Journal Pre-proof

Association of Cardiovascular Mortality and Deep Learning-Funduscopic Atherosclerosis Score derived from Retinal Fundus Images

Jooyoung Chang, PHD Candidate, Ahryung Ko, Research Fellow, Sang Min Park, Professor, Seulggie Choi, PHD Candidate, Kyuwoong Kim, PHD Candidate, Sung Min Kim, PHD Candidate, Jae Moon Yun, Professor, Uk Kang, Pres. of, Il Hyung Shin, VP, Joo Young Shin, Professor, Taehoon Ko, Professor, Jinho Lee, Research Fellow, Baek-Lok Oh, Professor, Ki Ho Park, Chair

PII: S0002-9394(20)30127-6

DOI: https://doi.org/10.1016/j.ajo.2020.03.027

Reference: AJOPHT 11278

To appear in: *American Journal of Ophthalmology*

Received Date: 5 December 2019

Revised Date: 16 March 2020

Accepted Date: 17 March 2020

Please cite this article as: Chang J, Ko A, Park SM, Choi S, Kim K, Kim SM, Yun JM, Kang U, Shin IH, Shin JY, Ko T, Lee J, Oh B-L, Park KH, Association of Cardiovascular Mortality and Deep Learning-Funduscopic Atherosclerosis Score derived from Retinal Fundus Images, *American Journal of Ophthalmology* (2020), doi: https://doi.org/10.1016/j.ajo.2020.03.027.

JChang: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing. **AKo**: Conceptualization, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Project administration. **SMPark**: Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition. **SChoi**: Conceptualization, Writing - Review & Editing. **KKim**: Conceptualization, Writing - Review & Editing. **SMKim**: Conceptualization, Writing - Review & Editing. **JMYun**: Conceptualization, Methodology, Data Curation, Writing - Review & Editing. **UKang**: Conceptualization, Writing - Review & Editing, Funding acquisition. **IHShin**: Conceptualization, Writing - Review & Editing, Project administration. **JYShin**: Conceptualization, Writing - Review & Editing. **TKo**: Conceptualization, Methodology, Software, Validation, Investigation, Resources, Data Curation, Writing - Review & Editing, Project administration, Funding acquisition. **JLee**: Conceptualization, Writing - Review & Editing. **BOh**: Conceptualization, Writing - Review & Editing. **KHPark**: Conceptualization, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Abstract

Purpose: The prediction of atherosclerosis using retinal fundus images and deep learning has not been shown possible. The purpose of this study is to develop a deep learning model which predicts atherosclerosis using retinal fundus images and to verify its clinical implications by conducting a retrospective cohort analysis.
Design: Retrospective cohort study.
Methods: The database at Health Promotion Center of Seoul National University Hospital (HPC-SNUH) was used. The deep learning model was trained on 15,408 images to predict carotid artery atherosclerosis, which we named the deep learning-funduscopic atherosclerosis score (DL-FAS). We constructed a retrospective cohort of participants aged 30-80 years who had completed elective health check-ups at HPC-SNUH. Using DL-FAS the as the main exposure, we followed participants for the primary outcome of death due to CVD until Dec. 31st, 2017.
Results: For predicting carotid artery atherosclerosis among testing-set subjects, the model achieved an AUROC, AUPRC, accuracy, sensitivity, specificity, positive predictive value, and negative predictive value of 0.713, 0.569, 0.583, 0.891, 0.404, 0.465, and 0.865 respectively. The cohort comprised of 32,227 participants, 78 CVD deaths, and 7.6-year median follow-up. Those with DL-FAS greater than 0.66 had an increased risk of CVD deaths compared to DL-FAS<0.33 (HR, 95%CI; 8.83, 3.16-24.7). Risk association was significant among intermediate and high Framingham risk score (FRS) subgroups. The DL-FAS improved the concordance by 0.0266 (95% CI, 0.0043-0.0489) over the FRS-only model. Relative integrated discrimination index (IDI) was 20.45% and net reclassification index (NRI) was 29.5%.
Conclusions: We developed a deep learning model which can predict atherosclerosis from retinal fundus images. The resulting DL-FAS was an independent predictor of CVD deaths when adjusted for FRS and added predictive value over FRS.

Association of Cardiovascular Mortality and Deep Learning-Funduscopic Atherosclerosis Score derived from Retinal Fundus Images

Authors:
Jooyoung Chang[a]; Ahryung Ko[b]; Sang Min Park*[a,b]; Seulggie Choi[a]; Kyuwoong Kim[a]; Sung Min Kim[a]; Jae Moon Yun[b]; Uk Kang[c,d]; Il Hyung Shin[c]; Joo Young Shin[e]; Taehoon Ko[f]; Jinho Lee[g]; Baek-Lok Oh[g]; Ki Ho Park[g]

[a]. Department of Biomedical Sciences, Seoul National University Graduate School, Seoul, South Korea
[b]. Department of Family Medicine, Seoul National University Hospital, Seoul, South Korea
[c]. InTheSmart Co., Ltd., Seoul, South Korea
[d]. Seoul National University Hospital Biomedical Research Institute, Seoul, South Korea
[e]. Department of Ophthalmology, Seoul Metropolitan Government Seoul National University Boramae Medical Center, Seoul, Korea
[f]. Office of Hospital Information, Seoul National University Hospital, Seoul, South Korea
[g]. Department of Ophthalmology, Seoul National University Hospital, Seoul, South Korea

Positions:
JC, PHD Candidate at [a]; AK, Research Fellow at [b]; SMP, Professor at [a], Chair at [b]; SC, PHD Candidate at [a]; KK, PHD Candidate at [a]; SMK, PHD Candidate at [a]; JMY, Professor at [b]; UK, Pres. of [c], Professor at [d]; IHS, VP at [d]; JYS, Professor at [e]; TK, Professor at [f]; JL, Research Fellow at [g]; BLO, Professor at [g]; KHP, Chair at [g].

