Rīga Stradiņš University
Faculty of Medicine
second-level professional higher education
study programme “Medicine”

Deep learning (DL) Applications in Cardiovascular Risk Prediction Using Retinal Fundus Imaging

RESEARCH PAPER-LITERATURE REVIEW

Author of the Research Paper:
Walid Dawoud
Student ID No 045816

_____
/signature/

Research Supervisor:
Dr. Elīna Pašuna

RSU Department of Ophthalmology

_____
/signature/

Consultant:
Dr. Vincenzo Rizzuto

RSU Department of Ophthalmology

_____
/signature/

Riga, 2025

LIST OF ABBREVIATIONS AND TERMS

In this research, diverse statistical metrics are used to assess the reliability and clinical usefulness of predictive models. Without context, these metrics can often be challenging to interpret, as each emphasises different aspects of model performance. To facilitate comprehension, this review provides a concise glossary of commonly cited measures—such as discrimination indices, calibration, and reclassification metrics—alongside standard diagnostic measures, such as sensitivity, specificity, and predictive values.

Table 1. List of Abbreviations

Abbreviations	Full Term / Definition	Context in This Review
AI	Artificial Intelligence	Broad computational approaches used in medical imaging and analytics
ML	Machine Learning	Data-driven AI methods learn feature–outcome associations
DL	Deep Learning	Multilayer neural networks for end-to-end image analysis
CNN	Convolutional Neural Network	Core DL architecture for fundus image tasks
Transformer / Swin	Self-attention DL architecture (e.g., Swin Transformer)	Captures long-range image context; stability in retinal CVD prediction
Reti-CVD	Retinal Cardiovascular Disease biomarker	DL risk enhancer predicting 10-year CVD risk/events
RetiCAC	Calcium in the coronary artery (equivalent)	DL model estimating CAC-equivalent risk from fundus images
Reti-WHO	Retinal WHO 10-year risk	DL reproduction of WHO 10-year CVD risk from fundus photos
XMACE	Explainable MACE model (two-stage)	DL first predicts 10 risk factors → and then 5-year MACE
rpCVD	CVD model of retinal photography (primary care)	Pragmatic screening workflow in clinics
DL-FAS	Deep-Learning Funduscopic Atherosclerosis Score	Fundus-based surrogate related to carotid atherosclerosis/mortality
CAC	Coronary Artery Calcium	CT biomarker of coronary atherosclerosis
CIMT	Carotid Intima-Media Thickness	Ultrasound atherosclerosis surrogate
PWV	Pulse Wave Velocity	Arterial stiffness marker (vascular ageing)
baPWV	Brachial-Ankle Pulse Wave Velocity	Peripheral arterial stiffness index
CRAE / CRVE	Equivalent Central Retinal Arteriolar / Venular	Quantitative vessel calibre metrics
FRS	Framingham Risk Score	Traditional 10-year risk calculator
SCORE / SCORE2	Systematic Coronary Risk Evaluation (2)	European CVD risk models
QRISK3	QResearch Cardiovascular Risk Algorithm v3	UK 10-year CVD risk calculator
PCE	Pooled Cohort Equations	US AHA/ACC 10-yr ASCVD risk equation

ASCVD	Atherosclerotic Cardiovascular Disease	Composite outcome (e.g., MI, stroke, CV death)
MACE	Major Adverse Cardiovascular Events	Composite clinical endpoint
MI	Myocardial Infarction	Event endpoint in several analyses
BP	Blood Pressure	Systemic risk factor (SBP/DBP)
SBP / DBP	Systolic / Diastolic Blood Pressure	Continuous targets in DL regression
HbA1c	Glycated haemoglobin	Glycaemic control marker
HDL / LDL	High- / Low-Density Lipoprotein	Lipid biomarkers
eGFR	Estimated Glomerular Filtration Rate	Kidney function marker (external validation context)
AUC	Area Under the Curve	Discrimination metric
ROC	Receiver Operating Characteristic	Curve relating sensitivity/specificity
C-index	Concordance Index	Time-to-event discrimination (survival models)
NRI / IDI	Net Reclassification / Integrated Discrimination Improvement	Incremental value over baseline scores
HR	Hazard Ratio	Relative risk of survival models
MAE	Mean Absolute Error	Accuracy for continuous predictions
Grad-CAM	Gradient-Weighted Class Activation Mapping	DL saliency visualisation
SHAP	Shapley Additive Explanations	Feature-importance explanation
IGOS	Integrated Gradients Occlusion Sensitivity	Saliency method used in XMACE
SIVA-DLS	Singapore I Vessel Assessment – DL System	Automated retinal calibre measurement
QUARTZ	Quantitative Analysis of the retinal vessel system	Automated vasculometry (tortuosity, width, area)
AI-SaMD / SaMD	AI Software as a Medical Device	Regulatory designation (e.g., RetiCVD)
UKB	UK Biobank	Major training/validation cohort
NHIS	National Health Insurance Service (Korea)	Population cohort context
CMERC-HI	Cardiovascular and Metabolic Disease Aetiology Research Centre—Health Imaging	Korean cohort used in validation
PLWH	People Living With HIV	Special subgroup for CVD risk evaluation
DXA	Dual-Energy X-ray Absorptiometry	Body composition modality in multimodal models
WHO	World Health Organisation	Source of guideline risk calculators

Table 2. Key Metrics and Technical Terms

Metric / Term	**What It Measures**	**How to Interpret**	**Why Does It Matter for This Review**
AUC / ROC	Discrimination of binary outcomes	0.5 = chance; 1.0 = perfect	Standard comparator for DL vs FRS/SCORE/QRISK3
C-index	Discrimination for time-to-event outcomes	Higher is better; similar to AUC for survival data	Needed for event prediction (MI, MACE, mortality)
Calibration (slope/intercept)	Agreement of predicted vs observed risk	Ideal: slope ≈ 1, intercept ≈ 0	Ensures DL risk is clinically reliable and usable
NRI / IDI	Improvement in baseline risk models	Positive values = better reclassification/discrimination	Shows DL adds value beyond traditional scores

Sensitivity / Specificity	True-positive / true-negative rates	Balance depends on screening vs diagnosis goals	Evaluate DL screening for hypertension, CAC, etc.
MAE	Average absolute prediction error	Lower = better	Accuracy for continuous targets (BP, age, HbA1c)
HR (Cox model)	Relative hazard per unit increase	HR > 1 indicates a higher risk	Quantifies prognostic value of DL biomarkers
External Validation	Testing in an independent cohort	Preserved AUC + good calibration	Demonstrates transportability/generalisability
Recalibration	Adjusting the model intercept/slope to cohort	Reduces systematic over/under-prediction	Critical when moving from UKB to EyePACS/Kenya/etc.
Transportability	Stability across populations/cameras	Minimal performance drop after transfer	Addresses bias, fairness, and domain shift
Reclassification Analysis	Movement between risk categories	Correct up/down-moves = clinical utility	Key for borderline/intermediate risk decisions
Two-Stage Modelling (e.g., XMACE)	Predict risk factors → and then events	Transparent intermediate features	Improves interpretability and calibration
Multimodal DL	Imaging + clinical variables	The larger variance is explained	Often, single-modality models outperform.
Vasculometry (calibre, tortuosity, area)	Quantitative geometry of vessels	Deviations indicate microvascular damage	Physiological bridge from retina to systemic risk
Retinal Age Gap	DL-predicted age – chronological age	Larger gap = accelerated vascular ageing	Linked to arterial stiffness and CVD events
Grad-CAM / SHAP / IGOS	Saliency/attribution methods	Highlight decision-relevant regions	Builds clinician trust; error analysis
Case-Control Sampling	Enrichment for scarce events	Balances classes when events are rare	Stabilises training for stroke/MI datasets
Cross-Validation	Internal resampling evaluation	Reduces variance in small datasets.	Common in early feasibility studies
Survival Analysis (Cox)	Time-to-event modelling	Accounts for follow-up time/censoring	Suited to longitudinal cohorts (UKB, NHIS)

TABLE OF CONTENTS

INTRODUCTION

Cardiovascular disease (CVD) is the main cause of morbidity and mortality worldwide, accounting for nearly one-third of all deaths worldwide (Al-Absi et al., 2022). The increasing prevalence of cardiovascular disease (CVD) has led to the development of multivariable risk assessment tools such as the Framingham Risk Score (FRS), Systematic Coronary Risk Evaluation (SCORE) and QRISK3. These tools are designed to pinpoint people who may benefit from early preventive measures. These traditional models are well-established in guidelines and are used in clinical settings. However, they are dependent on invasive techniques and require resources such as blood pressure readings, cholesterol analyses, and laboratory tests. This dependency restricts their scalability and accessibility in primary care settings and settings with limited(Hu et al., 2023; Qian et al., 2024)

Furthermore, conventional risk calculators are predominantly developed based on Western populations and frequently need adjustments when used for other groups to ensure precision and prevent incorrect classifications. Thus, there is a discrepancy in the efficacy of prediction between different ethnic and socioeconomic populations(Hu et al., 2025; Lui et al., 2023; Qian et al., 2024). These variations have increased the interest in developing more fair and non-invasive tools for assessing cardiovascular risk.

In this context, the retina—commonly referred to as a "window to the body,"—has become a significant site for evaluating systemic vascular health. Retinal fundus imaging gives the opportunity for direct observation of the microvascular network, which has important structural and physiological similarities to the coronary and cerebral circulations. Pathological features of the retina, such as arteriolar narrowing, venular dilation, reduced vessel density, increased tortuosity, and microaneurysms, indicate underlying systemic vascular damage and have been consistently linked to cardiometabolic disorders, including hypertension, diabetes, and atherosclerosis(Rim et al., 2021; Rudnicka et al., 2022; Wong et al., 2024; L. Zhang et al., 2020). The physiological connection between the morphology of the retina and cardiovascular outcome is further strengthened by emerging deep learning (DL) based biomarkers such as, the retinal age gap and the Reti-WHO score. These biomarkers are associated with arterial stiffness, the occurrence of myocardial infarction, stroke, and mortality (Nusinovici et al., 2022; W. Zhang et al., 2023; Zhu, Hu, et al., 2022).

Recent developments in artificial intelligence (AI), particularly deep learning (DL), have expanded our ability to analyse and integrate a complex set of multimodal data such as imaging, health records, and genomics to help detect, prognosis, and improve clinical decisions. DL is a branch of AI built upon multi-layered artificial neural networks, where computers use

inputs such as medical images and learn hierarchal data representations and interpret complex patterns. Thus, create an end-to-end model capable of making predictions without the need for manual engineering of the features(Grzybowski, 2025).

In the field of ophthalmology, it has facilitated an automation process to extract and predict individual cardiovascular risk factors directly from fundus images(Gerrits et al., 2020; Wang et al., 2024; White et al., 2024). It has led to development of models that predict surrogates including coronary artery calcium (CAC) and carotid intima-media thickness (CIMT), and more recently to directly assess composite cardiovascular disease (CVD) risk or predict future cardiovascular events(Chang et al., 2020; Rim et al., 2021; Tseng et al., 2023; Vaghefi et al., 2024). Further advances in interpretability and validation, such as the development of automated vasculometry pipelines and significant regulatory achievements such as the pivotal Reti-CVD trial, have advanced the clinical application of deep learning (DL) models of the retina. (Hu et al., 2025; C. J. Lee et al., 2023; Rudnicka et al., 2022).

This review of the literature aims to evaluate and synthesise existing evidence on deep learning (DL) applications in retinal imaging to predict cardiovascular risk. Review the physiological and methodological principles of the field, trace the development of DL techniques from estimating risk factors to predicting events, and critically assesses aspects such as interpretability, clinical benchmarking, and validation frameworks. It also investigates ongoing challenges, including generalisability, data quality, and clinical integration, that need to be addressed to ensure the safe, fair, and extensive use of retinal DL systems.

RESEARCH METHODOLOGY

1. STUDY DESIGN AND APPROACH

This review of the literature employs a thematic and narrative strategy to compile recent findings on the application of deep learning (DL) techniques to the imaging of the retinal fundus to predict cardiovascular risk. The review highlights methodological patterns, performance trends, and clinical significance. The thematic design was chosen due to the wide variety of existing studies, which differ greatly in terms of data sources, target outcomes, algorithms, and validation methods. The aim is to pinpoint common themes, methodological advancements, and limitations that define current research in this area rather than to aggregate effect sizes. This narrative method focuses on synthesizing main themes and emerging trends instead of offering an exhaustive collection of evidence. The metrics reported illustrate individual studies and should be understood within their specific context, as they are not drawn from pooled or meta-analytic estimates. This approach is particularly suitable for a field that is rapidly evolving and remains diverse in study design, outcome definitions, and validation frameworks. It facilitates the integration of methodological insights and clinical implications while acknowledging that formal quantitative generalization is limited by the variability present in current research.

1.1. Research Question

The review is centred on the main research question:
"Is the deep learning (DL) model used in retinal fundus imaging capable of providing reliable and clinically valuable predictions about cardiovascular risk"?
This study aims to evaluate the predictive reliability of retinal deep learning (DL) models by assessing their precision, calibration, and external validity, along with their clinical utility compared to or combined with conventional cardiovascular risk assessment methods.

1.2. Data Sources and Literature Research

A comprehensive electronic search was conducted using the PubMed and Scopus databases. This search covered the period from January 2020 to March 2025, with the intention of encompassing the recent increase in deep learning (DL) applications in the realm of medical imaging.

1.3. Inclusion and exclusion criteria

The process of selecting studies was organised into several steps. It began with an initial screening using titles and abstracts to exclude irrelevant articles, followed by an in-depth full-text review to verify eligibility. Each study obtained was assessed for methodological quality, relevance to the research goal, and reporting clarity.

