

Prediction of Cardiovascular Markers and Diseases Using Retinal Fundus Images and Deep Learning: A Systematic Scoping Review

Livie Yumeng Li^{1,2}, Anders Aasted Isaksen², Benjamin Lebiecka-Johansen², Kristian Funck², Vajira Thambawita³, Stine Byberg⁴, Tue Helms Andersen⁵, Ole Norgaard⁵, Adam Hulman^{1,2}

¹Department of Public Health, Aarhus University, Aarhus, Denmark

²Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus, Denmark

³Simula Metropolitan Center for Digital Engineering, Oslo, Norway

⁴Clinical Epidemiological Research, Copenhagen University Hospital – Steno Diabetes Center Copenhagen, Herlev, Denmark

⁵Department of Education, Danish Diabetes Knowledge Center, Copenhagen University Hospital – Steno Diabetes Center Copenhagen, Herlev, Denmark

Corresponding authors:

Livie Yumeng Li

Palle Juul-Jensens Boulevard 11, Entrance A, Steno Diabetes Center Aarhus

Aarhus N, DK-8200, Denmark

livie.ymli@ph.au.dk

Adam Hulman

Palle Juul-Jensens Boulevard 11, Entrance A, Steno Diabetes Center Aarhus

Aarhus N, DK-8200, Denmark

adahul@rm.dk

+45 23 70 74 81

Abstract

Background: Cardiovascular risk prediction models based on sociodemographic factors and traditional clinical measurements have received significant attention. With rapid development in deep learning for image analysis in the last decade and the well-known association between micro- and macrovascular complications, some recent studies focused on the prediction of cardiovascular risk using retinal fundus images. The objective of this scoping review is to identify and describe studies using retinal fundus images and deep learning to predict cardiovascular risk markers and diseases.

Methods: We searched MEDLINE and Embase for peer-reviewed articles on 17 November 2023. Abstracts and relevant full-text articles were independently screened by two reviewers. We included studies that used deep learning for the analysis of retinal fundus images to predict cardiovascular risk markers (e.g. blood pressure, coronary artery calcification, intima-media thickness) or cardiovascular diseases (prevalent or incident). Studies that used only predefined characteristics of retinal fundus images (e.g. tortuosity, fractal dimension) were not considered. Study characteristics were extracted by the first author and verified by the senior author. Results are presented using descriptive statistics.

Results: We included 24 articles in the review, published between 2018 and 2023. Among these, 21 (88%) were cross-sectional studies and eight (33%) were follow-up studies with outcome of clinical CVD. Five studies included a combination of both designs. Most studies (n=23, 96%) used convolutional neural networks to process images. We found nine (38%) studies that incorporated clinical risk factors in the prediction and four (17%) that compared the results to commonly used clinical risk scores in a prospective setting. Three of these reported improved discriminative performance. External validation of models was rare (n=5, 21%). Only four (17%) studies made their code publicly available.

Conclusions: There is an increasing interest in using retinal fundus images in cardiovascular risk assessment. However, there is a need for more prospective studies, comparisons of results to clinical risk scores and models augmented with traditional risk factors. Moreover, more extensive code sharing is necessary to make findings reproducible and more impactful beyond a specific study.

Keywords

cardiovascular disease; cardiovascular risk; retinal fundus image; deep learning; clinical prediction; systematic scoping review

Introduction

Cardiovascular diseases (CVDs) are the leading causes of mortality globally.¹ Clinical risk prediction models can help to identify individuals at high risk and target preventive efforts including lifestyle and pharmacological interventions.² There are some well-established and validated CVD prediction models based on sociodemographic factors and traditional clinical variables in the general population and in selected subpopulations, like the Framingham score,³ SCORE2,⁴ and QRISK models.⁵ A recent comprehensive systematic review called for validating existing tools, tailoring existing models to specific populations, and identifying new data sources to be included in models, instead of testing small alterations of the established clinical risk models.⁶

Developments in machine learning for image analysis in the last decade have made it feasible to include images as a data type in risk prediction models, potentially in combination with traditional risk factors (multimodal prediction models). Retinal fundus imaging is a relatively simple and the only non-invasive method used for assessing the state of blood vessels in the body. People with type 1 and type 2 diabetes are regularly invited to diabetic retinopathy screening with retinal fundus images, which makes it a potentially relevant predictor and allows us to follow the progression in microvascular disease. It's likely that the underlying disease causes damage to all blood vessels, resulting in both micro- and macrovascular complications. The state of the blood vessels in the retina could reflect that of blood vessels elsewhere. There is a well-established association between diabetic retinopathy and both micro- and macrovascular complications, including CVD.^{7,8} However, it remains an open question whether these associations can be translated into clinically relevant predictors. Instead of using disease status as the predictor, the actual images might carry more relevant information for CVD prediction.

