An Explainable Artificial Intelligence-Enabled Electrocardiogram Model for the Prediction of Coronary Artery Calcification

Changho Han

The Graduate School
Yonsei University
Department of Biomedical Systems Informatics

An Explainable Artificial Intelligence-Enabled Electrocardiogram Model for the Prediction of Coronary Artery Calcification

A Dissertation Submitted
to the Department of Biomedical Systems Informatics
and the Graduate School of Yonsei University
in partial fulfillment of the
requirements for the degree of
Doctor of Medicine

Changho Han

December 2024

This certifies that the Dissertation of Changho Han is approved

Thesis Supervisor	Dukyong Yoon
Thesis Committee Member	Jung-Sun Kim
Thesis Committee Member	Jin Young Park
Thesis Committee Member	Hyung-Chul Lee
Thesis Committee Member	Seng Chan You

The Graduate School
Yonsei University

December 2024

ACKNOWLEDGEMENTS

배움에는 졸업이 없습니다. 이번 박사 학위 졸업은 부족한 제가 앞으로 더 나은 연구자, 그리고 더 훌륭한 사람이 되라는 의미로 새기며, 겸손한 자세로 끊임없이 정진하겠습니다.

먼저, 저의 지도교수님이신 윤덕용 교수님께 깊은 존경과 감사의 마음을 전합니다. 교수님께서 앞서 길을 닦아주셨기에 저도 자신 있게 이 길을 걸을 수 있었습니다. 학위 과정 동안 보여주신 아낌없는 지원과 격려, 연구의 방향 설정뿐만 아니라 연구자로서의 올바른 자세와 태도에 대한 가르침은 앞으로도 제 연구와 삶에 큰 지침이 될 것입니다.

항상 곁에서 믿어주시고 조건 없는 사랑으로 보살펴 주신 부모님께 감사드립니다. 때로는 제가 말썽을 피워도 저의 중심을 잡아주시고, 언제나 든든한 버팀목이 되어 올바른 길로 인도해 주셨습니다. 아직 부모님의 깊은 뜻을 온전히 이해하려면 멀었지만, 부모님의 가르침 덕분에 삶의 방향을 찾을 수 있었습니다. 아직 제가 사람으로서 더 많이 성장하고 지혜로워져야 함을 느낍니다. 그럴 때마다 부모님께서 제게 해주셨던 말씀들을 되새기며 마음의 길잡이로 삼겠습니다. 항상 건강하셨으면 좋겠고, 앞으로 더 자랑스러운 아들이 되도록 노력하겠습니다.

연구실 동료 여러분께도 감사의 인사를 드립니다. 여러분 덕분에 함께 성장하며 많은 것을 배울 수 있었고 연구 과정에서 주고받은 조언과 격려가 큰 힘이 되었습니다. 연구실의 직원분들께도 감사드립니다. 가끔 제가 번거롭게 해드렸을 때에도 항상 웃으며 도와주시고 지원해 주셔서 정말 감사했습니다.

언제나 긍정적인 영감을 주며 제 곁을 지켜준 친구들에게도 감사의 마음을 전합니다. 어린 시절 여러분을 만나 함께 자라왔고, 이제는 사회인으로서 발돋움하는 모습을 보니 정말 좋고 뿌듯합니다. 여러분이 제 친구라는 것이 참으로 큰 힘이 됩니다. 앞으로도 서로 돕고 발전하며 좋은 관계를 이어갔으면 좋겠습니다.

2024년 말

한창호

TABLE OF CONTENTS

LIST OF FIGURES ····· iii
LIST OF TABLES ····· v
ABSTRACT ····· vi
1. INTRODUCTION ····· 1
1.1. Background ····· 1
1.1.1. Coronary artery calcification ····· 1
1.1.2. Potential of artificial intelligence in electrocardiogram analysis ····· 2
1.2. Related studies ····· 3
1.3. Objectives ····· 5
2. METHODS ····· 7
2.1. Data sources and labeling ····· 7
2.1.1. Data for model development and testing ····· 7
2.1.2. Data for external validations ····· 14
2.1.3. Data for multinational retrospective cohort analyses ····· 17
2.2. Data preprocessing ····· 19
2.2.1. ECG sampling rate and lead selection ····· 19
2.2.2. Data augmentation ····· 19
2.3. AI-enabled ECG model development ····· 19
2.4. Pre-trained VAE model development ····· 23
2.5. Outcomes ····· 35
2.5.1. AI-ECG model training objective ····· 35
2.5.2. XGBoost model using VAE features ····· 35
2.6. Performance evaluation ····· 36
2.6.1. Performance comparison ····· 36
2.6.2. Performance metrics and risk categorization ····· 36
2.7. Multinational retrospective cohort analyses ····· 37
2.8. Statistical analysis ····· 40
2.9. Software ····· 40
2.10. Ethics approval ····· 40
3. RESULTS ····· 42
3.1. Dataset characteristics ····· 42

3.1.1. Dataset sizes ····· 42
3.1.2. Dataset for model development, testing and external validations ····· 42
3.1.3. Dataset for multinational retrospective cohort analyses ····· 44
3.2. Model performance ····· 47
3.2.1. Model performance in the health screening test dataset ····· 47
3.2.2. Model performance in the external validations ····· 56
3.2.3. Performance comparison ····· 56
3.3. Multinational retrospective cohort analyses ····· 61
3.3.1. SH cohort analysis dataset ····· 61
3.3.2. UKB cohort analysis dataset ····· 67
3.3.3. Subgroup analyses ····· 70
3.4. Interpretation ····· 72
3.5. Compliance with reporting guidelines ····· 79
4. DISCUSSION ····· 84
4.1. Summary of key findings ····· 84
4.2. Implications ····· 84
4.3. Strengths ····· 87
4.4. Limitations ····· 87
5. CONCLUSIONS ····· 89
REFERENCES ····· 90
ABSTRACT IN KOREAN ····· 98
PUBLICATIONS ····· 100
CONFERENCE PRESENTATIONS ····· 102
AWARDS AND HONORS ····· 103
PATENTS ····· 104

LIST OF FIGURES

<Fig 1> Overview of the study. ····· 6
<Fig 2> Data flow diagram (SH, overview). ····· 9
<Fig 3> Data flow diagram (SH, model development dataset). ····· 10
<Fig 4> Data flow diagram (SH, health screening test dataset). ····· 11
<Fig 5> Data flow diagram (SH, cohort analysis dataset). ····· 12
<Fig 6> Data flow diagram (YSH). ····· 15
<Fig 7> Data flow diagram (AUMC). ····· 16
<Fig 8> Data flow diagram (UKB, cohort analysis dataset). ····· 18
<Fig 9> EfficientNet-B0 architecture. ····· 22
<Fig 10> VAE architecture (overview). ····· 25
<Fig 11> VAE architecture (specific). ····· 26
<Fig 12> Factor traversals of all the ECG factors (lead I). ····· 27
<Fig 13> Factor traversals of all the ECG factors (lead II). ····· 28
<Fig 14> Factor traversals of all the ECG factors (lead V1). ····· 29
<Fig 15> Factor traversals of all the ECG factors (lead V2). ····· 30
<Fig 16> Factor traversals of all the ECG factors (lead V3). ····· 31
<Fig 17> Factor traversals of all the ECG factors (lead V4). ····· 32
<Fig 18> Factor traversals of all the ECG factors (lead V5). ····· 33
<Fig 19> Factor traversals of all the ECG factors (lead V6). ····· 34
<Fig 20> ROC curves of the AI-ECG model (CACS ≥ 400). ····· 48
<Fig 21> ROC curves of the AI-ECG model (CACS > 0). ····· 49
<Fig 22> PR curves of the AI-ECG model (CACS ≥ 400).····· 50
<Fig 23> PR curves of the AI-ECG model (CACS > 0). ····· 51
<Fig 24> Boxplot of AI-ECG score by CACS group. ····· 52
<Fig 25> AI-ECG-derived reclassification within each PCE risk category (SH, health screening test dataset).····· 55
<Fig 26> ROC curves (XGBoost using traditional ECG features, CACS ≥ 400). ····· 57
<Fig 27> ROC curves (XGBoost using traditional ECG features, CACS > 0). ····· 58
<Fig 28> PR curves (XGBoost using traditional ECG features, CACS ≥ 400). ····· 59
<Fig 29> PR curves (XGBoost using traditional ECG features, CACS > 0). ····· 60
<Fig 30> AI-ECG-derived reclassification within each PCE risk category (SH, cohort analysis

dataset). ····· 62
<Fig 31> Kaplan-Meier curves (SH cohort analysis dataset). ····· 63
<Fig 32> AI-ECG-derived reclassification within each PCE risk category (UKB, cohort analysis dataset). ····· 68
<Fig 33> Kaplan-Meier curves (UKB cohort analysis dataset). ····· 69
<Fig 34> Hazard ratios of the AI-ECG model by subgroups. ····· 71
<Fig 35> SHAP summary plot of XGBoost model using VAE features. ····· 73
<Fig 36> Factor traversals of ECG factor 48. ····· 74
<Fig 37> Factor traversals of ECG factor 39. ····· 75
<Fig 38> Factor traversals of ECG factor 36. ····· 76
<Fig 39> Factor traversals of ECG factor 23. ····· 77
<Fig 40> Factor traversals of ECG factor 31. ····· 78

LIST OF TABLES

<Table 1> Examples of CT readings containing CACS. ····· 13
<Table 2> Neural network architecture summary (EfficientNet-B0). ····· 21
<Table 3> Neural network architecture summary (VAE). ····· 24
<Table 4> Variables extracted from the UKB ····· 39
<Table 5> Dataset characteristics (model development, testing and external validation datasets). ····· 43
<Table 6> Dataset characteristics (SH cohort analysis dataset). ····· 45
<Table 7> Dataset characteristics (UKB cohort analysis dataset). ····· 46
<Table 8> AI-ECG performance for predicting CACS ≥ 400. ····· 53
<Table 9> AI-ECG performance for predicting CACS > 0. ····· 54
<Table 10> Net reclassification improvement. ····· 64
<Table 11> Cox regression analysis results. ····· 65
<Table 12> C-index comparison. ····· 66
<Table 13> TRIPOD checklist (prediction model development and validation). ····· 80

ABSTRACT

An Explainable Artificial Intelligence-Enabled Electrocardiogram Model for the Prediction of Coronary Artery Calcification

The coronary artery calcium (CAC) score, assessed via computed tomography (CT) to quantify calcium deposits in the coronary arteries, is a marker of atherosclerosis and a robust predictor of coronary events. Current cardiovascular disease (CVD) risk prediction models, such as the ACC/AHA Pooled Cohort Equations (PCE), guide primary prevention but often yield borderline risk classifications, leaving decision-making uncertain. In such cases, CAC scoring is recommended by guidelines as an additional tool to guide decisions. The presence of CAC often favors initiating primary prevention measures like statin therapy, while its absence may lead to withholding statins.

However, routine CAC scoring is limited by high costs, radiation exposure, and lack of insurance coverage. In contrast, electrocardiograms (ECGs) are widely used, non-invasive, cost-effective, and radiation-free. Advances in deep convolutional neural networks have enabled artificial intelligence (AI) models to detect previously undetectable conditions from ECGs. An AI-ECG capable of predicting CAC could provide valuable CVD risk insights. Especially in routine health screenings, where ECGs are widely performed, this approach could enable opportunistic CAC screening in the general population, facilitating earlier detection of coronary artery calcification and timely implementation of primary prevention strategies.

We aimed to develop an AI-ECG model to predict CAC and validate its potential for opportunistic screening in health screening settings. To ensure broader applicability, we aimed to perform external validation in health screening settings at two separate institutions. Additionally, we aimed to evaluate the clinical implications and potential impact of our AI-ECG model on decision-making through multinational retrospective cohort analyses spanning two different countries. Finally, we aimed to provide visual morphological explainability of model predictions.

The AI-ECG model was trained on over 194,000 ECGs annotated with CAC scores from Severance Hospital (SH). It was tested on a health checkup dataset (SH, 14,242 ECGs) where participants had both ECG and CT measurements of CAC on the same visit. External validation used datasets from Yongin Severance Hospital (YSH, 729 ECGs) and Ajou University Medical Center (AUMC, 2056 ECGs). In multinational retrospective cohort analyses, 52,400 ECGs from SH health screenings (not matched with CAC measurements) and 30,623 ECGs from the United Kingdom Biobank (UKB) were utilized. A variational autoencoder (VAE) pre-trained on over 5 million ECGs was employed to enhance interpretability by providing visual explanations of ECG features influencing predictions.

Our AI-ECG model showed strong performance in predicting CAC, achieving an AUROC of 0.841 for CACS $\geq$ 400 and an AUROC of 0.720 for CAC > 0 in the health screening test dataset.

Our AI-ECG model demonstrated robust performance in external validation, with AUROCs of 0.784 and 0.814 in YSH and AUMC datasets for predicting CACS ≥ 400, and 0.691 and 0.701 for CACS > 0.

The AI-ECG model could screen individuals in the PCE low-risk group with the highest likelihood of having CAC, and those in the PCE moderate-risk group with the lowest likelihood of having CAC: In the PCE low-risk category, 24.9% had CAC > 0; among these individuals, the proportion increased to 45.7% when selecting those identified as high-risk by the AI-ECG model; In the PCE moderate-risk category, 35.4% had CACS = 0; among these individuals, the proportion increased to 60.9% when selecting those identified as low-risk by the AI-ECG model. Among PCE low-risk individuals who were reclassified as high-risk by AI-ECG, the incidence rate (IR) of major adverse cardiovascular event (MACE) was higher compared to those in the PCE moderate-risk category who were reclassified as low-risk by AI-ECG (SH cohort analysis dataset MACE IR per 1000 person-year [PY]: 6.0 vs. 3.3, P = 0.007, UKB cohort analysis dataset MACE IR per 1000 PY: 8.3 vs. 7.0, P = 0.360). Thus, it would be more reasonable to withhold statin therapy (down-risk) in individuals classified as PCE moderate risk but AI-ECG low risk, and to initiate statin therapy (up-risk) in those classified as PCE low risk but AI-ECG high risk. AI-ECG was an independent risk factor for MACE (adjusted hazard ratio [95% CI]: 1.087 [1.053–1.123] in the SH cohort analysis dataset and 1.117 [1.061–1.175] in the UKB cohort analysis dataset). AI-ECG provided additional predictive value beyond the PCE, with the combined PCE plus AI-ECG score outperforming the PCE alone in terms of C-index. The association between AI-ECG and MACE remained consistent across all demographic and PCE-based subgroups. We provided visual morphological interpretations of ECG factors associated with increased predicted risk, identifying potential changes such as upward shift of the ST segment in the anteroseptal leads with reciprocal downward shift in the inferolateral leads, downward shift of the ST segment in all leads, longer PR interval, and others, to be associated with CAC.

Our AI-ECG model proves to be an effective tool for predicting coronary artery calcification. We demonstrated the potential integration of our AI-ECG model into clinical workflow by showing its dual utility: it can either screen individuals who would benefit most from CACS measurement, or directly guide decisions regarding statin therapy initiation or withholding through patient reclassification. The ubiquitous availability of ECGs, combined with our finding that the AI-ECG model serves as an independent risk factor for cardiovascular events, suggests its potential for incorporation as a CVD risk prediction tool. Particularly in routine health screenings where ECGs are universally performed, AI-ECG-based CAC prediction could enable opportunistic CAC screening in the general population, paving the way for earlier detection and timely implementation of primary prevention strategies.

Key words: coronary artery calcification, coronary artery calcium score, artificial intelligence, electrocardiogram, AI-ECG, health screening, primary prevention

1. INTRODUCTION

1.1. Background

1.1.1. Coronary artery calcification

Atherosclerotic cardiovascular disease (ASCVD) remains a leading cause of morbidity and mortality worldwide, imposing a substantial burden on individuals and healthcare systems. The estimated number of individuals aged 20 and older affected by coronary heart disease in the United States is 18.2 million[1]. Each year, around 605,000 Americans experience their first myocardial infarction, while over 200,000 have a recurrent event[2]. In 2016, the health care costs associated with coronary heart disease amounted to $80 billion in the United States[3]. Early identification of individuals at high risk is critical for implementing preventive measures and reducing adverse cardiovascular events.

In current clinical practice, cardiovascular disease (CVD) risk prediction frameworks, such as the American College of Cardiology/American Heart Association (ACC/AHA) Pooled Cohort Equations (PCE), evaluate an individual's risk of developing CVD by incorporating clinical variables[4]. These frameworks guide recommendations for primary prevention strategies based on the calculated risk score. However, risk assessment often yields borderline or intermediate classifications where the decision to initiate treatment remains uncertain. In these cases, coronary artery calcium (CAC) scoring is recommended by several guidelines as an additional measure to guide primary prevention decisions[5-7]. The presence of CAC (i.e., CAC score > 0) in such individuals often tips the balance toward initiating primary prevention measures such as statin therapy, whereas the absence of CAC may sometimes lead to recommendations to withhold statins.

Coronary artery calcification (CAC) is associated with the progression of advanced atherosclerosis[8,9]. The coronary artery calcium score (CACS), assessed by computed tomography (CT) to quantify calcium deposits in the coronary arteries, serves as an excellent measure of atherosclerotic plaque burden[8-11]. Numerous studies validate that CAC is indicative of subclinical atherosclerosis[12-14]. The presence and burden of CAC provide direct evidence of the extent of coronary artery disease (CAD) and predict future cardiovascular events independently of traditional risk factors, with predictive power that has been widely corroborated and surpasses that of any other non-invasive biomarker for this condition[9,15-21]. These insights have positioned CAC scoring as a valuable tool for stratifying cardiovascular risk.

Despite its clinical value and inclusion in several guidelines for cardiovascular risk stratification, routine CAC measurement is limited by significant barriers. CT-based CAC scoring involves high costs, exposes patients to radiation, and is largely inaccessible in resource-limited settings. Moreover, insurance coverage for CAC screening is virtually nonexistent[22,23], further restricting its widespread adoption. These limitations underscore the need for alternative methods to identify the disease early in a cost-effective and accessible manner.

1.1.2. Potential of artificial intelligence in electrocardiogram analysis

In contrast, an electrocardiogram (ECG) stands as a sensitive, cost-effective, non-invasive, and radiation-free diagnostic tool commonly utilized in various health evaluations. The recent application of artificial intelligence (AI) techniques to ECGs has enabled the automatic classification or diagnosis of various cardiac diseases, such as arrhythmia and ischemia[24-28]. Moreover, with the leverage of deep convolutional neural networks on ECGs, numerous AI models proficiently identifying diseases and conditions that were previously undetectable through conventional ECG interpretation have emerged[27,29]. Importantly, many of these advanced AI models have demonstrated their effectiveness through rigorous prospective validation in real-world scenarios. For example, Attia et al. (2019) developed an AI-enabled ECG (AI-ECG) algorithm capable of identifying patients with atrial fibrillation during normal sinus rhythm, while Noseworthy et al. (2022) found in a prospective trial that the AI-guided targeted screening of atrial fibrillation with ECGs actually resulted in a significant increase in atrial fibrillation detection rates, particularly among those classified as high-risk by the algorithm[30,31]. Moreover, Attia et al. (2019) developed an AI-ECG algorithm capable of identifying patients at a high likelihood of low ejection fraction, while Yao et al. (2021) found in a pragmatic randomized clinical trial that the usage of this AI-powered clinical decision support tool significantly improved the early diagnosis of patients with low ejection fraction in routine primary care settings[32,33].

Furthermore, various techniques have been introduced to visually explain AI-ECG predictions, which is crucial as it transforms the AI from a black-box model to a transparent and interpretable tool[34,35]. For example, van de Leur et al. (2022) and Wouters et al. (2023) pre-trained a variational autoencoder (VAE) model to learn the intrinsic factors influencing median beat ECG morphology in an unsupervised manner[35,36]. They then used this pre-trained VAE model to explain ECG morphological features related to various ECG diagnosis statements, reduced ejection fraction, mortality, and outcomes following cardiac resynchronization therapy[35,36].

If an AI-ECG can predict CAC, individuals undergoing ECGs can gain valuable insights into potential ASCVD risks. Especially in routine health screenings, where ECGs are universally

performed, this approach could enable opportunistic screening for coronary artery calcification in the general population, paving the way for earlier detection and timely implementation of primary prevention strategies.

1.2. Related studies

In a study by Farjo et al. (2020)[37], a logistic regression machine learning model was developed to predict CACS of 400 or higher using ECG features from continuous wavelet transforms alongside various clinical features. To our knowledge, this was the first study to create a machine learning model for predicting CACS. The study used data from 534 subjects, split into training (80%) and testing (20%) sets. A second cohort of 87 patients undergoing invasive coronary angiography was used to validate the model. The machine learning models were developed to predict binary outcomes: CAC = 0 and CAC ≥ 400. The CAC = 0 model, based on clinical features, achieved an area under the receiver operating characteristics curve (AUROC) of 0.84, with sensitivity, specificity, and accuracy of 92%, 70%, and 75%, respectively. The CAC ≥ 400 model, using both ECG and clinical features, had an AUROC of 0.87, with sensitivity, specificity, and accuracy of 91%, 75%, and 81%, respectively. The CAC ≥ 400 model was also tested for its ability to predict outcomes such as coronary artery stenosis, revascularization needs, and major adverse cardiovascular events in patients. The CAC ≥ 400 model significantly predicted the need for revascularization ($P < 0.001$), and major adverse cardiovascular events during a two-year follow-up period. The machine learning models demonstrated the ability to use easily obtainable clinical and ECG data to predict CACS and stratify cardiovascular risk, potentially providing a low-risk, non-invasive alternative to current methods. However, none of the ECG features ranked among the top three most important factors. Instead, the top three were coronary artery disease, age, and sex, suggesting that clinical features played a more significant role than ECGs in predicting CACS of 400 or higher in this study. More advanced AI methods, such as deep convolutional neural networks, could be essential for extracting more complex features from ECG data to enhance predictive accuracy. While the study demonstrates the potential of using machine learning models based on ECG and clinical data for cardiovascular risk assessment, it emphasizes the need for further research with larger and more diverse populations to fully realize the clinical utility of these approaches.