* Corresponding Author:
Sang Min Park, Department of Family Medicine and Biomedical Sciences, College of Medicine, Seoul National University, 101 Daehak-ro, Jongno-gu, Seoul, South Korea
Tel.: 82-2-2072-3331
Fax: 82-2-766-3276
Email: smpark.snuh@gmail.com

Short title: Deep Learning-Funduscopic Atherosclerosis Score and CVD
Word Count: 3,559 (from Introduction to Conclusion)

Abstract

Purpose: The prediction of atherosclerosis using retinal fundus images and deep learning has not been shown possible. The purpose of this study is to develop a deep learning model which predicts atherosclerosis using retinal fundus images and to verify its clinical implications by conducting a retrospective cohort analysis.
Design: Retrospective cohort study.
Methods: The database at Health Promotion Center of Seoul National University Hospital (HPC-SNUH) was used. The deep learning model was trained on 15,408 images to predict carotid artery atherosclerosis, which we named the deep learning-funduscopic atherosclerosis score (DL-FAS). We constructed a retrospective cohort of participants aged 30-80 years who had completed elective health check-ups at HPC-SNUH. Using DL-FAS the as the main exposure, we followed participants for the primary outcome of death due to CVD until Dec. 31st, 2017.
Results: For predicting carotid artery atherosclerosis among testing-set subjects, the model achieved an AUROC, AUPRC, accuracy, sensitivity, specificity, positive predictive value, and negative predictive value of 0.713, 0.569, 0.583, 0.891, 0.404, 0.465, and 0.865 respectively. The cohort comprised of 32,227 participants, 78 CVD deaths, and 7.6-year median follow-up. Those with DL-FAS greater than 0.66 had an increased risk of CVD deaths compared to DL-FAS<0.33 (HR, 95%CI; 8.83, 3.16-24.7). Risk association was significant among intermediate and high Framingham risk score (FRS) subgroups. The DL-FAS improved the concordance by 0.0266 (95% CI, 0.0043-0.0489) over the FRS-only model. Relative integrated discrimination index (IDI) was 20.45% and net reclassification index (NRI) was 29.5%.
Conclusions: We developed a deep learning model which can predict atherosclerosis from retinal fundus images. The resulting DL-FAS was an independent predictor of CVD deaths when adjusted for FRS and added predictive value over FRS.

Table of Contents Statement:

The prediction of atherosclerosis using retinal fundus images and deep learning has not been shown possible. This study develops and validates a deep learning model for atherosclerosis prediction using retinal fundus images. A retrospective cohort analysis for cardiovascular mortality outcomes using this deep learning-funduscopic atherosclerosis score (DL-FAS) shows significantly added value beyond Framingham risk score. The DL-FAS may allow retinal fundus imaging to be used as a non-invasive screening tool for CVD.

Introduction

Cardiovascular disease (CVD) is the most common cause of death worldwide and accounts for 32% of all deaths[1]. In 2015, CVD affected more than 400 million patients and caused more than 17 million deaths worldwide[2]. Thus, the assessment of CVD risk and the prevention thereof is of clinical importance.

Retinal fundus imaging contains valuable information regarding vascular health[3 4 5 6], and with the emergence of computer assisted retinal imaging, various measurements using this modality has been shown to correlate with severity of heart failure[7], certain stoke subtypes[8], and hypertension[9].

Recent developments in deep learning has revealed that CVD risk factors such as age, sex, and smoking status could be predicted using retinal fundus images[10]. However, the prediction of atherosclerosis, a subclinical marker of CVD, using retinal fundus imaging and deep learning has not been shown possible yet. Furthermore, the clinical implications of prediction of cardiovascular-related risk factors by retinal fundus imaging have not been addressed in terms of reclassification of patients at risk of CVD and time-to-event analysis. A formal analysis of additional benefits to the risk stratification of patients and time-to-event analysis can provide clinicians evidence to consider retinal fundus imaging for assessing patients of borderline CVD-risk.

In this study, we developed and validated a deep model which uses retinal fundus images to predict whether a patient has atherosclerosis. We named the predicted value the deep learning-funduscopic atherosclerosis score (DL-FAS). Furthermore, we determined whether DL-FAS added value to the prediction of cardiovascular death above that of the Framingham Risk Score (FRS) and conducted a retrospective cohort of over 30,000 patients for incident cardiovascular deaths.

Methods

Study Population

Data of participants who had completed medical check-ups at the Health Promotion Center of Seoul National University Hospital (HPC-SNUH)[11], from January 2005 through December 2016, and received a retinal fundus image exam, were used for this study. HPC-SNUH offers elective medical health check-ups including a survey, physical examinations, laboratory testing, and medical imaging. Participants must subscribe to one of many packages available at HPC-SNUH and are not offered based on any indication. Retinal fundus imaging is included in all packages, however carotid artery sonography is only offered for more expensive packages. Participant data was merged with the National Death Certificate database to ascertain the death status and cause of death up to December 31st, 2017. Patients were anonymized before analysis was performed, and the need for patient consent was waived by the institutional review board at SNUH (IRB#: H-1703-044-837).

The deep learning model was validated using two-phase approach. Firstly, during the training phase, participants with retinal fundus examinations plus carotid artery sonography were used to develop the deep model for prediction of atherosclerosis (n=6,597). These patients were divided into the training, validation, and

testing sets, where the testing set patients were used to determine whether the model accurately predicted atherosclerosis. The deep model makes predictions for one eye image at a time, and the final averaged score of both eyes was named as the deep learning-funduscopic atherosclerosis score (DL-FAS). Secondly, during the cohort phase, those with only retinal fundus examinations but without carotid artery sonography (n=32,227) were used to validate whether the DL-FAS could predict future cardiovascular deaths. The two-phased validation approach is essential because (1) the ability to predict atherosclerosis must be verified with cardiovascular mortality studies to draw meaningful clinical implications and contribute to clinical decision making and (2) a one-phase approach to directly predict cardiovascular death would ignore time-to-event analysis and provide no meaningful etiological insights.

Carotid artery atherosclerosis measurement

The carotid artery intima-media thickness (CIMT) and existence of carotid artery plaque was used as the proxy marker for atherosclerosis. CIMT was measured through ultrasonography by averaging three measurements 10 mm proximal to the bifurcation [12]. The far wall IMT was identified as the region between the lumen-intima interface and the media-adventitia interface. Those with CIMT measurements of 0.9 mm or more [13] or carotid artery plaque [14] were considered to have atherosclerosis. While thresholds of higher than 0.9 mm are better correlated to CVD outcomes [15], a conservative threshold of 0.9 mm was used because atherosclerosis develops gradually over time and retinal features may be present in earlier stages of disease. Carotid artery plaque was defined as a focal increase in thickness of 0.5 mm or 50% of the surrounding CIMT value. Both the left and right carotid arteries were measured. The carotid artery findings were based on the sonography reports written by board-trained radiologists. For confirmation and data-cleaning, all ultrasonographic images and corresponding reports were reviewed by four board-trained family medicine physicians.

Acquisition of retinal fundus images

A Canon CR-2 (Tokyo, Japan), a digital non-mydriatic fundus camera[16], was used to obtain retinal fundus images. Patients were not pupil-dilated and took one color fundus photo per eye. The field of view included both the disc and macula and was limited to a 45° angle of view.

Deep learning model development and validation

Patients with both retinal fundus imaging and carotid artery sonography on the same health-checkup were used to train and tune the deep model. The training, validation, and testing sets were divided on a patient level as outlined in Supplementary Figure 2. A total of 15,408 images were used in the training process of the model. 5,296 patients were used for training, 647 patients were used for tuning, and 654 patients were used for testing (Supplementary Table 1). Each set had a similar proportion of atherosclerosis positive images (Supplementary Table 1).