1.3.1. Inclusion criteria:

Studies were included based on the following criteria:

- Original Peer-reviewed research studies or systematic reviews published in English between 2020 and 2025.
- Studies employing deep learning (DL) or machine learning (ML) methods applied to imaging of the retinal fundus.
- Outcomes that include the prediction of the risk of cardiovascular disease (CVD), hypertension, stroke, coronary artery disease, atherosclerosis, or composite CVD endpoints such as MACE or ASCVD.
- Studies that reported quantitative performance metrics (e.g. AUC, C-index, sensitivity/specificity, HR, NRI).
- Human data used with a clear description of the data set and the validation procedure (internal or external)

1.3.2. Exclusion criteria:

The exclusion was based on the following criteria:

- Non-original works such as abstracts, editorials, correspondences, commentaries, and conference summaries without full papers and peer reviewed publications.
- Studies focussing solely on non-cardiovascular outcomes (e.g., Alzheimer's, CKD) or non-retinal imaging (e.g., OCT, CT).
- Paediatric or animal studies.
- Studies without performance evaluation metrics.

1.4. Data collection

For each included study, relevant data was extracted into a structured summary table, capturing:

- Author(s), year, and country of origin.
- Dataset characteristics (sample size, population, source, imaging modality).
- Type of deep learning (DL) model (e.g., CNN, ResNet, ensemble, multimodal).
- Target outcome(s).
- Performance metrics (AUC, C-index, calibration, NRI/IDI, sensitivity, specificity).
- Validation strategy (internal/external).
- Key findings, limitations, and interpretability features.

1.5. Screening and selection of studies

The search results were organised in a reference management system, where duplicates were automatically removed. The selection was done in two stages:

1. Screening of the title and abstract to exclude studies irrelevant to retinal imaging, AI, or cardiovascular outcomes
2. Full-text review to confirm relevance and adherence to inclusion criteria.

1.6. Data Presentation:

The results are presented in narrative and tabular formats.

- The main text provides a summarised overview of methodological trends, datasets, model architectures, performance trends, and interpretability techniques.
- The appendices include a full summary table of the studies that include datasets, model types, validation design, and quantitative results of every study

1.7. Use of supportive tools

In this literature review, artificial intelligence tools were used to create an outline, organize and subdivide topics, analyze methodological areas in some studies where clarification was needed, help creating tables and charts and sometimes for translation purposes.

RESULTS

2. RETINA AS A BIOMARKER OF CARDIOVASCULAR HEALTH

2.1. The Retina as a Non-Invasive Window to Systemic Vascular Health

The retina provides a unique and non-invasive opportunity to observe human microcirculation directly. Retinal arterioles and venules reflect the microvascular networks found in the heart and brain, and pathological changes in the coronary or cerebral circulation are often mirrored in a similar manner in the retinal fundus. By studying, microvascular changes in retinopathy —such as, microaneurysms, haemorrhages and subtle changes in vessel metrics —we can gain insight into the cardiovascular health and predict potential future cardiovascular events(Rim et al., 2021; Zhu, Hu, et al., 2022). There is strong evidence from epidemiological studies that alterations in the systemic microvasculature are mirrored in the retinal vasculature. However, the process of assessing alterations in the retinal vasculature is time consuming and requires expertise, which has limited the wider use of these methods in primary care and non-ophthalmic settings (Coronado et al., 2021; Rudnicka et al., 2022).

The link between cardiovascular health and retinal microvascular abnormalities has been demonstrated by a substantial number of studies. Retinal fundus photographs taken with Topcon TRC-NW6S and Canon CR6-45 NM non-mydriatic retinal cameras primarily capturing the macular and optic disk regions of the central retina which contain vessels representing the systemic microvascular health, reveal indicators that are associated with cardiometabolic conditions such as hypertension, hyperglycaemia (diabetes), and dyslipidaemia(Gerrits et al., 2020; Son et al., 2020; L. Zhang et al., 2020) . In this context, Retinal microvascular measurements taken together with DL can extract characteristic vascular changes related to coronary atherosclerosis. An atherosclerosis score derived from fundus imaging using the canon CR-2 model independently predicted cardiovascular mortality and improved risk reclassification beyond the traditional risk score such as the Framingham model(Chang et al., 2020). Others DL models have used macula centred, paired colour retinal photographs of the participants obtained using the AFP-210 and TRC-NW8 non-mydriatic retinal cameras, to estimate coronary artery calcium (CAC) directly from fundus images, categorized future cardiovascular risk and improved risk classification in borderline or intermediate risk groups(Rim et al., 2021; Tseng et al., 2023).

Recent research shows that algorithmic evaluation of digital fundus images centred on the macula can be used to predict cardiovascular risk similarly to traditional methods. This analysis can improve conventional risk assessment, highlighting the importance of the eye as a

non-invasive window into broader cardiovascular health(Rim et al., 2021; Tseng et al., 2023; Vaghefi et al., 2024; Wang et al., 2024).

2.2. Microvascular Changes Visible in the Retina: (Classical retinal microvascular signs (narrowing, AV nicking, retinopathy)

Fundus imaging exposes microvascular features that are related to cardiovascular health, such as retinal microvascular abnormalities, vessel calibre, and vascular geometry. These attributes have been associated with cardiometabolic risk and macrovascular outcomes in extensive cohort studies (Rim et al., 2021). Traditional microvascular abnormalities such as arteriolar narrowing, arteriovenous nicking, and retinopathy lesions are recognised as markers linked to cardiovascular disease and its related risk factors. Studies have analysed multimodal evidence by integrating retinal and clinical data to provide additional evidence supporting these connections within disease-specific contexts, such as cardiometabolic risk associated with diabetes or HIV (Al-Absi et al., 2022)

2.3. Quantitative vasculometry: calibre, venule/arteriole ratios, tortuosity, fractal metrics

Quantitative measurements of vessel calibre serve as an important prognostic value for cardiovascular disease. For example, there is an independent relationship between the onset of myocardial infarction and the narrowing of the arterioles and the widening of the venules. This association has been shown in recent studies that deep-learning approaches to measure vessel calibre which further indicate the potential screening benefits when integrated into standard ophthalmic examinations(Wong et al., 2024). In addition to quantitative calibre measurement, characteristic changes in retinal geometry play a significant role in risk assessment. Tortuosity is one geometric value that has been used in AI-driven retinal vasculometry to predict large-scale occurrences of circulatory mortality, myocardial infarction, and stroke (Rudnicka et al., 2022). The difference in biological age and the overtime age gap of the retina is associated with arterial stiffness and the occurrence of cardiovascular disease (CVD). This suggests that ageing processes can be observed overtime in the fundus and are related to potential future cardiovascular health events (Nusinovici et al., 2022; Zhu, Chen, et al., 2022)

3. DEEP LEARNING (DL) IN MEDICAL IMAGING

3.1. Brief overview of AI/ML vs DL

Artificial intelligence (AI) in medical imaging includes a variety of computational methods that are designed to recognise patterns and assist decision making by using intricate data sets. In this context, machine learning (ML) refers to algorithms that can learn statistical connexions between features and outcomes, without the need for explicit programming. Deep learning (DL), is a subset of machine learning (ML) which employs multilayer neural networks to derive features directly from unprocessed images, facilitating end-to-end tasks like detection, classification, and regression in ophthalmic imaging (Grzybowski, 2025).

3.2. Common architectures: CNNs, transfer learning, and ensemble approaches

Within the field of deep learning (DL) architectures, convolutional neural networks (CNNs) form the foundation of medical image analysis due to their ability to capture spatial hierarchies and local patterns. In ophthalmology, widely used convolutional neural network architectures such as **Inception-V3**, **ResNet-50**, and **MobileNet-V2** have shown strong efficacy in disease classification and regression tasks within retinal imaging. (Gerrits et al., 2020; White et al., 2024)

Convolutional Neural Networks (CNNs) have been able to predict cardiovascular risk by estimating blood pressure, hypertension, vascular ageing, and cardiovascular events, especially when used alongside clinical variables in multimodal approaches. The technique of transfer learning, which involves initialising networks that are pre-trained on extensive natural-image datasets like ImageNet, has been crucial in tackling the challenge of limited medical datasets. This approach improves convergence speed and improves the generalizability of models when they are fine-tuned for retinal images(Wang et al., 2024; W. Zhang et al., 2023)

Recent research has adopted collective methods by merging results from multiple convolutional neural networks (CNNs) or combining deep learning (DL) predictions with conventional clinical variables, with the aim of improving prediction accuracy and reducing overfitting. In addition, current studies have concentrated on employing transformer-based encoders to stabilise long-range contexts and increase robustness. For example, the Swin Transformer has been designed to proficiently capture extensive contextual information in retinal images. These models have demonstrated significant stability in predicting cardiovascular risk compared to traditional CNNs(Vaghefi et al., 2024; W. Zhang et al., 2023).

4. EVOLUTION OF DL METHODOLOGIES FOR RETINAL CVD RISK ASSESSMENT

Incorporation of deep learning (DL) techniques into retinal fundus imaging to assess cardiovascular disease (CVD) risk has evolved through three distinct methodological phases. It began with establishing biological feasibility and has advanced to demonstrate practical clinical application. Initially, the focus was on simple proof-of-concept models, which have now evolved into sophisticated systems designed for direct clinical risk prediction. This section details the methodological journey from predicting individual risk factors to forecasting clinical outcomes.

4.1. Predicting Individual Risk Factors

Individual systemic risk factors such as age, sex, blood pressure, glycaemic status, lipid profile, and smoking habits are integral parts of cardiovascular risk assessments. The use of deep learning (DL) in the analysis of retinal fundus images focusses on successfully predicting these risk factors, which would boost the medical reasoning for using retinal imaging as a systemic indicator of vascular health (Gerrits et al., 2020; Qian et al., 2024; Rim et al., 2021; L. Zhang et al., 2020)

Hypertension is one of the most important adjustable risk factors behind cardiovascular illness and death. The use of scalable, non-invasive screening methods could improve early detection and management before end organ damage develops (Gerrits et al., 2020; L. Zhang et al., 2020). The retina is a compelling target for this purpose due to its microvascular network, which provides reliable reflection of cumulative vascular stress, rather than short-term hemodynamic fluctuations.(Gerrits et al., 2020; Wang et al., 2024).

The application of deep learning (DL) to retinal images has demonstrated **moderate accuracy** in estimation of blood pressure to classify hypertension, it has outperformed demographic only models but has fallen short of achieving full diagnostic precision(Gerrits et al., 2020; Wang et al., 2024; White et al., 2024; L. Zhang et al., 2020).

Early research in deep learning (DL) with retinal fundus images showed promise. In a rural Chinese population (n=625; 1,222 images), the **CNN Inception-v3 image** achieved an accuracy of 68.8% in detecting hypertension (L. Zhang et al., 2020). Similarly, In the Qatar biobank (n=3,000; 12,000 images), the **MobileNet-V2** models predicted systolic and diastolic blood pressure with a mean absolute error (MAE) 8.96 and 6.84 mmHg, respectively (Gerrits et al., 2020).

More recently, a prospective validation cohort study in Kenya, using algorithms trained in UK Biobank, reported systolic and diastolic pressure readings with MAE values of 12.32 and 7.98 mmHg, respectively, demonstrating practicality across populations but also showing the need for recalibration when models are transferred across different ethnic and imaging domains (White et al., 2024)Large scale biobank analyses have further supported this potential, by predicting blood pressure within a relevant clinical margins of error and simultaneously identifying correlated risk factors such as age, sex, smoking and BMI directly from fundus images (Wang et al., 2024). However, predictions have been significantly confounded by age and sex, which, unless adjusted, can raise the risk of overly optimistic predictions (Gerrits et al., 2020).

However, demographic factors such as age and sex have significantly impacted both vascular morphology and predictions derived from deep learning (DL). This has confounded performance, which, if not adequately controlled, can potentially lead to inflated perceived accuracy(Gerrits et al., 2020). Most studies utilise convolutional neural networks (CNNs) that are pre-trained on ImageNet, such as Inception-V3 and ResNet-50, and commonly integrate data from both eyes while primarily relying on internal cross-validation. External validations are infrequent and depend on clinical reference measures, which themselves are subject to variations in age, sex, and image quality (Gerrits et al., 2020; White et al., 2024; L. Zhang et al., 2020).

In conclusion, these models demonstrate a consistent biological association between retinal microvasculature patterns and the regulation of systemic blood pressure. Their primary benefits include being non-invasive, scalable, and providing reproducible signals across different cohorts. However, they also have drawbacks, such as limited effect sizes, inconsistent labelling, and the possibility that algorithms may predict demographic proxies instead of actual hemodynamic states. Thus, due to its current performance and limited generalisability, it is best used as an additional tool for assessing cardiovascular risk rather than as a replacement for diagnostic assessments in systems predicting cardiovascular risk.

4.2. Other Systemic Risk factors

In addition to blood pressure, deep learning (DL) models have demonstrated that demographic characteristics such as age and sex are prominently encoded in retinal images. Systems developed using extensive biobank datasets can estimate chronological age with mean absolute errors as low as 2.8 years and can determine sex with AUCs nearing 0.97, indicating the existence of rich vascular and textural signals(Gerrits et al., 2020).

Retinal imaging similarly captures certain indicators of glycaemic control: models designed to predict glycated haemoglobin (HbA1c) reach mean absolute error (MAE) values of approximately 0.6 to 0.8%. These values are adequate for differentiating between normoglycemia and diabetes, yet they are not sufficient to substitute for laboratory testing (Gerrits et al., 2020; Qian et al., 2024).