In a preliminary search, we identified a review on the topic of our scoping review.⁹ However, there are some limitations of this work. Firstly, the search procedure was described only very briefly. Secondly, the authors only considered open-access articles in the review. Thirdly, they had a focus on diabetic retinopathy grading and image segmentation unrelated to cardiovascular disease. Since the review was published, we have identified several relevant articles on the topic.^{10,11} Finally, we aim to evaluate the studies from a more clinical perspective (e.g. performance comparison with or added value to existing risk scores).

We chose to conduct a systematic scoping review instead of a classical systematic review because scoping reviews serve to scope a body of literature, examine research practices, and clarify concepts, which fit well with our goals.¹² We aimed to describe studies, with an emphasis on clinical perspectives, that used retinal fundus images and deep learning for the prediction of cardiovascular markers and diseases.

Methods

The scoping review was conducted in accordance with the Joanna Briggs Institute (JBI) methodology for scoping reviews and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses extension for scoping review (PRISMA-ScR) guidelines.¹³ The scoping review protocol was published on Figshare on December 19, 2023.¹⁴

Eligibility Criteria

Participants

We considered studies analysing data from human participants regardless of their health status (e.g. population-based studies or cohorts with a specific condition).

Context

This scoping review focuses on studies in the clinical research context, regardless of the geographical location, ethnicity, and gender composition of the study populations. Methodological articles comparing different methods for CVD risk prediction based on retinal fundus images were considered if any of the methods used deep learning.

Types of Sources

We included peer-reviewed articles (e.g. original articles, brief reports; peer-reviewed full-length articles in conference proceedings). Included studies needed to be human clinical studies (including methodological studies with examples using clinically relevant outcomes and measures). Furthermore, we only included studies that used deep learning to predict cardiovascular markers, or presence of CVD, or CVD incidence based on retinal fundus images. Studies could also use a deep learning-derived score from retinal fundus images (not pre-defined retinal features), to predict cardiovascular markers or diseases.

We excluded review articles, editorials, conference abstracts and preprints. Studies were excluded if they only included optical coherence tomography (OCT) or other eye images for prediction, if they extracted pre-defined features of the vessel network (e.g. tortuosity, fractal dimension) and then associated them with CVD markers, or if they predicted factors that were not based on measurements of the cardiovascular system, but merely risk factors of CVD (e.g. age, sex, cholesterol, HbA1c). We also excluded association studies, even if they investigated the association between a deep learning-derived score and cardiovascular markers or diseases, since such studies do not report any predictive performance metrics.

Articles in languages not understood by the review team (English, Danish, Swedish, Norwegian, Hungarian, German, Polish, and Chinese) that were considered eligible by title and abstract were not included in the synthesis but would have been included in an appendix for others to analyse.

Search Strategy

We considered MEDLINE and Embase as databases for the search. Most of the data science literature focuses on diabetic retinopathy grading, most likely due to the availability of open-access datasets from this domain. As this task is not directly relevant from a cardiovascular perspective and it is unlikely that such datasets have detailed phenotyping of cardiovascular health or follow-up for hard endpoints, we did not consider technical databases.

Information specialist THA conducted the search on November 17, 2023. The search comprised three key concepts: retina, cardiovascular diseases and artificial intelligence/machine learning. Each concept was searched using Medical Subject Headings (MeSH) and free-text words, and no limits were applied. The search string was developed in MEDLINE and subsequently translated to Embase. The search string was tested against eight key articles within the field and reviewed by another information specialist (ON). The full search string in both databases is available on Figshare.¹⁵

After the selection process, we used the software tool citationchaser¹⁶ to retrieve all references within and all articles citing the included articles and previous reviews within the same topic. Retrieved articles were screened to find all relevant articles.

Study Selection

Following the search, all identified citations were collated and uploaded into EPPI-Reviewer 6 and duplicates were removed.¹⁷

In the screening phase, two independent reviewers screened titles and abstracts to assess eligibility. At the beginning, we had a pilot screening workshop examine the in- and exclusion

of about 25 abstracts. In the official screening, the reviewers met to assess alignment in the process after screening 10% of the abstracts. After all titles and abstracts had been screened, full-text versions of relevant articles were retrieved and assessed in detail against the eligibility criteria by two or more independent reviewers. Reasons for the exclusion of articles at full-text screening were recorded and reported in the Results section. Any disagreements between the reviewers at any stage of the selection process were resolved through discussion. If there was no consensus after this, the senior author (AH) made the final decision. The study selection process is presented using a flow diagram.

Data Extraction

Research questions (Table 1) were pre-specified and published in the scoping review protocol.¹⁴ Data was extracted from included articles by the first author and verified by the senior author. Data was collected using a data extraction instrument developed based on the research questions. The extracted data included specific details about the study methods and characteristics relevant to the review questions listed in the protocol (e.g. first author and year of publication; study population and design; CVD outcomes; deep learning model used; predictive performance; comparison to clinical risk scores if included). The identified studies described in the data extraction table are published on Figshare.¹⁵

Data Analysis

Study characteristics are aggregated using descriptive statistics (frequencies and percentages). A narrative summary accompanies the tabulated and charted results and describes how the results relate to the review's objective and questions.