Training a deep learning model comes in three basic steps. Firstly, the retinal fundus image is input to the deep learning model which produces a prediction score between 0 and 1. Secondly, the prediction score is compared to the correct label of either atherosclerosis (1) or no atherosclerosis (0) measured by carotid artery sonography. Thirdly, the model's weights are adjusted to minimize its prediction error.

The Xception model [17] was used as the feature extractor followed by two fully connected layers. To speed up training, we used transfer learning with the Keras

software library (version 2.3.1), where the model weights are initialized with pretrained weights learned from ImageNet[18]. Image pre-processing such as random zoom or horizontal flip were used to prevent over-fitting.

During validation, the deep learning model outputs a prediction ranging between 0 and 1 for each eye which results in two prediction values for each patient per visit. To prevent duplicate patients, the DL-FAS was calculated for each patient-visit and defined as the average of predictions for both eyes. The area under receiver operating curve (AUROC) and area under the precision-recall curve (AUPRC) were calculated. The accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated using a threshold resulting in the maximum F1 score. The F1 score is calculated as 2 * (Precision * Recall) / (Precision + Recall). We identified saliency features using guided backpropagation[19, 20] (Supplementary Figure 1).

Cohort Construction

A retrospective cohort study was performed using patients aged 30 to 80 years with the index date set to their first medical health examination between 2005 through 2016. Participants who were used for training the deep model were excluded in the cohort study. Those with missing non-survey covariates were excluded. These patients are referred to as the cohort set and mutually exclusive to the training, validation, and test sets.

The primary exposure was the DL-FAS determined using retinal fundus images taken on the day of enrollment. Because the DL-FAS ranges between 0 and 1, the model thresholds were chosen so that each score category had equal range, namely 0 to 0.33, 0.33 to 0.67, and 0.67 to 1.0. For the stratified analysis, we used score categories resulting in an equal number of people per category, i.e. terciles. The methods for development and validation of the deep model is described in, “Deep learning model development and validation”.

The primary endpoint of the study was cardiovascular mortality defined according to the International Classification for Diseases 10th Revision codes I00 to I99. Patients were followed until Dec. 31st, 2017. The secondary endpoint of the study was all-cause mortality.

Exposure and Covariates

For adjustment of conventional risk factors, FRS was calculated using age, sex, high-density lipoprotein cholesterol, systolic blood pressure, and current smoker [21]. For stratified analysis, high, intermediate, and low risk patients were defined as FRS≥20%, FRS 10-19%, and FRS<10%, respectively.

For additional adjustment, body mass index, alcohol consumption, exercise frequency, diabetes, hypertension, and dyslipidemia were considered. Diabetes was defined as self-reported diabetes, self-reported diabetic medication history, fasting blood glucose > 126 mg/dL, or HbA1c ≥ 6.5%; hypertension as self-reported antihypertensive medication history; and dyslipidemia as low density lipoprotein >160 mg/dL or self-reported medication history. Missing survey variables were considered as a category of their own.

Statistical Analysis

The discrimination, calibration, and reclassification of predicted atherosclerosis score over FRS risk levels was assessed for predicting cardiovascular deaths among cohort-phase patients. Logistic regression was used to model cardiovascular deaths

using FRS risk levels (1) or FRS risk levels plus DL-FAS (2). The relative integrated discrimination improvement (IDI)[22], category-free net reclassification improvement (NRI)[22], and Hosmer-Lemeshow Chi-Square Test with 10 groups [23] was performed by comparing the logits of these two models.

For a time-to-event analysis, a multivariable Cox regression model was used to estimate the hazard ratio and 95% confidence intervals. To show whether the DL-FAS significantly added to the CVD mortality prediction of FRS risk scores[21], we calculated the c-statistic estimates for Cox regression models using FRS risk levels versus FRS risk levels plus DL-FAS. The difference of these concordances was calculated to see whether the addition of the DL-FAS significantly improved the concordance estimate. Concordance estimates and difference of concordance estimates were calculated using Uno's method [24]. All statistical analysis was performed using Statistical Analysis System 9.0 (SAS Institute North Carolina, United States of America).

Results

Among 751 testing set patient visits with 276 tested positive for carotid artery atherosclerosis (prevalence, 0.368), the model was able to predict the sonography-confirmed carotid artery atherosclerosis with an AUROC of 0.713 and AUPRC of 0.569. The model achieved an F1 score of 0.611 at the optimal DL-FAS threshold of 0.368. The accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were 0.583, 0.891, 0.404, 0.465, and 0.865 respectively, indicating a predictive model driven by low specificity. Saliency maps positively identified the retinal vessels contributing to positive atherosclerosis prediction as well as pathologic findings including disc rim narrowing, increased cup-to-disc ratio, peripapillary atrophy, and cotton wool spots (Supplementary Figure 1).

For the validation of the atherosclerosis score to predict cardiovascular mortality, we used the deep model to predict the atherosclerosis of cohort set patients and conducted prediction analysis and a time-to-event analysis. The cohort consisted of 32,227 patients (Figure 2) with varying characteristics with respect to their DL-FAS (Table 1). Of 32,227 patients in the cohort, only 0.23% (74) patients had only one eye image and 99.8% had images for both eyes. The resulting DL-FAS variances were similar—0.025 and 0.021, respectively. There was a strong association between the DL-FAS and age, sex, and FRS risk level. The median follow-up for the study was 7.6 years, with 78 incident CVD deaths.

The prediction modelling was performed for cardiovascular mortality of a logistic model using only FRS risk levels versus a model using FRS risk levels plus DL-FAS. The IDI analysis showed that the model using FRS risk levels plus DL-FAS had IDI of 0.0007 (*p*-value=0.008) and relative IDI of 20.45% over the model using FRS risk levels alone. Category-free NRI was 29.5% (*p*-value=0.009). Hosmer Lemeshow Chi Square Test showed *p*-value of 0.427 for the FRS model and *p*-value=0.609 for the FRS plus DL-FAS model, indicating no evidence for poor fitting.

For the time-to-event analysis, the hazard ratios for each DL-FAS score group were calculated for CVD mortality, adjusting for FRS and other baseline risk factors (Table 2). Compared to the lowest atherosclerosis score group, those with scores 0.33-0.67 and 0.67-1.00 had significantly higher risk of CVD mortality (HR, 95%CI; 2.94, 1.41-6.15; 8.83, 3.16-24.7; respectively). This positive association showed significant

trend (p for trend <0.001). DL-FAS was associated with CVD mortality when adjusted for all baseline covariates in their single covariate forms (Supplementary Table 2). Among individual covariates in the multivariable model, only age, systolic blood pressure, and smoking status were significantly associated with CVD mortality (Supplementary Table 2). The DL-FAS was also associated with all-cause mortality (Supplementary Table 3).