Smoking status can be predicted with about 80% precision by analysing vessel calibre and peripapillary pigmentation(Gerrits et al., 2020). On the other hand, the models have consistently shown low performance in predicting lipid biomarkers such total, LDL, HDL, triglyceride(Gerrits et al., 2020; Qian et al., 2024).

Contemporary models either directly incorporate these variables or utilise interpretable two-phase processes. For instance, the XMACE model, which initially calculates ten risk factors—age, sex, BMI, SBP, DBP, HbA1c, HDL, LDL, triglycerides, and smoking—from a single retinal image and subsequently uses these calculations to forecast major adverse cardiovascular events (MACE) over a five-year period, thus correlating image-derived features with established clinical predictors(Qian et al., 2024).

In general, the results indicate that deep-learning algorithms can identify several traditional risk factors directly from retinal images, with varying accuracy levels. The most robust and widely applicable signals are found for age, sex, and blood pressure, while predictions for lipid and glucose levels are limited by biological complexity and diverse labelling. These investigations establish the methodological and physiological foundations for the more sophisticated surrogate and composite-risk models described in subsequent sections.

4.3. Retinal Biomarker Predicting Surrogates of Atherosclerosis

The application of deep learning (DL) to retinal fundus images has been used to estimate coronary artery calcium (CAC) scores, which is an acknowledged imaging biomarker for assessing cardiovascular risk.

The RetiCAC algorithm was trained to predict the CAC score from retinal photographs and categorise individuals according to future cardiovascular risk. RetiCAC outperformed individual clinical parameters, including age, blood glucose level, or smoking habits, in predicting the presence of calcium in the coronary artery. When RetiCAC predictions were combined with pooled cohort equations (PCE), the outcomes in risk stratification improved, particularly for individuals in the intermediate and borderline risk groups. (Rim et al., 2021).

Based on this work, a clinically orientated extension of RetiCAC was developed, termed RetiCVD, which is a deep learning-based retinal biomarker system that has received regulatory

approval from the Korean Ministry of Food and Drug Safety (K-MFDS) as a Software-as-a-Medical Device (AI-SaMD).). This approval was granted following a pivotal multi-centre validation study conducted by (C. J. Lee et al., 2023) which illustrated its ability to predict cardiovascular disease in both Korean and UK populations.

When recalibrated for the UK Biobank population, this model effectively categorised individuals into three distinct risk groups and showed clinical value, particularly in the borderline QRISK subgroup (7.5-10% 10-year risk) where it accurately reclassified patients and aligned risk with treatment criteria. This shows the role of DL as risk enhancer alongside traditional risk scores (Tseng et al., 2023).

The efficacy of Reti-CVD was further validated in a large Korean cohort where Reti-CVD was used as an AI Software as a Medical Device (AI-SaMD). The model predicted cardiovascular risk events with performance that was non-inferior to CAC and better than CIMT and baPWV for prognostic risk stratification(C. J. Lee et al., 2023). Together, these findings show that Reti-CVD can be used as a credible and scalable method in combination with a traditional risk score for assessment, particularly for patients in the borderline or intermediate risk group.

Other research teams have focused on various atherosclerotic indicators. For example, (Son et al., 2020) developed a deep learning (DL) model designed to identify unilateral and bilateral abnormal coronary artery calcification (CAC) using fundus images, where the model achieved an Area under the curve (AUC) of 0.823, and Another research group (Chang et al., 2020) developed a model capable of predicting carotid-ultrasound-determined atherosclerosis (CIMT) with an AUC of 0.713, demonstrating the link between vascular characteristics derived from fundus images and accumulation of systemic plaque. Although the study is waiting for validation, the investigations exemplify how deep learning (DL) models can be used to derive surrogate biomarkers from non-invasive imaging techniques(Chang et al., 2020; Son et al., 2020).

In addition to CAC substitutes, various retinal biomarkers have been investigated as indicators of systemic atherosclerosis and vascular ageing. The traditional method of assessing the diameter of the retinal vessel is resource intensive and requires highly skilled professionals, limiting its application in primary clinical settings. In response to this, the Singapore I Vessel Assessment deep-learning system (SIVA-DLS) was introduced, which was able to use heat maps from colour fundus photographs and provide the width of the retinal vessel. SIVA-DLS achieved high intra-class correlation coefficients (0.82–0.95) compared to manual measurements, demonstrating reliable standardisation for vascular metrics. The SIVA DLS system was

later used to validate the association of retinal vessel measurements with myocardial infarction (Wong et al., 2024).

4.4. Direct Prediction of Composite Risk Scores & Clinical Events

The subsequent methodological progression shifts the focus from surrogate vascular biomarkers to direct clinical outcomes. At this stage, deep learning (DL) frameworks are trained in a comprehensive way to predict cardiovascular events such as major adverse cardiovascular events (MACE), myocardial infarction (MI) and stroke or replicate validated composite risk scores used in clinical settings. The transition from predicting individual risk factors to predicting meaningful outcomes represents an important advancement towards prevention and management strategies.

4.4.1. Coronary Artery Disease and Myocardial Infarction

Coronary artery disease (CAD) and myocardial infarction (MI) continue to be among the leading causes of early death worldwide; therefore, early detection of high-risk individuals is important for prevention and specific treatments. Assessment of retinal fundus images offers valuable information about microvascular health on a systemic level, allowing DL models to identify patterns linked to coronary atherosclerosis and forecast possible future cardiac incidents(Rim et al., 2021; Wong et al., 2024).

In extensive biobank cohorts, automated retinal vasculometry has shown that narrower arterioles and wider venules are associated with increased atherosclerotic burden and the occurrence of cardiac events (Rudnicka et al., 2022; Wong et al., 2024). In the UK Biobank study (N≈35,900), arteriolar narrowing derived from DL independently predicted MI, produced modest gains in discrimination when combined with clinical models (Area under the curve (AUC) from 0.782 to 0.787; continuous NRI ≈18%)(Wong et al., 2024). These findings reveal that subtle changes in retinal blood vessels, which might not be visible to the naked eye, have the potential to represent systemic cardiovascular risk.

In addition to specific vessel attributes, multimodal deep learning (DL) techniques that integrate fundus imaging, with demographic and clinical variables such as, visualising specific regions linked with vascular tortuosity and calibre have demonstrated strong effectiveness in predicting ischemic heart disease (AUROC 0.78 to 0.87)(C. J. Lee et al., 2023; W. Zhang et al., 2023). These combined retinal imaging methods with traditional risk factors, validated in European and East Asian cohorts, show consistent associations between predicted vascular risk

patterns of DL and major adverse cardiovascular events (MACE)(C. J. Lee et al., 2023; Rim et al., 2021; Tseng et al., 2023; Vaghefi et al., 2024; Wong et al., 2024).

In the Korean NHIS cohort, the DL derived funduscopic atherosclerosis score (DL-FAS) was trained on the characteristics of the carotid ultrasound, which predicted cardiovascular mortality beyond the Framingham Risk Score (FRS) (HR 1.63, 95% CI 1.15–2.31, IDI 20.45%, NRI 29.5%) (Chang et al., 2020). Building upon these findings **Reti-CVD** was developed and validated, an imaging biomarker trained to deduce coronary-artery-calcium-equivalent risk from fundus photos. In its pivotal trial (n ≈ 70 000), participants stratified by Reti-CVD risk groups showed a graded association with composite CVD events (adjusted HR 2.02; 95% CI 1.26–3.24) (C. J. Lee et al., 2023; Tseng et al., 2023). This was subsequently validated in a multi-ethnic Singaporean cohort of Chinese, Malay, and Indian participants(Yi et al., 2023)

In conclusion, deep learning (DL) analysis of retinal fundus images can identify microvascular signs of coronary atherosclerosis and the risk of myocardial infarction. It is marginally effective when compared to imaging-based models such as CAC scoring, it offers a promising, non-invasive method to complement existing risk prediction tools for cardiovascular risk identification (C. J. Lee et al., 2023; Rudnicka et al., 2022; Wong et al., 2024; W. Zhang et al., 2023).

4.4.2. Stroke Risk Prediction

Stroke contributes significantly to long term death and disability worldwide, and an accurate risk prediction has a significant impact on prevention. Retinal imaging non-invasively captures small altercations related to stroke pathophysiology, and the application of DL models enables quantifying these and giving predictions. (Nusinovici et al., 2022; Rudnicka et al., 2022; Zhu, Hu, et al., 2022).

Coronado et al. (2021) evaluated various deep learning (DL) architectures, including VGG19 and Inception-V3, along with a vascular embossed method applied to UK Biobank images. Among 412 stroke patients and 2,060 controls, the vascular-embedding model integrated with gradient boost exhibited superior performance (Area under the curve (AUC) 0.626), surpassing Inception-V3 (AUC 0.499) and VGG19 (AUC 0.548). In an age-specific subgroup VGG19 demonstrated improved performance (AUC 0.714), showing that vascular embeddings can improve the interpretability and efficiency of microvasculature feature learning (Coronado et al., 2021).

In a separate investigation conducted by Zhu et al. in 2022 involving 35,000 UK Biobank participants, it showed that each additional year difference between predicted DL and

chronological age was correlated with a 3% increase in arterial stiffness and a 55% higher risk of cardiovascular events including stroke. Furthermore, these associations were found to be independent of conventional risk factors (Zhu, Chen, et al., 2022). Similar observations were made in Asian cohorts, demonstrating that accelerated retinal ageing is related to increased vascular fragility and cerebrovascular risk (Nusinovici et al., 2022).

Together, these studies illustrate that the use of deep learning (DL) in fundus images can identify microvascular alterations related to age and structure, which are indicative of future potential stroke and cerebrovascular conditions. However, the generalizability is constrained by moderate effect sizes, retrospective study designs, and the absence of extensive, prospective, and multi-ethnic validations(Coronado et al., 2021; Nusinovici et al., 2022; Rudnicka et al., 2022; Zhu, Hu, et al., 2022).

4.4.3. Replication of Established Risk Scores and Composite CVD

Assessing composite cardiovascular risk is crucial in clinical settings, as treatment choices—like starting statins or antihypertensive medications—are informed by multivariable risk assessments rather than individual factors. Enhancing risk classification, particularly for those with borderline or intermediate risk, significantly affects preventive healthcare. (Rim et al., 2021; Yi et al., 2023). Applying deep learning (DL) to retinal images addresses this requirement by predicting established surrogates such as CAC or directly forecasting aggregated risk scores and composite cardiovascular outcomes. This approach, therefore, provides a scalable and non-invasive tool for risk evaluation in primary care settings. (C. J. Lee et al., 2023; Vaghefi et al., 2024)

In the realm of event-based forecasting, numerous studies have used deep learning (DL) models to replicate or supplement the composite risk calculators recommended by the guidelines. The Reti-WHO, capable of forecasting the 10-year CVD outcomes directly from fundus images, achieved comparable performance to traditional WHO scores (Area und The Reti-WHO(AUC) 0.69 vs 0.68). Additionally, Reti-WHO demonstrated enhanced longitudinal stability (W. Zhang et al., 2023).

Similarly, the **XMACE model** uses a two-stage, interpretable design: the first network estimates ten conventional risk factors (age, sex, BMI, SBP, DBP, HbA1c, HDL, LDL, triglycerides, smoking) from a single retinal image, and the second predicts 5-year MACE risk. XMACE achieved a C-statistic of 0.725, outperforming SCORE (+8.2 %) and FRS (+7.1 %) while preserving interpretability (Qian et al., 2024)

Longitudinal studies have established a connection between retinal biomarkers and cardiovascular outcomes. Models that were trained using atherosclerosis markers derived from fundus images, such as CIMT, CAC, and vessel calibre, demonstrated independent associations with the onset of cardiovascular disease and mortality, even after adjusting for the Framingham Risk Score (FRS)(Chang et al., 2020; Rudnicka et al., 2022). Additionally, the retinal age gap, which is the difference between the predicted age of deep learning (DL) and the actual chronological age, has been identified as a marker of vascular ageing and an indicator of future risk of cardiovascular disease (Nusinovici et al., 2022; Zhu, Hu, et al., 2022).

In conclusion, the application of deep learning (DL) to retinal imaging in various datasets can predict notable cardiovascular events and reproduce established composite risk scores with moderate but reliable precision. Key predictors include deep learning-derived measurements of vessel diameter and tortuosity, as well as surrogate scores for atherosclerosis such as DL-FAS and Reti-CVD (Chang et al., 2020; C. J. Lee et al., 2023; Tseng et al., 2023)

The advancement of prediction from individual variables to clinical events has created a future platform for the integration and validation of these models in real world clinical practice. These advances validate the application of retinal deep learning (DL) as a non-invasive tool for assessing risk at the population level and as a potential tool for enhancing risk evaluation in individuals with borderline or intermediate risk in clinical environments.

5. METHODOLOGICAL FRAMEWORKS: SURROGATE DRIVEN VS END POINT DRIVEN MODELS

The use of deep learning (DL) in the imaging of the retinal fundus has resulted in the development of two primary strategies to predict cardiovascular risk: surrogate-driven models and end-point-risk driven models. These approaches differ in their targets, validation methods, and possible impact on clinical decision-making.

5.1. Surrogate-driven models:

Surrogate-driven models are used to train deep learning (DL) systems to identify well-known imaging or physiological indicators of vascular health, such as coronary artery calcium (CAC), carotid atherosclerosis, vessel diameter, or pulse wave velocity. Once identified, these markers are related to cardiovascular outcomes or integrated into risk tools based on guidelines.