Results

The search resulted in 1,990 records, of which 172 were duplicates (Fig 1). After screening the titles and abstracts of the remaining 1,818 records, 1,790 were excluded as irrelevant. Of the 28 records included for full-text screening, 10 were excluded because they did not use deep learning (n=5), were not original articles (n=2), did not focus on CVD (n=2) or did not use fundus retinal images (n=1). Another six records were identified in citations and references of the included studies and previous reviews on similar topics.^{9,18} In total, 24 studies were included in this scoping review, all of which were published after 2018.

Among the identified studies, 11 out of 24 (46%) predicted markers of subclinical CVD as the outcome, 18 predicted clinical CVD, and five studies included both types of outcomes. Ten out of 18 studies were cross-sectional using retinal fundus images to predict prevalent CVD, and eight were cohort studies including clinical CVD.

Markers of subclinical CVD included systolic and diastolic blood pressure, left ventricular characteristics, brachial-ankle pulse-wave velocity, coronary artery calcium (CAC) score, and carotid intima-media thickness (Table 2). Clinical CVD included hypertension, peripheral arterial disease, coronary heart disease, cerebrovascular disease, stroke, myocardial infarction, heart failure, and CVD mortality. Apart from the desired outcomes listed in the studies, one study had an outcome described as CVD without further details, and one study had the outcome described as major adverse cardiovascular events (MACE) without further details included.

Most studies were conducted in the general population or without the mention of specific patient groups (22 out of 24). Two studies had more specific study populations, one included people with atrial fibrillation,¹⁹ and another included people with type 1 diabetes or type 2 diabetes.²⁰ We used the first authors' first affiliation as a proxy for the geographical location of the studies. More than half of the identified studies were conducted in Asia (13 out of 24,

54%): five from South Korea, four from China, and four from Singapore. We did not identify any studies from Africa, South America or Australia.

Most studies (n=23, 96%) used convolutional neural networks to process images. One study used the vision transformer deep learning architecture.²¹ No studies used a deep learning framework specifically developed for time-to-event data. Several performance metrics were used in the studies. The most often used predictive performance metric was discrimination, characterised by the area under the receiver operating characteristic curve (AUROC), also referred to as C-statistic or concordance index in time-to-event settings (n=18 out of 24, 75%). For continuous outcomes, the most often used metrics were accuracy (n=6, 25%), mean absolute error (MAE) and the coefficient of determination also known as r-squared (n=3).

We found nine studies that combined clinical risk factors (in tabular data form) with images as predictors in their studies. Five out of nine included clinical risk factors in the deep learning models, whereas the other four applied a two-step approach extracting new variables using deep learning and then including them in Cox regression or Poisson regression models. Five studies compared the performance with established clinical risk scores such as Framingham Risk Score,²² Systematic Coronary Risk Assessment (SCORE),²³ QRISK3,²⁴ and the Pooled Cohort Equation for atherosclerotic CVD (PCE-ASCVD) (Table 3).²⁵ One study presented a comparison with a customised clinical risk prediction model developed as part of the same study. Four studies reported incremental improvement in the predictive performance after adding retinal fundus images or scores derived from them to the clinical risk score or model of their choice.

External validation of models was rare (n=5 out of 24, 21%), while internal validation was a common practice. Eight studies used cross-validation, and 12 split and set part of their study sample aside for validation. No study used a fully open-access dataset. Eight studies used data from the UK Biobank,²⁶ which is available to anyone for a data access fee. Only four studies made their code publicly available on GitHub. None of the studies reported randomised clinical trials to evaluate the implementation of the developed algorithms and scores.

Discussion

We identified and described studies exploring the integration of retinal fundus images in the prediction of cardiovascular markers and diseases. The majority of these studies were cross-sectional, lacking examination of predictive utility for hard clinical endpoints in prospective settings. Few studies compared their models with established cardiovascular risk scores, evaluating the potential of the multimodal approaches. In most of these cases, including retina images led to small improvements in discriminative performance. However, other clinically relevant metrics, like calibration, were often overlooked.

Clinical risk prediction models are traditionally developed using regression-based statistical methods that can only handle categorical and numerical variables i.e. tabular data. Recent advancements in deep learning allowed the analysis and integration of images with promising results in various medical domains, including clinical risk assessment of cardiovascular health.²⁷ A multimodal approach offers the opportunity to extract novel insights from new, and often routinely collected data types, expanding the utilisation of available clinical data.