Because DL-FAS were highly correlated with FRS, age, and sex, a stratified analysis was performed for subgroups based on age, sex, and FRS risk levels (Table 3). For subgroups of low, intermediate, and high risk, the highest tercile of DL-FAS had significantly higher risk of CVD mortality compared to the lowest tercile (HR, 95%CI; 4.76, 1.05-21.63; 3.14,1.04-9.47; 5.11, 1.94-13.5; respectively) with significant trend (p for trend; 0.038; 0.032; 0.001; respectively).

For subgroups of patients aged 30-49 years, DL-FAS was not significantly associated with CVD mortality (HR, 95%CI; 1.69, 0.39-7.25; 0.34, 0.03-4.00; respectively). Among those 50 years or more, the highest and middle terciles had significantly higher risk of CVD mortality compared to the lowest tercile (HR, 95%CI; 2.66, 1.15-6.17; 5.09, 2.25-11.6; respectively) with significant trend (p for trend, <0.001). Among both male and female participants, the highest tercile of DL-FAS was associated with a higher risk of CVD mortality compared to the lowest tercile (HR, 95% CI; 3.03, 1.14-8.09; 9.61, 1.95-47.3; respectively).

The concordance estimates for the Cox regression model fitted on FRS only had a concordance of 0.78 (0.73-0.82); the model fitted on FRS plus DL-FAS had a concordance of 0.81 (0.76-0.85). The improvement in concordance was 0.0266 with p-for-difference of 0.020 (Supplementary Table 4).

Discussion

The purpose of our study was to develop a measurement using retinal fundus images which could predict atherosclerosis and stratify the cardiovascular risk of patients over conventional risk factors such as FRS, diabetes, hypertension, dyslipidemia, and health habits. We trained a deep model which predicted atherosclerosis with moderate predictive performance, and the resulting DL-FAS was significantly associated with an increased hazard for CVD mortality among a cohort of otherwise-healthy participants after adjustment for FRS. Furthermore, the significant risk association was evident in the stratified analysis of intermediate-risk participants, and the addition of the DL-FAS significantly improved the concordance estimate over the FRS only model.

Our work is novel in several points. First, the prediction of atherosclerosis using retinal fundus images and deep learning has not been done . Second, our work provides validation using not only a cross-sectional analysis, but also a longitudinal retrospective cohort for CVD mortality outcomes. Using a cohort to verify newly developed deep learning measurements is uncommon. Third, our work shows that the DL-FAS is an independent predictor of CVD mortality over conventional risk estimates such as the FRS. Some have used deep learning to predict cardiovascular risk factors such as age, sex, and blood pressure, but did not show the added diagnostic value of their models over conventional risk-estimate models[10].

The Predictive Value of DL-FAS

Our results show not only that retinal fundus images may be used to predict the

atherosclerosis of the carotid arteries but also that this prediction may add to conventional risk-stratification scores such as the Framingham Risk Score for longitudinal outcomes of cardiovascular mortality. Previous meta-analysis and cohort studies have analyzed the added predictive value of c-reactive protein(CRP)[25], carotid artery intima-media thickness(CIMT)[26], CT coronary artery calcium score (CTCS)[27, 28], and ankle-brachial index(ABI)[29, 30] beyond FRS. Improvement of c-statistics above conventional risk factors and Framingham risk score were insignificant for CIMT (0.00-0.002) [26, 27], ABI (0.00-0.002) [27, 30], and CRP (0.00) [27] while improvements were mostly significant for CTCS (0.02-0.13) [27, 28]. In our work, the improved concordance estimate of DL-FAS for cardiovascular mortality was small but significant (0.027; 95% CI, 0.004-0.049), and relative IDI and NRI measures were also significant at 20.45% and 29.5%, respectively.

One important result of this study is that the DL-FAS showed significant association even among intermediate risk patients. The risk-stratification of such patients has been an area of considerable research, as conventional risk estimates may underestimate risk of patients with evidence of asymptomatic preclinical atherosclerosis[31]. The American College of Cardiology Foundation/American Heart Association (ACC/AHA)[32] and the European Society of Cardiology (ESC)[33] guidelines recommend further investigation for intermediate/moderate risk patients to search for target organ damage including coronary artery calcium score, ankle-brachial index, and atherosclerotic plaque detection by carotid artery scanning. Conducting a retinal fundus image exam for all such patients may be premature at this this stage, but our results show an added benefit of using the DL-FAS among intermediate risk patients. Because retinal fundus imaging is non-invasive compared to blood tests, it may find new possibilities for utilization in stratifying intermediate risk patients.

Mechanism and features

The mechanism by which the deep learning model predicts atherosclerosis is not clear, but we think that the deep model recognized features of the retinal microvasculature to predict atherosclerosis. Many studies have verified significant relationships between findings of retinal fundus images and cardiovascular disease. Retinal vascular pathology is associated with cerebral small vessel disease[34]. Arteriovenous nicking is associated with increased odds of cardiovascular mortality[35]. Retinal vascular caliber is associated with greater risk of death due to coronary heart disease[36]. Retinal microvascular hemorrhage, microaneurysms, soft exudates, and arteriovenous nicking are associated with an increased risk of stroke[37]. Tortuosity is associated with death due to ischemic heart disease[38]. These studies show that the retinal fundus images hold valuable information regarding cardiovascular health, which we presume our model was able to extract.

The saliency maps using guided backpropagation showed that retinal vessels were making positive contributions to the atherosclerosis prediction. Previous works have used retinal fundus images to predict anemia [20] or other cardiovascular risk factors [10] using deep learning and have provided similar saliency maps which identify vascular anatomy. Our work suggests that certain changes in the retinal vasculature may be a biomarker for atherosclerosis. Furthermore, some saliency maps identified pathologic findings related to glaucoma, such as increased cup-to-disc ratio, disc rim narrowing, peripapillary atrophy, and cotton wool spots, which suggests the possible association

between glaucomatous and atherosclerotic change via overlapping mechanisms of hypertension.

Limitations and strengths

Our study must be interpreted considering the following limitations. Firstly, our database constructed using single-center data and consists entirely of Korean nationals. The generalizability of our results may be limited because CVD risk is dependent on ethnicity[39]. Further research is warranted on the development of DL-FAS using multi-ethnic populations. Secondly, while the threshold-independent metric of DL-FAS for prediction of atherosclerosis was greater than baseline, the DL-FAS had low accuracy and specificity at the designated threshold. At the current threshold, the DL-FAS may not be fit for specific detection of atherosclerosis. Thirdly, the DL-FAS did not significantly increase the risk of cardiovascular death among those under 50 years of age. While this was most likely due to the small number of cardiovascular deaths, the application of the DL-FAS among younger age groups may not be appropriate. Fourthly, incident cardiovascular disease, myocardial infarction, or stroke information was not available for the current study, and the primary endpoint of our study was death due to cardiovascular diseases. Hence, such low event rates limit the interpretability of event classification. Our results may not accurately estimate risk of incident sudden cardiovascular diseases like stroke, myocardial infarction, or heart failure. Finally, we did not have access to medical charts to verify the CVD mortality outcomes in the death certificate database. While, by law, only medical professionals can issue death certificates, the lack of a robust chart review may cause misclassification bias.

