kearns, V. R., Geraghty, B., and Steel, D. H. W. (2021). Biomechanical properties of retina and choroid: a comprehensive review of techniques and translational relevance. *Eye* 35, 1818–1832. doi: 10.1038/s41433-021-01437-w
- Fry, A., Littlejohns, T. J., Sudlow, C. L. M., Doherty, N., Adamska, L., Sprosen, T., et al. (2017). Comparison of sociodemographic and health-related characteristics of uk biobank participants with those of the general population. *Am. J. Epidemiol.* 186, 1026–1034. doi: 10.1093/aje/kwx246
- Germanese, C., Anwer, A., Eid, P., Steinberg, L.-A., Guenancia, C., Gabrielle, P.-H., et al. (2024). Artificial intelligence-based prediction of neurocardiovascular risk score from retinal swept-source optical coherence tomography-angiography. *Sci. Rep.* 14:27089. doi: 10.1038/s41598-024-78587-w
- Ghosal, S., and Sinha, B. (2022). Using machine learning to assess cardiovascular risk in type 2 diabetes mellitus patients without established cardiovascular disease: the mark-2 analysis. *Int. J. Diabet. Technol.* 1, 114–118. doi: 10.4103/ijdt.ijdt_8_23
- Hippisley-Cox, J., Coupland, C. A. C., and Brindle, P. M. (2017). Development and validation of qrisk3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 357:j2099. doi: 10.1136/bmj.j2099
- Ichikawa, Y., and Hukushima, K. (2024). “Learning dynamics in linear vae: posterior collapse threshold, superfluous latent space pitfalls, and speedup with KL annealing,” in *International Conference on Artificial Intelligence and Statistics* (New York: PMLR), 1936–1944.
- Keane, P. A., Grossi, C. M., Foster, P. J., Yang, Q., Reisman, C. A., Chan, K., et al. (2016). Optical coherence tomography in the uk biobank study – rapid automated analysis of retinal thickness for large population-based studies. *PLoS ONE* 11:e0164095. doi: 10.1371/journal.pone.0164095
- Khozeimeh, F., Sharifrazi, D., Izadi, N. H., Joloudari, J. H., Shoeibi, A., Alizadehsani, R., et al. (2022). RF-CNN-F: random forest with convolutional neural network features for coronary artery disease diagnosis based on cardiac magnetic resonance. *Sci. Rep.* 12:11178. doi: 10.1038/s41598-022-15374-5
- Kingma, D. P., and Welling, M. (2013). Auto-encoding variational bayes. *CoRR*, abs/1312.6114. doi: 10.48550/arXiv.1312.6114
- Langner, S., Terheyden, J. H., Geerling, C. F., Kindler, C., Keil, V. C. W., Turski, C. A., et al. (2022). Structural retinal changes in cerebral small vessel disease. *Sci. Rep.* 12:9315. doi: 10.1038/s41598-022-13312-z
- Li, Y., Sperrin, M., and van Staa, T. P. (2019). R package “QRISK3”: an unofficial research purposed implementation of ClinRisk’s QRISK3 algorithm into r. *F1000Research* 8:2139. doi: 10.12688/f1000research.21679.1
- Lindstrom, M. (2023). *Deaths from Cardiovascular Disease Surged 60% Globally over the last 30 Years: Report*. Geneva: World Heart Federation.
- Livingstone, S., Morales, D. R., Donnan, P. T., Payne, K., Thompson, A. J., Youn, J.-H., et al. (2021). Effect of competing mortality risks on predictive performance of the QRISK3 cardiovascular risk prediction tool in older people and those with comorbidity: external validation population cohort study. *Lancet. Healthy Longevity* 2, e352–e361. doi: 10.1016/S2666-7568(21)00088-X
- Lucas, B. D., and Kanade, T. (1981). “An iterative image registration technique with an application to stereo vision,” in *International Joint Conference on Artificial Intelligence* (Vancouver, BC: International Joint Conferences on Artificial Intelligence Organization).