For instance, **(RetiCAC)** predicted future cardiovascular events across several cohorts and enhanced reclassification when incorporated with pooled cohort equations, especially for individuals at borderline and intermediate risk levels(Rim et al., 2021; Son et al., 2020). Similarly, Carotid markers and DL-based vasculometry, which assess vessel calibre, tortuosity, and area, were used to train retinal atherosclerosis scores. These scores showed an independent correlation with cardiovascular mortality and myocardial infarction (Chang et al., 2020).

Other studies have evaluated alternative markers such as Pulse wave velocity (PWV), which is a marker of arterial stiffness, from fundus images and demonstrate its association with hypertension and the development of carotid plaques over time in longitudinal studies (Wong et al., 2024).The strength of these models is their interpretability and capacity to consistently exhibit hazard trends. However, their improvements in discrimination are typically slight and often depend on indirect proxies instead of direct outcomes. (Chang et al., 2020; Rim et al., 2021; Son et al., 2020; Wang et al., 2024; Wong et al., 2024)

5.2. End Point and Risk Driven Models

On the other hand, End point-based DL models can directly predict clinical events or actionable risk categories such as incident myocardial infarction, stroke, or evaluate if a person's measurements exceed the guideline based thresholds (e.g., a 10-year ASCVD risk of ≥7.5%)(Qian et al., 2024).

For example, XMACE effectively identified individuals who exceeded the ASCVD risk threshold with high precision and maintained its performance when recalibrated within external diabetic populations (Vaghefi et al., 2024). In certain instances, models employ a two-step pipeline; first, they deduce conventional risk factors from retinal images, which is then followed by estimating major adverse cardiovascular events (MACE). This approach has demonstrated interpretability preservation and prediction performance relative to traditional risk calculators such as SCORE, SCORE2 and the Framingham Risk Score (FRS). (Y. C. Lee et al., 2023; Mellor et al., 2023) .

The Multimodal approach, integrating fundus imaging with traditional calculators has achieved higher AUROCs and better calibration compared to traditional risk scores alone. In particular, for patients classified as intermediate risk, which shows that it can play the role of a risk enhancer rather than replacing traditional risk assessing tools. (Tseng et al., 2023).

Together, surrogate-driven methodologies underscore the biological feasibility of using retinal imaging to reflect systemic vascular health. However, end-point-driven models stress the clinical importance and utility in decision-making frameworks.

Understanding this methodological separation has resulted in the creation of strategies for interpretability and validation. The subsequent section explores these translational themes—namely interpretability, clinical validation, and generalizability—that are fundamental to the secure and effective integration of retinal DL systems in predicting cardiovascular risk.

6. KEY RESEARCH THEMES AND CLINICAL APPLICATIONS

In addition to advances in methodology, a primary concern in retinal deep learning (DL) research is the conversion of model performance into clinical trustworthiness. Key objectives include improving interpretability, guaranteeing external validity, and ensuring strong generalizability across varied populations. To address this, two complementary pathways have emerged:

1. Transparency enhanced end to end models where explainability is directly incorporated into the models.
2. Biomarker driven models, where already established markers such as vessel calibre, tortuosity, or retinal age is quantified.

This section provides a synthesis of the findings related to these approaches, starting with model interpretability and explainability.

6.1. Interpretability and Explainability in retinal DL models

A major barrier to the clinical application of DL systems is the "Black Box" nature of the models, where there is a lack of transparency behind the predictions. Several frameworks have been developed for AI models to focus on explainability such as quantifying already established vascular markers, interpretable and explainable clinical reasoning steps, and visualisation of decision pathways.

6.1.1. Features Specific DL models

An approach includes the development of deep learning (DL) algorithms that explicitly assess physiologically interpretable characteristics of vascular systems instead of generating abstract risk scores. The Singapore I Vessel Assessment (SIVA-DLS) system is an entirely automated algorithm designed to quantify the calibre of retinal vessels. Specifically, it focusses on the central retinal arteriolar equivalent (CRAE) and the central retinal venular equivalent (CRVE), both recognised indicators of cardiovascular health. It demonstrated a link between narrower arterioles and the prediction of future clinical models, by generating heat maps from fundus images and highlighting areas used to estimate vessel diameter. Thus, connecting computational output directly to vascular anatomy that is interpretable for humans (Wong et al., 2024).

Similarly, QUARTZ software, an automated vasculometry pipeline, is designed to quantify the tortuosity, width variation, and arterioles and venule area. QUARTZ-derived metrics independently predicted mortality due to circulatory diseases and stroke, demonstrating predictive values comparable to those of the Framingham Risk Score (FRS) (Rudnicka et al., 2022). A major advantage of these models is that they provide predictions in units, μm for vessel width and tortuosity index, which enhances direct pathophysiological interpretation and reproduction across different datasets.

These feature-centric methodologies collectively demonstrate how deep learning (DL) can maintain interpretability by replicating conventional quantitative ophthalmic indicators, while enhancing scale and precision through automation.

6.1.2. Two Stage Explainable Models

XMACE, an alternative design approach, focusses on enhancing interpretability in end-to-end models. It uses a clear two-step process:

Stage 1: estimation of ten conventional risk factors (age, sex, BMI, SBP, DBP, HbA1c, HDL, LDL, triglycerides, smoking) directly from fundus images.

Stage 2: Calculation of the 5-year risk of major adverse cardiovascular events (MACE) based on these intermediate estimates.

This dual methodology produces a comprehensible clinical profile showing the rationale behind the model's prediction of increased risk. To enhance explainability, an adapted version of the Integrated Gradient Occlusion Sensitivity (IGOS) saliency technique was used, which demonstrated that the model focus was on the retinal arterioles and venules, aligned with vascular pathophysiology (Qian et al., 2024).

6.1.3. Novel interpretable Biomarkers

Deep learning (DL) output has facilitated the development of novel, intuitive biomarkers that condense intricate vascular data into concise, easily understandable metrics.

One such understandable concept is the "retinal age gap", which is defined as the difference between an individual's deep learning (DL)-predicted retinal age and their actual chronological age, serving as an indicator of overall vascular ageing. According to (Zhu, Chen, et al., 2022) an increase of one year in the retinal age gap is associated with a 3% rise in the risk of cardiovascular events. Furthermore, those in the highest quartile of this gap exhibited a 55% increased risk of cardiovascular disease (CVD) events, even after accounting for confounding

factors. In another study, (Nusinovici et al., 2022) verified the effectiveness of this biomarker in forecasting all-cause mortality and cardiovascular disease (CVD), thus solidifying its role as a functional indicator of biological aging.

Another example is the Reti-WHO model, which directly derives the World Health Organisation (WHO) 10-year cardiovascular disease (CVD) risk score from retinal images. With an Area under the curve (AUC) of 0.69 (compared to 0.68 for the conventional WHO model) and improved stability over years, it indicates that retinal vascular structure might more reliably reflect long-term cumulative risk compared to variable clinical measures (W. Zhang et al., 2023)

Collectively, these biomarkers demonstrate interpretability by simplifying complex image data into singular and intuitive health indicators that clinicians and patients can easily comprehend.

6.1.4. Visualization Techniques

Researchers have integrated visual explanation techniques that correlate deep learning (DL) with anatomical features to provide insight into how models arrive at their decisions.

Gradient-weighted Class Activation Mapping (Grad-CAM) and Shapley Additive Explanations (SHAP) are the most widely used methods. In the study by (Al-Absi et al., 2022) Grad-CAM heat maps of fundus images were used to diagnose cardiovascular disease through integration of DXA and retinal imaging. The visualisation showed that the model was concentrated mainly on vascular regions in the centre of the retina —such as the main retinal vessels and the optic disc, and peripapillary regions, areas that correspond to established vascular indicators of systemic health. In a similar study, (Y. C. Lee et al., 2023) applied saliency mapping to the Reti-CVD algorithm and found that its predictions were mainly affected by the retinal vascular arcades instead of the background areas, thus reinforcing its physiological relevance. (Qian et al., 2024)demonstrated comparable findings by applying gradient analysis within the XMACE model, showing that the model focusses on retinal microvasculature that are linked with systemic changes and vascular changes due to atherosclerosis.

Examples of Grad-CAM heatmaps (shown below) illustrates the visual connection between DL outputs and retinal anatomy using interpretability techniques. These attention maps demonstrate that the model focusses on physiological structures related to systemic vascular integrity—such as the optic and vascular arcades—thus improving clinical plausibility and transparency.

Clinical plausibility indicates how well the learnt features and predictions of a model align with established physiological or pathological mechanisms**.** Here, the alignment between model focus and retinal blood vessels emphasises that predictions are based on biologically relevant information instead of imaging artefacts.

However, **transparency** refers to the degree to which a model's internal decision-making process is understandable and verifiable by humans. Visualization tools, such as Grad-CAM, enhance transparency by enabling clinicians to understand where and why the model concentrates, thereby increasing interpretability and trust in AI-driven diagnostic systems(Al-Absi et al., 2022; Rudnicka et al., 2022; Wang et al., 2024).

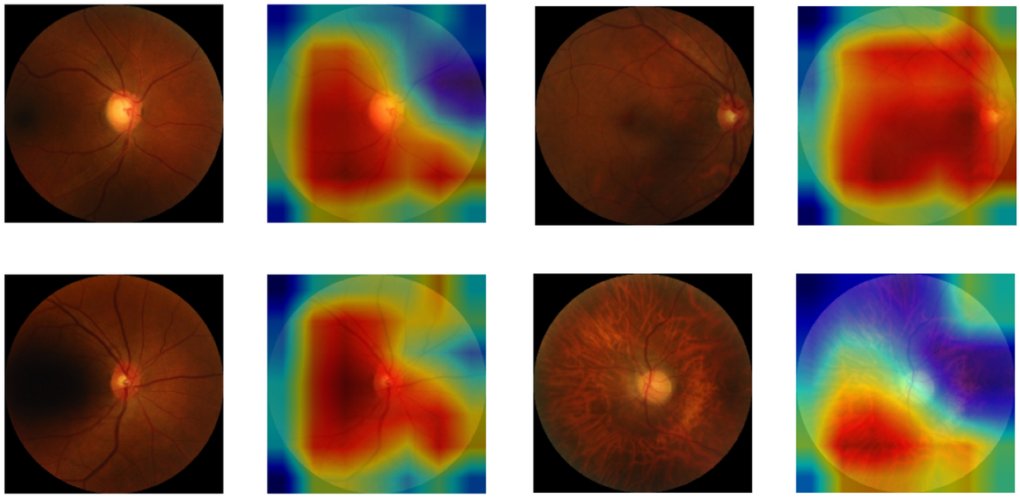

Figure 1. **Heatmaps generated from a multimodal model with overlaid GradCam**. The Red-ish region shows strong attention to the model, and the blue-ish region shows less attention.

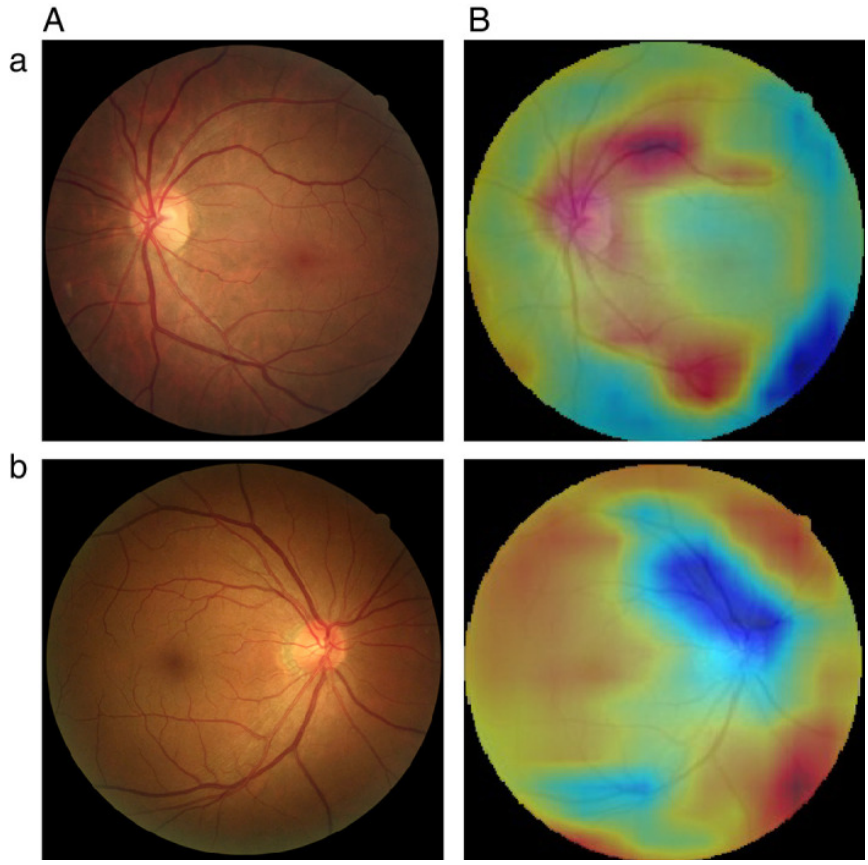

Figure 2: Saliency maps of the Reti-ageing score mode highlighting dense capillary beds related to vascular ageing.
(A) Examples of original retinal fundus images. (B) Examples of saliency maps.
(a) Normal sample; (b) Vascular ageing sample

By visually correlating deep learning (DL) attention with anatomical structures, these visualisation techniques increase clinician confidence, aid in error assessment, and bridge the gap between algorithmic reasoning and clinical interpretation. Together, these visualisation frameworks form the basis for how DL can be applied to retinal imaging. It provides insight into the next phase of development— advancing clinical validation, generalisability, and large-scale application. This next section examines the main themes that have recently surfaced, focussing on how improvements in interpretability, clinical validation, and generalisability are closing the gap between algorithmic breakthroughs and practical cardiovascular prevention.