A study by Han et al. (2022)[38] explored the potential of applying deep learning to ECGs to predict CACS. The research developed binary classification models using deep convolutional neural networks to predict CACS (≥ 100, ≥ 400, and ≥ 1000) solely from ECG waveform data. The model development and internal validation dataset included 8,178 ECGs from 5,765 patients, and the external validation dataset included 1,745 ECGs from 877 patients. The ECG data were paired with CACS measured within 60 days of the ECG recording to ensure that the CAC values remained

relevant to the corresponding ECG data. In the internal validation, the models achieved AUROC scores of 0.753, 0.802, and 0.835 for CACS ≥ 100, ≥ 400, and ≥ 1000, respectively. Similarly, in the external validation, the models showed AUROCs of 0.718, 0.777, and 0.803, indicating that the models were generalizable across different populations. The AI models outperformed traditional logistic regression models that relied on conventional ECG features. This suggests that deep learning techniques are better at detecting subtle signals in ECGs that may correlate with coronary artery calcification, making them more effective for predicting CAC. However, the study had several limitations. However, a notable limitation of this study is the lack of exclusion of individuals with clinical ASCVD. According to guidelines[5-7], CAC scoring is not recommended for individuals with clinical ASCVD, as they are already candidates for primary prevention or more advanced treatments. Consequently, AI-ECG-based CAC prediction holds no utility for this group. Without excluding individuals with clinical ASCVD, it is unclear whether the model's performance is biased toward those with an existing diagnosis or symptoms, and whether it would be equally effective for those who are the intended targets of opportunistic detection. Another limitation was the lack of demonstration of an association between the AI model's predictions and actual cardiovascular disease outcomes. Furthermore, the impact of the AI model on decision-making remains unclear, as it was not demonstrated how the predictions could guide clinical care or influence therapeutic strategies. Additionally, the study lacked interpretability in the AI model, as no techniques were incorporated to explain how the model made its predictions. This "black box" nature is a common challenge in deep learning models. Additionally, the study's dataset was relatively small, with only 8178 ECGs used for training and validation, potentially limiting the model's ability to generalize to broader, more diverse populations.

A study by Awasthi et al. (2023)[39] developed and tested AI models designed to detect CAD using ECGs. This research aimed to enhance the detection of CAD and improve risk stratification for acute coronary events and mortality. The study utilized a large dataset from over 7 million patients across more than 70 hospitals and clinics in the United States, focusing on individuals without a prior history of ASCVD. The study developed separate AI models to detect three specific markers of CAD: elevated CACS, obstructive coronary artery disease, and regional left ventricular akinesis, which could indicate possible prior myocardial infarction. These models achieved high levels of accuracy, with AUROC scores of 0.88 for detecting CACS of 300 or greater, 0.85 for identifying obstructive CAD, and 0.94 for detecting regional left ventricular akinesia. This level of accuracy suggests the AI models are highly effective at detecting underlying coronary disease from ECG data alone. One of the study's key findings was that the AI models could predict the risk of acute coronary events and all-cause mortality over time periods as short as three years, even in patients who had no known history of ASCVD. Patients who tested positive on one, two, or all three AI models were found to have progressively higher hazard ratios for acute coronary events, including myocardial infarctions, compared to those who tested negative. For instance, patients testing positive on all three models had a significantly higher risk of acute coronary events over a

three-year period, with a hazard ratio of 11.75, compared to those with no positive results. The study also highlighted the potential clinical utility of AI-ECGs in providing more timely risk assessments than conventional methods. However, one limitation of the study was the lack of interpretability in the AI models, as no techniques were employed to explain how the models made their predictions. Additionally, regarding the CACS prediction model, it was unclear whether the cohort used in the study included individuals with clinical ASCVD for whom CAC scoring is not recommended according to guidelines[5-7], as they are already candidates for primary prevention or advanced treatments. As a result, AI-ECG-based CAC prediction has no utility for individuals with clinical ASCVD. Without excluding this group, it is unclear whether the model's performance is biased toward those with existing diagnoses or symptoms and whether it would be similarly effective for the intended population targeted for opportunistic detection.

1.3. Objectives

In this study, we developed an AI-ECG model to predict coronary artery calcification and validated its potential for opportunistic screening in health screening settings. To ensure broader applicability, we performed external validation in health screening settings at two separate institutions. We evaluated the clinical implications and potential impact of our AI-ECG model on decision-making through multinational retrospective cohort analyses spanning two different countries. We also provided visual morphological explainability of model predictions.

A more detailed overview of the study is provided below (Figure 1). We developed our AI-ECG model predicting coronary artery calcification using an extensive dataset comprising over 194,000 ECGs annotated with CACS. We then tested the model on a health screening dataset comprising individuals with both ECG and CT-derived CAC measurements obtained during their evaluations, validating its potential for opportunistic screening. We then conducted external validations using health screening datasets from separate institutions to evaluate the model's applicability in distinct yet potentially analogous populations. Additionally, we validated the clinical implications of our AI-ECG model through multinational retrospective cohort analyses using the UK Biobank (UKB), a dataset representative of the general population in the United Kingdom, and a health screening dataset from South Korea, focusing on cardiovascular events. We assessed how effectively the AI-ECG model reclassifies individuals for initiating primary prevention decisions, beyond traditional risk assessment tools, to evaluate its potential impact on decision-making. Moreover, we examined whether the AI-ECG model serves as an independent risk factor for predicting cardiovascular events. Finally, we integrated a pre-trained VAE model to provide visual morphological explainability, as described in previous studies[35,36].

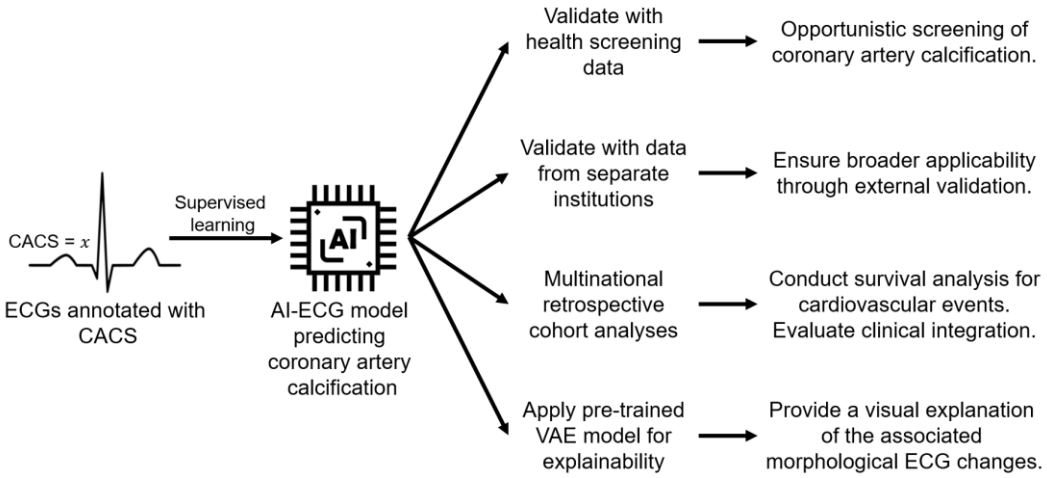


Figure 1. Overview of the study.

CACS: coronary artery calcium score; ECG: electrocardiogram; AI: artificial intelligence; AI-ECG: artificial intelligence-enabled electrocardiogram; VAE: variational autoencoder.

2. METHODS

2.1. Data sources and labeling

2.1.1. Data for model development and testing

The standard 12-lead ECG data and electronic medical records (EMR) from Severance Hospital (SH) were utilized for AI-ECG model development and internal validation (Figures 2-5). The 12-lead ECG database from SH, sourced from the General Electric (GE) Healthcare MUSE™ system, encompasses approximately 5.6 million ECG records from 1.4 million individuals, spanning from 1993 to 2022. The database incorporates data from health screenings. This database consists of raw waveforms (one-dimensional ECG signal), measurement metrics like heart rate, PR interval, and QT interval, along with automatic ECG interpretations generated by the GE ECG machine. Each ECG recording has a duration of 10 seconds with sampling rates of either 500 Hz or 250 Hz. The GE ECG algorithm constructs a median waveform for each ECG recording, spanning 1.2 seconds. This is achieved by aligning all QRS complexes of identical shape and deriving a representative QRS complex using the median voltage. This median waveform is also incorporated into the database.

From the EMR database, we retrieved CT readings of heart-related scans conducted between November 2005 and August 2022 for individuals aged 18 and above. Examples of CT readings containing CACS are shown in Table 1. Notably, CT scans undertaken during health screenings were available from December 2010 to August 2022. To extract the CACS from these CT readings, we employed regular expressions, using a comprehensive range of search terms such as “calcium score” and “CAC score” to ensure all relevant CACS were extracted.

Data from individuals with both ECG recordings and CACS from CT readings were used to develop the AI-ECG model for predicting CACS (Figures 3 and 4). Specifically, ECGs recorded during health screenings were extracted if a corresponding CAC measurement via CT was performed during the same visit (Figure 4). These ECGs were subsequently labeled with the corresponding CACS. These ECGs were designated as the health screening hold-out test dataset. ECGs not recorded during health screenings were extracted if their recordings fell within a 30-day period surrounding the CAC measurements, either preceding or following them, and these ECGs were subsequently labeled with the respective CACS (Figure 3). We chose this 30-day window since CAC is a gradually progressing condition, making it unlikely for significant changes to be observed

within a month[40,41]. While a longer time window might provide more samples, it also increases the potential for substantial variation in the CACS. Thus, we chose a time window of 30 days to balance between these trade-offs. In instances where multiple CAC measurements were taken within this 30-day period relative to an ECG, the ECG was labeled with the CACS from the closest date. These ECGs constituted the model development dataset. ECGs bearing automatic interpretations that included any of the following phrases were excluded: “lead reversal,” suggesting potential lead misplacement; “poor quality,” signifying the presence of artifacts; and “pacemaker,” indicating the potential presence of an artificial pacemaker. To prevent data leakage and overestimation of performance, ECGs from individuals present in the health screening test dataset were additionally excluded from the model development dataset. The model development dataset was then randomly partitioned into training and validation datasets at an 80:20 ratio, stratified by CACS ≥ 400, while ensuring no overlap of individuals between the two.

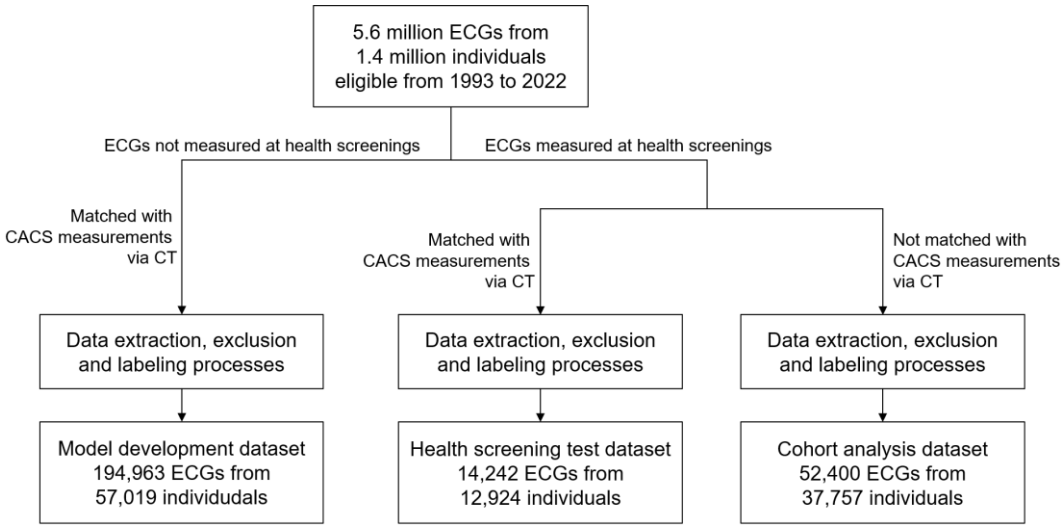


Figure 2. Data flow diagram (SH, overview).

SH: Severance Hospital; ECG: electrocardiogram; CACS: coronary artery calcium score; CT: computed tomography.

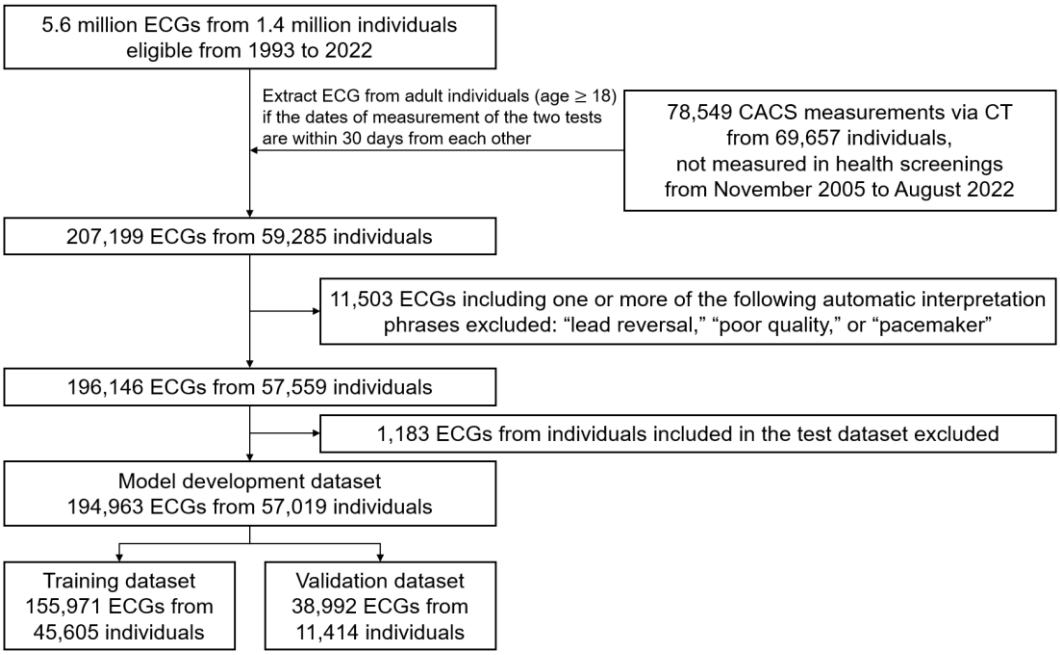


Figure 3. Data flow diagram (SH, model development dataset).

SH: Severance Hospital; ECG: electrocardiogram; CACS: coronary artery calcium score; CT: computed tomography

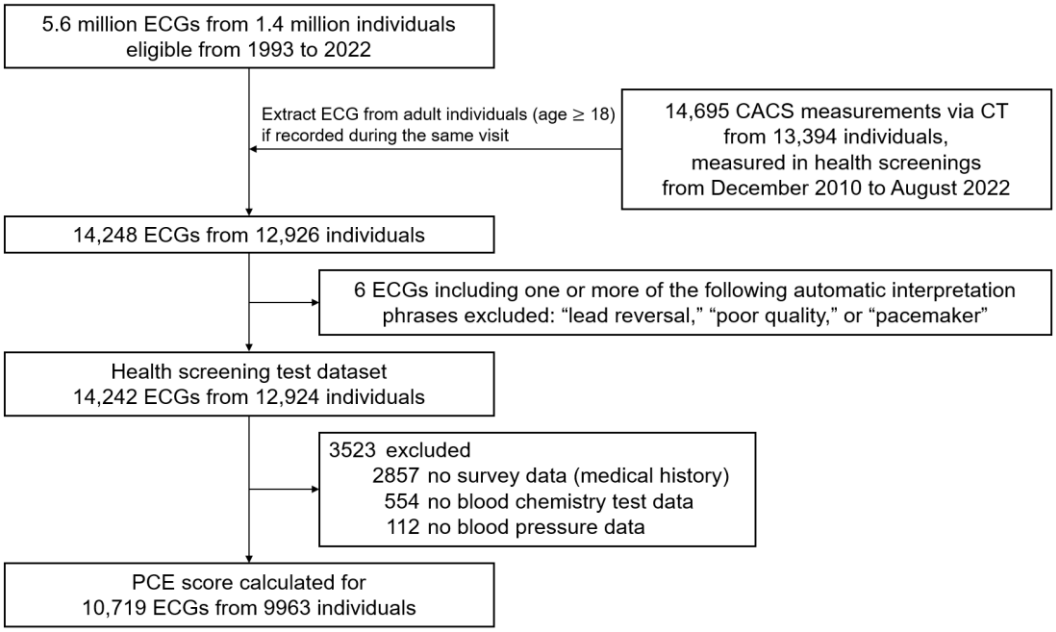


Figure 4. Data flow diagram (SH, health screening test dataset).

SH: Severance Hospital; ECG: electrocardiogram; CACS: coronary artery calcium score; CT: computed tomography

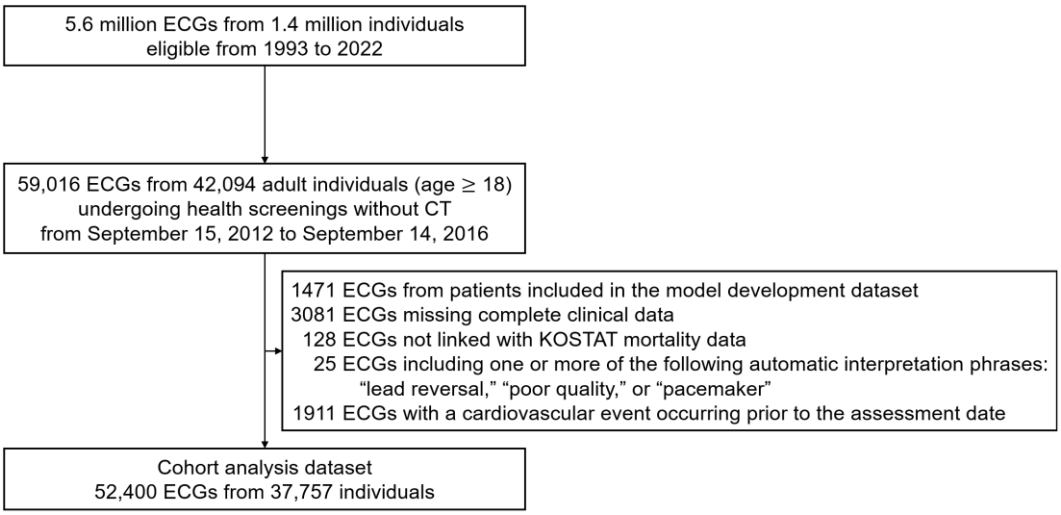


Figure 5. Data flow diagram (SH, cohort analysis dataset).

SH: Severance Hospital; ECG: electrocardiogram; CT: computed tomography, KOSTAT: Statistics Korea

Table 1. Examples of CT readings containing CACS. The table shows two examples of CT readings in which regular expression was applied to extract CACS. We have bolded the sentences including the CACS.

	CT reading
Example 1	Average heart rate : 58 bpm, Reconstruction cardiac phase : 74% Extensive calcified plaques at all three coronary arteries. The maximum diameter stenosis of LAD is 43%. Moderate stenosis (57%) of 1st diagonal branch. The maximum diameter stenosis of LCX is 39%. Discrete near total occlusion at distal PL branch. **Agatston calcium score is 1385.** Global LV systolic function is within normal range. (Indexed values, LVEF: 77%, LVEDV: 69 mL/m2) There is no evidence of intracardiac mass in LV and LA. Valve calcification is not seen. No evidence of pericardial thickening or calcification or effusion. Diffuse atherosclerosis at thoracic aorta. Normal lung parenchyma on covered scan area. Conclusion) 1. Extensive calcified plaques at all three coronary arteries. - with moderate stenosis of 1st diagonal branch. - with discrete near total occlusion at PL branch. 2. Diffuse atherosclerosis at thoracic aorta. -- Two vessel disease adv) coronary angiography
Example 2	Average heart rate : 49 bpm, Reconstruction cardiac phase : 74% Coronary arteries are well pacified without significant stenosis or calcified plaque. **Agatston calcium score is zero.** LV function is within normal range. (LVEF: 70%, LVEDV: 86 mL) There is no evidence of intracardiac mass in LV and LA. Valve calcification is not seen. No evidence of pericardial thickening or calcification or effusion. Normal lung parenchyma and thoracic cage on covered scan area. Conclusion) Normal coronary CT angiogram (Ca. score=0)

CT: computed tomography; CACS: coronary artery calcium score; LAD: left anterior descending artery; LCX: left circumflex artery; LVEF: left ventricular ejection fraction; LVEDV: left ventricular end-diastolic volume; LA: left atrium; LV: left ventricle; PL: posterolateral

2.1.2. Data for external validations

We conducted external validations of our AI-ECG model using health screening data from two separate institutions: Yongin Severance Hospital (YSH) and Ajou University Medical Center (AUMC) (Figures 6 and 7). The standard 12-lead ECG database from YSH, sourced from the GE Healthcare MUSE™ system, encompasses approximately 222,000 ECG records from 102,000 individuals, spanning from March 2020 to July 2023 (Figure 6). The database incorporates data from health screenings. We retrieved CACS from CT reports produced during health screenings between April 2020 and August 2022, utilizing the same search criteria as applied to the SH database. The standard 12-lead ECG database from AUMC, also sourced from the GE Healthcare MUSE™ system, encompasses approximately 1.7 million ECG records from 740,000 individuals, spanning from 1993 to July 2020 (Figure 7). The database incorporates data from health screenings. We retrieved CACS from CT reports produced during health screenings between October 2003 and September 2012, utilizing the same search criteria as applied to the SH database. ECGs recorded during these health screenings were selected and labeled with the respective CACS if a corresponding CT scan from the same visit was available. Each ECG recording had a duration of 10 seconds, with sampling rates of either 500 Hz or 250 H. Each database entry also includes a median waveform, lasting 1.2 seconds. The exclusion criteria based on automatic interpretation phrases, as used for the SH database, were also applied to the ECGs from YSH and AUMC.