- Maldonado-Garcia, C., Bonazzola, R., Ravikumar, N., and Frangi, A. F. (2022). “Predicting myocardial infarction using retinal oct imaging,” in *Annual Conference on Medical Image Understanding and Analysis* (Cham: Springer International Publishing).
- Matuleviciūtė, I., Sidaraitė, A., Tatarunas, V., Veikutienė, A., Dobilienė, O., and Zaliuniene, D. (2022). Retinal and choroidal thinning—a predictor of coronary artery occlusion? *Diagnostics* 12:2016. doi: 10.3390/diagnostics12082016
- Nusinovi, S., Rim, T. H., Yu, M., Lee, G., Tham, Y.-C., Cheung, N., et al. (2022). Retinal photograph-based deep learning predicts biological age, and stratifies morbidity and mortality risk. *Age Ageing* 51:afac065. doi: 10.1093/ageing/afac065
- Park, H.-Y. L., Park, Y. G., Cho, A. H., and Park, C. K. (2013). Transneuronal retrograde degeneration of the retinal ganglion cells in patients with cerebral infarction. *Ophthalmology* 120, 1292–1299. doi: 10.1016/j.ophtha.2012.11.021
- Poplin, R., Varadarajan, A. V., Blumer, K., Liu, Y., McConnell, M. V., Corrado, G. S., et al. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* 2, 158–164. doi: 10.1038/s41551-018-0195-0
- Rais, K., Amroune, M., Benmachiche, A., and Haouam, M. Y. (2024). Exploring variational autoencoders for medical image generation: a comprehensive study. *arXiv [preprint]* arXiv:2411.07348. doi: 10.48550/arXiv.2411.07348

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frai.2025.1624550/full#supplementary-material>

- Rakusiewicz, K., Kanigowska, K., Hautz, W., and Ziolkowska, L. (2021). The impact of chronic heart failure on retinal vessel density assessed by optical coherence tomography angiography in children with dilated cardiomyopathy. *J. Clin. Med.* 10:2659. doi: 10.3390/jcm10122659
- Runsewe, O. I., Srivastava, S. K., Sharma, S., Chaudhury, P., and Tang, W. W. (2024). Optical coherence tomography angiography in cardiovascular disease. *Prog. Cardiovasc. Dis.* 87, 60–72. doi: 10.1016/j.pcad.2024.10.011
- Stein, D. M., Ishikawa, H., Hariprasad, R., Wollstein, G., Noecker, R. J., Fujimoto, J. G., et al. (2006). A new quality assessment parameter for optical coherence tomography. *Br. J. Ophthalmol.* 90, 186–190. doi: 10.1136/bjo.2004.059824
- Wong, D. Y., Lam, M. C., Ran, A., and Cheung, C. Y. (2022). Artificial intelligence in retinal imaging for cardiovascular disease prediction: current trends and future directions. *Curr. Opin. Ophthalmol.* 33, 440–446. doi: 10.1097/ICU.0000000000000886
- Yang, L., Wu, H., Jin, X., Zheng, P., Hu, S., Xu, X., et al. (2020). Study of cardiovascular disease prediction model based on random forest in eastern china. *Sci. Rep.* 10:5245. doi: 10.1038/s41598-020-62133-5
- Yeung, S. C., You, Y., Howe, K. L., and Yan, P. (2020). Choroidal thickness in patients with cardiovascular disease: a review. *Surv. Ophthalmol.* 65, 473–486. doi: 10.1016/j.survophthal.2019.12.007
- Yu, W., Yang, B., Xu, S., Gao, Y., Huang, Y., and Wang, Z. (2023). Diabetic retinopathy and cardiovascular disease: a literature review. *Diabet. Metabol. Syndrome Obesity* 16, 4247–4261. doi: 10.2147/DMSO.S438111
- Zhou, Y., Chia, M. A., Wagner, S. K., Ayhan, M. S., Williamson, D. J., Struyven, R. R., et al. (2023). A foundation model for generalizable disease detection from retinal images. *Nature* 622, 156–163. doi: 10.1038/s41586-023-06555-x