We found a series of studies on prediction of CAC from one research group, where they developed and validated a retinal fundus image-based score referred to as Reti-CVD or RetiCAC, and its utility was compared to existing CVD risk scores.^{28–30} The score was proposed as a non-invasive and cost-efficient alternative to the computer tomography-derived CAC score. Multiple validation studies investigated the incremental

predictive value of RetiCAC on top of established clinical risk prediction models. Rim et al. added the score to PCE-ASCVD to predict SCORE CVD events and Tseng et al. added the score to QRISK3 to predict 10-year CVD respectively.^{28,29} They both reported small improvements in predictive performance after adding retinal image-derived scores to the models. In another study, Reti-CVD was used to identify people at intermediate and high risk of CVD according to established clinical risk scores (PCE-ASCVD, QRISK3 and Framingham Risk Score), and reported that it could effectively identify these groups.¹¹ Lee et al. conducted a regulated pivotal trial (single-centre conformity design, confirmatory retrospective analysis) to validate the efficacy of Reti-CVD for stratification of CVD risk.³⁰ The trial concluded superior performance in risk stratification compared to some subclinical CVD markers (carotid intima media thickness, pulse wave velocity), and non-inferiority to CAC score-based risk stratification. The commercialised product based on the RetiCAC/Reti-CVD score obtained CE approval as a class IIa medical device in the European Union and several countries in Asia.³¹ However, our search has not identified any randomised controlled trial conducted to evaluate the actual clinical impact of using the score.

In addition to the studies focusing on CAC prediction, three other studies compared their predictive performance with established clinical risk scores. Poplin et al. used a deep learning model to predict the 5-year MACE from retinal images.³² They compared the predictive performance of the original SCORE model to its image-augmented version. They did not find evidence for improved predictive performance by including retinal images. Chang et al. had a similar approach to the outcome of CVD mortality using the Framingham Risk Score. This study reported minor improvements by integrating images into the model.³³ In a case-control study, Lee et al. predicted the prevalence of CVD using a multimodal deep learning model integrating retinal images and traditional risk factors and compared the results to the predictive ability of PCE-ASCVD.¹⁰ They found major performance improvement by using the deep learning model.

When examining performance evaluations, we found a strong focus on model discrimination, which describes how well a model ranks predicted probabilities between individuals with and without an event. However, in a clinical setting, CVD risk scores are often used with a specific threshold (e.g. 10%) to make decisions on interventions, e.g. treatment intensification. Therefore, good discrimination alone is not sufficient, but calibration is highly relevant as well, which was overlooked by most of the studies. A miscalibrated model can lead to resource misallocation. Individuals incorrectly predicted with a high risk of CVD may be administered unnecessary tests or treatments, while those who could benefit from interventions may be overlooked.

Apart from the use of clinically relevant performance metrics for model evaluation, fair comparisons between existing and new models are also of high importance. Some of the identified studies reported that retinal images can be used to predict demographic factors (e.g. age, sex) and traditional clinical risk factors (e.g. blood pressure, body mass index).³² Many of these factors are associated with CVD risk, therefore, good performance of a prediction model based only on images does not necessarily mean that retinal images have additional value for CVD risk prediction, which is the underlying motivation of most identified studies. Models might achieve good performance by indirectly predicting risk factors that are easier to collect than retinal images.³² This question can be partly addressed by stratifying participants by clinical risk factors possible to predict based on retinal images (e.g. age, sex).³⁴ Another solution is to conduct incremental comparisons integrating images with traditional clinical risk factors, however, one should be careful with interpreting the results if important, unobserved risk factors are not included in the models.

Although studies demonstrated the potential of deep learning to contribute to improvements in cardiovascular risk assessment, methodological challenges, like insufficient validation, and poor reproducibility due to the lack of openness about data and code, must be

addressed by future studies. Moreover, diverse study populations, both for development and validation, would improve the equality and fairness of potential clinical applications.^{35,36} No impact study was identified in our scoping review, which is a crucial missing link to measure the actual clinical utility of retinal image-based prediction models in CVD risk assessment.

Although most deep learning models were originally not designed to analyse time-to-event data, recent developments enabled them to account for this special data structure.^{37,38} However, none of the identified studies applied such a time-to-event framework. The most common approach to circumvent this limitation was a two-step approach combining deep learning with classical survival analysis methods.^{20,28–30} In step one, the deep learning model takes one or a pair of retinal fundus images as input and outputs an intermediate score or abstract feature representation. In step two, the score is used, often in combination with clinical risk factors, as input for a Cox or Poisson regression model, that then outputs predicted probabilities. Since these statistical models are consistent with what is used in classical survival analyses, the comparison and validation are analogous to established risk scores and risk prediction models. Some studies ignored the time-to-event nature of data, which can lead to underestimation of CVD risk especially in high-risk groups.³⁹

Strengths and Limitations

As a systematic scoping review, the strength of our study is the comprehensive and systematic exploration of the scope. The search strategy was thoroughly developed and documented contributing to transparency and reproducibility. Our review was written with a broad target audience addressing study characteristics relevant from both a clinical and a data science perspective. One limitation of our study is that we searched only medical databases, and might have overlooked studies from the technical sciences and engineering communities. To address this, we screened all references and citing articles of the included studies and reviews similar to ours. This process resulted in six additional articles, five of which were not indexed in the medical databases we searched.