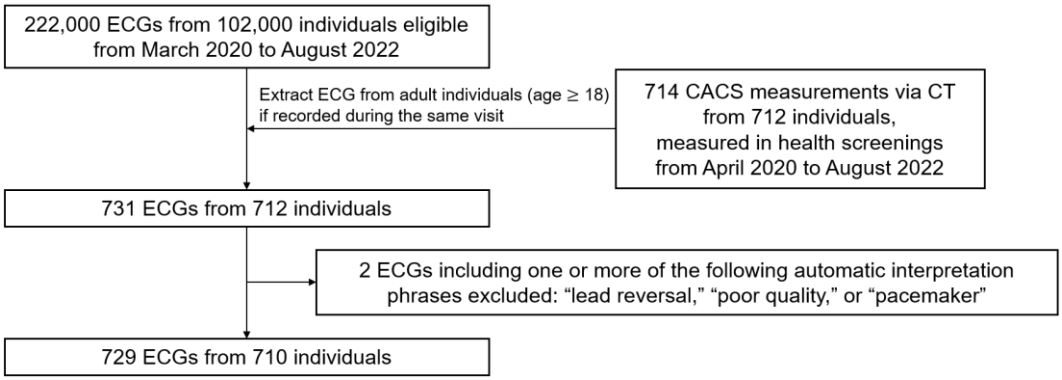


Figure 6. Data flow diagram (YSH).

YSH: Yongin Severance Hospital; ECG: electrocardiogram; CACS: coronary artery calcium score; CT: computed tomography.

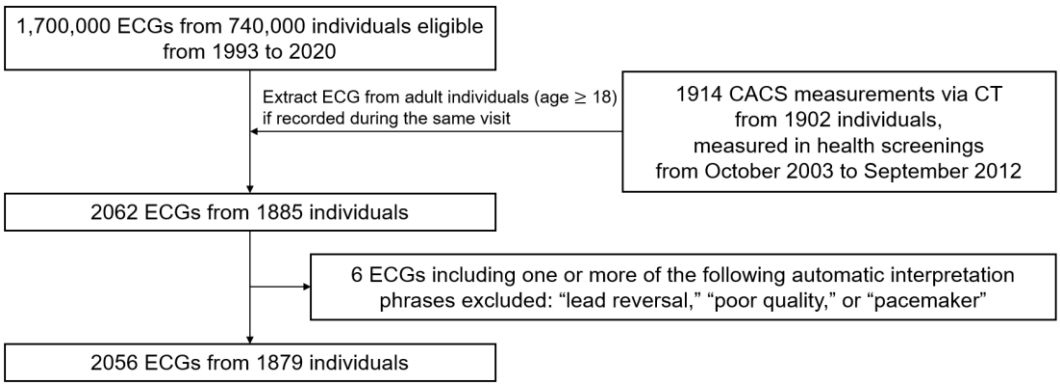


Figure 7. Data flow diagram (AUMC).

AUMC: Ajou University Medical Center; ECG: electrocardiogram; CACS: coronary artery calcium score; CT: computed tomography.

2.1.3. Data for multinational retrospective cohort analyses

We conducted multinational retrospective cohort analyses using datasets from two different countries to validate whether the AI-ECG model could predict future cardiovascular events (Figures 5 and 8). This cohort analysis is essential for verifying the AI-ECG model's potential clinical relevance and its impact on decision-making.

First, data from individuals undergoing health screenings at SH, with only ECG recordings, but without CACS from CT readings, from September 15, 2012 to September 14, 2016, were used (Figure 5). ECGs from patients included in the model development dataset, those missing clinical data, not linked with mortality data, those with automatic interpretation phrases including “lead reversal,” “poor quality,” or “pacemaker,” and those with prior cardiovascular events were excluded.

Secondly, we utilized data from the UKB, a large-scale biomedical database representing the general population of the United Kingdom[42]. Established in 2006, the UKB cohort is a significant international health resource that has gathered extensive data and biological samples from approximately half a million participants aged 40 to 69 years at the time of enrollment. In 2015, the UKB launched its imaging study, aiming to scan 20% of the original cohort. Resting 12-lead ECG data were collected during this imaging study, which we used for our analysis (Figure 8).

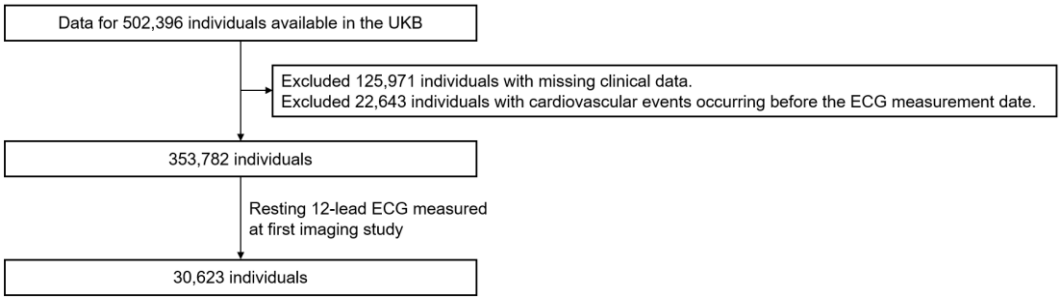


Figure 8. Data flow diagram (UKB, cohort analysis dataset).

UKB: United Kingdom Biobank; ECG: electrocardiogram

2.2. Data preprocessing

2.2.1. ECG sampling rate and lead selection

All SH, YSH, AUMC and UKB data underwent identical preprocessing methods. ECGs with a sampling rate of 250 Hz underwent upsampling to 500 Hz using linear interpolation, ensuring a uniform 500 Hz rate across all ECGs. Each waveform was standardized with z-score normalization, bringing the mean to 0 and the standard deviation to 1. According to the Einthoven law and Goldberger equation, only two of the six limb leads (leads I, II, III, aVR, aVL, aVF) are needed to calculate the other four[43]. Therefore, using any two limb leads provides the same information as all six. We thus used eight leads (leads I, II, V1-V6) from the 12 available as input.

2.2.2. Data augmentation

During the training phase, we employed a data augmentation strategy. Although ECG changes linked to coronary artery calcification are not well-studied, prior knowledge indicates that coronary heart disease-related ECG findings are more often morphological than rhythm-based[44]. Thus, we hypothesized that calcification-related ECG changes would similarly involve morphological features and reasoned that analyzing a 2.5-second segment, rather than the full 10-second ECG, would suffice. Thus, for every training epoch, we randomly chose a distinct 2.5-second segment from the 10-second ECG, introducing slight variations in the data across epochs to emulate data augmentation effectively. For the internal validation, internal testing, external validation, and retrospective cohort analyses datasets, the 10-second ECGs were segmented into four non-overlapping 2.5-second intervals, and all segments were evaluated for consistency.

2.3. AI-enabled ECG model development

We utilized the raw waveforms of the ECGs as input and adopted the 1-dimensional variant of EfficientNet-B0 for our AI-ECG model architecture (Table 2, Figure 9)[45]. We trained our EfficientNet model without leveraging any pre-trained weights. Given that our dataset exhibited class imbalance—a known factor that can adversely affect classification performance—we

implemented widely recognized techniques to counterbalance its effects[46]. Specifically, we employed both oversampling of the minority class and undersampling of the majority class. In each training epoch, we adjusted the training dataset by randomly oversampling the minority class and randomly undersampling the majority class, so that both classes were of equal size while preserving the original training dataset's total size. Hyperparameter optimization was achieved through comprehensive empirical tests and grid search, leading us to select a batch size of 512, a learning rate of 0.01, and the Adam optimizer. The choice to deploy the EfficientNet-B0 architecture arose from these hyperparameter optimization trials: Among various network scales, kernel sizes, and strides of EfficientNet explored, the default 1-dimensional version of EfficientNet-B0 demonstrated superior performance. To guard against over-fitting, we implemented early stopping during training, contingent upon observed validation loss.

Ensemble learning refers to a methodology that combines multiple individual models to achieve better generalization performance[47,48]. We built 25 individual EfficientNet-B0 models using the same training strategies. For each of these models, we generated outputs from four non-overlapping 2.5-second intervals within 10-second ECGs, resulting in a total of $25 \times 4 = 100$ outputs. This approach was applied across the internal validation, internal testing, external validation, and retrospective cohort analysis datasets. We then adopted a soft voting ensemble method, averaging these 100 outputs to obtain the final result. Consequently, the average area under the receiver operating characteristics curve (AUROC) on the test set improved from 0.706 (standard deviation 0.004) for individual EfficientNet-B0 models to 0.720 for CAC > 0 and from 0.822 (standard deviation 0.003) for individual models to 0.841 for CAC ≥ 400 after applying the soft voting ensemble method.

Table 2. Neural network architecture summary (EfficientNet-B0).

EfficientNet-B0			
Stage	**Operator**	**Output shape**	**Layers**
Input		8 × 1250	
1	Conv1d (k=3)	32 × 625	1
2	SepConv (k=3)	16 × 625	1
3	MBConv (k=3)	24 × 313	2
4	MBConv (k=5)	40 × 157	2
5	MBConv (k=3)	80 × 79	3
6	MBConv (k=5)	112 × 79	3
7	MBConv (k=5)	192 × 40	4
8	MBConv (k=3)	320 × 40	1
9	Conv1d (k=1)	1280 × 40	1
10	AvgPool	1280	1
11	Linear	2	1

Conv1d: 1-dimensional convolution; SepConv: depthwise separable convolution; MBConv: mobile inverted bottleneck convolution; AvgPool: average pooling

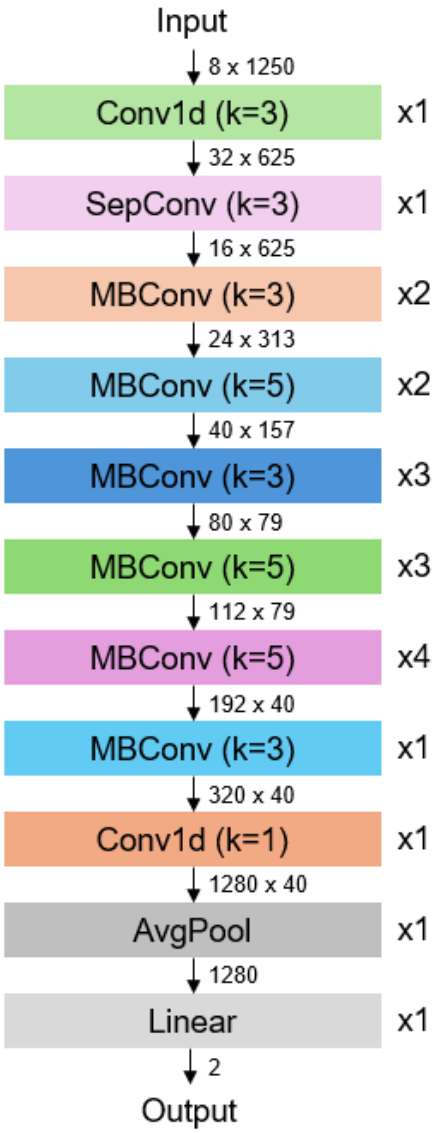


Figure 9. EfficientNet-B0 architecture.

Conv1d: 1-dimensional convolution; SepConv: depthwise separable convolution; MBConv: mobile inverted bottleneck convolution; AvgPool: average pooling

2.4. Pre-trained VAE model development

We integrated a pre-trained VAE model to provide visual morphological explainability, as described in previous studies (Table 3, Figures 10 and 11)[35,36,49]. This methodology employs a VAE architecture to learn the intrinsic factors influencing median beat ECG morphology in an unsupervised manner. The VAE comprises two primary components: the encoder, which translates the input ECG data into a condensed latent space, termed ECG factors, and the decoder, which interprets points from this latent space (ECG factors) to approximate the initial data space, aiming to reconstruct the original input data as closely as possible[35,36,49]. The VAE's training objective encompasses two loss metrics. The first, known as the reconstruction loss, measures how well the decoded data matches the original data. The second, the Kullback-Leibler Divergence loss, quantifies the deviation of the encoded distribution (ECG factors) from a predetermined distribution, typically a standard Gaussian. The aggregate loss constitutes a balanced summation of these metrics with an appropriate ratio. By decoding the ECG factors and delineating their impact on median beat ECG morphology, individual ECG factor interpretability becomes feasible. The unsupervised training nature of VAEs allows for capitalizing on expansive datasets and provides an automated method to unveil inherent data structures efficiently. In essence, the VAE model efficiently compresses any ECG to a set number of descriptive, independent factors and can also reproduce or create ECGs using these factors.

We pre-trained the VAE model using the entire set of 5.6 million median waveforms from the standard 12-lead ECG database of SH. We randomly divided this dataset in a 9:1 ratio (while ensuring no individual overlap) to create the training and validation sets. We explored the essential hyperparameters outlined by previous studies[35,36]: the summation ratio (β) between the two loss components and the number of ECG factors. In our current experiment settings, we found that 48 ECG factors and a β value of 16 yield the most optimal VAE model during factor traversal assessments. Consequently, we adopted the model trained with these hyperparameters. To guard against over-fitting, we implemented early stopping during training, contingent upon observed validation loss. We found that 29 ECG factors were significant, and the factor traversals for these are reported in Figures 12-19.

Table 3. Neural network architecture summary (VAE).

Pre-trained variational autoencoder			
Stage	**Operator**	**Output shape**	**Layers**
Input		8 × 600	
1	CausalConvolutionBlock	128 × 600	7
2	CausalConvolutionBlock	64 × 600	1
3	AvgPool	64	1
4: latent space	Linear, Softplus	μ: 48 σ: 48	1
5	Reparameterization	48	1
6	Linear	64	1
7	Linear	38400	1
8	Reshape	64 × 600	1
9	CausalConvolutionBlock	128 × 600	7
10	CausalConvolutionBlock	8 × 600	1
11: output	Flatten, Linear, Softplus, Reshape	μ: 8 × 600, σ: 8 × 600	1

VAE: variational autoencoder; AvgPool: average pooling

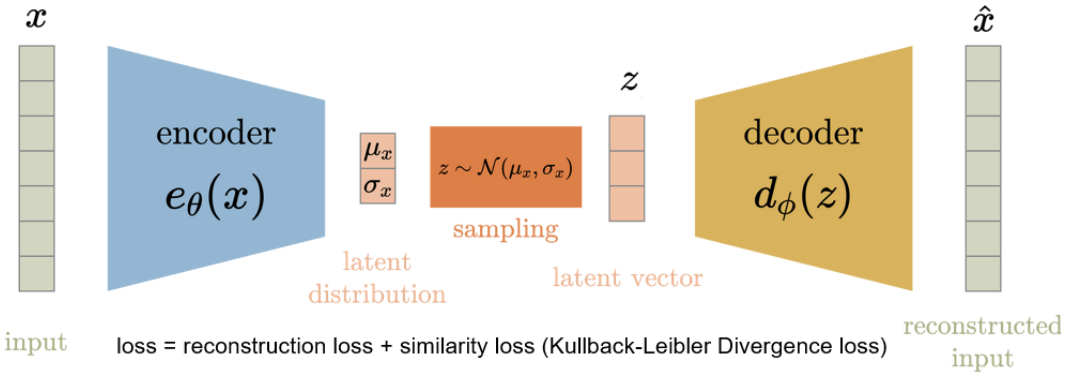


Figure 10. VAE architecture (overview).

VAE: variational autoencoder

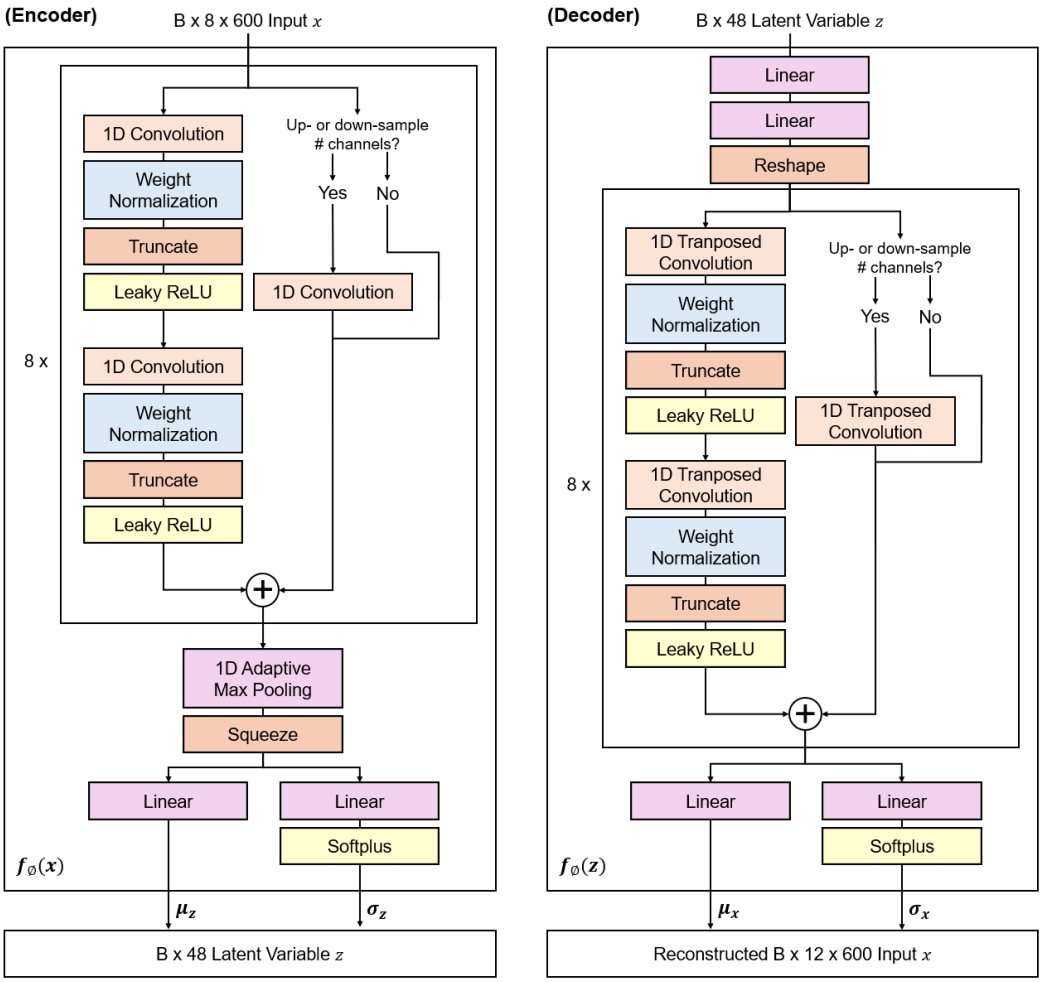


Figure 11. VAE architecture (specific).

VAE: variational autoencoder

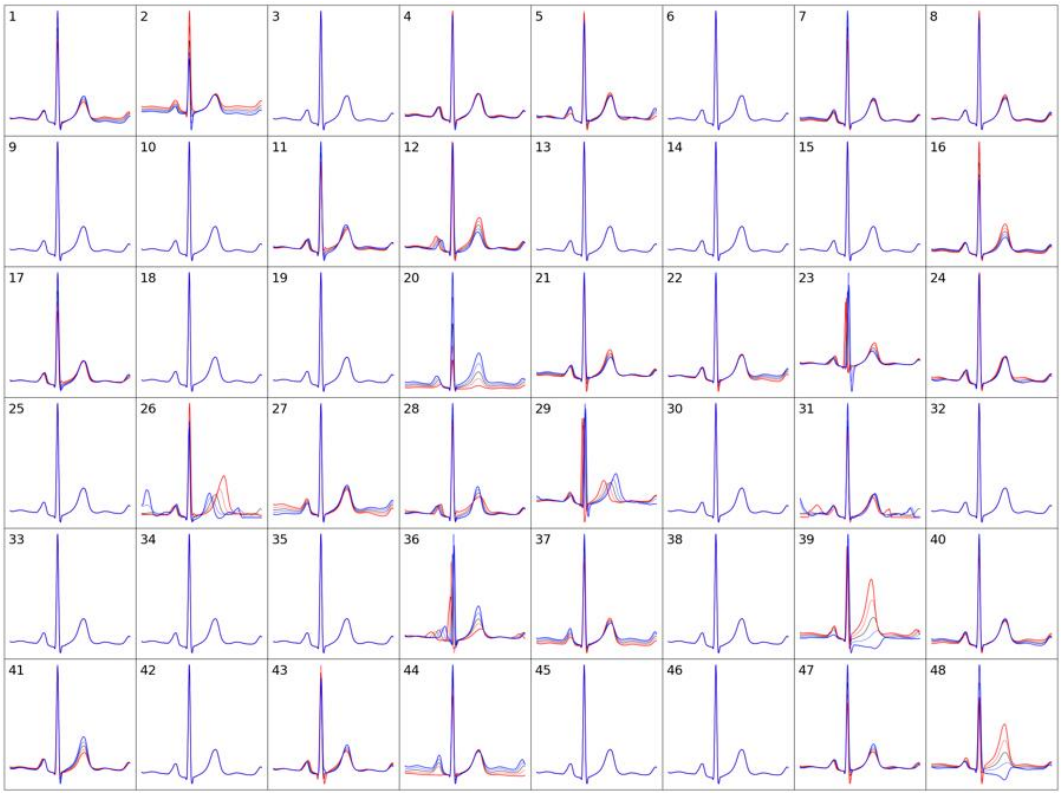


Figure 12. Factor traversals of all the ECG factors (lead I). 29 ECG factors (factor numbers 1, 2, 4, 5, 7, 8, 11, 12, 16, 17, 20, 21, 22, 23, 24, 26, 27, 28, 29, 31, 36, 37, 39, 40, 41, 43, 44, 47, 48) were significant.