6.2. Pathways to Clinical Validations and Integration

The practical application of deep learning (DL) systems in real-world scenarios is currently being assessed through benchmarking studies, pragmatic trials, and regulatory approval processes. The initial step in the clinical application of retinal DL models is to benchmark their predictive capabilities against established cardiovascular risk calculators. This comparative analysis serves as a critical benchmark for evaluating clinical relevance, ensuring that the risk

estimates produced by the DL models match or exceed the discrimination and calibration of traditional multivariable tools such as the Framingham Risk Score (FRS), SCORE/SCORE2, QRISK3, and Pooled Cohort Equations (PCE). Benchmarking helps determine whether retinal biomarkers derived from DL can act as a standalone predictor or add a prognostic value to existing scores. The subsequent subsection reviews the main benchmarking initiatives documented in the literature and examines how these image-based models compare to conventional, guideline-supported frameworks.

6.2.1. Benchmarking versus traditional risk scores

Cardiovascular risk SCORE such as the Framingham Risk Score (FRS), SCORE/SCORE2, QRISK3, and the pooled cohort equations (PCE) have long been the cornerstone of the guideline-based evaluation of ASCVD. The results of these scores estimate the 10-year risk of major cardiovascular incidents and help direct preventive strategies and clinical interventions. (Lee et al., 2023; Vaghefi et al., 2024; Zhang et al., 2023). However, the predictions are limited by calibration drift across ethnicities and low discrimination in groups that are in the intermediate or borderline risk (Yi et al., 2023).

In various studies, retinal deep learning (DL) methodologies have demonstrated discrimination capabilities on the same level as traditional risk calculators, and in certain cases exceeding the traditional calculators. Studies using models such as, Reti-CVD, Reti-CAC, SIVA-DLS and QUARTZ show that measurements of features of the vessels in the fundus such as, vessel calibre, tortuosity, and width variation have significant value for prognosing cardiovascular outcomes (Rudnicka et al., 2022; Wong et al., 2024).

In extensive population cohort studies such as the UK Biobank, EyePACS 10K, and CMERC, retinal deep learning (DL) biomarkers showed discrimination abilities that were on par with or exceeded those of traditional scores such as, FRS, QRISK3, SCORE, and PCE, while enhancing risk reclassification in intermediate groups. For example, surrogate modals and multimodal approaches, such as the DL Funduscopic Atherosclerosis Score (DL-FAS), the dual-stage XMACE model, and Reti-CVD, demonstrated improved calibration and net reclassification within subgroups by offering additional value when integrated with conventional risk scores. These results were consistent in internal and external validations, highlighting the robustness and generalisability of the retinal AI biomarker in the prediction of cardiovascular risk. (Chang et al., 2020; Qian et al., 2024; Vaghefi et al., 2024).

In general, benchmarking research reveals two key findings. First, retinal deep learning (DL) models show discrimination metrics that are comparable to or somewhat superior to

conventional risk calculators, improving individual risk estimation, particularly in borderline risk groups. Secondly, they provide a scalable and non-invasive approach to enhance guideline-based risk stratification.

Details of the performance metrics such as, 95% confidence intervals, *p*-values, and comparators are summarised in Table 3 (shown below).

Table 3. Benchmarking of Retinal Deep learning (DL) Models against Traditional Cardiovascular Risk Scores

Deep learning (DL) Model / Approach	**Comparative Performance vs. Traditional Scores**	**References**
SIVA-DLS	Automated vessel calibre improved MI prediction when added to clinical models (**AUC 0.782 → 0.787; Δ = 0.005; *p* = 0.010; NRI 18.35 [95 % CI 6.27 – 32.61]**). Reference model: multivariable clinical risk factors (age, BP, DM, cholesterol)	(Wong et al., 2024)
QUARTZ Vasculometry	Arteriolar/venular width and tortuosity predicted circulatory mortality (**C-statistics 0.68–0.71; 95 % CI as reported**), comparable to FRS; externally validated in EPIC-Norfolk.	(Rudnicka et al., 2022)
DL-FAS (Funduscopic Atherosclerosis Score)	Independent of FRS; improved reclassification (**IDI ≈ 20 %, NRI ≈ 30 %; *p<0.001***) for all-cause and CVD mortality.	(Chang et al., 2020)
XMACE (Two-Stage Explainable Model)	Outperformed SCORE (+8.2 %) and FRS (+7.1 %) for 5-year MACE prediction; (**C-stat 0.725; 95 % CI 0.701 – 0.749; *p* < 0.001**).	(Qian et al., 2024)
Reti-CVD (AI-SaMD)	Hazard-ratio trend = 2.02 (95 % CI 1.26 – 3.24; $p < 0.001$); independent predictive value HR 3.56 (1.34 – 9.51; $p = 0.011$); non-inferior to CAC (ΔAUC = –0.01; $p = 0.41$), superior to CIMT and baPWV ($p < 0.001$); validated in UK Biobank (ΔC = 0.014 [0.010 – 0.017]; $p < 0.001$). Reference models: QRISK3 / CAC	(Y. C. Lee et al., 2023; Tseng et al., 2023)
rpCVD	Predictive precision comparable to the WHO CVD risk score (**AUC 0.672 vs 0.693; Δ = –0.021; *p* = 0.23; 95 % CI as reported**) in Australian primary care; high real-world feasibility.	(Hu et al., 2025)
Reti-WHO	Replicated WHO 10-year risk classification (**AUC 0.69 [95 % CI 0.66 – 0.72] vs 0.68; *p* = 0.842**); greater year-to-year longitudinal stability.	(W. Zhang et al., 2023)
Retinal-only ASCVD Classifier	(95 % CI 0.87 – 0.91; $p < 0.001$) in UK Biobank and 0.88 in EyePACS 10K; correlated with observed ASCVD events. Reference: PCE.	(Vaghefi et al., 2024)
RetiCAC / Surrogate CAC Models	Retinal DL model predicted CAC presence (**AUC 0.83 [0.80 – 0.86] vs FRS 0.77 [0.74 – 0.80]; *p* < 0.001**) — enhanced risk stratification in borderline / intermediate groups.	(Rim et al., 2021)

6.2.2. Risk-Enhancer Role (Borderline/Intermediate Risk Re-Stratification)

A common theme across recent studies is the use of DL biomarkers as risk enhancers that augment guideline based clinical scores, rather than replacing them. In the field of preventive cardiology, there is uncertainty in determining whether to start statin or antihypertensive therapy for individuals with a 10-year risk of ASCVD ranging from 5% to 7.5% (borderline) or 7.5% to 20% (intermediate). Traditional risk assessment tools, including the Framingham Risk Score (FRS), QRISK3, and Pooled Cohort Equations (PCE), may not adequately detect subtle microvascular or inflammatory issues within these categories, potentially resulting in improper treatment decisions (Y. C. Lee et al., 2023; Rim et al., 2021)

An example is The Reti-CVD algorithm, a clinically orientated extension of RetiCAC, which, after recalibrated in the UK Biobank population, effectively categorised individuals into three distinct risk groups and showed clinical value in the borderline QRISK subgroup (7.5-10% 10-year risk), where it accurately reclassified patients and aligned risk with treatment thresholds. This shows the role of DL as risk enhancer alongside traditional risk scores (Tseng et al., 2023).

In a subsequent critical trial, newly approved AI Software as a Medical Device (AI-SaMD) demonstrated that the Reti-CVD risk groups exhibited a trend of grading the hazard ratio of around 2.0 (95% CI 1.26–3.24) for cardiovascular events, independent of conventional risk factors. In comparison to established imaging biomarkers, Reti-CVD's performance was non-inferior to coronary calcium scoring and exceeded that of carotid intima-media thickness (CIMT) and brachial-ankle pulse-wave velocity (baPWV)(Y. C. Lee et al., 2023). Comparable examples based on similar principles are found in other models. The rpCVD model, assessed in primary-care clinics in Australia, showed predictive accuracy comparable to the WHO 10-year cardiovascular disease risk chart (Area under the curve (AUC) 0.672 vs 0.693) and fit naturally into standard workflows (Hu et al., 2025). In the context of borderline patients, these objective microvascular indicators provide clinicians with concrete evidence to validate either the intensification or reduction of preventive treatment.

Collectively, these studies suggest that retinal deep learning (DL) systems are currently more effective as supplementary risk enhancers within multivariable frameworks, rather than replacing them. Their primary clinical utility is to reclassify people with uncertain intermediate risk by detecting underlying microvascular damage that traditional variables may not capture. With expanded regulatory validation and improved workflow integration, retinal DL could play a role similar to that of coronary artery calcium (CAC) scoring—offering a non-invasive, radiation-free, and globally accessible enhancement to cardiovascular risk assessment.

6.2.3. Pragmatic and Pivotal Clinical trials

An important step towards applying the DL algorithm in real world clinical settings is to provide evidence through pragmatic studies and obtaining regulatory-level clinical validation via pivotal trials. Pragmatic trials assess the ability of these models to operate within existing healthcare settings, while pivotal trials determine their safety, validity, and reproducibility for clinical use.

The rpCVD system represents one of the initial practical applications of retinal deep learning (DL) screening in the context of primary care. An Australian multicentre study by (Hu et al., 2025)revealed that its predictive accuracy is on level with the WHO 10-year cardiovascular disease (CVD) risk chart. In addition, it showed high imaging success rates and was well accepted by patients and clinicians. This pragmatic assessment validated that non-mydriatic fundus photography can be seamlessly incorporated into routine procedures, facilitating rapid, radiation-free, and laboratory-independent risk assessment directly at the point of care. These results highlight the operational viability of population-wide screening in settings where conventional laboratory-based methods are either unavailable or require extensive resources(Hu et al., 2025).

At the regulatory level, the Reti-CVD algorithm represents a significant advance in the progression of DL from experimental validation to clinical application. Based on the earlier RetiCAC model, it underwent a prospective multicentre pivotal trial in South Korea and was later approved by the Korean Ministry of Food and Drug Safety (K-MFDS) as an AI Software as a Medical Device (AI-SaMD) (C. J. Lee et al., 2023)

This key study evaluated 1106 participants, including those without history of cardiovascular events, from the Cardiovascular and Metabolic Diseases Etiology Research Centre – High Risk (CMERC-HI) cohort. Over the course of 5 years, 3% of these individuals developed new cardiovascular disease. The study's main goal was to prospectively assess the primary endpoint of cardiovascular disease occurrence, while the secondary endpoints involved comparing the predictive accuracy with established imaging markers— such as, coronary artery calcium (CAC), carotid intima-media thickness (CIMT), and brachial-ankle pulse wave velocity (baPWV).

Reti-CVD demonstrated a substantial hazard-ratio gradient among three risk categories (HR trend = 2.02; 95% CI 1.26–3.24; $p < 0.001$). Adjusted analyses indicated independent predictive strength (HR 3.56 [1.34–9.51]; $p = 0.011$), proving non-inferior to CAC ($\Delta AUC = -0.01$; $p = 0.41$) and outperforming CIMT and baPWV ($p < 0.001$). This evidence served as

the foundation for its regulatory approval by K-MFDS, marking it as the first clinically authorized retinal biomarker for predicting CVD risk.

In a study by Tseng et al. (2023) involving the UK Biobank with 48,260 participants and a median follow-up period of 11 years, further validation demonstrated the broad applicability of Reti-CVD. Reti-CVD was independently related to incident CVD with a hazard ratio of 1.41 (1.30-1.52) and a p-value of less than 0.001. Moreover, when incorporated into the QRISK3 model, it notably improved discrimination with a ΔC of 0.014 (95% CI 0.010–0.017) and a highly significant p-value of less than 0.001 (Tseng et al., 2023).

These crucial and practical validations together establish the clinical reliability and regulatory approval of Reti-CVD as a consistent and applicable tool for cardiovascular risk stratification, on par with traditional imaging biomarkers like CAC and CIMT(Hu et al., 2023; C. J. Lee et al., 2023; Tseng et al., 2023).

6.3. Generalizability Across Populations

6.3.1. Cross-ethnic transport and model recalibration

Research indicates that retinal deep learning (DL) models can be applied across different populations but often require adjustments. A model trained using the UK Biobank to assess BP, HbA1c, eGFR, and various diagnoses underwent prospective validation in Kenya. The results showed that the performance was "comparable or slightly lower" compared to the development data. There was a particular degradation in performance to predict HbA1c (MAE increased from 0.34% to 1.96%). The researchers noted that recalibrating the models might be necessary, with the difference likely due to differences in demographic composition and imaging equipment (TOPCON in the UKB versus Canon CR-2AF in Kenya)(White et al., 2024).

On the other hand, Reti-CVD system underwent successful validation within a diverse Singaporean cohort, including Chinese, Malay, and Indian individuals, the Reti-CVD biomarker maintained its independent prognostic relationship with the onset of cardiovascular disease, thereby affirming its applicability across different ethnic groups(Yi et al., 2023).

In conclusion, the findings of the studies show the potential of retinal microvasculature for risk assessment, it is essential to adopt the models to local populations to guarantee fair and reliable data.

6.3.2. Application in special subgroups (e.g., diabetes, PLWH, others)

Deep learning (DL) models are also being assessed in clinical populations where conventional cardiovascular risk assessment tools often do not perform or incorporate specific pathophysiological characteristics.