In our work, we trained a classifier for atherosclerosis prediction then used it for a time-to-event Cox analysis, but several alternative methods may improve the results. The use of Cox directly as a loss function to train the deep model or the incorporation of covariate factors as auxiliary inputs may further improve the predictive value of the deep model. Though these methods provide an opportunity for improved performance, our purpose was not to produce the best model possible but to 1) predict and screen atherosclerosis using retinal fundus images via deep learning and 2) to validate its clinical implications using analyses of risk stratification, cardiovascular mortality association, and improvement beyond FRS. Our purpose was achievable with a deep model using retinal fundus images alone. However, technical optimizations for improvements in predictive power merits future work.

The strengths of our work include the cross-sectional and longitudinal cohort study design, the analysis of a novel measurement over conventional risk factors, and the comparison of concordance estimates of the cohort analysis. The DL-FAS was not only a good predictor of atherosclerosis at one point in time but also associated with incident cardiovascular deaths in the cohort analysis of 245,900 person-years. The adjustment for multiple conventional risk factors namely FRS, body mass index, diabetes, hypertension, dyslipidemia, and other health habits, indicates the DL-FAS is an independent predictor, even within the intermediate risk patient strata. Furthermore, the comparison of concordance estimates between FRS-only and FRS plus DL-FAS shows the added benefit of adjusting for atherosclerosis score derived from retinal fundus images.

Conclusion

In this single-center retrospective cohort study of Koreans, we showed that a deep model could be used to predict atherosclerosis using retinal fundus images. Furthermore, we showed that the resulting DL-FAS was associated with CVD mortality after adjustment of conventional risk factors including FRS and increased the c-statistic of the Cox model beyond FRS.

Acknowledgements

a. Funding/Support: InTheSmart, Co. Ltd., Seoul, South Korea (Grant Number: 0620180650) via Seoul National University Hospital Research Fund and Seoul National University Hospital Research Fund (Grant Number: 0320190160).
b. Financial Disclosures: Uk Kang, Il Hyung Shin are employees of InTheSmart, Co. Ltd.; all other authors have no financial disclosures. All authors declare no conflict of interest.
c. No other acknowledgements.

406 ## References

407 1. Wang H, Naghavi M, Allen C, et al. Global, regional, and national life expectancy,
408 all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a
409 systematic analysis for the Global Burden of Disease Study 2015. The Lancet
410 2016;388(10053):1459-1544.
411 2. Roth GA, Johnson C, Abajobir A, et al. Global, Regional, and National Burden of
412 Cardiovascular Diseases for 10 Causes, 1990 to 2015. J Am Coll Cardiol 2017;70(1):1-
413 25.
414 3. Bhargava M, Ikram M, Wong T. How does hypertension affect your eyes? J Hum
415 Hypertens 2012;26(2):71.
416 4. Flammer J, Konieczka K, Bruno RM, Virdis A, Flammer AJ, Taddei S. The eye
417 and the heart. Eur Heart J 2013;34(17):1270-1278.
418 5. Grassi G, Buzzi S, Dell'Oro R, et al. Structural alterations of the retinal
419 microcirculation in the “prehypertensive” high-normal blood pressure state. Curr Pharm
420 Des 2013;19(13):2375-2381.
421 6. Wong TY. Is retinal photography useful in the measurement of stroke risk?
422 Lancet Neurol 2004;3(3):179-183.
423 7. Nagele MP, Barthelmes J, Ludovici V, et al. Retinal microvascular dysfunction in
424 heart failure. Eur Heart J 2018;39(1):47-56.
425 8. Doubal FN, MacGillivray TJ, Hokke PE, Dhillon B, Dennis MS, Wardlaw JM.
426 Differences in retinal vessels support a distinct vasculopathy causing lacunar stroke.
427 Neurology 2009;72(20):1773.
428 9. Cheung CY, Ikram MK, Sabanayagam C, Wong TY. Retinal microvasculature as
429 a model to study the manifestations of hypertension. Hypertension 2012;60(5):1094-103.
430 10. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk
431 factors from retinal fundus photographs via deep learning. Nat Biomed Eng
432 2018;2(3):158-164.
433 11. Yoon C, Goh E, Park SM, Cho B. Effects of smoking cessation and weight gain
434 on cardiovascular disease risk factors in Asian male population. Atherosclerosis
435 2010;208(1):275-279.
436 12. Stein JH, Korcarz CE, Hurst RT, et al. Use of carotid ultrasound to identify
437 subclinical vascular disease and evaluate cardiovascular disease risk: a consensus
438 statement from the American Society of Echocardiography Carotid Intima-Media
439 Thickness Task Force endorsed by the Society for Vascular Medicine. J Am Soc
440 Echocardiogr 2008;21(2):93-111.
441 13. Mancia G, De Backer G, Dominiczak A, et al. 2007 Guidelines for the
442 Management of Arterial Hypertension: The Task Force for the Management of Arterial
443 Hypertension of the European Society of Hypertension (ESH) and of the European
444 Society of Cardiology (ESC). J Hypertens 2007;25(6):1105-1187.
445 14. Members ATF, Mancia G, Fagard R, et al. 2013 ESH/ESC Guidelines for the
446 management of arterial hypertension: The Task Force for the management of arterial
447 hypertension of the European Society of Hypertension (ESH) and of the European
448 Society of Cardiology (ESC). Eur Heart J 2013;34(28):2159-2219.
449 15. O'Leary DH, Polak JF, Kronmal RA, Manolio TA, Burke GL, Wolfson SK, Jr.
450 Carotid-artery intima and media thickness as a risk factor for myocardial infarction and