ECG: electrocardiogram

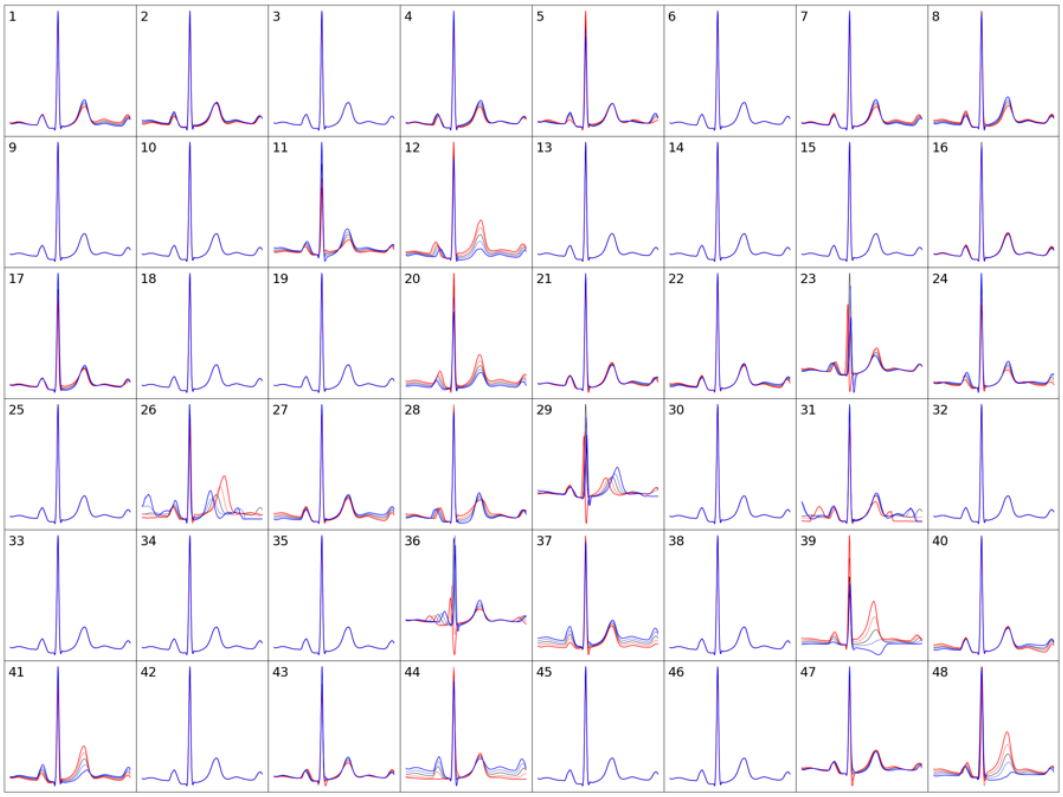


Figure 13. Factor traversals of all the ECG factors (lead II). 29 ECG factors (factor numbers 1, 2, 4, 5, 7, 8, 11, 12, 16, 17, 20, 21, 22, 23, 24, 26, 27, 28, 29, 31, 36, 37, 39, 40, 41, 43, 44, 47, 48) were significant.

ECG: electrocardiogram

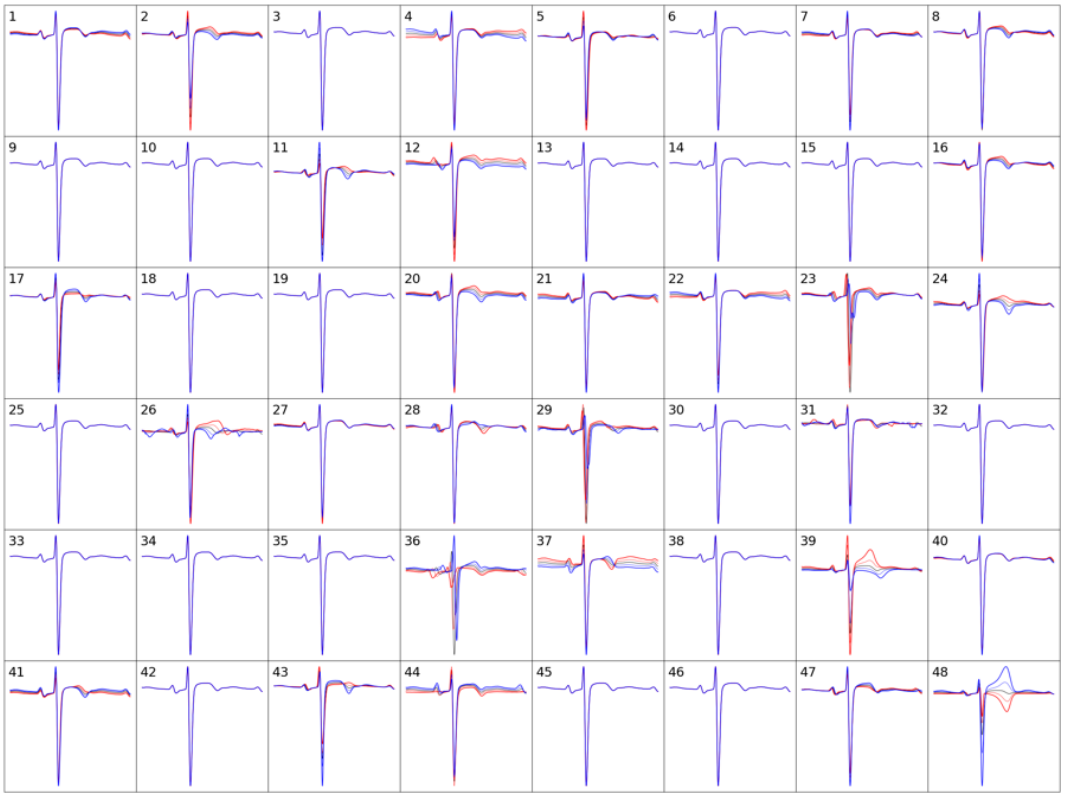


Figure 14. Factor traversals of all the ECG factors (lead V1). 29 ECG factors (factor numbers 1, 2, 4, 5, 7, 8, 11, 12, 16, 17, 20, 21, 22, 23, 24, 26, 27, 28, 29, 31, 36, 37, 39, 40, 41, 43, 44, 47, 48) were significant.

ECG: electrocardiogram

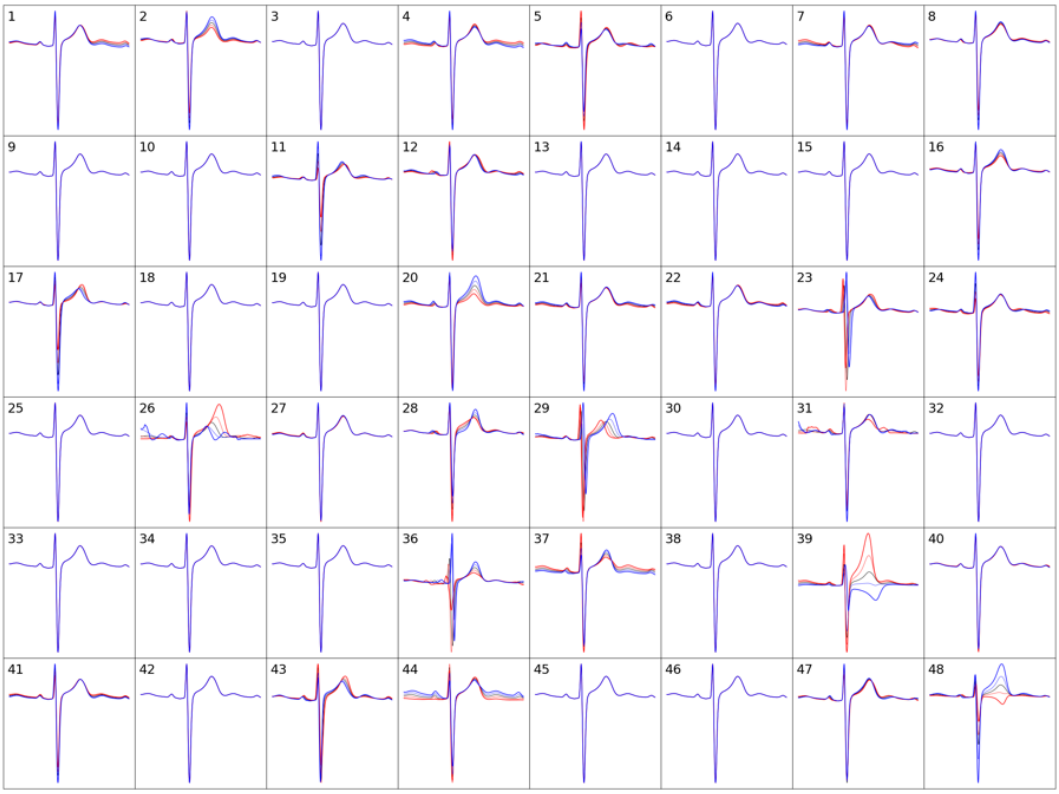


Figure 15. Factor traversals of all the ECG factors (lead V2). 29 ECG factors (factor numbers 1, 2, 4, 5, 7, 8, 11, 12, 16, 17, 20, 21, 22, 23, 24, 26, 27, 28, 29, 31, 36, 37, 39, 40, 41, 43, 44, 47, 48) were significant.

ECG: electrocardiogram

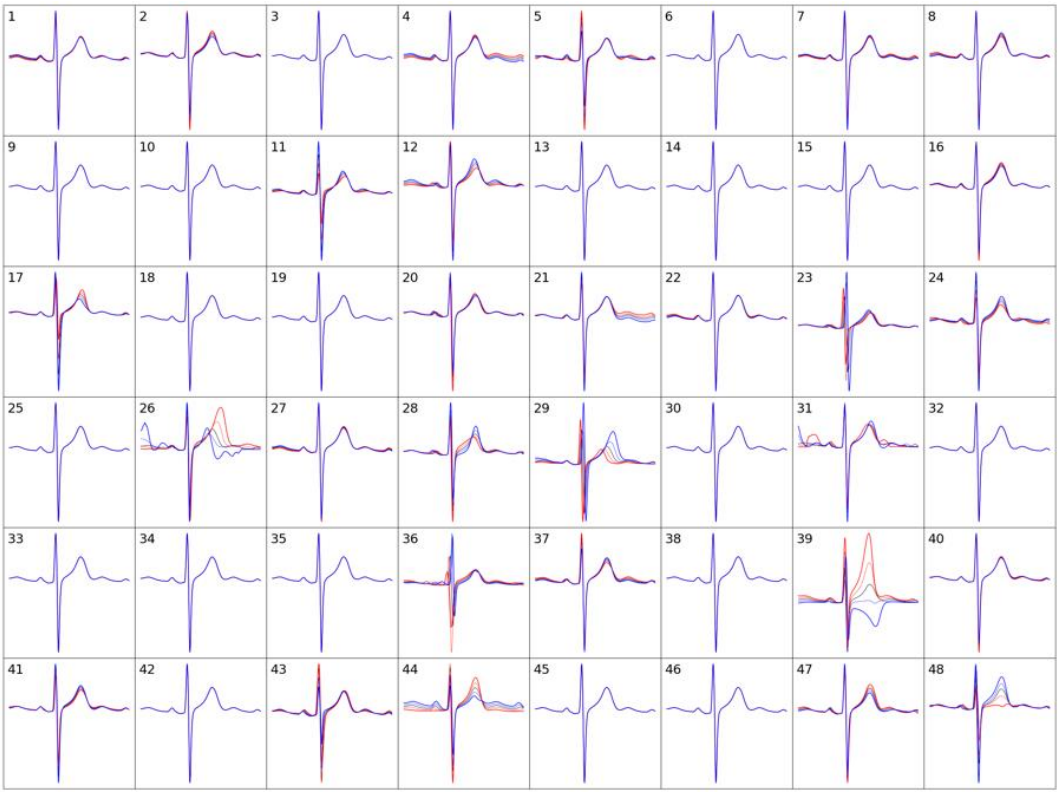


Figure 16. Factor traversals of all the ECG factors (lead V3). 29 ECG factors (factor numbers 1, 2, 4, 5, 7, 8, 11, 12, 16, 17, 20, 21, 22, 23, 24, 26, 27, 28, 29, 31, 36, 37, 39, 40, 41, 43, 44, 47, 48) were significant.

ECG: electrocardiogram

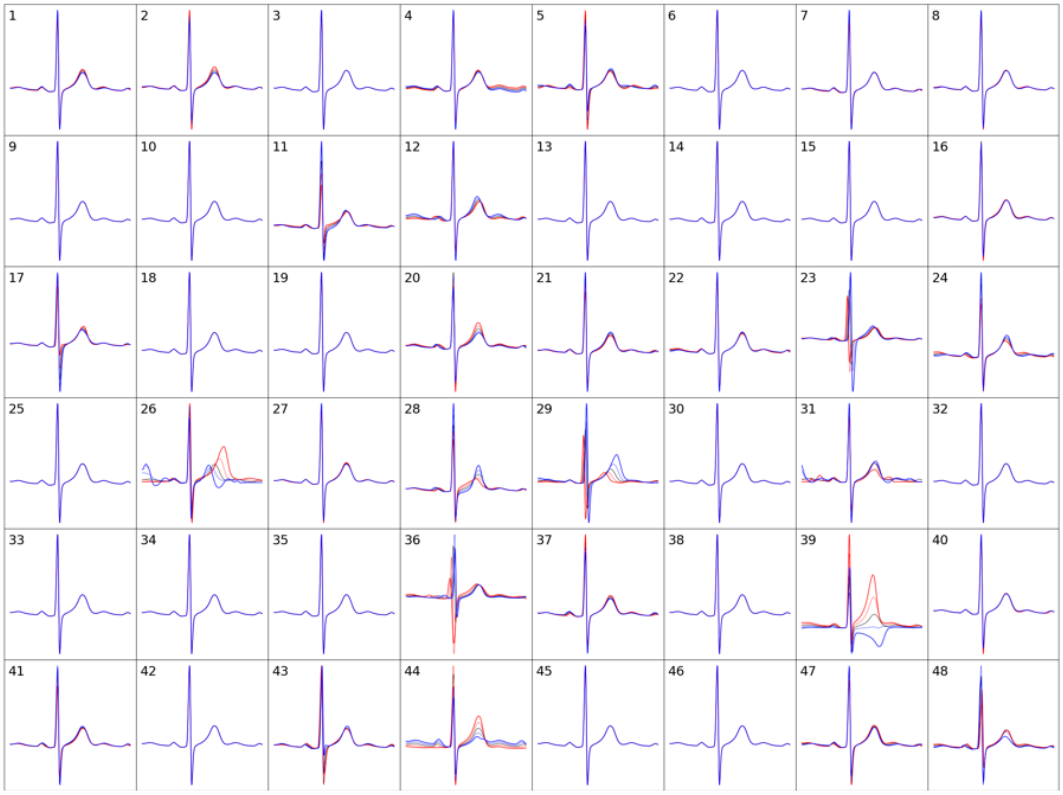


Figure 17. Factor traversals of all the ECG factors (lead V4). 29 ECG factors (factor numbers 1, 2, 4, 5, 7, 8, 11, 12, 16, 17, 20, 21, 22, 23, 24, 26, 27, 28, 29, 31, 36, 37, 39, 40, 41, 43, 44, 47, 48) were significant.

ECG: electrocardiogram

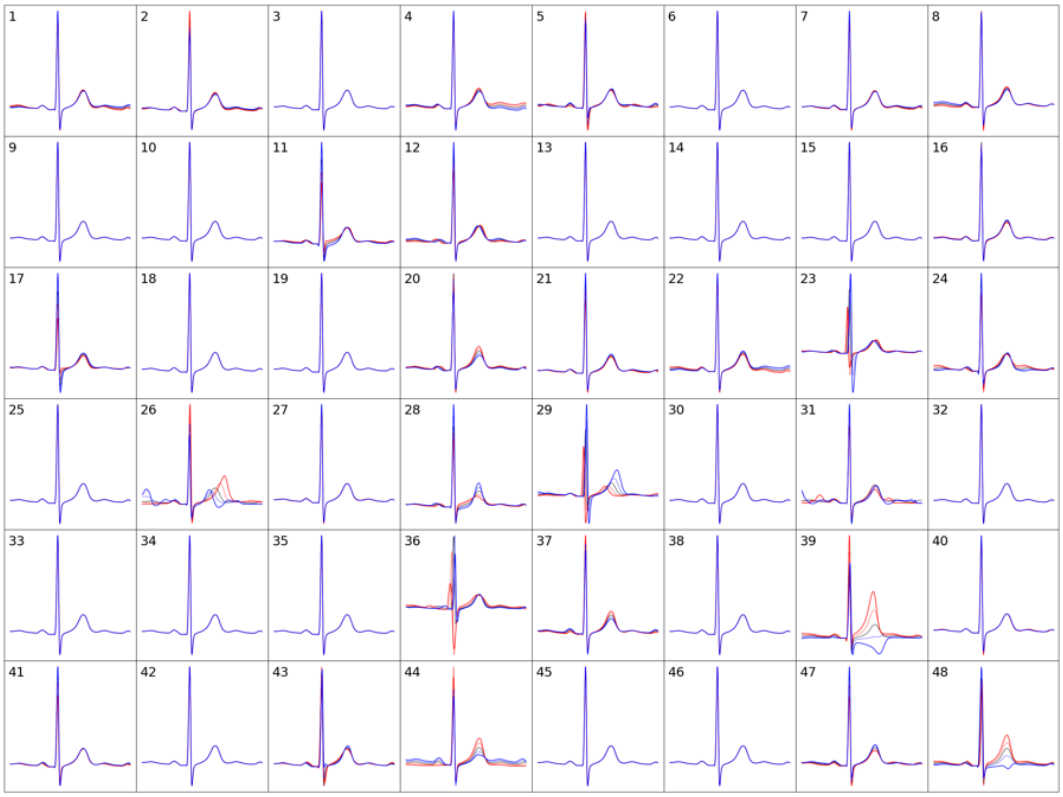


Figure 18. Factor traversals of all the ECG factors (lead V5). 29 ECG factors (factor numbers 1, 2, 4, 5, 7, 8, 11, 12, 16, 17, 20, 21, 22, 23, 24, 26, 27, 28, 29, 31, 36, 37, 39, 40, 41, 43, 44, 47, 48) were significant.

ECG: electrocardiogram

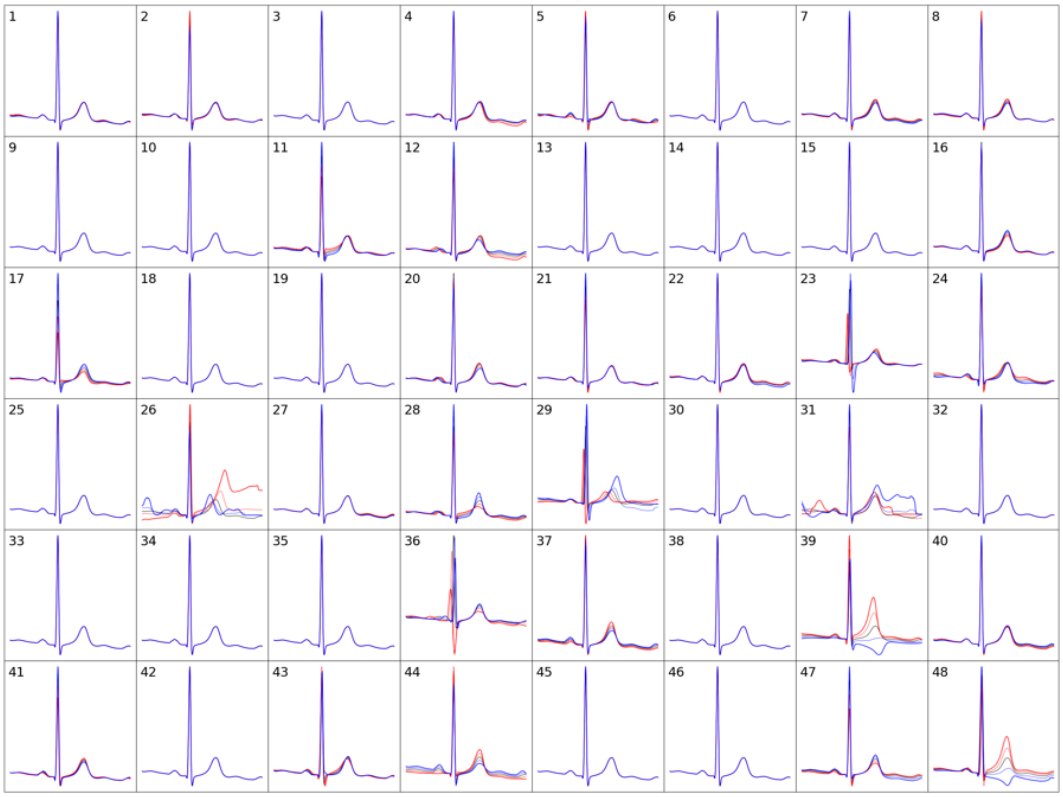


Figure 19. Factor traversals of all the ECG factors (lead V6). 29 ECG factors (factor numbers 1, 2, 4, 5, 7, 8, 11, 12, 16, 17, 20, 21, 22, 23, 24, 26, 27, 28, 29, 31, 36, 37, 39, 40, 41, 43, 44, 47, 48) were significant.