- **People with Diabetes:** In a nationwide cohort of people with diabetes, retinal imaging offered independent prognostic information on cardiovascular outcomes, even after considering established clinical and biochemical variables. Nevertheless, the additional predictive value was marginal (C-statistic increase ≈ 0.002), indicating that in high-risk groups already subjected to intensive metabolic monitoring, retinal biomarkers provide limited additional benefit. (Mellor et al., 2023)

- **People Living with HIV (PLWH):** In Asian individuals living with HIV, a deep learning (DL) model tailored for retinal analysis in this group surpassed conventional risk calculations in detecting coronary atherosclerosis and obstructive coronary artery disease. This emphasises the usefulness of retinal biomarkers in scenarios where chronic inflammation and antiretroviral treatment modify vascular risk profiles (Lui et al., 2023).

Advancements in external validation and targeted subgroup analysis show a maturing field, but also highlight the ongoing challenges that must be addressed to ensure impartial and responsible clinical implementation.

7. DATA, PIPELINE, AND EVALUATION CONSIDERATIONS

Before assessing the performance of models, it is important to consider how preprocessing decisions affect the quality and interpretability of retinal features examined by deep learning (DL) systems.

7.1. Preprocessing strategies (segmentation vs. whole-image; normalization)

Effective preprocessing is important for the performance of DL algorithms. **Segmentation-based** and **whole-image** analysis are the two common strategies. Segmentation-based approaches focus on isolating specific anatomical structures, such as the optic disc, macula, or retinal blood vessels, by employing algorithms like **U-Net**. This technique evaluates the characteristics of vessels that are associated with cardiovascular risk, such as calibre, tortuosity, and branching complexity. The extraction of these specific biomarkers (such as vessel calibre and tortuosity) has facilitated the extraction of CVD-relevant endpoints, including myocardial infarction and mortality (Grzybowski, 2025; Rudnicka et al., 2022).

On the other hand, whole-image analysis enables convolutional neural networks (CNNs) to independently learn both local and global patterns without the need for predefined regions of interest, thereby identifying texture or colour signals linked to systemic diseases. This has allowed predicting the risk of BP and CVD from the retina of a single fundus image (Grzybowski, 2025; Rim et al., 2021).

Preprocessing strategies, in both segmentation and whole image analysis, typically integrate illumination/contrast normalisation, contrast enhancement, and data augmentation techniques (such as rotations, flips, and colour modifications) to improve robustness against the variability found in fundus cameras and imaging protocols (C. J. Lee et al., 2023; Vaghefi et al., 2024; White et al., 2024). The choice between the two approaches often depends on the intended application: the segmentation approach is applied for pathophysiological interpretability and vessel biomarker extraction, whereas the whole-image approach is applied for CNNs for end-to-end prediction tasks such as screening cardiovascular event risk (Grzybowski, 2025; Rim et al., 2021).

7.2. Evaluation Metrics, Reliability, and Calibration

To evaluate the performance of DL-algorithms in prognostic modelling, the terms reliability and accuracy are used to represent how consistently, and precisely predictive models

classify individuals into appropriate cardiovascular risk categories. These aspects are assessed by examining four interrelated components:

1. **Discrimination**, indicating the model's ability to differentiate between individuals at high and low risk (evaluated via the Area under the curve (AUC) or C-statistic)
2. **Calibration,** which signifies the match between predicted and observed risks
3. **Reclassification,** assessing the shifts of individuals into more suitable risk categories
4. **Robustness or transportability**, reflecting the model's consistent performance across various cohorts, populations, and imaging modalities.

The summarised trends across these metrics are synthesised below, while the detailed metrics are presented in the table.

7.2.1. Quantitative Performance Trends

The application of DL models to retinal imaging in various datasets has shown moderate to strong ability to predict hypertension, coronary artery disease, stroke, and general cardiovascular outcomes. Improvements in Area under the curve (AUC) and net reclassification are frequently observed by internal and external validation on different continents and imaging systems. This includes testing across continents and crucial clinical data, which underscores the reliability of DL's predictive capabilities(Chang et al., 2020; Y. C. Lee et al., 2023; Qian et al., 2024; Rim et al., 2021; Rudnicka et al., 2022; Tseng et al., 2023; Vaghefi et al., 2024).

DL Systems that use surrogate biomarkers such as RetiCAC/Reti-CVD, DL-FAS, and AI-enhanced vasculometry can efficiently differentiate risk categories and enhance the accuracy of event predictions, surpassing traditional scoring systems such as pooled cohort equations or Framingham Risk Score (FRS) (Chang et al., 2020; Rim et al., 2021; Rudnicka et al., 2022; Tseng et al., 2023). End-point and multimodal and two-stage configurations exhibit AUC/C-statistics that are on par with or exceed those of SCORE, SCORE2 and FRS in specific groups (Y. C. Lee et al., 2023; Qian et al., 2024). The improvement is most significant among the risk groups classified as borderline or intermediate, indicating where clinical uncertainty is highest (Rim et al., 2021; Tseng et al., 2023).

7.2.2. Calibration and Transportability

In general, the calibration quality is mostly satisfactory but frequently needs recalibration for specific cohorts. The models developed using the UK Biobank maintained their performance after recalibration in external datasets such as EyePACS, highlighting the adjustment

for different populations. (Vaghefi et al., 2024). Two-stage frameworks that first evaluate intermediate risk factors (such as blood pressure and cholesterol) and then estimate events have increased calibration transparency, facilitating a more straightforward understanding of prediction inaccuracies(Qian et al., 2024). These recalibration methods increase the predictive accuracy of DL algorithms when applied across diverse populations and imaging areas.

7.2.3. Reliability and Longitudinal Robustness

The reliability in these deep learning (DL) models is improved by external validation across diverse populations in Asia, the United Kingdom, and the United States, and, in some cases, supported by multi-year longitudinal follow-up studies(Chang et al., 2020; Tseng et al., 2023). Within the high-risk clinical group assessed using Reti-CVD, as AI Software as a Medical Device (AISaMD)—there is a clear hazard gradient is observed, while maintaining independent prognostic significance along with established vascular markers such as coronary artery calcium score (CAC), carotid intima-media thickness (CIMT) and brachial-ankle pulse-wave velocity (baPWV)(C. J. Lee et al., 2023). Similarly, deep learning-powered retinal vasculometry has consistently shown a correlation with myocardial infarction and cardiovascular mortality in various external cohorts, thus reinforcing its validity and long term use for risk monitoring. (Rudnicka et al., 2022; Wong et al., 2024).

7.2.4. Summary of Evaluation Insights

Collective evidence reviewed suggests that deep learning retinal (DL) models demonstrate strong discrimination capabilities, reasonable calibration, and consistent reliability when properly validated and calibrated. The most significant contribution is improving traditional multivariable risk assessments by improving the classification of intermediate-risk categories and identifying microvascular damage that traditional variables do not capture. As regulatory approvals increase and workflow integration enhances, retinal DL systems could play a role similar to CAC scoring, offering a non-invasive, radiation-free, and globally accessible enhancement to cardiovascular risk evaluation(Hu et al., 2025; Y. C. Lee et al., 2023; Rim et al., 2021; Tseng et al., 2023).

7.3. Handling Class imbalance and Event Scarcity

In retinal imaging cohorts, event endpoints such as myocardial infarction (MI), major adverse cardiovascular events (MACE), stroke, and cardiovascular (CVD) mortality are relatively rare in imaging cohorts, creating a class imbalance that may distort model optimisation, exaggerate perceived discrimination, and destabilise prediction thresholds. The reviewed literature suggests various approaches to address these challenges, including adaptations in study design, selection of surrogate endpoints, and calibration of models.

7.3.1. Design Approaches to Increase Informative Positives

Extensive population studies have been used in conjunction with long-term follow-up durations to gather an adequate number of events for survival analysis. For example, Within the UK Biobank study, automated retinal vasculometry was used to forecast circulatory mortality and the occurrence of myocardial infarction (MI) through time-to-event modelling(Rudnicka et al., 2022; Wong et al., 2024). In instances where events were limited, case-control sampling is used to enhance class balance. For example, in the stroke research conducted by(Coronado et al., 2021), 412 stroke cases and 2,060 controls were used, illustrating that vascular-embedding features surpassed traditional CNNs in datasets with few events.

7.3.2. Use of Surrogate and Composite Endpoints

To address the issue of limited frequency of events, several studies have used surrogate markers of atherosclerosis, which are more prevalent than clinical outcomes, to train the models. These surrogates include retinal CAC indicators such as RetiCAC and Reti-CVD, carotid-atherosclerosis models, and deep learning-based funduscopic atherosclerosis scores. Each of these has demonstrated an independent prognostic value and has contributed to improved reclassification(Chang et al., 2020; Y. C. Lee et al., 2023; Rim et al., 2021; Son et al., 2020; Tseng et al., 2023). Various groups have expanded endpoints to include composite outcomes such as MACE, leading to an increase in the number of positive cases and XMACE, leading to a more reliable survival estimate (Qian et al., 2024).

7.3.3. Two-Stage and Threshold-Based Frameworks

Due to the interaction between **imbalance**, **label noise**, and **proxy effects**, **multi-stage model designs** have been implemented to make the training more stable. For example, the XMACE model initially deduced standard CV risk factors from fundus images, followed by predicting five-year MACE, this two-step process increases the reliability of the model when there is a lack of actual events available(Qian et al., 2024). Likewise, classifiers based on threshold criteria, models that predict if the pooled-cohort-equation (PCE) risk is ≥ 7.5%, generated a more balanced and clinical interpretable data (Vaghefi et al., 2024)

7.3.4. Calibration and Threshold Adjustment

Because event prevalence differs between study populations, models often require local recalibration to maintain performance, an issue discussed in Section 9.2.2

7.3.5. Contextual Dependence in High-Risk Subgroups

Performance consistency can also depend on the underlying population, as shown in disease-specific cohorts such as diabetes and HIV, where event prevalence and baseline risk influence the model's added value (see Section 7.3.2)

7.3.6. Reporting and Evaluation Under Imbalance

In the majority studies, the most significant enhancements were observed in the models' ability to accurately reclassify individuals into appropriate risk categories, rather than substantial increases in Area under the curve (AUC) values Models using survival-analysis techniques, like QUARTZ, SIVA-DLS, and Reti-CVD, were particularly effective on datasets with a limited number of events due to their consideration of the time duration until an event occurs(Chang et al., 2020; C. J. Lee et al., 2023; Y. C. Lee et al., 2023; Rudnicka et al., 2022; Wong et al., 2024).

7.3.7. Summary

Together, the findings of these studies indicate that the management of event scarcity can be achieved by increasing data sets, using surrogates or combined outcomes, and two-step models to stabilise training. The accuracy is further increased with the recalibration of the local population. Improvements in performance are most notable in borderline-risk categories where the clearest benefit is observed with reclassification and time-to-event(hazard) analyses(Qian et al., 2024; Rim et al., 2021; Rudnicka et al., 2022; Tseng et al., 2023).

8. CHALLENGES, LIMITATIONS, AND FUTURE DIRECTIONS

There is significant potential in applying deep learning (DL) to retinal images to predict cardiovascular risk. However, several ongoing limitations hinder its reliability and clinical application. These challenges include clinical usefulness, data quality, and translational areas, highlighting the gap between algorithmic performance and practical clinical utility.

8.1. Demonstrating Clinically Meaningful Utility

A major challenge involves demonstrating that statistically significant improvements in model accuracy can lead to clinically relevant improvements in patient care. Although deep learning (DL) algorithms often outdo conventional risk scores in terms of discrimination and reclassification metrics, it is still unclear whether these improvements result in significant decreases in cardiovascular events or influence treatment choices (C. J. Lee et al., 2023; Mellor et al., 2023; W. Zhang et al., 2023). In certain populations, particularly among diabetes patients who already receive extensive clinical follow-up, the additional predictive benefit of retinal biomarkers seems limited. This raises doubts about the real-world clinical use in settings where detailed medical information is already available (Mellor et al., 2023). To address this limitation, prospective, outcome-based studies are needed to evaluate whether DL based risk reclassification can meaningfully translate to clinical use such as, preventive strategies or therapeutic interventions(Y. C. Lee et al., 2023; W. Zhang et al., 2023).

8.2. Generalizability and Algorithmic Bias

Deep learning (DL) systems that are trained on extensive, yet demographically uniform datasets frequently show domain shift when applied to novel populations. For example, models created with the predominantly Caucasian UK Biobank were shown to be less reliable in ethnically diverse groups, highlighting the need for fair validation and recalibration (White et al., 2024). The Reti-CVD model showed similar inconsistency, achieving optimal performance within its training domain but needing modifications for application to other ethnic groups, including East Asian or multi-ethnic Singaporean populations(Tseng et al., 2023; Vaghefi et al., 2024; Yi et al., 2023). These differences highlight the need for population-specific calibration, along with broader issues regarding algorithmic bias and fairness in predicting cardiovascular risk.

8.3. Data Quality and Standardization

The technical quality of the retinal images is a significant factor in determining the reliability of the models. Images of substandard quality or ungradable images caused by cataracts, poor focus, or inadequate illumination can lead to inaccurate predictions or even model malfunction (Qian et al., 2024; Wong et al., 2024). Moreover, reproducibility is further compromised by differences in fundus camera design, variations in field of view, and inconsistencies in colour normalization (Rim et al., 2021). Even within extensive biobank datasets, selection bias is a concern, where the participants tend to be healthier than the broader population, necessitating the need for external validation and local recalibration to preserve predictive accuracy(Zhu, Chen, et al., 2022).