stroke in older adults. Cardiovascular Health Study Collaborative Research Group. N Engl J Med 1999;340(1):14-22.
16. Panwar N, Huang P, Lee J, et al. Fundus Photography in the 21st Century—A Review of Recent Technological Advances and Their Implications for Worldwide Healthcare. Telemed J E Health 2016;22(3):198-208.
17. Chollet F. Xception: Deep learning with depthwise separable convolutions. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit, 2017:1251-1258.
18. Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. International journal of computer vision 2015;115(3):211-252.
19. Springenberg JT, Dosovitskiy A, Brox T, Riedmiller M. Striving for Simplicity: The All Convolutional Net. arXiv e-prints, 2014.
20. Mitani A, Huang A, Venugopalan S, et al. Detection of anaemia from retinal fundus images via deep learning. Nat Biomed Eng 2020;4(1):18-27.
21. D'Agostino RB, Sr., Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008;117(6):743-53.
22. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27(2):157-72; discussion 207-12.
23. HOSMER DW, HOSMER T, LE CESSIE S, LEMESHOW S. A COMPARISON OF GOODNESS-OF-FIT TESTS FOR THE LOGISTIC REGRESSION MODEL. Stat Med 1997;16(9):965-980.
24. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med 2011;30(10):1105-17.
25. Kaptoge S, Di Angelantonio E, Lowe G, et al. C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: an individual participant meta-analysis. Lancet 2010;375(9709):132-40.
26. Den Ruijter HM, Peters SA, Anderson TJ, et al. Common carotid intima-media thickness measurements in cardiovascular risk prediction: a meta-analysis. JAMA 2012;308(8):796-803.
27. Ferket BS, van Kempen BJH, Hunink MGM, et al. Predictive value of updating Framingham risk scores with novel risk markers in the U.S. general population. PLoS One 2014;9(2):e88312-e88312.
28. Peters SA, den Ruijter HM, Bots ML, Moons KG. Improvements in risk stratification for the occurrence of cardiovascular disease by imaging subclinical atherosclerosis: a systematic review. Heart 2012;98(3):177-84.
29. Fowkes FG, Murray GD, Butcher I, et al. Ankle brachial index combined with Framingham Risk Score to predict cardiovascular events and mortality: a meta-analysis. JAMA 2008;300(2):197-208.
30. Murphy TP, Dhangana R, Pencina MJ, D'Agostino RB, Sr. Ankle-brachial index and cardiovascular risk prediction: an analysis of 11,594 individuals with 10-year follow-up. Atherosclerosis 2012;220(1):160-7.
31. Reiner Z, Catapano AL, De Backer G, et al. ESC/EAS Guidelines for the management of dyslipidaemias: the Task Force for the management of dyslipidaemias of the European Society of Cardiology (ESC) and the European Atherosclerosis Society

(EAS). Eur Heart J 2011;32(14):1769-818.
32. Greenland P, Alpert JS, Beller GA, et al. 2010 ACCF/AHA guideline for assessment of cardiovascular risk in asymptomatic adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2010;56(25):e50-103.
33. Perk J, De Backer G, Gohlke H, et al. European Guidelines on cardiovascular disease prevention in clinical practice (version 2012). The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts). Eur Heart J 2012;33(13):1635-701.
34. Kwa VI, van der Sande JJ, Stam J, Tijmes N, Vrooland JL. Retinal arterial changes correlate with cerebral small-vessel disease. Neurology 2002;59(10):1536-40.
35. Wong TY, Klein R, Nieto FJ, et al. Retinal microvascular abnormalities and 10-year cardiovascular mortality: a population-based case-control study. Ophthalmology 2003;110(5):933-40.
36. Rochtchina E, Burlutsky G, Liew G, et al. Retinal vessel diameter and cardiovascular mortality: pooled data analysis from two older populations. Eur Heart J 2007;28(16):1984-1992.
37. Wong TY, Klein R, Couper DJ, et al. Retinal microvascular abnormalities and incident stroke: the Atherosclerosis Risk in Communities Study. Lancet 2001;358(9288):1134-40.
38. Witt N, Wong TY, Hughes AD, et al. Abnormalities of retinal microvascular structure and risk of mortality from ischemic heart disease and stroke. Hypertension 2006;47(5):975-81.
39. DeFilippis AP, Young R, Carrubba CJ, et al. An analysis of calibration and discrimination among multiple cardiovascular risk scores in a modern multiethnic cohort. Ann Intern Med 2015;162(4):266-75.

Figures Captions

Central Illustration. Overview Schematic of Deep Learning Model Development and Clinical Validation using Retrospective Cohort Study.

Abbreviations: IDI, integrated discrimination index; NRI, net reclassification index.

Caption: The central illustration shows the overview of the development and clinical validation of the deep learning-funduscopic atherosclerosis score. The model is developed using participants with existing carotid artery sonography data along with retinal fundus images. Patients are randomized patient-wise into training, validation, and testing sets. For clinical validation, participants of age 30 to 80 years with retinal fundus images only are selected and used for calculation of the DL-FAS using the developed model. The DL-FAS is then used as the primary exposure for the prediction analysis and time-to-event analysis of cardiovascular mortality.

Figure 1. The performance metric curves and saliency map of Deep Learning-Funduscopic Atherosclerosis Score for prediction of atherosclerosis and Cox model-fit-statistics beyond Framingham Risk Score.

Caption: [a] The ROC curve for prediction of atherosclerosis using DL-FAS among the test set patients. [b] The PRC curve for prediction of atherosclerosis using DL-FAS among the test set patients. [c] The time-dependent AUROC of FRS only and FRS plus DL-FAS Cox models among cohort set patients calculated by Uno's method based on 50 perturbed samples. [d] Saliency map representation using guided backpropagation for contributions toward positive DL-FAS.

Figure 2. Design of cohort

Caption: No caption.

Table 1. Characteristics of cohort set participants aged 30-80 years whose retinal fundus photographs were not used for training the deep model.

	Total	DL-Funduscopic Atherosclerosis Score		
		0-0.33	0.33-0.67	0.67-1.0
Total, n	32,227	13,057	18,310	860
Age, mean (std)	52.6 (10.6)	44.2 (7.0)	57.6 (8.3)	71.7 (6.3)
Male Sex, %	49.5	44.8	52.6	57.0
Body Mass Index, %				
<23 kg/m^2	41.9	46.6	38.6	40.9
23-25 kg/m^2	25.9	23.9	27.3	25.0
≥25 kg/m^2	32.2	29.5	34.1	34.1
FRS Risk Level [a], %				
Low	58.9	82.5	44.1	15.0
Intermediate	26.3	14.8	34.2	32.4
High	14.8	2.6	21.8	52.6
Cigarette Smoking, %				
Never	47.1	48.9	46.1	43.5
Past	20.3	15.8	23.1	29.0
Current	16.6	19.4	14.9	9.8
Missing	16.0	16.0	15.9	17.8
Alcohol Consumption, %				
Non-drinker	41.3	41.8	40.4	51.6
Current-Drinker	37.4	35.4	39.2	29.5
Missing	21.3	22.8	20.3	18.8
Exercise Frequency, %				
Regular	31.6	35.9	28.7	27.6
None	39.2	35.0	42.4	36.4
Missing	29.2	29.1	29.0	36.0
Diabetes[b], %				
Yes	19.1	12.2	23.2	37.1
Hypertension [c], %				
Yes	17.3	6.1	23.7	49.7
Dyslipidemia [d], %				
Yes	31.8	33.0	30.8	33.8

Abbreviations: DL, deep learning; FRS, Framingham Risk Score.

[a] calculated using age, sex, high-density lipoprotein cholesterol, systolic blood pressure, smoker.