ECG: electrocardiogram

2.5. Outcomes

2.5.1. AI-ECG model training objective

We set the training objective of the EfficientNet model to predict CACS ≥ 400 as a binary classification task. This threshold was selected because a CAC score above 400 is clinically recognized as signifying a high risk of a cardiovascular event and is classified as "severe disease", providing an imperative benchmark for early intervention and risk stratification[15,16,50-52]. While this threshold guided the model's training, its utility is not limited to this specific threshold. Coronary artery calcification represents a diverse disease process, characterized by variability in histological features and degrees of atherosclerosis progression[53]. The model generates outputs on a continuous scale, inherently capturing patterns relevant to various levels of calcification across different CACS.

2.5.2. XGBoost model using VAE features

Subsequently, ECG factors from the pre-trained VAE model were employed to construct an XGBoost model to predict CACS ≥ 400. This model was trained, validated, and tested using the same corresponding datasets as the EfficientNet model. Using Shapley Additive exPlanations (SHAP) analysis[54], we determined which ECG factors had the greatest impact on the prediction. We applied the SHAP method to the test dataset. For the interpretation of the top contributing ECG factors, we utilized a method termed "factor traversals"[35,36]: By modulating the values of an individual ECG factor from -3 (represented in blue) to 3 (represented in red), advancing in increments of 1.5 units, and then using the decoder part of the VAE model to reconstruct the ECG, we were able to overlay these reconstructed ECGs on a single plot. The reconstructed ECG corresponding to an ECG factor value of 0 was shown in grey. This visualization allowed us to comprehend the variations in ECG morphology attributable to each individual ECG factor. The factor traversals of all the significant ECG factors from our pre-trained VAE model are provided in Figures 12-19.

2.6. Performance evaluation

2.6.1. Performance comparison

For performance comparison, we constructed an XGBoost model that uses traditional ECG features (ventricular rate, atrial rate, PR interval, QRS duration, QT interval, QT interval corrected, P axis, R axis, T axis) provided by a built-in software in the ECG machine (GE MUSE^{TM}) with the training objective of predicting CACS ≥ 400.

2.6.2. Performance metrics and risk categorization

We evaluated the performance of the AI-ECG model in predicting CACS ≥ 400 and also CACS > 0. We generated receiver operating characteristic (ROC) curves and precision-recall (PR) curves, subsequently evaluating the area under the ROC curve (AUROC) and the area under the PR curve (AUPRC).

We classified PCE risk scores into low, moderate, and high-risk categories using the 7.5% and 20% thresholds, in line with current cholesterol management guidelines, which recommend initiation of primary prevention with statins for those in the moderate or high-risk categories[4]. However, for borderline or intermediate classifications, where the decision to initiate treatment remains uncertain, guidelines recommend that the presence of CAC supports initiating primary prevention measures, while its absence suggests refraining from such measures[5-7].

Therefore, in the PCE low-risk category, it would be beneficial for the AI-ECG to increase the likelihood of identifying the presence of CAC. Conversely, in the PCE moderate-risk category, it would be advantageous for the AI-ECG to improve the detection of individuals without CAC. Accordingly, in the validation set of the model development dataset, the threshold at which the positive predictive value (PPV) for a CACS > 0 is 0.800 was defined as AI-ECG high risk. Similarly, the threshold at which the negative predictive value (NPV) for a CACS > 0 is 0.800 was defined as AI-ECG low risk. Individuals falling between these thresholds were categorized as AI-ECG moderate risk. At the determined thresholds, we calculated performance metrics for CAC ≥ 400 and > 0, including accuracy, sensitivity, specificity, PPV, NPV, and the F1 score.

2.7. Multinational retrospective cohort analyses

We conducted multinational retrospective cohort analyses using datasets from two different countries. First, we used SH health screening data. Diagnostic records (in International Classification of Disease, 10th Revision [ICD-10] codes) from individuals included in the SH health screening data for the cohort analysis were extracted from their EMR database. In South Korea, a government organization called Statistics Korea (KOSTAT) offers a service that links researchers' data with mortality data, based on resident registration numbers, following specific ethical and application procedures. Data from SH health screenings for the cohort analysis was linked with KOSTAT mortality data. The KOSTAT mortality data includes information on whether the individual is deceased, the date of death, and the cause of death (in ICD-10 codes).

Survival analyses were conducted to assess the occurrence of major adverse cardiovascular event (MACE). MACE was defined as an aggregate of fatal or non-fatal myocardial infarction (I21 - I25), ischemic stroke (I63 and I64), or cardiovascular death[55], with cardiovascular death also defined using the same ICD-10 codes for fatal or non-fatal myocardial infarction and ischemic stroke in the mortality data. Individuals who experienced a MACE before the health screening date or within 90 days after the health screening date were excluded from the analysis. The reason for setting a 90-day washout period after the health screening date was to exclude cases where pre-existing cardiovascular disease was detected through additional tests following abnormal findings during the health screening. The survival analysis observation period began after the 90-day washout. For the remaining individuals, data were censored at the date of the first MACE, 10 years from the health screening date, or August 10, 2024, whichever came first.

Kaplan-Meier curves for MACE were plotted to compare the AI-ECG risk groups. The net reclassification improvement (NRI) was evaluated to assess up-risking or down-risking of individuals within the PCE low- or moderate-risk categories based on the AI-ECG-derived risk categories[56].

We assessed whether the AI-ECG model serves as an independent risk factor for MACE: Various risk factors measured during health screenings that are included in the PCE — including age, sex, diabetes mellitus, hypertension, smoking status, total cholesterol, high-density lipoprotein (HDL) cholesterol and systolic blood pressure — were extracted from the EMR database and used along with the AI-ECG output as independent variables, and MACE as the dependent variable in a Cox proportional hazards regression to evaluate the association between AI-ECG output and MACE, with appropriate adjustments.

We evaluated whether the AI-ECG model adds predictive value beyond the PCE for MACE: The predictive performance of the PCE, measured by Harrell’s concordance index (C-index), was

compared with that of the combined PCE plus AI-ECG score. The PCE plus AI-ECG score for each individual was calculated as follows: “PCE score + AI-ECG score * 20”.

We also conducted a retrospective cohort analysis using the UKB data. Clinical and outcome variables corresponding to those extracted from the SH dataset were also obtained from the UKB, with the codes used for extraction detailed in Table 4. Similarly, the AI-ECG output was derived from the UKB data, and the same analyses conducted for the SH dataset were performed on the UKB dataset. Individuals who experienced a MACE before the ECG measurement date were excluded from the analysis. Data were censored at the date of MACE, 6 years from the ECG measurement date, or November 2021 (final point where participants’ diagnostic codes were followed up), whichever came first.

We also conducted subgroup analyses within the cohort analysis datasets. Within subgroups defined by sex, age (under 60 years and 60 years or older), and PCE risk category (low-risk group and moderate or higher-risk group), Cox proportional hazards regression was performed to evaluate the hazard ratio of the AI-ECG model's output, adjusting for the variables included in the PCE.

Table 4. Variables extracted from the UKB

Description	Category	Coding
Age at recruitment	Baseline characteristics	Field: 21022
Sex	Baseline characteristics	Field: 31
Diabetes diagnosed by doctor	Medical conditions	Field: 2443
Medication for cholesterol, blood pressure or diabetes	Medication	Field: 6177
Medication for cholesterol, blood pressure or diabetes, or exogenous hormones	Medication	Field: 6153
Smoking status	Smoking	Field: 20116
Systolic blood pressure	Blood pressure	Field: 4080
Cholesterol	Blood chemistry	Field: 30690
HDL cholesterol	Blood chemistry	Field: 30760
Ethnic background	Ethnicity	Field: 21000
ECG datasets	ECG at rest, 12 lead	Field: 20205
Acute myocardial infarction ICD-10: I21	First occurrence	Field: 131298
Subsequent myocardial infarction ICD-10: I22	First occurrence	Field: 131300
Certain current complications following acute myocardial infarction ICD-10: I23	First occurrence	Field: 131302
Other acute ischemic heart disease ICD-10: I24	First occurrence	Field: 131304
Chronic ischemic heart disease ICD-10: I25	First occurrence	Field: 131306
Cerebral infarction ICD-10: I63	First occurrence	Field: 131366
Stroke, not specified as hemorrhage or infarction ICD-10: I64	First occurrence	Field: 131368
Date of death	Death register	Field: 40000
Underlying (primary) cause of death: ICD-10	Death register	Field: 40001

UKB: United Kingdom Biobank; HDL: high density lipoprotein; ECG: electrocardiogram, ICD-10: International Classification of Disease, 10th revision

2.8. Statistical analysis

We compared the characteristics between the datasets. We evaluated the normality of continuous data using the Shapiro-Wilk test. Normally distributed continuous variables were compared using the independent samples t-test, while non-normally distributed variables were compared using the Mann–Whitney U test for two-group comparisons. For comparisons among three or more groups, analysis of variance (ANOVA) and Kruskal–Wallis tests were used, respectively. Categorical data were analyzed using the chi-square test, while comparisons of AUROCs utilized the Delong test[57].

For Harrell's C-index and the NRI, the 95% confidence intervals (CIs) were determined based on 2,000 bootstrapping (resampling with replacement) runs, with the 2.5th and 97.5th percentile borders reported[58]. The Kaplan–Meier method was used to plot survival curves for the low-, moderate-, and high-risk groups based on the risk scoring methods. The pairwise log-rank test with post-hoc Bonferroni correction was used to compare the survival functions across these risk groups statistically. The p-values for the incidence rate differences and hazard ratio differences were calculated using the z-test. Statistical significance was set at $P < 0.05$ for all tests.

2.9. Software

Neural network models were developed in Python (version 3.8.5) utilizing the "PyTorch" framework (version 1.11.0). SHAP analysis was conducted using the "shap" library (version 0.43.0) in Python. For model evaluation and further statistical analyses, we used the "Scikit-learn" library (version 0.23.2) in Python. The Delong test was conducted using the "pROC" library (version 1.18.4) in R (version 4.2.0). The Cox proportional hazards regression was conducted using the "survival" library (version 3.2.7) in R.

2.10. Ethics approval

The Institutional Review Boards (IRB) of SH, YSH and AUMC approved this study and waived the requirement for informed consent because only anonymized data were used retrospectively (IRB no. 4-2022-1299 and 4-2022-1506 [SH], 9-2024-0032 [YSH], AJOUIRB-DB-

2024-207 [AUMC]). The UKB was approved by the North West Multi-centre Research Ethics Committee as a Research Tissue Bank (RTB) approval (approval number: 21/NW/0157)[42]. This approval means that researchers do not require separate ethical clearance and can operate under RTB approval. All participants provided informed consent for participation.

3. RESULTS

3.1. Dataset characteristics

3.1.1. Dataset sizes

The model development dataset from SH included 194,964 ECGs from 57,019 individuals and the health screening test dataset from SH included 14,242 ECGs from 12,924 individuals (Figures 2-4). The external validation dataset from YSH included 729 ECGs from 710 individuals and the external validation dataset from AUMC included 2056 ECGs from 1879 individuals (Figures 6 and 7). The cohort analysis dataset from SH included 52,400 ECGs from 37,757 individuals (Figure 5). The cohort analysis dataset from UKB included 30,623 ECGs from 30,623 individuals (Figure 8).

3.1.2. Dataset for model development, testing and external validations

Table 5 shows the characteristics of the datasets used in the AI-ECG model development, testing and external validations. The test dataset and the external validation datasets, having been extracted from health screening data, represented a healthier spectrum of individuals, characterized by younger ages (mean [standard deviation, SD]: 53.1 [10.0] years, 57.1 [10.7] years and 51.1 [8.2] years vs. 61.8 [13.3] years, $P < 0.001$), lower CACS (mean [SD]: 59.0 [212.2] agatston units [AU], 81.5 [274.8] AU and 36.4 [152.3] AU vs. 295.1 [920.9] AU, $P < 0.001$), and lower proportions of CACS $\geq$ 400 (3.8 %, 6.0 % and 2.5% vs. 17.8 %, $P < 0.001$) and CACS > 0 (37.9 %, 46.8 % and 29.5 % vs. 62.0 %, $P < 0.001$), relative to the model development dataset.

Table 5. Dataset characteristics (model development, testing and external validation datasets).

	Model development dataset (SH) N = 194,963	Health screening test dataset (SH) N = 14,242	Health screening external validation dataset (YSH) N = 729	Health screening external validation dataset (AUMC) N = 2056	P-Value
Number of patients	57,019	12,926	710	1879	
Sex, male	108,828 (55.8%)	8,502 (59.7%)	427 (58.6%)	1540 (74.9%)	<0.001
Age	61.8 ± 13.3	53.1 ± 10.0	57.1 ± 10.7	51.1 ± 8.2	<0.001
CACS	295.1 ± 920.9	59.0 ± 212.2	81.5 ± 274.8	36.4 ± 152.3	<0.001
CACS > 0	120,837 (62.0%)	5394 (37.9%)	341 (46.8%)	606 (29.5%)	<0.001
CACS ≥ 400	34,637 (17.8%)	542 (3.8%)	44 (6.0%)	52 (2.5%)	<0.001

SH: Severance Hospital; YSH: Yongin Severance Hospital; AUMC: Ajou University Medical Center; CACS: coronary artery calcium score.

3.1.3. Dataset for multinational retrospective cohort analyses

Tables 6 and 7 show the characteristics of the datasets used in the retrospective cohort analyses. Individuals who experienced a MACE within 10 years in the SH cohort analysis dataset (Table 6, N = 1611) were older (median [interquartile range]: 59 [52 – 65] years vs. 47 [38 – 55] years) and had a higher proportion of males (74.1% vs. 56.0%) compared to those without events (N = 50,789). They also had more comorbidities, including higher rates of diabetes mellitus (13.3% vs. 3.9%) and hypertension (32.0% vs. 12.0%). Their median systolic blood pressure was higher (median [IQR]: 125 [116–134] mmHg vs. 119 [109–129] mmHg), and they had elevated AI-ECG scores (median [IQR]: 0.283 [0.149–0.432] vs. 0.105 [0.038–0.235]) and PCE scores (median [IQR]: 8 [4–14.8] vs. 1.9 [0.5–2.5]) compared to individuals without events.

Individuals who experienced a MACE within 6 years in the UKB cohort analysis dataset (Table 7, N = 699) were older (median [IQR]: 61 [55–64] years vs. 55 [49–61] years) compared to those without events (N = 29,924). A higher proportion of males was observed (71.0% vs. 46.7%), and they had more comorbidities, including higher rates of diabetes mellitus (6.7% vs. 2.4%) and hypertension (24.5% vs. 12.1%). The proportion of current smokers was also higher (8.6% vs. 6.2%). They had elevated systolic blood pressure (median [IQR]: 140.5 [129–154.5] mmHg vs. 133 [122.5–145.5] mmHg), lower HDL cholesterol (median [IQR]: 49.7 [42.4–58.2] mg/dL vs. 55.7 [46.8–66.4] mg/dL), higher AI-ECG scores (median [IQR]: 0.359 [0.233–0.479] vs. 0.202 [0.105–0.328]), and higher PCE scores (median [IQR]: 10.2 [5.6–15.5] vs. 4.3 [1.8–9.1]).

Table 6. Dataset characteristics (SH cohort analysis dataset).

	Positive for MACE N = 1611	Negative for MACE N = 50,789	P-value
Age, years [IQR]	59 [52 – 65]	47 [38 – 55]	<0.001
Sex (male) (%)	1193 (74.1%)	28,423 (56.0%)	<0.001
DM (%)	215 (13.3%)	1958 (3.9%)	<0.001
HTN (%)	516 (32.0%)	6071 (12.0%)	<0.001
Current smoker (%)	319 (19.8%)	9273 (18.3%)	0.122
SBP, mmHg [IQR]	125 [116 – 134]	119 [109 – 129]	<0.001
Total cholesterol, mg/dL [IQR]	192 [169 – 215]	192 [170 – 215]	0.550
HDL cholesterol, mg/dL [IQR]	48 [42 – 57]	52 [44 – 62]	<0.001
AI-ECG score [IQR]	0.283 [0.149 – 0.432]	0.105 [0.038 – 0.235]	<0.001
PCE score [IQR]	8 [4 – 14.8]	1.9 [0.5 – 2.5]	<0.001

SH: Severance Hospital; MACE: major adverse cardiovascular event; IQR: interquartile range; DM: diabetes mellitus; HTN: hypertension; SBP: systolic blood pressure; HDL: high-density lipoprotein; AI: artificial intelligence; ECG: electrocardiogram; PCE: Pooled Cohort Equations

Table 7. Dataset characteristics (UKB cohort analysis dataset).

	Positive for MACE N = 699	Negative for MACE N = 29,924	P-value
Age, years [IQR]	61 [55 – 64]	55 [49 – 61]	<0.001
Sex (male) (%)	496 (71.0%)	13,977 (46.7%)	<0.001
DM (%)	47 (6.7%)	724 (2.4%)	<0.001
HTN (%)	171 (24.5%)	3618 (12.1%)	<0.001
Current smoker (%)	60 (8.6%)	1863 (6.2%)	0.014
SBP, mmHg [IQR]	140.5 [129 – 154.5]	133 [122.5 – 145.5]	<0.001
Total cholesterol, mg/dL [IQR]	218.3 [190.2 – 245.8]	220.4 [194 – 248.5]	0.100
HDL cholesterol, mg/dL [IQR]	49.7 [42.4 – 58.2]	55.7 [46.8 – 66.4]	<0.001
AI-ECG score [IQR]	0.359 [0.233 – 0.479]	0.202 [0.105 – 0.328]	<0.001
PCE score [IQR]	10.2 [5.6 – 15.5]	4.3 [1.8 – 9.1]	<0.001

UKB: United Kingdom Biobank; MACE: major adverse cardiovascular event; IQR: interquartile range; DM: diabetes mellitus; HTN: hypertension; SBP: systolic blood pressure; HDL: high-density lipoprotein; AI: artificial intelligence; ECG: electrocardiogram; PCE: Pooled Cohort Equations

3.2. Model performance

3.2.1. Model performance in the health screening test dataset

Figures 20 and 21 display the ROC curves, and Figures 22 and 23 display the PR curves of the models. Our AI-ECG model exhibited AUROCs of 0.841 and 0.720 in predicting CACS ≥ 400 and > 0, respectively, in the test dataset. The respective AUPRCs were 0.289 and 0.603. Tables 8 and 9 show the performance metrics at the thresholds used to define the AI-ECG risk categories. AI-ECG model output was correlated with CACS (Figure 24), demonstrating that the model is not confined to a single threshold but instead encodes information applicable to diverse CACS levels.

Figure 25 illustrates the AI-ECG-derived reclassification within each PCE risk category in the health screening test dataset. Among the 10,719 individuals, 7519, 2652, and 548 were categorized as low, moderate, and high risk, respectively, based on the PCE. Among individuals categorized as low risk by the PCE, 45.7% of those classified as high risk by the AI-ECG were found to have CAC, which is higher than the overall proportion of CAC > 0 (24.9%) in the PCE low-risk category. Among individuals categorized as moderate risk by the PCE, 60.9% of those classified as low risk by the AI-ECG were CAC-free, which is higher than the overall proportion of CAC = 0 (35.4%) in the PCE moderate-risk category.

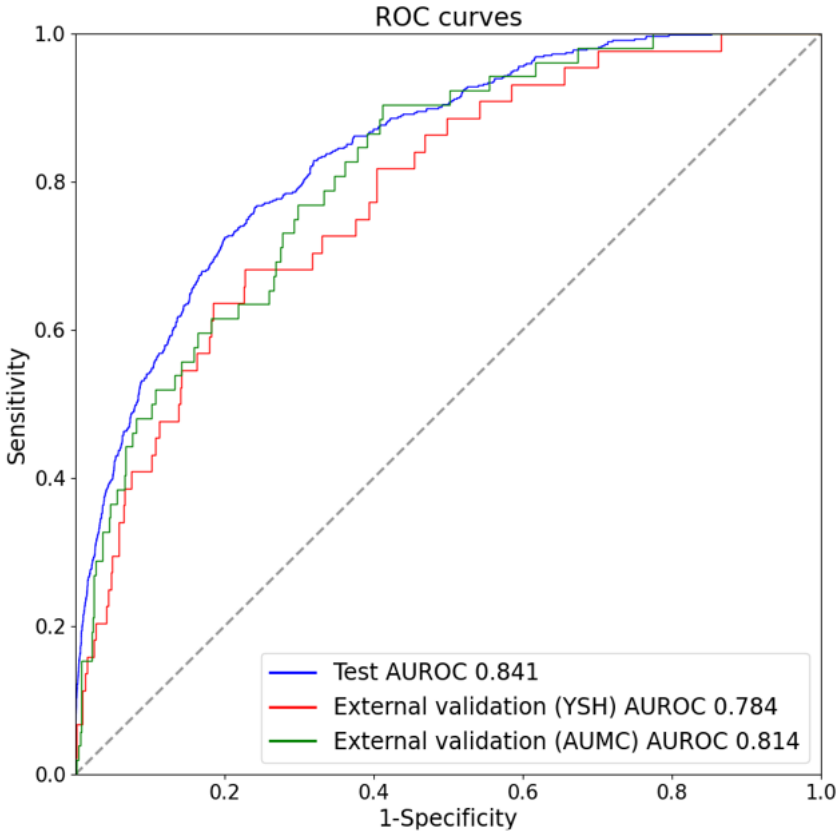


Figure 20. ROC curves of the AI-ECG model (CACS ≥ 400).