In addition to these technical challenges, high clinical skills and experience are required to acquire high quality fundus photographs. Multiple studies have identified the role of image quality and clinician expertise as a key factor that affects the performance and reliability of DL models. (Zhang et al., 2023) developed an automated quality-control pipeline designed to philtre out low-quality fundus images using a vessel-to-background ratio and quadrant-level screening, thus preventing model compromise due to poor data input. (White et al., 2024) highlighted the importance of operator skill and environmental lighting on image quality, by excluding ungradable and low-quality images from the UK Biobank and Kenyan validation cohort to preserve the model's predictive accuracy. The collective findings were supported by(Hu et al., 2025) in the practical real-world validation of the Reti-CVD screening system, where the success rates of non-mydriatic imaging varied considerably among operators, highlighting the critical role of clinical training for consistent implementation.

Together, these studies emphasise that the technical quality of image acquisition and the expertise of clinicians are crucial factors in influencing the robustness and generalisability of cardiovascular predictions based on deep learning (DL) of the retina.

8.4. Multi-modal Data Fusion

Future advances in cardiovascular risk modelling are expected to be based on multimodal integration, which involves merging retinal characteristics with other physiological or imaging data sources. An important example is the multimodal model developed by (Al-Absi et al., 2022), which integrated fundus imagery with dual-energy X-ray absorptiometry (DXA) measurements to forecast coronary artery disease. This integrated strategy outperformed single-modality models, demonstrating how retinal imaging can enhance existing diagnostic

techniques. Other multimodal studies aimed at predicting cardiovascular risk have similarly suggested integrating retinal fundus imaging with clinical and demographic data(Lui et al., 2023; Vaghefi et al., 2024).

Recent developments in oculomic research underscore the possibility of merging retinal imaging with genetic or polygenic data. Integration of oculomics and genomics can reveal common biomarkers linked to vascular conditions and aneurysms, highlighting the potential to combine hereditary and imaging-based predictors in future cardiovascular models (Zhang et al., 2024).

Collectively, these hybrid systems indicate a promising path towards comprehensive, non-invasive cardiovascular profiling that utilizes multiple biomarkers together.

8.5. Longitudinal Analysis and Stability

A promising area of research is longitudinal modelling. Retinal biomarkers have the potential to capture stable and cumulative indicators of vascular health, in contrast to short-term clinical measures such as blood pressure or lipid levels that can vary daily. Longitudinal analyses reveal that metrics derived from deep learning, especially the retinal age gap and Reti-WHO score, show greater consistency from year to year compared to conventional risk assessments, indicating that they may better represent the underlying microvascular health(W. Zhang et al., 2023; Zhu, Chen, et al., 2022).

Based on predictive capabilities and cross-sectional findings derived from AI-driven analysis of colour fundus photographs (CFPs), research into retinal biomarkers has been developed to predict future cardiovascular events and long-term mortality. Ty. Numerous studies have used survival-analysis methodologies, including Cox proportional-hazards modelling, by applying deep features extracted from CFPs to assess the risk of developing cardiovascular disease (Chang et al., 2020; Rim et al., 2021).

This advancement is supported by evidence from extensive perspective cohort studies. Chang et al. (2020) illustrated that a deep learning (DL) model, which was trained using retinal images to predict carotid atherosclerosis and plaque burden, demonstrated a significant association with cardiovascular mortality (Chang et al., 2020). Similarly, the studies conducted by Rudnicka et al. (2022) and Wong et al. (2024) revealed that a decrease in central retinal arteriolar equivalents and an increase in venular calibre, as assessed by automated vasculometry systems such as SIVA-DLS and QUARTZ, were linked to a greater risk of significant adverse cardiovascular events and overall mortality(Rudnicka et al., 2022; Wong et al., 2024).

Moreover, Zhu et al. (2022) used data from the UK Biobank to demonstrate that the age gap in the retina was independently related to mortality and cardiovascular outcomes. A wider gap was associated with an increased risk of developing incident CVD. Subsequent research confirmed that larger age gaps also predicted stroke and various systemic diseases, highlighting the potential of retinal ageing metrics as comprehensive biomarkers of cumulative vascular damage (Zhu et al., 2022a).

These longitudinal and survival analyses collectively indicate that retinal imaging can advance beyond cross-sectional screening to enable ongoing risk assessment. The long-term application of deep learning (DL) retinal biomarkers is useful in personalised monitoring and evaluating the efficacy of preventive strategies in cardiovascular healthcare (Chang et al., 2020; Rudnicka et al., 2022; Wong et al., 2024; Zhang et al., 2023; Zhu et al., 2022b).

8.6. Implementation and Clinical Translation

Even the most precise algorithms must prove their practicality and worth within actual healthcare settings. The primary obstacle to clinical implementation is external validation. Various systems show outstanding internal outcomes, but are affected by a decrease in predictive power in practical and resource limited settings (Vaghefi et al., 2024). A significant achievement towards clinical implementation is the approval of Reti-CVD as an AI software Medical Device (AI-SaMD), supported by pivotal clinical trial data(Hu et al., 2025; Y. C. Lee et al., 2023). To apply deep retinal learning from experimental validation to widespread clinical use, it requires establishing consistent imaging standards and ensuring fair population coverage. By overcoming these technical and clinical challenges, we can ensure responsible implementation of AI systems to provide substantial benefits(Lui et al., 2023; Qian et al., 2024; Tseng et al., 2023; Vaghefi et al., 2024)

CONCLUSION

This review of the literature shows the evolution of deep-learning analysis of retinal fundus images, from early work on surrogate markers to developed models that have become an effective and non-invasive method for assessing cardiovascular risk. The field has evolved from models that determined individual risk factors to building systems that can estimate both short-term and long-term cardiovascular risks. Multiple cohort and prospective studies have validated the clinical application, where a major milestone has been the development of a clinically approved AI tool for the stratification of cardiovascular risk (Reti-CVD). This, along with the development of novel biomarkers such as the "retinal age gap" and consistent demonstrations of DL systems in improving conventional risk scores (e.g., improving reclassification in the borderline group), has allowed earlier and targeted prevention of CVDs at the population level.

However, three main significant challenges are central to clinical application. First, it is crucial to demonstrate that DL can demonstrate clear clinical use beyond performance measures, that it can demonstrate net benefit in patient outcomes and effect treatment decisions. Second, it must ensure generalisability and fairness in algorithms through extensive validation studies in different demography, multimorbidity, imaging devices, and healthcare settings. It requires adopting calibration, drift monitoring, and routine surveillance as standard procedure.

Third, it is important to improve the interpretability of the models for clinical integration into workflows. Models should clearly indicate identifiable vascular characteristics and the features that lead to predictions. It must provide an explanation that is clearly communicated to both clinicians and patients. It is of utmost importance that AI models display transparency and compatibility to build trust and ensure effective use of AI based retinal analysis in clinical settings. Lastly, how widespread the use will depend on addressing practical issues related to cost, image quality, central vs peripheral retinal imaging, and incorporation with existing medical records.

In conclusion, the development of explainable, thoroughly validated and fairly implemented deep-learning systems for retinal analysis marks a significant change in the assessment of cardiovascular risk. Deep learning (DL) applications in Retinal imaging can improve cardiovascular risk evaluation to an accessible and highly scalable screening method. It has the potential to improve current risk assessment models, prioritise patients earlier in disease progression, and ultimately reduce the impact of cardiovascular illness and death rates.

BIBLIOGRAPHY

1. Al-Absi, H. R. H., Islam, M. T., Refaee, M. A., Chowdhury, M. E. H., & Alam, T. (2022). Cardiovascular Disease Diagnosis from DXA Scan and Retinal Images Using Deep Learning. *Sensors*, *22*(12), 4310. https://doi.org/10.3390/s22124310
2. Chang, J., Ko, A., Park, S. M., Choi, S., Kim, K., Kim, S. M., Yun, J. M., Kang, U., Shin, I. H., Shin, J. Y., Ko, T., Lee, J., Oh, B.-L., & Park, K. H. (2020). Association of Cardiovascular Mortality and Deep Learning-Funduscopic Atherosclerosis Score derived from Retinal Fundus Images. *American Journal of Ophthalmology*, *217*, 121–130. https://doi.org/10.1016/j.ajo.2020.03.027
3. Coronado, I., Abdelkhaleq, R., Yan, J., Marioni, S. S., Jagolino-Cole, A., Channa, R., Pachade, S., Sheth, S. A., & Giancardo, L. (2021). Towards Stroke Biomarkers on Fundus Retinal Imaging: A Comparison Between Vasculature Embeddings and General Purpose Convolutional Neural Networks. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, *2021*, 3873–3876. https://doi.org/10.1109/EMBC46164.2021.9629856
4. Gerrits, N., Elen, B., Craenendonck, T. V., Triantafyllidou, D., Petropoulos, I. N., Malik, R. A., & De Boever, P. (2020). Age and sex affect deep learning prediction of cardiometabolic risk factors from retinal images. *Scientific Reports*, *10*(1), 1–9. https://doi.org/10.1038/s41598-020-65794-4
5. Grzybowski, A. (Ed.). (2025). *Artificial Intelligence in Ophthalmology*. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-83756-2
6. Hu, W., Lin, Z., Clark, M., Henwood, J., Shang, X., Chen, R., Kiburg, K., Zhang, L., Ge, Z., van Wijngaarden, P., Zhu, Z., & He, M. (2025). Real-world feasibility, accuracy and acceptability of automated retinal photography and AI-based cardiovascular disease risk assessment in Australian primary care settings: A pragmatic trial. *NPJ Digital Medicine*, *8*, 122. https://doi.org/10.1038/s41746-025-01436-1
7. Lee, C. J., Rim, T. H., Kang, H. G., Yi, J. K., Lee, G., Yu, M., Park, S.-H., Hwang, J.-T., Tham, Y.-C., Wong, T. Y., Cheng, C.-Y., Kim, D. W., Kim, S. S., & Park, S. (2023). Pivotal trial of a deep-learning-based retinal biomarker (Reti-CVD) in the prediction of cardiovascular disease: Data from CMERC-HI. *Journal of the American Medical Informatics Association: JAMIA*, *31*(1), 130–138. https://doi.org/10.1093/jamia/ocad199
8. Lee, Y. C., Cha, J., Shim, I., Park, W.-Y., Kang, S. W., Lim, D. H., & Won, H.-H. (2023). Multimodal deep learning of fundus abnormalities and traditional risk factors for cardiovascular risk prediction. *Npj Digital Medicine*, *6*(1), 1–10. https://doi.org/10.1038/s41746-023-00748-4
9. Lui, G., Leung, H. S., Lee, J., Wong, C. K., Li, X., Ho, M., Wong, V., Li, T., Ho, T., Chan, Y. Y., Lee, S. S., Lee, A. P., Wong, K. T., & Zee, B. (2023). An efficient approach to estimate the risk of coronary artery disease for people living with HIV using machine-learning-based retinal image analysis. *PloS One*, *18*(2), e0281701. https://doi.org/10.1371/journal.pone.0281701
10. Mellor, J., Jiang, W., Fleming, A., McGurnaghan, S. J., Blackbourn, L., Styles, C., Storkey, A. J., McKeigue, P. M., Colhoun, H. M., & Scottish Diabetes Research Network Epidemiology Group. (2023). Can deep learning on retinal images augment known risk factors for cardiovascular disease prediction in diabetes? A prospective cohort study from the national screening programme in Scotland. *International Journal of Medical Informatics*, *175*, 105072. https://doi.org/10.1016/j.ijmedinf.2023.105072
11. Nusinovici, S., Rim, T. H., Yu, M., Lee, G., Tham, Y.-C., Cheung, N., Chong, C. C. Y., Da Soh, Z., Thakur, S., Lee, C. J., Sabanayagam, C., Lee, B. K., Park, S., Kim, S. S., Kim, H. C., Wong, T.-Y., & Cheng, C.-Y. (2022). Retinal photograph-based deep learning predicts biological age, and stratifies morbidity and mortality risk. *Age and Ageing*, *51*(4), afac065. https://doi.org/10.1093/ageing/afac065
12. Qian, Y., Li, L., Nakashima, Y., Nagahara, H., Nishida, K., & Kawasaki, R. (2024). Is cardiovascular risk profiling from UK Biobank retinal images using explicit deep learning estimates of traditional risk factors equivalent to actual risk measurements? A prospective cohort study design. *BMJ Open*, *14*(10), e078609. https://doi.org/10.1136/bmjopen-2023-078609
13. Rim, T. H., Lee, C. J., Tham, Y.-C., Cheung, N., Yu, M., Lee, G., Kim, Y., Ting, D. S. W., Chong, C. C. Y., Choi, Y. S., Yoo, T. K., Ryu, I. H., Baik, S. J., Kim, Y. A., Kim, S. K., Lee, S.-H., Lee, B. K., Kang, S.-M., Wong, E. Y. M., … Wong, T. Y. (2021). Deep-learning-based

cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs. *The Lancet. Digital Health*, *3*(5), e306–e316. https://doi.org/10.1016/S2589-7500(21)00043-1
14. Rudnicka, A. R., Welikala, R., Barman, S., Foster, P. J., Luben, R., Hayat, S., Khaw, K.-T., Whincup, P., Strachan, D., & Owen, C. G. (2022). Artificial intelligence-enabled retinal vasculometry for prediction of circulatory mortality, myocardial infarction and stroke. *British Journal of Ophthalmology*, *106*(12), 1722–1729. https://doi.org/10.1136/bjo-2022-321842
15. Son, J., Shin, J. Y., Chun, E. J., Jung, K.-H., Park, K. H., & Park, S. J. (2020). Predicting High Coronary Artery Calcium Score From Retinal Fundus Images With Deep Learning Algorithms. *Translational Vision Science & Technology*, *9*(6), 28. https://doi.org/10.1167/tvst.9.2.28
16. Tseng, R. M. W. W., Rim, T. H., Shantsila, E., Yi, J. K., Park, S., Kim, S. S., Lee, C. J., Thakur, S., Nusinovici, S., Peng, Q., Kim, H., Lee, G., Yu, M., Tham, Y.-C., Bakhai, A., Leeson, P., Lip, G. Y. H., Wong, T. Y., & Cheng, C.-Y. (2023). Validation of a deep-learning-based retinal biomarker (Reti-CVD) in the prediction of cardiovascular disease: Data from UK Biobank. *BMC Medicine*, *21*, 28. https://doi.org/10.1186/s12916-022-02684-8
17. Vaghefi, E., Squirrell, D., Yang, S., An, S., Xie, L., Durbin, M. K., Hou, H., Marshall, J., Shreibati, J., McConnell, M. V., & Budoff, M. (2024). Development and validation of a deep-learning model to predict 10-year atherosclerotic cardiovascular disease risk from retinal images using the UK Biobank and EyePACS 10K datasets. *Cardiovascular Digital Health Journal*, *5*(2), 59–69. https://doi.org/10.1016/j.cvdhj.2023.12.004
18. Wang, R., Tan, Y., Zhong, Z., Rao, S., Zhou, Z., Zhang, L., Zhang, C., Chen, W., Ruan, L., & Sun, X. (2024). Deep Learning-Based Vascular Aging Prediction From Retinal Fundus Images. *Translational Vision Science & Technology*, *13*(7), 10. https://doi.org/10.1167/tvst.13.7.10
19. White, T., Selvarajah, V., Wolfhagen-Sand, F., Svangård, N., Mohankumar, G., Fenici, P., Rough, K., Onyango, N., Lyons, K., Mack, C., Nduba, V., Noorali Saleh, M., Abayo, I., Siddiqui, A., Majdanska-Strzalka, M., Kaszubska, K., Hegelund-Myrback, T., Esterline, R., Manzur, A., & Parker, V. E. R. (2024). Prediction of cardiovascular risk factors from retinal fundus photographs: Validation of a deep learning algorithm in a prospective non-interventional study in Kenya. *Diabetes, Obesity and Metabolism*, *26*(7), 2722–2731. https://doi.org/10.1111/dom.15587
20. Wong, Y. L., Yu, M., Chong, C., Yang, D., Xu, D., Lee, M. L., Hsu, W., Wong, T. Y., Cheng, C., & Cheung, C. Y. (2024). Association between deep learning measured retinal vessel calibre and incident myocardial infarction in a retrospective cohort from the UK Biobank. *BMJ Open*, *14*(3), e079311. https://doi.org/10.1136/bmjopen-2023-079311
21. Yi, J. K., Rim, T. H., Park, S., Kim, S. S., Kim, H. C., Lee, C. J., Kim, H., Lee, G., Lim, J. S. G., Tan, Y. Y., Yu, M., Tham, Y.-C., Bakhai, A., Shantsila, E., Leeson, P., Lip, G. Y. H., Chin, C. W. L., & Cheng, C.-Y. (2023). Cardiovascular disease risk assessment using a deep-learning-based retinal biomarker: A comparison with existing risk scores. *European Heart Journal - Digital Health*, *4*(3), 236–244. https://doi.org/10.1093/ehjdh/ztad023
22. Zhang, L., Yuan, M., An, Z., Zhao, X., Wu, H., Li, H., Wang, Y., Sun, B., Li, H., Ding, S., Zeng, X., Chao, L., Li, P., & Wu, W. (2020). Prediction of hypertension, hyperglycemia and dyslipidemia from retinal fundus photographs via deep learning: A cross-sectional study of chronic diseases in central China. *PLOS ONE*, *15*(5), e0233166. https://doi.org/10.1371/journal.pone.0233166
23. Zhang, W., Tian, Z., Song, F., Xu, P., Shi, D., & He, M. (2023). Enhancing stability in cardiovascular disease risk prediction: A deep learning approach leveraging retinal images. *Informatics in Medicine Unlocked*, *42*, 101366. https://doi.org/10.1016/j.imu.2023.101366
24. Zhang, Y., Li, S., Wu, W., Zhao, Y., Han, J., Tong, C., Luo, N., & Zhang, K. (2024). Machine-learning-based models to predict cardiovascular risk using oculomics and clinic variables in KNHANES. *BioData Mining*, *17*(1), 12. https://doi.org/10.1186/s13040-024-00363-3
25. Zhu, Z., Chen, Y., Wang, W., Wang, Y., Hu, W., Shang, X., Liao, H., Shi, D., Huang, Y., Ha, J., Tan, Z., Kiburg, K. V., Zhang, X., Tang, S., Yu, H., Yang, X., & He, M. (2022). Association of Retinal Age Gap With Arterial Stiffness and Incident Cardiovascular Disease. *Stroke*, *53*(11), 3320–3328. https://doi.org/10.1161/STROKEAHA.122.038809

APPENDICES:

Appendix 1 Master Summary Table of Included Studies

Author (Year, Country)	Dataset Characteristics	Deep learning (DL) Model	Target Outcome(s)	Performance Metrics	Validation Strategy	Key Findings	Limitations / Interpretability
Zhang et al., 2020 (China)	625 people (1,222 images), rural China; fundus photos	Inception-V3 CNN	Hypertension, hyperglycaemia, dyslipidaemia	Acc ≈ 69% (HTN); AUROC modest for metabolic traits	Internal split	Feasibility: single-image prediction of systemic risk factors	Small, single-region cohort; confounding by age/sex
Gerrits et al., 2020 (Qatar)	3,000 Qatar Biobank (≈12k images)	MobileNet-V2 CNN	Age, sex, BP, HbA1c, lipids, testosterone	MAE age ≈ 2.8 y; AUC sex ≈ 0.97	Internal	Retina encodes multiple cardiometabolic traits	Limited clinical outcome linkage
Chang et al., 2020 (Korea)	32,227 SNUH; fundus + carotid US	DL-FAS (funduscopic atherosclerosis score)	Carotid atherosclerosis; CVD mortality	AUC 0.713; ΔC-index +0.0266; HR ≈ 1.6–8.3	Internal + external	The Fundus-derived atherosclerosis score adds to the FRS	Single ethnicity; retrospective
Son et al., 2020 (Korea)	20,130 persons; 44k fundus images	Inception-V3 CNN	High CAC (>100)	AUC ≈ 0.83	5-fold CV	Fundus DL detects elevated CAC	No event outcomes; single centre
Coronado et al., 2021 (USA/UK)	2,472 UK Biobank; stroke cases/controls	VGG19, Inception-V3, vasculature embeddings	Prior stroke classification	AUC up to ≈ 0.71 (age strata); ~0.63 overall	Internal split	Vessel embed features outperform generic CNN	Case–control; moderate discrimination
Rim et al., 2021 (Korea/UK)	Multi-cohort; fundus with CAC labels	DL predicting CAC (RetiCAC concept)	CAC presence; future risk stratification	AUC high for CAC; improves risk reclassification	Internal/External (by cohort)	DL CAC-from-retina can act as a risk enhancer	Surrogate endpoint; needs event validation

Nusinovici et al., 2022 (SG/KR/UK)	~40k train; ~56k UKB test	CNN ("retinal age")	Biological age; mortality/CVD	C-index ≈ 0.70; HR ~2.4 (CVD mort)	External (UKB)	Retinal "biological age" predicts mortality/CVD	Black-box; screening utility
Rudnicka et al., 2022 (UK)	88,052 UKB + 7,411 EPIC-Norfolk	AI vasculometry (QUARTZ)	Circulatory mortality, MI, stroke	C ≈ 0.75–0.77	External (EPIC-Norfolk)	Vasculometry ≈ FRS without clinical input	Needs deployment studies
Al-Absi et al., 2022 (Qatar)	500 Qatar Biobank	ResNet; multimodal (DXA + fundus)	CVD diagnosis	Accuracy ≈ 78%	Internal	Fusion (DXA+retina) > single modal	Small sample; class balance
Zhu et al., 2022 (UK)	46,969 UK Biobank	CNN (retinal age gap)	Arterial stiffness; incident CVD/stroke	HR ~1.03 per 1 year gap; Q4 ≈ +55% risk	External (UKB)	Retinal age gap = biomarker of vascular ageing	Observational; confounding by age
Lee et al., 2023 (KR/SG)	1,106 CMERC-HI	CNN (Reti-CVD; AI-SaMD)	5-yr composite CVD events	Adjusted HR ≈ 2.0–3.6 in strata	Internal; pivotal design;	First regulated retinal AI biomarker; non-inferior to CAC	Few events; Korean cohort
Tseng et al., 2023 (UK)	48,260 UK Biobank (no prior CVD)	Reti-CVD (thresholded)	10-yr CVD risk; events	ΔC-stat ~+0.014; HR ~1.4	Internal + subgroups	Enhances QRISK3, reclassifies borderline	Black-box; UK-only
Yi et al., 2023 (Singapore)	Multi-ethnic (Chinese/Malay/Indian)	Reti-CVD biomarker	Incident CVD events	Independent HR gradient across groups	External to the dev domain	Confirms generalisability of Reti-CVD	Ethnic subgroup sizes; device variation
Lee et al., 2023 (Korea/UK)	3,518 SMC train/val; UKB external	Multimodal DL (fundus + risk factors)	Current CVD (ICD) & event association	AUROC 0.78 (SMC); 0.87 (UKB); HR ~6.3	External (UKB)	Multimodal > single modality; strong association	Prospective utility pending
Zhang et al., 2023 (HK)	3,765 civil servants; 55,540	Swin Transformer	WHO 10-year CVD score replication	R^2 0.503; MAE 1.58; sens 0.81; spec 0.66	Internal longitudinal	Greater stability than WHO	False-negatives in

	images; 6-yr FU	(Reti-WHO)				score year-to-year	high-risk tails
Mellor et al., 2023 (UK)	UK Biobank	DL in retinal images + risk factors	Augmenting CVD risk prediction	Modest Δ C index; improved NRI in subgroups	External checks within UKB	Retinal DL adds incremental value	Gains are modest; calibration sensitivity
Lui et al., 2023 (HK)	115 PLWH; CTA + fundus	ML (retina + clinical)	Coronary atherosclerosis; obstructive coronary artery disease	AUC up to 0.99 (combined)	Internal	Retina + clinical > clinical alone in PLWH	Small, niche cohort
White et al., 2024 (UK/Kenya)	UKB-trained; prospective Kenya validation	DL risk-factor predictor	SBP/DBP, HbA1c, etc.	MAE SBP ~12.3, DBP ~8.0 (Kenya)	Prospective external	Feasible in the clinic; domain shift manageable with recalibration	Device/population shift; metric drift
Wang et al., 2024 (China)	4,376 pts; 8,865 images; 2 centres	CNN ("Reti-Ageing")	Vascular ageing (PWV ≥1400 cm/s)	AUC 0.83 / 0.78	Internal + external	DL predicts arterial stiffness better than clinicians	Limited diversity
Vaghefi et al., 2024 (NZ/US)	44,176 UKB + 8,969 EyePACS	CNN + demographics	10-yr ASCVD (PCE ≥7.5%)	AUC 0.89 (UKB); 0.84 (EyePACS)	External (EyePACS)	Generalizable thresholding of ASCVD risk	EyePACS diabetes bias; cross-sectional
Wong et al., 2024 (UK)	>30k UK Biobank; long FU	DL vasculometry (SIVA)	Incident MI	ΔAUC +0.005; NRI ~18%; HR ~1.64	Internal	Arteriolar narrowing independently predicts MI	Predominantly Caucasian; treatment effects unclear
Zhang et al., 2024 (Korea)	32,122 KNHANES	Ensemble ML (oculomics + clinical, 25 models)	TyG, AIP (CVD proxies)	AUC 0.91 (int), 0.90 (ext)	5-fold + external	Oculomics + survey data are highly predictive	Surrogate endpoints; cross-sectional

Qian et al., 2024 (UK)	UK Biobank	XMACE two-stage (10 RFs → 5-yr MACE)	5-yr MACE; ASCVD thresholding	C-stat ≈ 0.73; > FRS/SCORE; interpretable	Internal/External (within UKB)	Transparent pipeline; IGOS aligns with vessels	Needs multi-ethnic, prospective tests
Hu et al., 2025 (Australia)	361 primary-care participants	Proprietary DL (rpCVD)	10-year CVD risk (clinical workflow)	AUC 0.672 vs WHO 0.693	Pragmatic clinical study	Feasible, acceptable, clinic-ready imaging	Small sample; relies on external training

Declaration

I, Walid Dawoud, certify that this Research Paper on the topic

"Dziļās mācīšanās pielietojumi sirds un asinsvadu slimību riska prognozēšanā, izmantojot tīklenes fundus attēlus: literatūras apskats"

(Title of the Research Paper in Latvian)

"Deep learning (DL) Applications in Cardiovascular Risk Prediction Using Retinal Fundus Imaging: A literature Review"

(Title of the Research Paper in English)

is my own independent research. All other sources of data, definitions and quotations in my Research Paper are referenced accordingly. The text of this Research Paper has never been submitted in any way to any other National Examination Commission for evaluation in its entirety or parts and has never been published in its entirety.

_____ /___Walid__Dawoud__/

(Student's signature) (Name, surname)

Riga, 07.11.2025