Perspectives

In this scoping review, we identified a gap between the proposed methods in clinical research and the latest developments in the field of deep learning. A fully deep learning-driven multimodal time-to-event prediction model could bring new insight to cardiovascular disease risk analysis and risk progression by learning from a series of images in longitudinal settings. Our findings support that improvements are needed in the recently proposed five critical quality criteria for artificial intelligence-based prediction model development and validation studies.⁴⁰ Research in this highly interdisciplinary field must strive to report adequate details from both the clinical and data science aspects. Further collaboration and better communication between the two fields are crucial to raising the quality of studies. More validation studies with a focus on diversity regarding geographical location and ethnicity should be conducted in the future. At the same time, efforts should be spent to promote open-access datasets and algorithms to align with the FAIR principles.⁴¹ With the new AI Act approved in the European Parliament and similar proposals introduced globally, it is expected that research utilising AI technology will be forced to adhere to stricter regulations regarding transparency, accountability, and ethical considerations throughout the developmental and implementation lifecycle.

Conclusion

Technological advancements in the field of artificial intelligence offer the potential to integrate new data types in analyses that were traditionally based on tabular data. Our scoping review presented the landscape of how deep learning methods integrate retinal fundus images in

cardiovascular risk prediction. Some evidence shows improvements in predictive performance when adding images to clinical risk factors, but similar to prediction research in general, methodological weaknesses have to be overcome before the full potential can be exploited. There is a need for conducting more prospective studies, promoting performance metrics other than discrimination, increasing efforts on external validation of findings in datasets from diverse settings, and initiating impact studies to measure the clinical value of new multimodal risk assessment tools while improving the reproducibility of research. Working with images and deep learning models requires new methodological skills and close collaboration between the clinical research and data science communities so that developed models are described and tested adequately, and most importantly that they fulfil a validated clinical purpose.

Author Contributions

LYL, AAI, BLJ, THA, ON, and AH conceptualised the study. THA and ON developed the search strategy with feedback from LYL, AAI, BLJ and AH. THA conducted the search. LYL, AAI, BLJ and AH screened abstracts. LYL and AH screened full-text articles. LYL extracted data from the identified articles. The data extraction forms were verified by AH. LYL analysed the data and presented the results. All authors contributed to the interpretation of the results. LYL wrote the original draft of the manuscript with support from AH. All authors read, edited and approved the final version of the manuscript. AH was responsible for the supervision of the project and funding acquisition.

Acknowledgements

The authors are grateful to Helene Bei Thomsen (Steno Diabetes Center Aarhus) for her valuable comments on the Discussion.

Funding

LYL, AAI, BLJ, KF, and AH are employed at Steno Diabetes Center Aarhus which is partly funded by a donation from the Novo Nordisk Foundation (NNF17SA0031230). LYL, BLJ, and AH are supported by a Data Science Emerging Investigator grant (NNF22OC0076725) by the Novo Nordisk Foundation. SB, THA and ON are employed at the Steno Diabetes Center Copenhagen, a public hospital and research institution under the Capital Region of Denmark, which is partly funded by the Novo Nordisk Foundation. The funders had no role in the design of the study.

Conflict of Interests

None.

Data Availability

The study protocol and the completed data extraction forms were published elsewhere (see Methods section).