ROC: receiver operating characteristics; AI: artificial intelligence; ECG: electrocardiogram; CACS: coronary artery calcium score; AUROC: area under the receiver operating characteristics curve; YSH: Yongin Severance Hospital; AUMC: Ajou University Medical Center

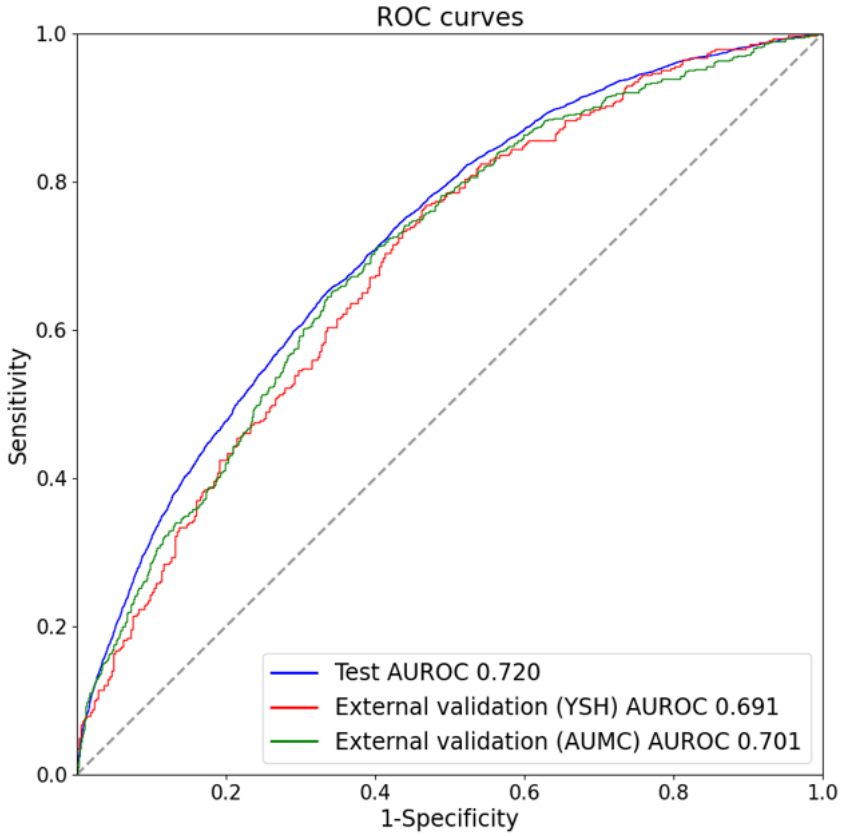


Figure 21. ROC curves of the AI-ECG model (CACS > 0).

ROC: receiver operating characteristics; AI: artificial intelligence; ECG: electrocardiogram; CACS: coronary artery calcium score; AUROC: area under the receiver operating characteristics curve; YSH: Yongin Severance Hospital; AUMC: Ajou University Medical Center

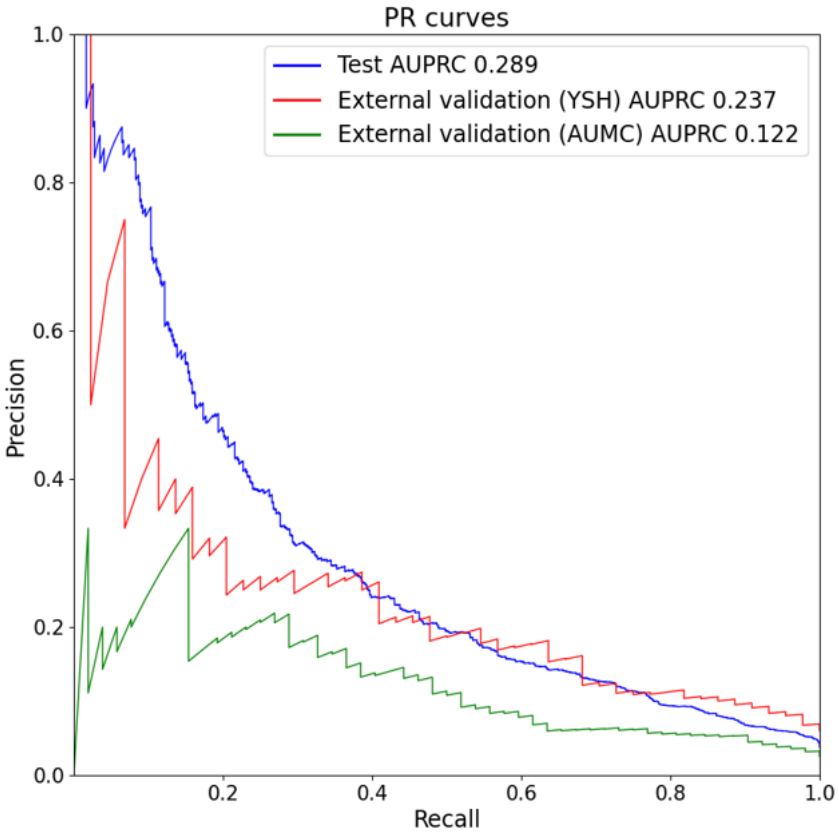


Figure 22. PR curves of the AI-ECG model (CACS ≥ 400).

PR: precision-recall; AI: artificial intelligence; ECG: electrocardiogram; CACS: coronary artery calcium score; AUPRC: area under the precision-recall curve; YSH: Yongin Severance Hospital; AUMC: Ajou University Medical Center

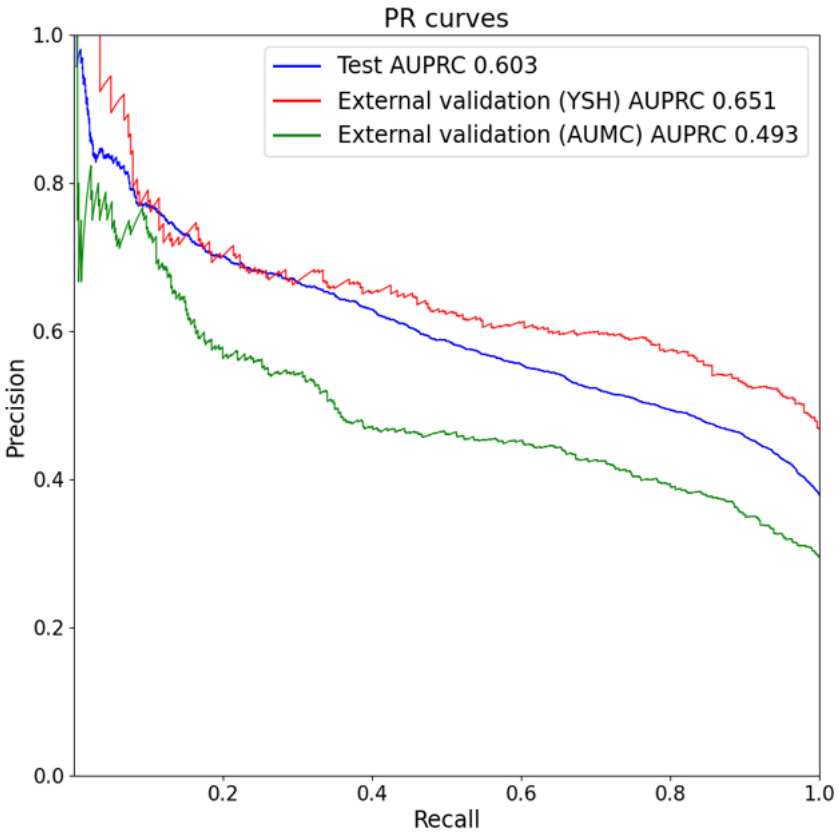


Figure 23. PR curves of the AI-ECG model (CACS > 0).

PR: precision-recall; AI: artificial intelligence; ECG: electrocardiogram; CACS: coronary artery calcium score; AUPRC: area under the precision-recall curve; YSH: Yongin Severance Hospital; AUMC: Ajou University Medical Center

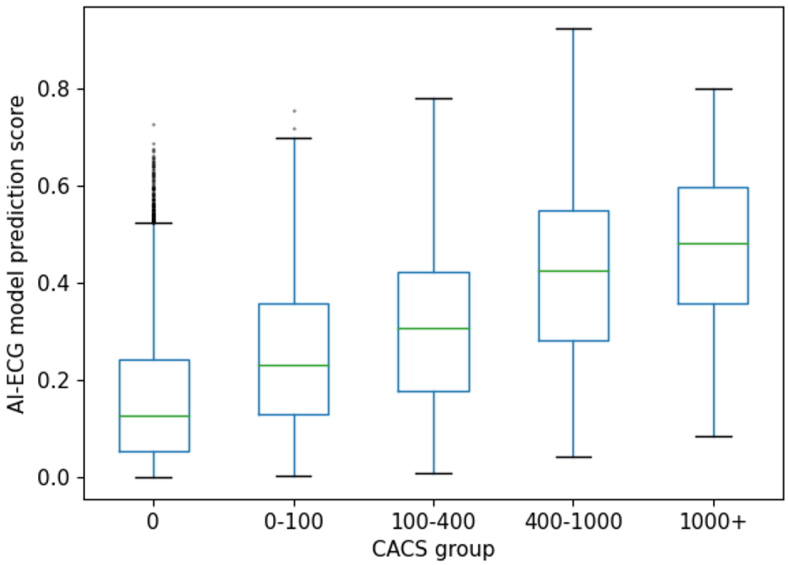


Figure 24. Boxplot of AI-ECG score by CACS group. AI-ECG model output was correlated with CACS. The Pearson correlation coefficient between the AI-ECG model outputs and CACS was 0.318 ($P < 0.001$).

AI: artificial intelligence; ECG: electrocardiogram; AI-ECG: artificial intelligence-enabled electrocardiogram; CACS: coronary artery calcium score.

Table 8. AI-ECG performance for predicting CACS ≥ 400. The table presents performance metrics at the thresholds used to define the AI-ECG risk categories in the health screening test dataset (SH) and two external health screening validation datasets (YSH and AUMC).

	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 score
High risk vs. low/moderate risk category threshold for AI-ECG						
SH	0.903	0.498	0.920	0.196	0.979	0.281
YSH	0.831	0.545	0.850	0.189	0.967	0.281
AUMC	0.875	0.519	0.884	0.104	0.986	0.174
High/moderate risk vs. low risk category threshold for AI-ECG						
SH	0.267	0.993	0.238	0.049	0.999	0.093
YSH	0.211	0.977	0.162	0.070	0.991	0.130
AUMC	0.183	1.000	0.162	0.030	1.000	0.058

AI: artificial intelligence, ECG: electrocardiogram, CACS: coronary artery calcium score; SH: Severance Hospital; YSH; Yongin Severance Hospital; AUMC: Ajou University Medical Center; PPV: positive predictive value; NPV: negative predictive value.

Table 9. AI-ECG performance for predicting CACS > 0. The table presents performance metrics at the thresholds used to define the AI-ECG risk categories in the health screening test dataset (SH) and two external health screening validation datasets (YSH and AUMC).

	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 score
High risk vs. low/moderate risk category threshold for AI-ECG						
SH	0.661	0.180	0.954	0.707	0.656	0.287
YSH	0.594	0.252	0.894	0.677	0.576	0.368
AUMC	0.721	0.241	0.922	0.564	0744	0.338
High/moderate risk vs. low risk category threshold for AI-ECG						
SH	0.544	0.915	0.318	0.450	0.861	0.603
YSH	0.569	0.944	0.240	0.522	0.830	0.672
AUMC	0.416	0.939	0.198	0.329	0.886	0.487

AI: artificial intelligence, ECG: electrocardiogram, CACS: coronary artery calcium score; SH: Severance Hospital; YSH; Yongin Severance Hospital; AUMC: Ajou University Medical Center; PPV: positive predictive value; NPV: negative predictive value.

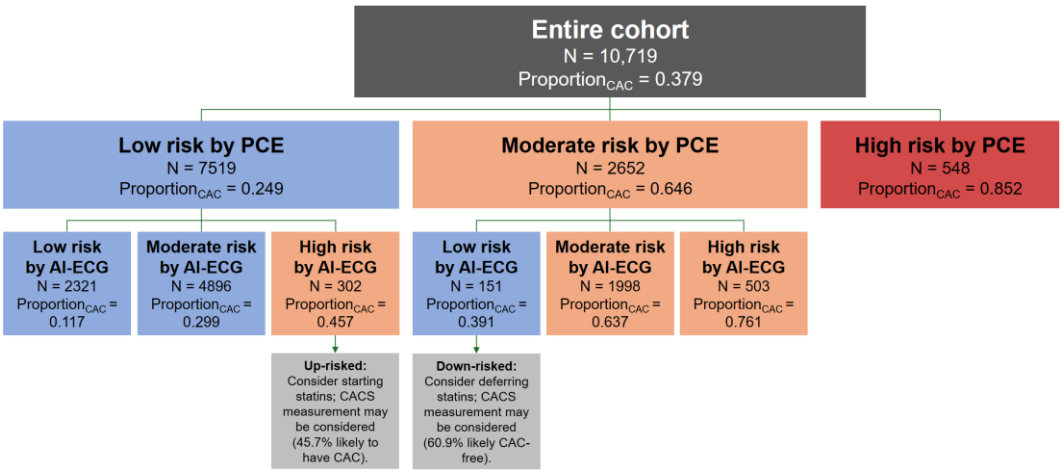


Figure 25. AI-ECG-derived reclassification within each PCE risk category (SH, health screening test dataset). "Proportion$_{CAC}$" refers to the proportion of individuals with CAC > 0.

AI: artificial intelligence; ECG: electrocardiogram; AI-ECG: artificial intelligence-enabled electrocardiogram; PCE: Pooled Cohort Equations; SH: Severance Hospital; CAC: coronary artery calcification; CACS: coronary artery calcium score.

3.2.2. Model performance in the external validations

In the external validations, the AI-ECG model maintained its efficacy, achieving AUROCs of 0.784 and 0.814, and AUPRCs of 0.237 and 0.122 in the YSH and AUMC datasets, respectively, for predicting CACS ≥ 400 (Figures 20 and 22), and AUROCs of 0.691 and 0.701, and AUPRCs of 0.651 and 0.493 in the YSH and AUMC datasets, respectively, for predicting CACS > 0 (Figures 21 and 23). Comparing the AUROC between the test dataset and the external validation datasets showed no significant difference (DeLong test [unpaired, two-sided], SH vs. YSH P = 0.112, SH vs. AUMC P = 0.350 for CACS ≥ 400, and SH vs. YSH P = 0.151, SH vs. AUMC P = 0.152 for CACS > 0). This underscores the model's generalizability to external environments.

3.2.3. Performance comparison

The XGBoost model, constructed with traditional ECG features for performance comparison, exhibited AUROCs of 0.668 and 0.600 in predicting CACS ≥ 400 and > 0, respectively, in the test dataset (Figures 26-29). The AI-ECG model outperformed the XGBoost model using traditional ECG features in all comparisons (Delong test, P < 0.001).

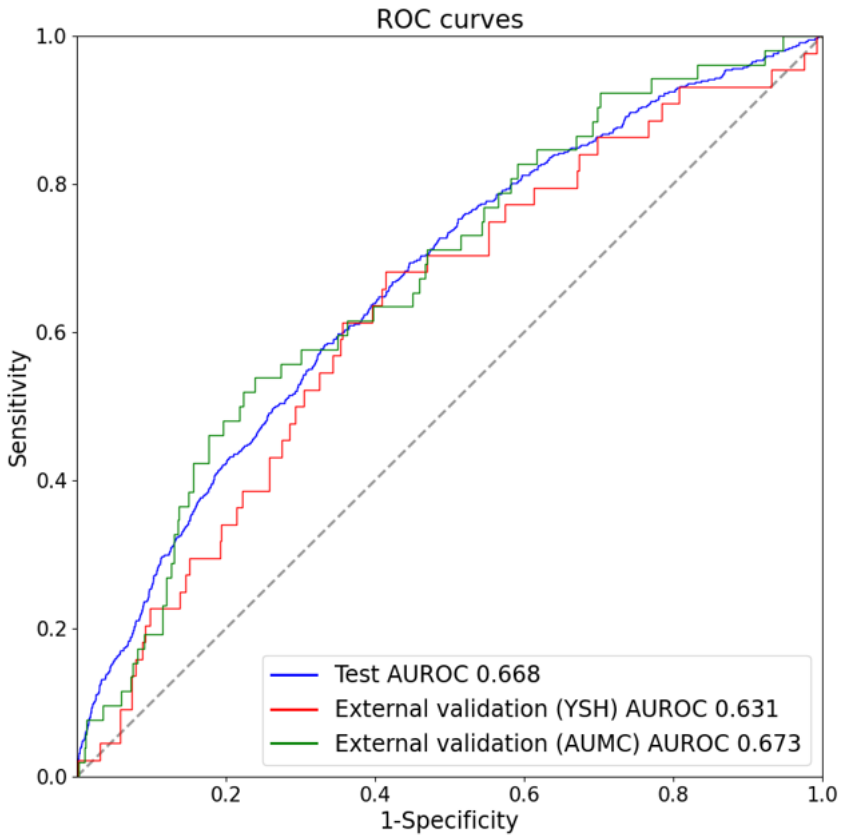


Figure 26. ROC curves (XGBoost using traditional ECG features, CACS ≥ 400).

ROC: receiver operating characteristics; ECG: electrocardiogram; CAC: coronary artery calcium; AUROC: area under the receiver operating characteristics curve; YSH: Yongin Severance Hospital; AUMC: Ajou University Medical Center

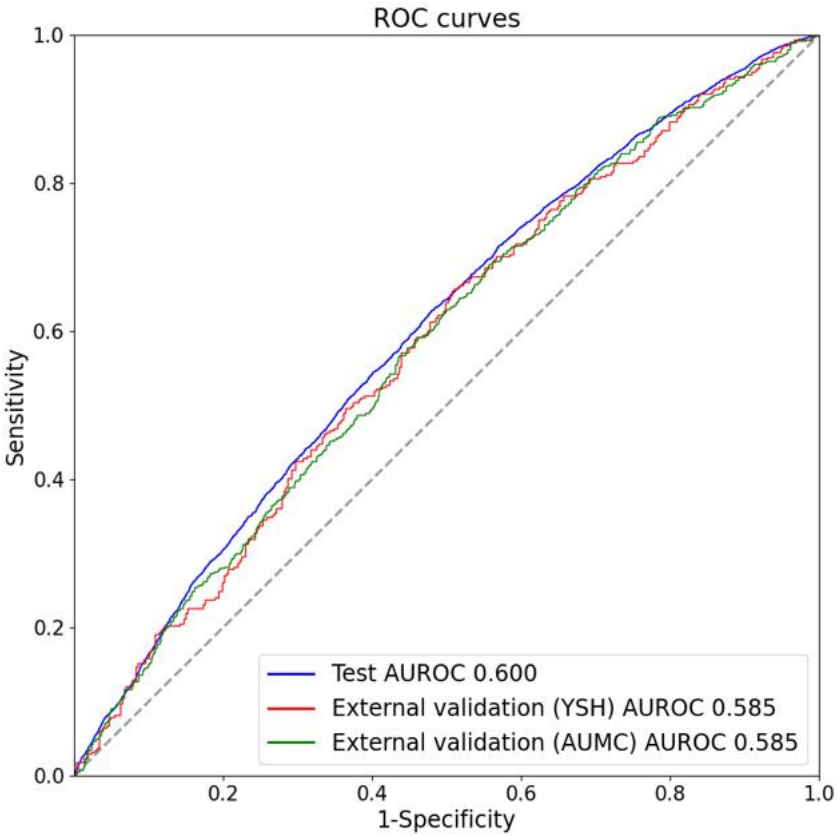


Figure 27. ROC curves (XGBoost using traditional ECG features, CACS > 0).

ROC: receiver operating characteristics; ECG: electrocardiogram; CAC: coronary artery calcium; AUROC: area under the receiver operating characteristics curve; YSH: Yongin Severance Hospital; AUMC: Ajou University Medical Center

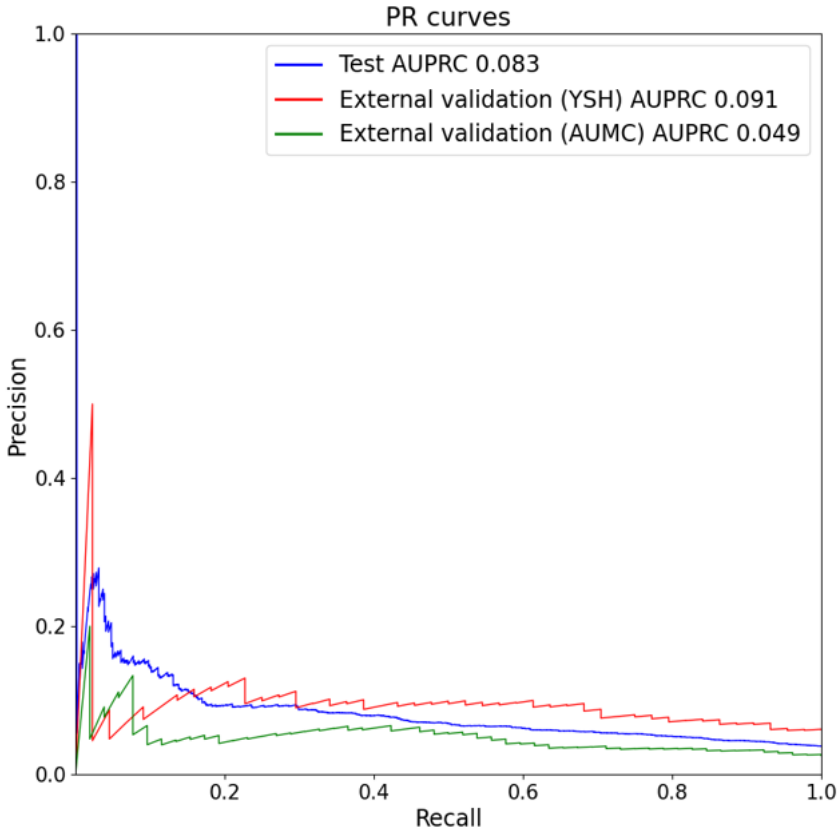


Figure 28. PR curves (XGBoost using traditional ECG features, CACS ≥ 400).