[b] Self-reported diabetes, self-reported diabetic medication history, fasting blood glucose > 126

mg/dL, or HbA1c ≥ 6.5% [c] Self-reported antihypertensive medication history. [d] Low density lipoprotein >160 mg/dL or self-reported medication history

Table 2. Risk of CVD mortality according to predicted atherosclerosis score using multivariable Cox regression model.

	DL-Funduscopic Atherosclerosis Score			
	0-0.33	0.33-0.67	0.67-1	P-trend
CVD mortality				
Cases, n	10	59	9	
Person-years, 10^3	118.7	123.2	4.0	
Incidence Rate, /10^3 PY	0.08	0.48	2.25	
aHR[a] (95% CI)	1 (ref.)	**2.94 (1.41-6.15)**	**8.83 (3.16-24.7)**	**<0.001**

Abbreviations: aHR, adjusted hazards ratio; CVD, cardiovascular disease; DL, deep learning; Q, quantile.

[a]Adjusted for Framingham Risk Score 10-year CVD risk (including age, sex, high-density lipoprotein cholesterol, total cholesterol, systolic blood pressure, smoker), body mass index, alcohol consumption, exercise frequency, diabetes, hypertension, and dyslipidemia.

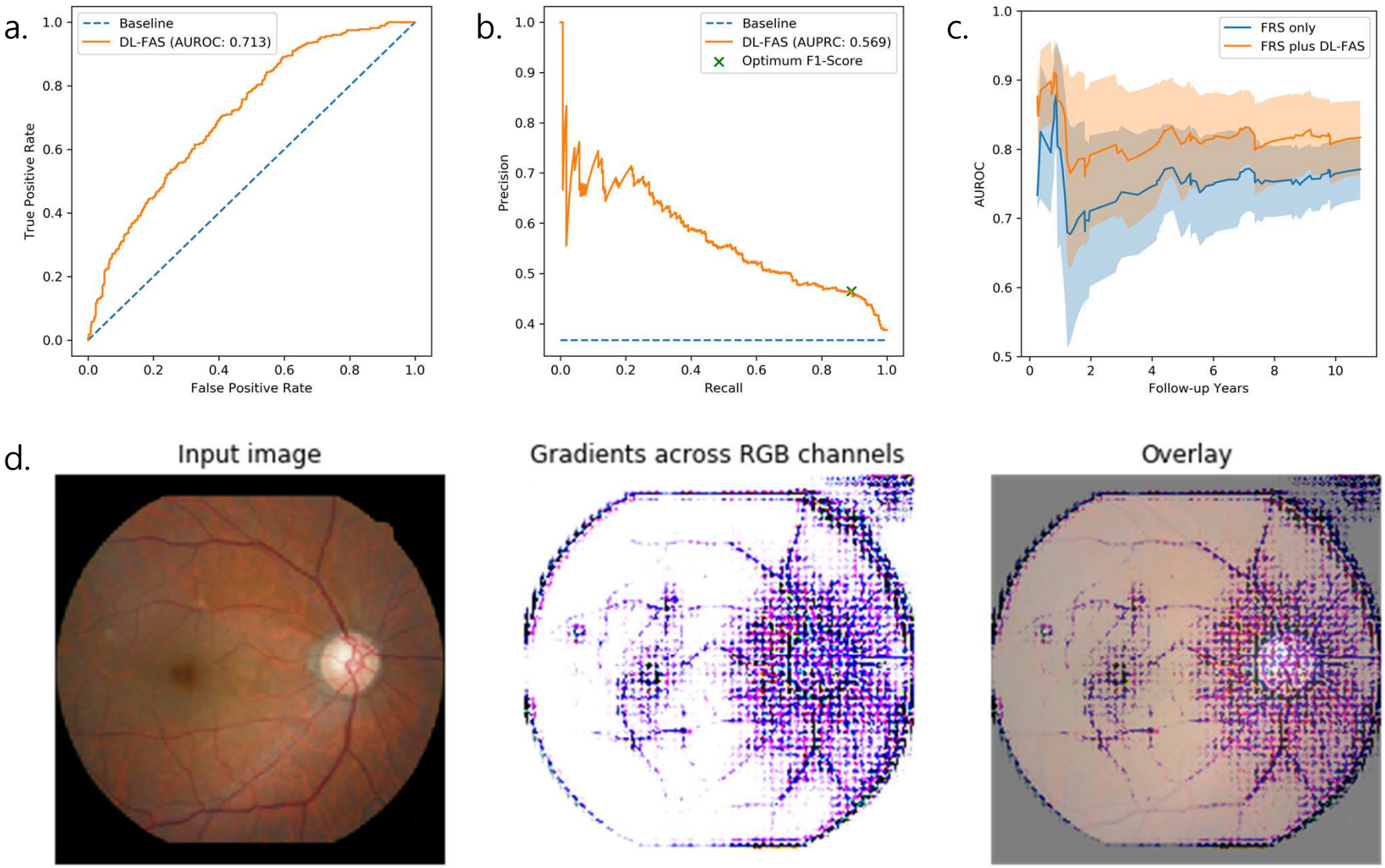
Table 3. Risk of CVD mortality by predicted atherosclerosis score stratified by FRS risk level and age.