References

1. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018;392(10159):1736-1788.
2. Bruninx A, Scheenstra B, Dekker A, Maessen J, Hof A van 't, Kietselaer B, et al. Using clinical prediction models to personalise lifestyle interventions for cardiovascular disease prevention: A systematic literature review. *Prev Med Rep* 2022;25:101672.
3. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General Cardiovascular Risk Profile for Use in Primary Care. *Circulation* 2008;117(6):743-53.
4. SCORE2 working group and ESC Cardiovascular risk collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J* 2021;42(25):2439-2454.
5. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ* 2007;335(7611):136.
6. Damen JAAG, Hooft L, Schuit E, Debray TPA, Collins GS, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* 2016;353:i2416.
7. Pearce I, Simó R, Lövestam-Adrian M, Wong DT, Evans M. Association between diabetic eye disease and other complications of diabetes: Implications for care. A systematic review. *Diabetes Obes Metab* 2019;21(3):467-478.
8. Xie J, Ikram MK, Cotch MF, Klein B, Varma R, Shaw JE, et al. Association of Diabetic Macular Edema and Proliferative Diabetic Retinopathy With Cardiovascular Disease. *JAMA Ophthalmol* 2017;135(6):586–593.
9. Barriada RG, Masip D. An Overview of Deep-Learning-Based Methods for Cardiovascular Risk Assessment with Retinal Images. *Diagnostics* 2023;13(1):68.
10. Lee YC, Cha J, Shim I, Park WY, Kang SW, Lim DH, et al. Multimodal deep learning of fundus abnormalities and traditional risk factors for cardiovascular risk prediction. *Npj Digit Med* 2023;6(1):14.
11. Yi JK, Rim TH, Park S, Kim SS, Kim HC, Lee CJ, et al. Cardiovascular disease risk assessment using a deep-learning-based retinal biomarker: a comparison with existing risk scores. *Eur Heart J - Digit Health* 2023;4(3):236-244.
12. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol* 2018;18(1):143.
13. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med* 2018;169(7):467-473.
14. [Online resource] Li LY, Isaksen AA, Lebiecka-Johansen B, Funck KL, Thambawita V, Byberg S, et al. Prediction of cardiovascular markers and diseases using retinal fundus images and deep learning: a systematic scoping review protocol. *Figshare* 2023. Available at <https://doi.org/10.6084/m9.figshare.24793203.v1>
15. [Online resource] Li LY, Isaksen AA, Lebiecka-Johansen B, Funck KL, Thambawita V, Byberg S, et al. Supplementary Materials for Prediction of Cardiovascular Markers and Diseases Using Retinal Fundus Images and Deep Learning: A Systematic Scoping Review. *Figshare* 2024. Available at <https://doi.org/10.6084/m9.figshare.25610088.v1>

16. Haddaway NR, Grainger MJ, Gray CT. Citationchaser: A tool for transparent and efficient forward and backward citation chasing in systematic searching. *Res Synth Methods* 2022;13(4):533-545.
17. Thomas J, Graziosi S, Brunton J, Ghouze Z, O'Driscoll P, Bond M, et al. EPPI-Reviewer: advanced software for systematic reviews, maps and evidence synthesis. London: EPPI-Centre, UCL Social Research Institute, University College London; 2022.
18. Hu W, Yii FSL, Chen R, Zhang X, Shang X, Kiburg K, et al. A Systematic Review and Meta-Analysis of Applying Deep Learning in the Prediction of the Risk of Cardiovascular Diseases From Retinal Images. *Transl Vis Sci Technol* 2023;12(7):14.
19. Li H, Gao M, Song H, Wu X, Li G, Cui Y, et al. Predicting ischemic stroke risk from atrial fibrillation based on multi-spectral fundus images using deep learning. *Front Cardiovasc Med* 2023;10:1185890.
20. Mellor J, Jiang W, Fleming A, McGurnaghan SJ, Blackbourn L, Styles C, et al. 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. *Int J Med Inf* 2023;175:105072.
21. Zhou Y, Chia MA, Wagner SK, Ayhan MS, Williamson DJ, Struyven RR, et al. A foundation model for generalizable disease detection from retinal images. *Nature* 2023;622(7981):156-163.
22. Ankle Brachial Index Collaboration. Ankle Brachial Index Combined With Framingham Risk Score to Predict Cardiovascular Events and Mortality: A Meta-analysis. *JAMA* 2008;300(2):197.
23. Conroy R. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 2003;24(11):987-1003.
24. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017;j2099.
25. Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, et al. 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk. *J Am Coll Cardiol* 2014;63(25):2935-2959.
26. Chua SYL, Thomas D, Allen N, Lotery A, Desai P, Patel P, et al. Cohort profile: design and methods in the eye and vision consortium of UK Biobank. *BMJ Open* 2019;9(2):e025077.
27. Muse ED, Topol EJ. Transforming the cardiometabolic disease landscape: Multimodal AI-powered approaches in prevention and management. *Cell Metab* 2024;36(4):670-683.
28. Rim TH, Lee CJ, Tham YC, Cheung N, Yu M, Lee G, et al. Deep-learning-based cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs. *Lancet Digit Health* 2021;3(5):e306-e316.
29. Tseng RMWW, Rim TH, Shantsila E, Yi JK, Park S, Kim SS, et al. Validation of a deep-learning-based retinal biomarker (Reti-CVD) in the prediction of cardiovascular disease: data from UK Biobank. *BMC Med* 2023;21(1):28.
30. Lee CJ, Rim TH, Kang HG, Yi JK, Lee G, Yu M, et al. Pivotal trial of a deep-learning-based retinal biomarker (Reti-CVD) in the prediction of cardiovascular disease: data from CMERC-HI. *J Am Med Inform Assoc* 2023;31(1):130-138.
31. Mediwhale. Journey. Available from: <https://mediwhale.com/journey/> (accessed 17 Apr 2024).

32. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng* 2018;2(3):158-164.
33. Chang J, Ko A, Park SM, Choi S, Kim K, Kim SM, et al. Association of Cardiovascular Mortality and Deep Learning-Fundusoscopic Atherosclerosis Score derived from Retinal Fundus Images. *Am J Ophthalmol* 2020;217:121-130.
34. Gerrits N, Elen B, Craenendonck TV, Triantafyllidou D, Petropoulos IN, Malik RA, et al. Age and sex affect deep learning prediction of cardiometabolic risk factors from retinal images. *Sci Rep* 2020;10(1):9432.
35. Varma T, Jones CP, Oladele C, Miller J. Diversity in clinical research: public health and social justice imperatives. *J Med Ethics* 2023;49(3):200-203.
36. [Preprint] Dwork C, Hardt M, Pitassi T, Reingold O, Zemel R. Fairness Through Awareness. *ArXiv* 2011;1104.3913.
37. Lee C, Yoon J, Schaar M van der. Dynamic-DeepHit: A Deep Learning Approach for Dynamic Survival Analysis With Competing Risks Based on Longitudinal Data. *IEEE Trans Biomed Eng* 2020;67(1):122-133.
38. Vale-Silva LA, Rohr K. Long-term cancer survival prediction using multimodal deep learning. *Sci Rep* 2021;11(1):13505.
39. Li Y, Sperrin M, Ashcroft DM, Van Staa TP. Consistency of variety of machine learning and statistical models in predicting clinical risks of individual patients: longitudinal cohort study using cardiovascular disease as exemplar. *BMJ* 2020;371:m3919.
40. Van Royen FS, Asselbergs FW, Alfonso F, Vardas P, Van Smeden M. Five critical quality criteria for artificial intelligence-based prediction models. *Eur Heart J* 2023;44(46):4831-4834.
41. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;3(1):160018.
42. Rim TH, Lee G, Kim Y, Tham YC, Lee CJ, Baik SJ, et al. Prediction of systemic biomarkers from retinal photographs: development and validation of deep-learning algorithms. *Lancet Digit Health* 2020;2(10):e526-e536.
43. Diaz-Pinto A, Ravikumar N, Attar R, Suinesiaputra A, Zhao Y, Levelt E, et al. Predicting myocardial infarction through retinal scans and minimal personal information. *Nat Mach Intell* 2022;4(1):55-61.
44. Nagasato D, Tabuchi H, Masumoto H, Kusuyama T, Kawai Y, Ishitobi N, et al. Prediction of age and brachial-ankle pulse-wave velocity using ultra-wide-field pseudo-color images by deep learning. *Sci Rep* 2020;10(1):19369.
45. Barriada RG, Simó-Servat O, Planas A, Hernández C, Simó R, Masip D. Deep Learning of Retinal Imaging: A Useful Tool for Coronary Artery Calcium Score Prediction in Diabetic Patients. *Appl Sci* 2022;12(3):1401.
46. Cho S, Song SJ, Lee J, Song J, Kim MS, Lee M, et al. Predicting Coronary Artery Calcium Score from Retinal Fundus Photographs Using Convolutional Neural Networks. In: Rutkowski L, Scherer R, Korytkowski M, Pedrycz W, Tadeusiewicz R, Zurada JM, eds. *Artificial Intelligence and Soft Computing* 2020:599-612.
47. Son J, Shin JY, Chun EJ, Jung KH, Park KH, Park SJ. Predicting High Coronary Artery Calcium Score From Retinal Fundus Images With Deep Learning Algorithms. *Transl Vis Sci Technol* 2020;9(2):28.

48. Dai G, He W, Xu L, Pazo EE, Lin T, Liu S, et al. Exploring the effect of hypertension on retinal microvasculature using deep learning on East Asian population. *PLOS ONE* 2020;15(3):e0230111.
49. Zhang L, Yuan M, An Z, Zhao X, Wu H, Li H, et al. 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* 2020;15(5):e0233166.
50. Mueller S, Wintergerst MWM, Falahat P, Holz FG, Schaefer C, Schahab N, et al. Multiple instance learning detects peripheral arterial disease from high-resolution color fundus photography. *Sci Rep* 2022;12(1):1389.
51. Ding YD, Zhang Y, He LQ, Fu M, Zhao X, Huang LK, et al. A deep-learning model for the assessment of coronary heart disease and related risk factors via the evaluation of retinal fundus photographs. *Zhonghua Xin Xue Guan Bing Za Zhi* 2022;50(12).
52. Coronado I, Abdelkhaleq R, Yan J, Marioni SS, Jagolino-Cole A, Channa R, et al. Towards Stroke Biomarkers on Fundus Retinal Imaging: A Comparison Between Vasculature Embeddings and General Purpose Convolutional Neural Networks. *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* 2021:3873-3876.
53. Jeena RS, Shiny G, Sukesh Kumar A, Mahadevan K. A Comparative analysis of stroke diagnosis from retinal images using hand-crafted features and CNN. *Journal of Intelligent and Fuzzy Systems* 2021;41(5):5327-5335.
54. Lim G, Lim ZW, Xu D, Ting DSW, Wong TY, Lee ML, et al. Feature Isolation for Hypothesis Testing in Retinal Imaging: An Ischemic Stroke Prediction Case Study. *Proceedings of the AAAI Conference on Artificial Intelligence* 2019;33(01):9510-9515.
55. Al-Absi HRH, Islam MT, Refaee MA, Chowdhury MEH, Alam T. Cardiovascular Disease Diagnosis from DXA Scan and Retinal Images Using Deep Learning. *Sensors* 2022;22(12):4310.