PR: receiver operating characteristics; ECG: electrocardiogram; CAC: coronary artery calcium; AUPRC: area under the precision-recall curve; YSH: Yongin Severance Hospital; AUMC: Ajou University Medical Center

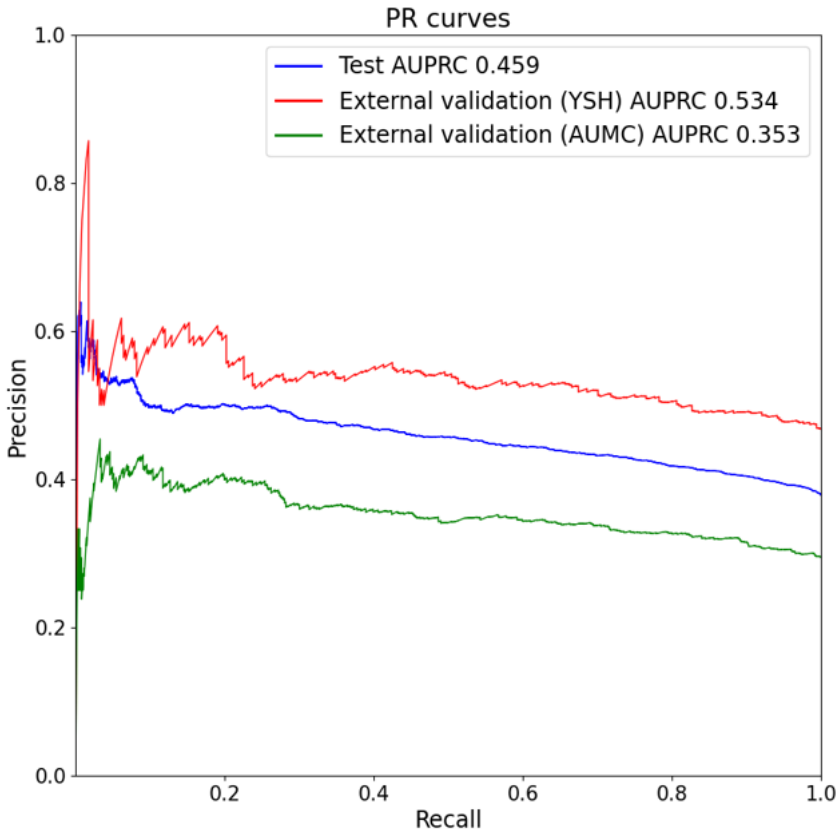


Figure 29. PR curves (XGBoost using traditional ECG features, CACS > 0).

PR: receiver operating characteristics; ECG: electrocardiogram; CAC: coronary artery calcium; AUPRC: area under the precision-recall curve; YSH: Yongin Severance Hospital; AUMC: Ajou University Medical Center

3.3. Multinational retrospective cohort analyses

3.3.1. SH cohort analysis dataset

Figure 30 illustrates the AI-ECG-derived reclassification within each PCE risk category in the SH cohort analysis dataset. Among the 52,400 individuals, 43,142, 7712, and 1546 were categorized as low, moderate, and high risk, respectively, based on the PCE. The MACE incidence rate (IR) per 1000 person-years (PY) was 1.9 for the PCE low-risk category and 9.2 for the PCE moderate-risk category, respectively. Among individuals classified as low risk by PCE but high risk by AI-ECG, the MACE IR per 1000 PY was 6.0, which was higher than the 3.3 observed in those classified as moderate risk by PCE but low risk by AI-ECG ($P = 0.007$). Within both the PCE low-risk and moderate-risk groups, there was a significant difference in Kaplan-Meier curves based on AI-ECG risk categories (Figure 31A, log-rank test $P < 0.001$; post-hoc pairwise comparisons: Low vs. Moderate $P < 0.001$, Moderate vs. High $P < 0.001$; adjusted significance level with Bonferroni correction: $0.5/2 = 0.025$. Figure 31B, log-rank test $P < 0.001$; post-hoc pairwise comparisons: Low vs. Moderate $P < 0.001$, Moderate vs. High $P < 0.001$; adjusted significance level with Bonferroni correction: $0.5/2 = 0.025$). The NRI for up-risking or down-risking based on AI-ECG within the PCE low- or moderate-risk categories (Figure 30) was 4.6% (95% CI: 3.2% – 5.9%) among events and -1.8% (95% CI: -2.0% – -1.6%) among non-events, resulting in a total NRI of 2.8% (95% CI: 1.4% – 4.1%) (Table 10A).

Table 11 presents the results of the Cox proportional hazards regression analysis for MACE. In the SH cohort analysis dataset, during a median follow-up of 7.9 years (interquartile range: 6.9 – 8.9 years), 1,110 individuals (1,611 ECGs) experienced a MACE. To facilitate a more intuitive interpretation of the data, we adjusted the scale of the AI-ECG output in the analysis. The AI-ECG output, originally presented on a scale from 0 to 1, was rescaled to a new range of 0 to 10 by multiplying it by a factor of ten. Consequently, the adjusted hazard ratios (HRs) for the AI-ECG output, as presented in Table 11, now reflect the change in hazards associated with a 10% absolute increase in the AI-ECG output. The analysis revealed significant positive associations between the AI-ECG output and the likelihood of MACE. For every 10% absolute increase in the AI-ECG output, the adjusted hazard for MACE was 1.087 (95% CI: 1.053 – 1.123) after adjusting for clinical variables. This indicates that the AI-ECG output is an independent risk factor for MACE. The PCE plus AI-ECG score, which yielded a C-index of 0.796 (Table 12, 95% CI: 0.782 – 0.802), surpassed the PCE by a difference of 0.010 (95% CI: 0.007 – 0.013) in the C-index.

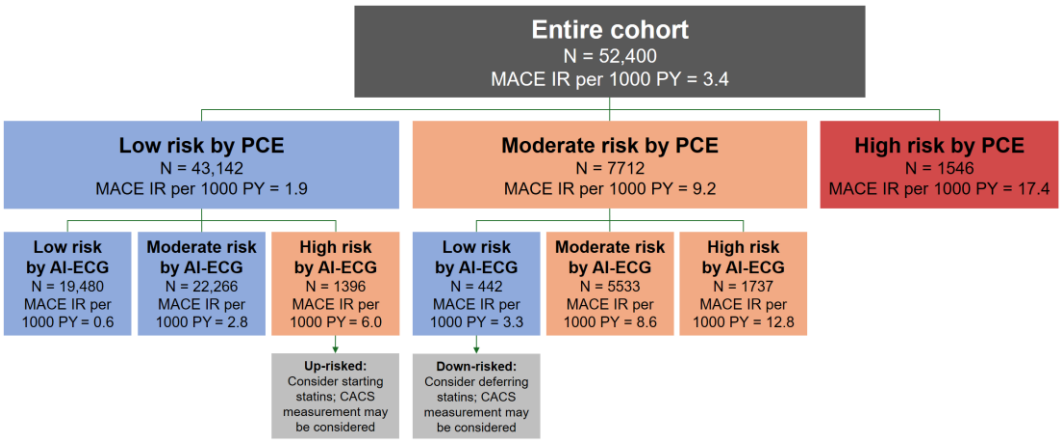


Figure 30. AI-ECG-derived reclassification within each PCE risk category (SH, cohort analysis dataset).

AI: artificial intelligence; ECG: electrocardiogram; AI-ECG: artificial intelligence-enabled electrocardiogram; PCE: Pooled Cohort Equations; SH: Severance Hospital; MACE: major adverse cardiovascular event; IR: incidence rate; PY: person-year; CAC: coronary artery calcification; CACS: coronary artery calcium score.

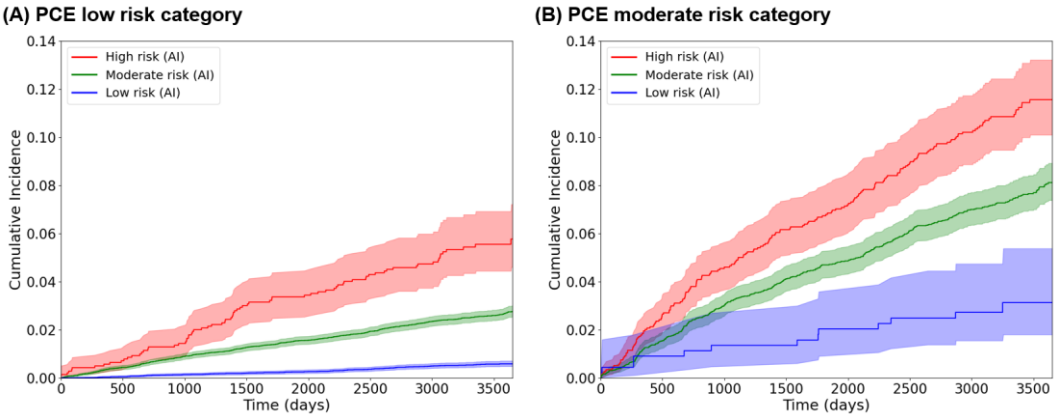


Figure 31. Kaplan-Meier curves (SH cohort analysis dataset). (A) Kaplan-Meier curves stratified by AI-ECG risk categories within the PCE low-risk group. (B) Kaplan-Meier curves stratified by AI-ECG risk categories within the PCE moderate-risk group.

SH: Severance Hospital; AI: artificial intelligence; ECG: electrocardiogram; AI-ECG: artificial intelligence-enabled electrocardiogram; PCE: Pooled Cohort Equations.

Table 10. Net reclassification improvement. The NRI for up-risking or down-risking based on AI-ECG within the PCE low- or moderate-risk categories.

(A) SH cohort analysis dataset

	Up-risked, % (95% CI)	**Down-risked, % (95% CI)**		
Number of cases (N = 1384)	5.5 (4.3 – 6.7)	0.9 (0.5 – 1.5)	Event NRI, % (95% CI)	4.6 (3.2 – 5.9)
Number of controls (N = 49,470)	2.7 (2.5 – 2.8)	0.9 (0.8 – 1.0)	Non-event NRI, % (95% CI)	-1.8 (-2.0 – -1.6)
			NRI, % (95% CI)	2.8 (1.4 – 4.1)

(B) UKB cohort analysis dataset

	Up-risked, % (95% CI)	**Down-risked, % (95% CI)**		
Number of cases (N = 604)	5.5 (3.7 – 7.4)	1.7 (0.7 – 2.7)	Event NRI, % (95% CI)	3.8 (1.7 – 6.0)
Number of controls (N = 28,679)	3.6 (3.4 – 3.8)	1.3 (1.2 – 1.5)	Non-event NRI, % (95% CI)	-2.3 (-2.5 – -2.0)
			NRI, % (95% CI)	1.5 (-0.6 – 3.7)

NRI: net reclassification improvement; SH: Severance Hospital; UKB: United Kingdom Biobank; CI: confidence interval.

Table 11. Cox regression analysis results. The AI-ECG output, originally presented on a scale from 0 to 1, was rescaled to a new range of 0 to 10 by multiplying it by a factor of ten.

	SH cohort analysis dataset		UKB cohort analysis dataset	
	Adjusted HR (95% CI)	**P-Value**	**Adjusted HR (95% CI)**	**P-Value**
Age, years	1.074 (1.068 – 1.080)	<0.001	1.064 (1.052 – 1.077)	<0.001
Sex (male)	1.926 (1.709 – 2.172)	<0.001	1.729 (1.441 – 2.075)	<0.001
DM	1.536 (1.321 – 1.785)	<0.001	1.690 (1.232 – 2.319)	0.001
HTN	1.407 (1.255 – 1.577)	<0.001	1.300 (1.076 – 1.571)	0.006
Current smoker	1.221 (1.071 – 1.392)	0.003	1.397 (1.069 – 1.824)	0.014
SBP, mmHg	1.003 (1.000 – 1.007)	0.078	1.008 (1.003 – 1.012)	<0.001
Total cholesterol, mg/dL	1.002 (1.001 – 1.004)	0.007	1.002 (1.000 – 1.004)	0.047
HDL cholesterol, mg/dL	0.994 (0.990 – 0.998)	0.005	0.979 (0.972 – 0.986)	<0.001
AI-ECG score * 10	1.087 (1.053 – 1.123)	<0.001	1.117 (1.061 – 1.175)	<0.001

SH: Severance Hospital; UKB: United Kingdom Biobank; HR: hazard ratio; CI: confidence interval; DM: diabetes mellitus; HTN: hypertension; SBP: systolic blood pressure; HDL: high-density lipoprotein; AI: artificial intelligence; ECG: electocardiogram

Table 12. C-index comparison. The C-index of the PCE for predicting MACE was 0.786 (95% CI: 0.777 – 0.796) in the SH cohort analysis dataset and 0.724 (95% CI: 0.705 – 0.741) in the UKB cohort analysis dataset.

	C-index of the PCE plus AI-ECG score (95% CI)	Differences in C-index with the PCE (95% CI)
SH cohort analysis dataset	0.796 (0.786 – 0.806)	0.010 (0.007 – 0.013)
UKB cohort analysis dataset	0.735 (0.716 – 0.754)	0.011 (0.004 – 0.019)

SH: Severance Hospital; UKB: United Kingdom Biobank; PCE: Pooled Cohort Equations; MACE: major adverse cardiovascular event; CI: confidence interval; AI: artificial intelligence; ECG: electrocardiogram

3.3.2. UKB cohort analysis dataset

Figure 32 illustrates the AI-ECG-derived reclassification within each PCE risk category in the UKB cohort analysis dataset. Among the 30,623 individuals, 20,744, 8539, and 1340 were categorized as low, moderate, and high risk, respectively, based on the PCE. The MACE IR per 1000 PY was 3.1 for the PCE low-risk category and 11.5 for the PCE moderate-risk category, respectively. Among individuals classified as low risk by PCE but high risk by AI-ECG, the MACE IR per 1000 PY was 8.3, which was higher than the 7.0 observed in those classified as moderate risk by PCE but low risk by AI-ECG, although the difference did not reach statistical significance ($P = 0.360$). Within the PCE low-risk group, there was a significant difference in Kaplan-Meier curves between individuals categorized as high and moderate risk by AI-ECG (Figure 33A, log-rank test: $P < 0.001$; post-hoc pairwise comparisons: low vs. moderate, $P = 0.035$; moderate vs. high, $P < 0.001$; adjusted significance level using Bonferroni correction: 0.025). However, within the PCE moderate-risk group, no significant difference was observed in Kaplan-Meier curves between the AI-ECG low- and moderate-risk groups (Figure 33B, log-rank test: $P < 0.001$; post-hoc pairwise comparisons: low vs. moderate, $P = 0.361$; moderate vs. high, $P < 0.001$; adjusted significance level using Bonferroni correction: 0.025). The NRI for up-risking or down-risking based on AI-ECG within the PCE low- or moderate-risk categories (Figure 32) was 3.8% (95% CI: 1.7% – 6.0%) among events and -2.3% (95% CI: -2.5% – -2.0%) among non-events, resulting in a total NRI of 1.5% (95% CI: -0.6% – 3.7%) (Table 10B).

Table 11 presents the results of the Cox proportional hazards regression analysis for MACE. In the UKB cohort analysis dataset, during a median follow-up of 3.5 years (interquartile range: 2.5 – 5.0 years), 699 individuals experienced a MACE. For every 10% absolute increase in the AI-ECG output, the adjusted hazard for MACE was 1.117 (95% CI: 1.061 – 1.175) after adjusting for clinical variables (Table 11). This indicates that the AI-ECG output is an independent risk factor for MACE. The PCE plus AI-ECG score, which yielded a C-index of 0.735 (Table 12, 95% CI: 0.716 – 0.754), surpassed the PCE by a difference of 0.011 (95% CI: 0.004 – 0.019) in the C-index.

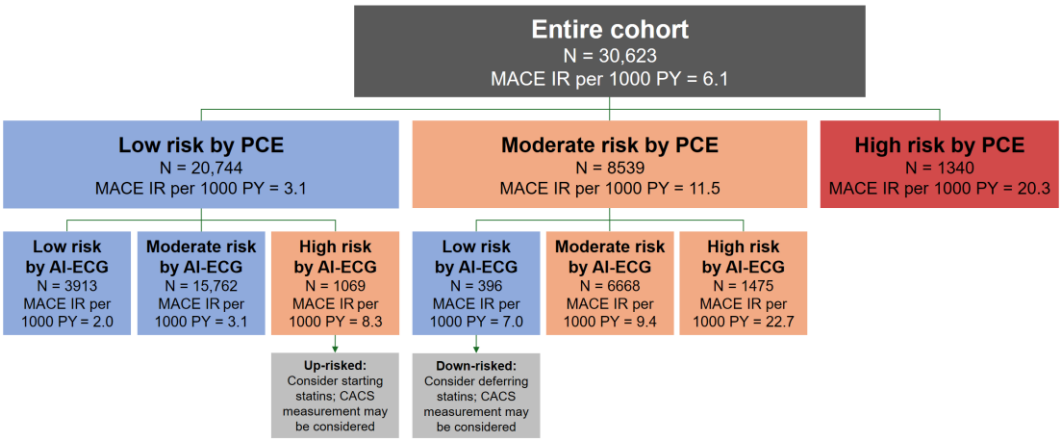


Figure 32. AI-ECG-derived reclassification within each PCE risk category (UKB, cohort analysis dataset).

AI: artificial intelligence; ECG: electrocardiogram; AI-ECG: artificial intelligence-enabled electrocardiogram; PCE: Pooled Cohort Equations; UKB: United Kingdom Biobank; MACE: major adverse cardiovascular event; IR: incidence rate; PY: person-year; CAC: coronary artery calcification; CACS: coronary artery calcium score.

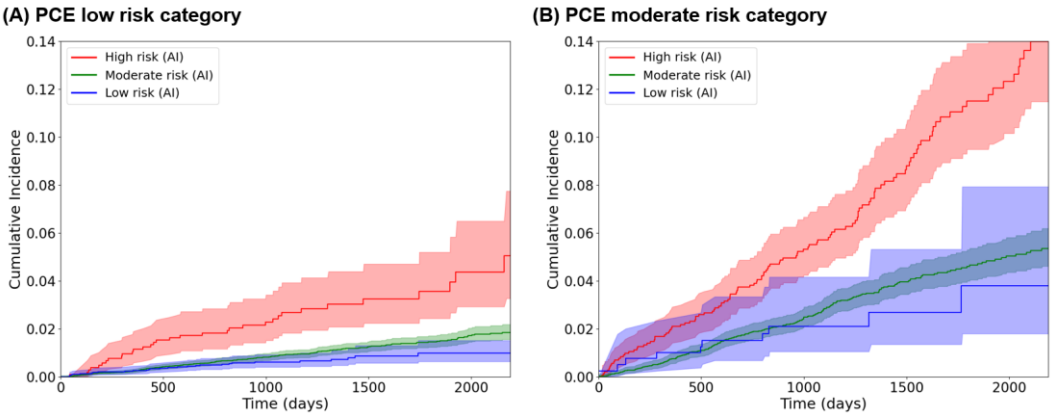


Figure 33. Kaplan-Meier curves (UKB cohort analysis dataset). (A) Kaplan-Meier curves stratified by AI-ECG risk categories within the PCE low-risk group. (B) Kaplan-Meier curves stratified by AI-ECG risk categories within the PCE moderate-risk group.

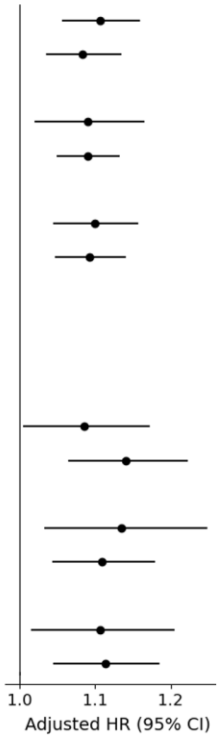
UKB: United Kingdom Biobank; AI: artificial intelligence; ECG: electrocardiogram; AI-ECG: artificial intelligence-enabled electrocardiogram; PCE: Pooled Cohort Equations.