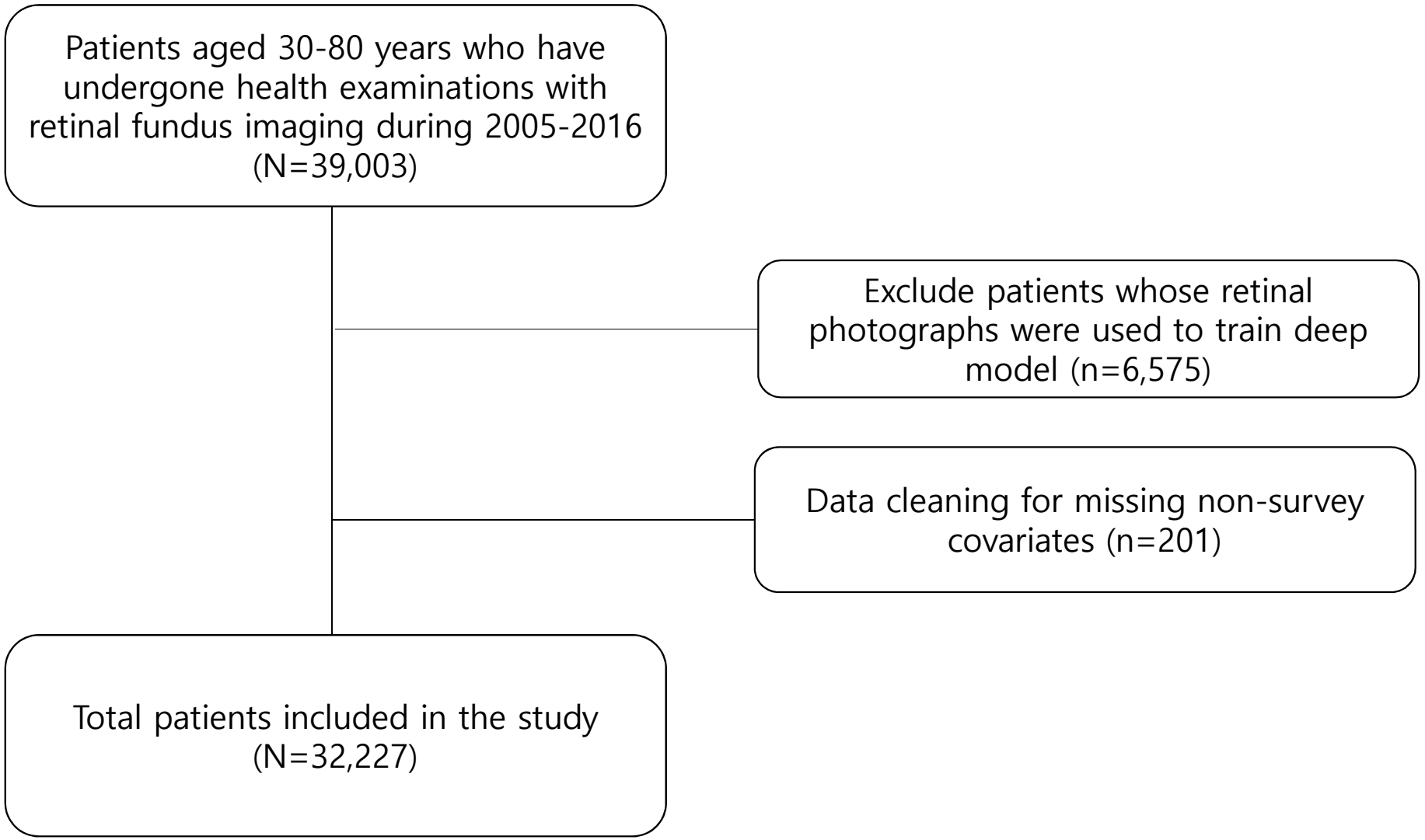
	DL-Funduscopic Atherosclerosis Score			
	Q1	Q2	Q3	P-trend
FRS Low risk				
Quantile limits	0.05-0.25	0.25-0.38	0.38-0.8	
Cases, n	3	5	7	
Person-years, 10^3	61.1	47.9	35.3	
Incidence Rate, /10^3 PY	0.05	0.10	0.20	
aHR[a] (95% CI)	1 (ref.)	2.25 (0.51-9.96)	**4.76 (1.05-21.6)**	**0.038**
FRS Intermediate risk				
Quantile limits	0.12-0.38	0.38-0.5	0.5-0.81	
Cases, n	5	7	12	
Person-years, 10^3	27.1	21.4	16.3	
Incidence Rate, /10^3 PY	0.18	0.33	0.74	
aHR[a] (95% CI)	1 (ref.)	1.50 (0.46-4.84)	**3.14 (1.04-9.47)**	**0.032**
FRS High risk				
Quantile limits	0.17-0.46	0.46-0.57	0.57-0.81	
Cases, n	6	15	18	
Person-years, 10^3	15.0	12.4	9.4	
Incidence Rate, /10^3 PY	0.40	1.21	1.91	
aHR[a] (95% CI)	1 (ref.)	**3.09 (1.18-8.06)**	**5.11 (1.94-13.5)**	**0.001**
Age, 30-49 years				
Quantile limits	0.05-0.22	0.22-0.3	0.3-0.73	
Cases, n	3	6	1	
Person-years, 10^3	42.1	36.1	23.2	
Incidence Rate, /10^3 PY	0.07	0.17	0.04	
aHR[a] (95% CI)	1 (ref.)	1.69 (0.39-7.25)	0.34 (0.03-4.00)	0.5342
Age, 50 years or more				
Quantile limits	0.13-0.4	0.4-0.51	0.51-0.81	
Cases, n	8	19	35	
Person-years, 10^3	60.9	45.7	37.2	
Incidence Rate, /10^3 PY	0.13	0.42	0.94	
aHR[a] (95% CI)	1 (ref.)	**2.66 (1.15-6.17)**	**5.09 (2.25-11.6)**	**<0.001**
Male				
Quantile limits	0.07-0.32	0.32-0.46	0.46-0.81	
Cases, n	7	10	37	
Person-years, 10^3	48.7	40.5	32.9	
Incidence Rate, /10^3 PY	0.14	0.25	1.12	
aHR[a] (95% CI)	1 (ref.)	0.99 (0.36-2.75)	**3.03 (1.14-8.09)**	**0.003**
Female				
Quantile limits	0.05-0.29	0.29-0.44	0.44-0.81	
Cases, n	2	5	17	
Person-years, 10^3	51.3	40.7	31.7	
Incidence Rate, /10^3 PY	0.04	0.12	0.54	

aHR[a] (95% CI)	1 (ref.)	2.63 (0.49-14.0)	**9.61 (1.95-47.3)**	**0.001**

Abbreviations: aHR, adjusted hazards ratio; CVD, cardiovascular disease; DL, deep learning; FRS, Framingham Risk Score; Q, quantile.

[a]Adjusted for Framingham Risk Score 10-year CVD risk (including age, sex, high-density lipoprotein cholesterol, total cholesterol, systolic blood pressure, smoker), body mass index, alcohol consumption, exercise frequency, diabetes, hypertension, and dyslipidemia.


a.
Baseline
DL-FAS (AUROC: 0.713)
True Positive Rate
False Positive Rate
b.
Baseline
DL-FAS (AUPRC: 0.569)
Optimum F1-Score
Precision
Recall
c.
FRS only
FRS plus DL-FAS
AUROC
Follow-up Years
d.
Input image
Gradients across RGB channels
Overlay



Highlights

- Retinal fundus imaging and deep learning may be used for stratification of CVD risk
- Deep learning added predictive value over conventional CVD risk scoring methods
- The developed model was verified on a large cohort of 30,000 Koreans

Dr. Jooyoung Chang is a third-year MD-PhD candidate at the Health System Data Science Laboratory in Seoul National University, Department of Biomedical Sciences. His studies include the application of deep learning and data science to medicine. Chang received his MD degree from Seoul National University and his undergraduate degree in Bio and Brain Engineering from KAIST.

Dr. Sang Min Park leads the Health System Data Science Laboratory at Seoul National University (SNU), Korea. He serves as Department Chair of Family Medicine at Seoul National University Hospital (SNUH), as Professor at Department of Biomedical Sciences, SNU, and as Professor at College of Medicine, SNU. Park earned his PHD from Harvard School of Public Health and his MPH and MD degrees from SNU. He completed his family medicine residency and fellowships at SNUH.