Tables

Table 1 - A priori defined research questions

1	To what extent is deep learning used for predicting cardiovascular risk using retinal fundus images and what characterises the studies?
2	Which cardiovascular markers and diseases are predicted in the studies?
3	What are the study populations (e.g. general population, type 2 diabetes)?
4	How diverse are study populations (age, sex, ethnicity, geographic location)?
5	How do prediction models perform? Which clinical prediction performance metrics are used?
6	Are deep learning models compared to classical clinical risk prediction models?
7	Are the models validated internally or externally?
8	Are the models tested or implemented in a clinical setting?
9	What kind of deep learning architectures are used (e.g. convolutional neural networks)?
10	In the case of incident CVD as the outcome (follow-up studies), how is the time-to-event nature of data handled?
11	Are there studies combining other data modalities (e.g. tabular/structured clinical data) with retinal fundus images?
12	Are there any publicly available datasets (open / restricted by payment / private)?
13	Are the models and the code publicly available (e.g. on Hugging Face or GitHub)?

Table 2 – Study design and outcome of included studies

	Markers of subclinical CVD	Clinical CVD
Cross-sectional studies	systolic and diastolic blood pressure ^{20,32,34,42} left ventricular characteristics ⁴³ brachial-ankle pulse-wave velocity ⁴⁴ coronary artery calcium score ^{28–30,45–47} carotid intima-media thickness ³³	hypertension ^{48,49} peripheral arterial disease ⁵⁰ coronary heart disease ^{10,51} cerebrovascular disease ¹⁰ stroke ^{19,52–54} diagnosis of CVD ⁵⁵
Follow-up studies		major adverse cardiovascular events ³² peripheral arterial disease ²⁰ coronary heart disease ^{20,29} stroke ^{20,21,29,30} myocardial infarction ^{20,21,30,43} major heart failure ^{21,30} CVD mortality ^{28–30,33}

Table 3 – Characteristics of studies that compared retina-based prediction models with established clinical cardiovascular risk scores

First author, year	Validation, model	Outcome, main results
Follow-up studies		
Poplin, 2018 ³²	internal split CNN* (Inception-v3)	5-year major adverse cardiovascular events Systematic Coronary Risk Assessment (SCORE): AUROC=0.72 (0.67,0.76) SCORE+retinal images: AUROC=0.72 (0.67,0.76) no improvement with adding algorithm
Chang, 2020 ³³	internal split CNN (Xception model) + Cox regression	CVD mortality Framingham Risk Score (FRS): AUROC=0.78 (0.73,0.82) FRS+retinal images: AUROC=0.81 (0.76,0.85) improvement with adding images: Δ AUROC = 0.0266 (0.0043,0.0489) $p = 0.02$
Rim, 2021 ²⁸	internal split & external CNN (EfficientNet)	cardiovascular disease events in SCORE outcome Pooled Cohort Equation (PCE): AUROC=0.595 (0.572,0.618) PCE + RetiCAC: AUROC=0.626 (0.595,0.657) improvement with adding images: Δ AUROC=0.031 (0.010,0.051) $p=0.0036$
Tseng, 2023 ²⁹	Not applicable (validation study) CNN + Cox regression	10-year CVD (non-statin group) QRISK3: AUROC=0.682 (0.672, 0.692) QRISK3+RetiCAC: AUROC=0.696 (0.686, 0.706) improvement with adding images: Δ AUROC=0.014 (0.010,0.017) $p<0.001$ (stage 1 hypertension cohort) QRISK3: AUROC=0.639 (0.620,0.658) QRISK3+RetiCAC: AUROC=0.652 (0.633,0.671) improvement with adding images: Δ AUROC=0.013 (0.007,0.019) $p<0.001$ (middle-aged cohort) QRISK3: AUROC=0.650 (0.638,0.662) QRISK3+RetiCAC: AUROC=0.674 (0.661,0.686) improvement with adding images: Δ AUROC=0.023 (0.018,0.029) $p<0.001$
Cross-sectional studies		
Lee, 2023 ¹⁰	internal split & external CNN (DenseNet-169)	prevalent CVD (case-control) Pooled Cohort Equation: AUROC=0.677 (0.658, 0.696) clinical risk factors + retinal images: AUROC=0.872 (0.857, 0.886)
*CNN : convolutional neural network		

Figures

Fig 1. PRISMA flow chart