3.3.3. Subgroup analyses

Figure 34 shows the hazard ratios of the AI-ECG model by subgroup. The association between the AI-ECG model and MACE remained consistent across all demographic and PCE-based subgroups. In the SH cohort analysis dataset, for every 10% absolute increase in the AI-ECG output, the adjusted hazard for MACE was 1.106 (95% CI: 1.057–1.158) in the age < 60 subgroup, 1.083 (95% CI: 1.036–1.133) in the age ≥ 60 subgroup, 1.090 (95% CI: 1.021–1.164) in the female subgroup, 1.090 (95% CI: 1.050–1.131) in the male subgroup, 1.100 (95% CI: 1.046–1.156) in the low-risk (PCE) subgroup, and 1.092 (95% CI: 1.048–1.139) in the moderate or high-risk (PCE) subgroup, after adjusting for clinical variables. There were no differences in adjusted HR comparison between subgroups (P-value for hazard ratio comparison: 0.518, 0.988, and 0.844, respectively). In the UKB cohort analysis dataset, for every 10% absolute increase in the AI-ECG output, the adjusted hazard for MACE was 1.085 (95% CI: 1.006–1.171) in the age < 60 subgroup, 1.141 (95% CI: 1.065–1.222) in the age ≥ 60 subgroup, 1.135 (95% CI: 1.034–1.247) in the female subgroup, 1.109 (95% CI: 1.044–1.178) in the male subgroup, 1.106 (95% CI: 1.016–1.204) in the low-risk (PCE) subgroup, and 1.113 (95% CI: 1.046–1.184) in the moderate or high-risk (PCE) subgroup, after adjusting for clinical variables. There were no differences in adjusted HR comparison between subgroups (P-value for hazard ratio comparison: 0.342, 0.677, and 0.912, respectively).

Figure 34. Hazard ratios of the AI-ECG model by subgroups. The hazard ratios were adjusted for the variables included in the PCE (age, sex, diabetes mellitus, hypertension, smoking status, total cholesterol, high-density lipoprotein cholesterol and systolic blood pressure). The AI-ECG output, originally presented on a scale from 0 to 1, was rescaled to a new range of 0 to 10 by multiplying it by a factor of ten.

SH cohort analysis dataset			
	Adjusted HR (95% CI)	**P-value**	**P-value for HR comparison**
Age			0.518
< 60	1.106 (1.057 – 1.158)	<0.001	
≥ 60	1.083 (1.036 – 1.133)	<0.001	
Sex			0.988
Female	1.090 (1.021 – 1.164)	0.010	
Male	1.090 (1.050 – 1.131)	<0.001	
PCE			0.844
Low risk	1.100 (1.046 – 1.156)	<0.001	
≥ Moderate risk	1.092 (1.048 – 1.139)	<0.001	
UKB cohort analysis dataset			
	Adjusted HR (95% CI)	**P-value**	**P-value for HR comparison**
Age			0.342
< 60	1.085 (1.006 – 1.171)	0.035	
≥ 60	1.141 (1.065 – 1.222)	<0.001	
Sex			0.677
Female	1.135 (1.034 – 1.247)	0.008	
Male	1.109 (1.044 – 1.178)	<0.001	
PCE			0.912
Low risk	1.106 (1.016 – 1.204)	0.019	
≥ Moderate risk	1.113 (1.046 – 1.184)	<0.001	



AI: artificial intelligence; ECG: electrocardiogram; AI-ECG: artificial intelligence-enabled electrocardiogram; PCE: Pooled Cohort Equations; HR: hazard ratio; CI: confidence interval; SH: Severance Hospital; UKB: United Kingdom Biobank

3.4. Interpretation

The XGBoost model constructed with ECG factors derived from the pre-trained VAE model demonstrated an AUROC of 0.734 in the test dataset. Figure 35 shows the SHAP summary plot of the top 10 important features of the XGBoost model using ECG factors, and Figures 36–40 provide visualizations of the important ECG factors: Upward shift of the ST segment in the anteroseptal leads with reciprocal downward shift of the ST segment in the inferolateral leads (lower values of ECG factor 48), downward shift of the ST segment in all leads (lower values of ECG factor 39), longer PR interval (higher values of ECG factor 36), increased QRS amplitude in the anterolateral leads with decreased QRS amplitude in the inferoseptal leads (lower values of ECG factor 23), and shorter TP interval (lower values of ECG factor 31) were associated with increased predicted risk.

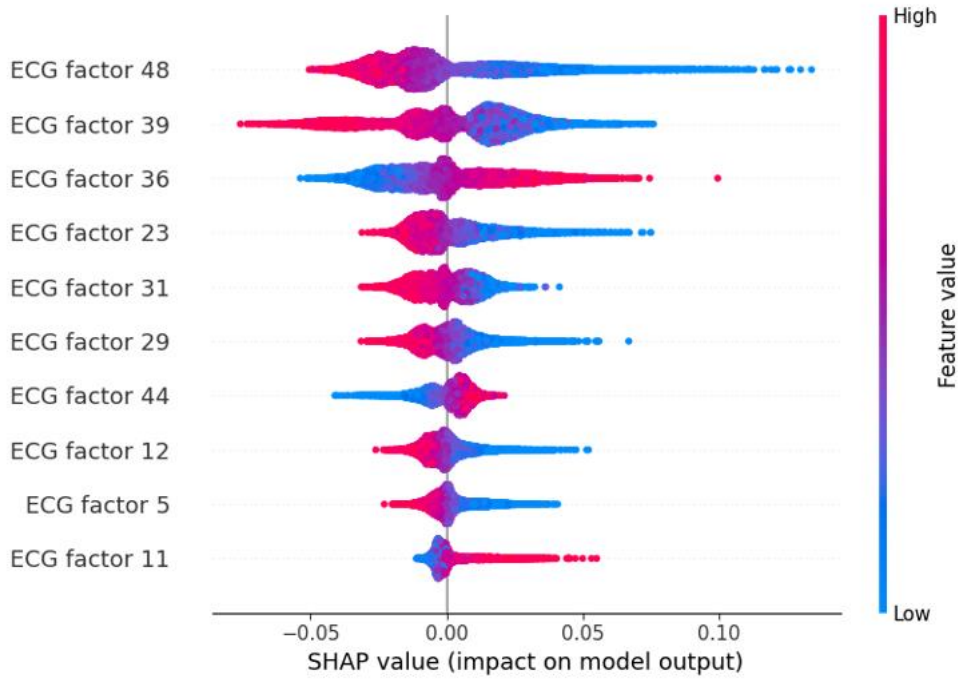


Figure 35. SHAP summary plot of XGBoost model using VAE features. Top 10 important features as shown by SHAP analysis.

SHAP: Shapley Additive exPlanations; VAE: variational autoencoder; ECG: electrocardiogram

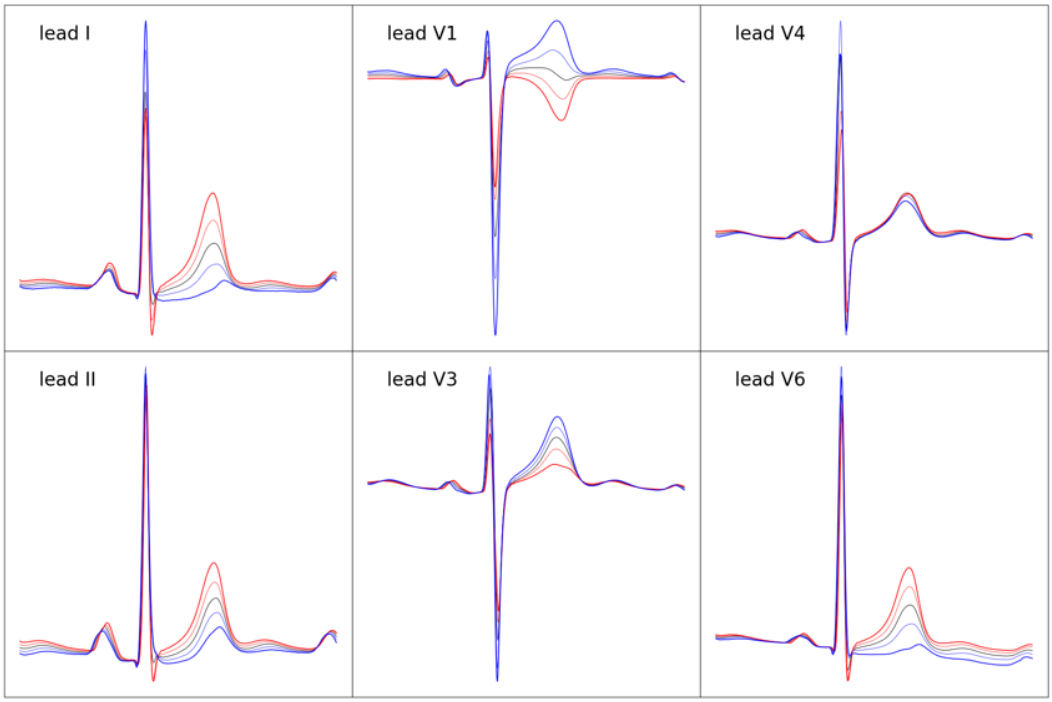


Figure 36. Factor traversals of ECG factor 48. Upward shift of the ST segment in the anteroseptal leads with reciprocal downward shift of the ST segment in the inferolateral leads (lower values of ECG factor 48), downward shift of the ST segment in all leads (lower values of ECG factor 39), longer PR interval (higher values of ECG factor 36), increased QRS amplitude in the anterolateral leads with decreased QRS amplitude in the inferoseptal leads (lower values of ECG factor 23), and shorter TP interval (lower values of ECG factor 31) were associated with increased predicted risk.

ECG: electrocardiogram

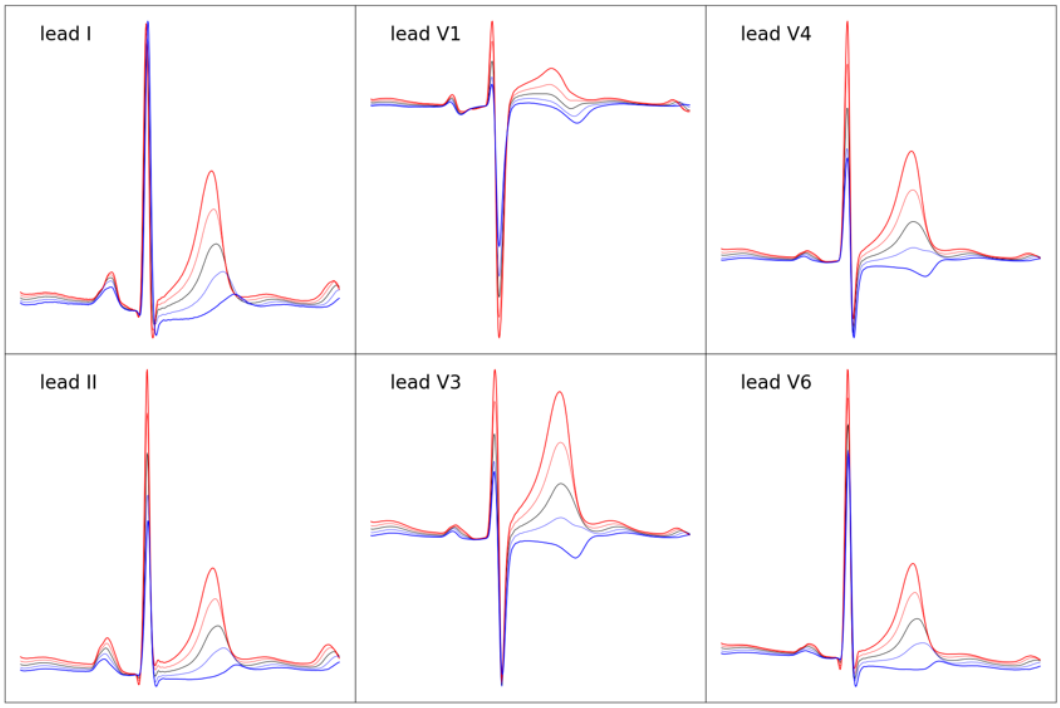


Figure 37. Factor traversals of ECG factor 39. Upward shift of the ST segment in the anteroseptal leads with reciprocal downward shift of the ST segment in the inferolateral leads (lower values of ECG factor 48), downward shift of the ST segment in all leads (lower values of ECG factor 39), longer PR interval (higher values of ECG factor 36), increased QRS amplitude in the anterolateral leads with decreased QRS amplitude in the inferoseptal leads (lower values of ECG factor 23), and shorter TP interval (lower values of ECG factor 31) were associated with increased predicted risk.

ECG: electrocardiogram

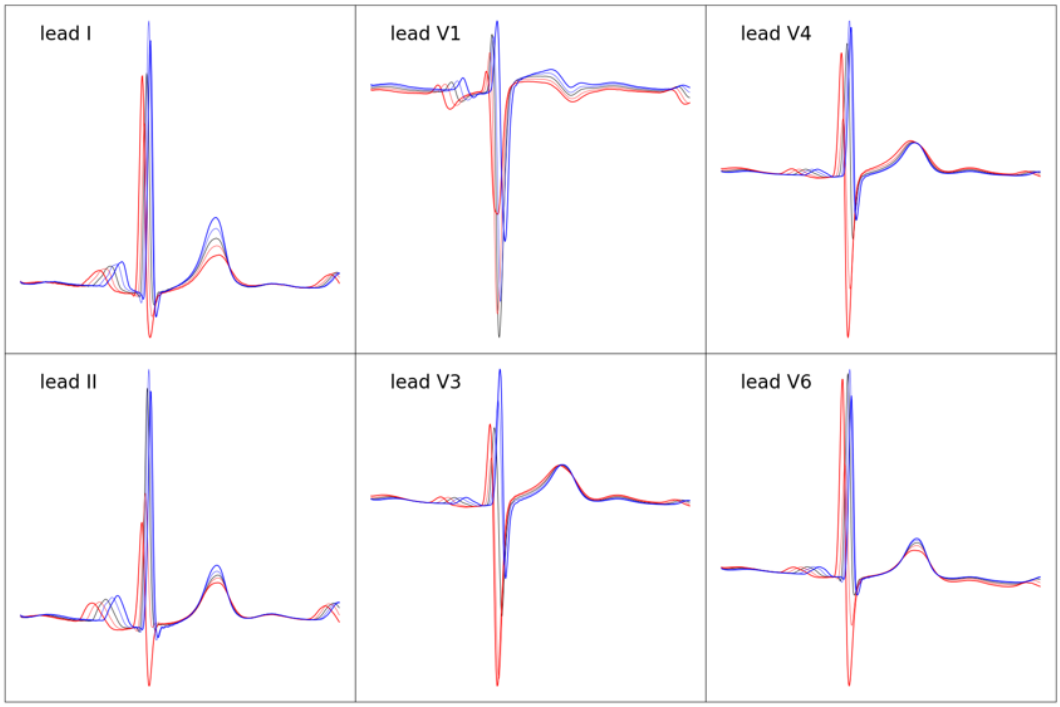


Figure 38. Factor traversals of ECG factor 36. Upward shift of the ST segment in the anteroseptal leads with reciprocal downward shift of the ST segment in the inferolateral leads (lower values of ECG factor 48), downward shift of the ST segment in all leads (lower values of ECG factor 39), longer PR interval (higher values of ECG factor 36), increased QRS amplitude in the anterolateral leads with decreased QRS amplitude in the inferoseptal leads (lower values of ECG factor 23), and shorter TP interval (lower values of ECG factor 31) were associated with increased predicted risk.

ECG: electrocardiogram

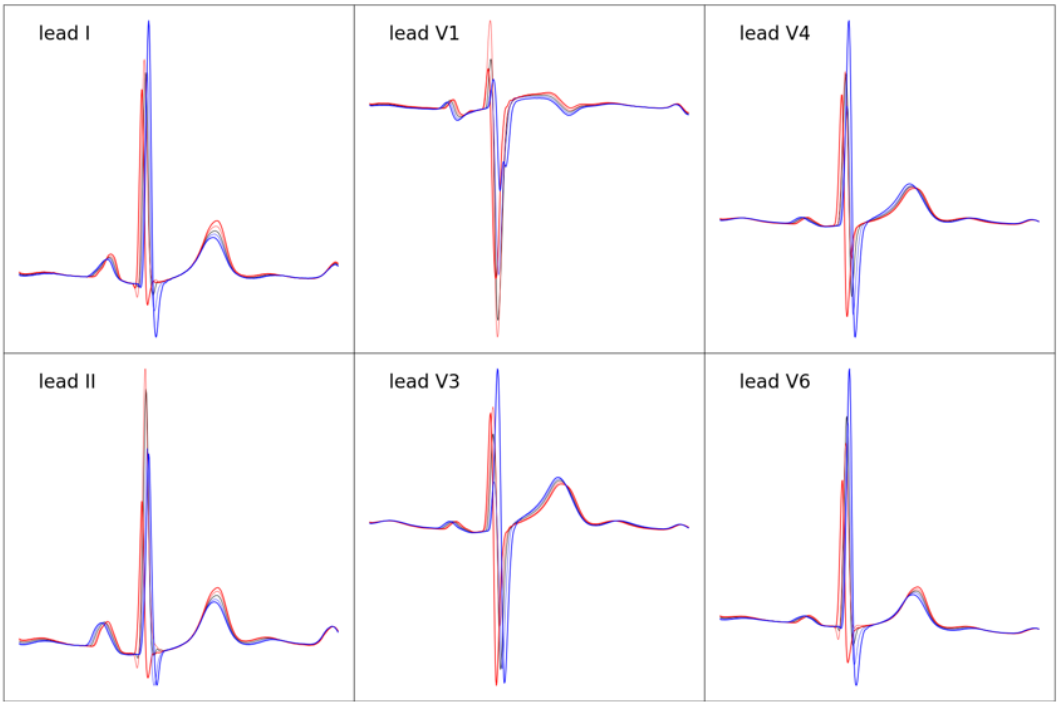


Figure 39. Factor traversals of ECG factor 23. Upward shift of the ST segment in the anteroseptal leads with reciprocal downward shift of the ST segment in the inferolateral leads (lower values of ECG factor 48), downward shift of the ST segment in all leads (lower values of ECG factor 39), longer PR interval (higher values of ECG factor 36), increased QRS amplitude in the anterolateral leads with decreased QRS amplitude in the inferoseptal leads (lower values of ECG factor 23), and shorter TP interval (lower values of ECG factor 31) were associated with increased predicted risk.

ECG: electrocardiogram

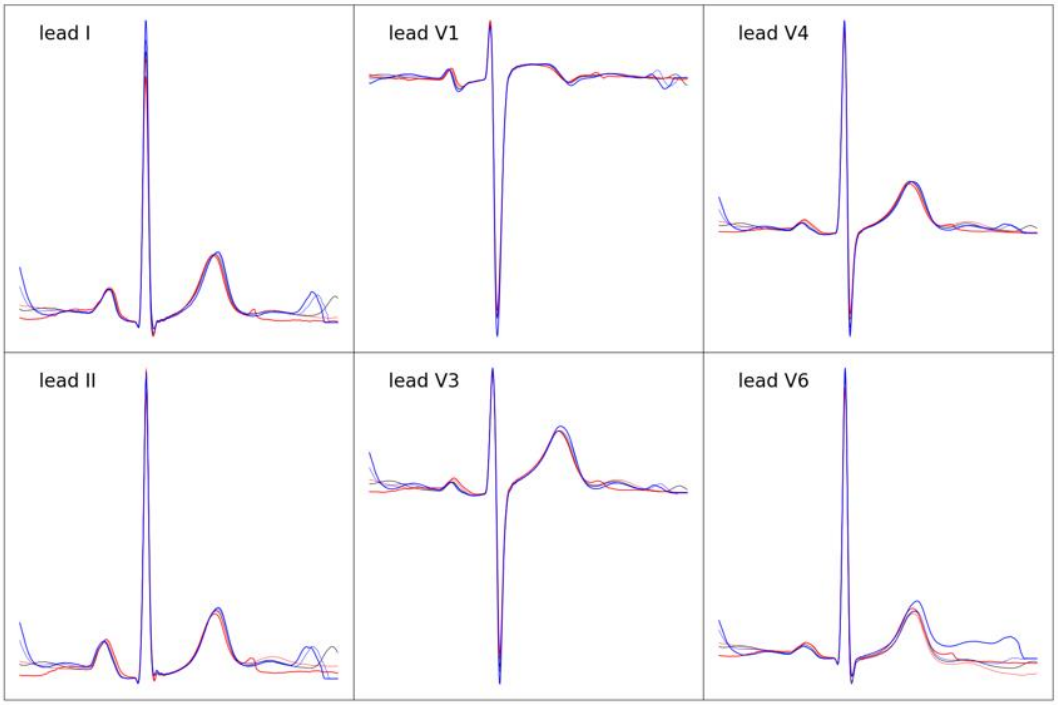


Figure 40. Factor traversals of ECG factor 31. Upward shift of the ST segment in the anteroseptal leads with reciprocal downward shift of the ST segment in the inferolateral leads (lower values of ECG factor 48), downward shift of the ST segment in all leads (lower values of ECG factor 39), longer PR interval (higher values of ECG factor 36), increased QRS amplitude in the anterolateral leads with decreased QRS amplitude in the inferoseptal leads (lower values of ECG factor 23), and shorter TP interval (lower values of ECG factor 31) were associated with increased predicted risk.

ECG: electrocardiogram

3.5. Compliance with reporting guidelines

The Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) Checklist for Prediction Model Development and Validation was followed (Table 13)[59].

Table 13. TRIPOD checklist (prediction model development and validation). Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V[59].

Section/Topic	Item		Checklist Item	Location
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	Appropriate title
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	Appropriate abstract
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	1.1. Background
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	1.3. Objectives
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	2.1. Data sources and labeling 2.7. Multinational retrospective cohort analyses
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	2.1. Data sources and labeling 2.7. Multinational retrospective cohort analyses
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	2.1. Data sources and labeling 2.7. Multinational retrospective cohort analyses
	5b	D;V	Describe eligibility criteria for participants.	2.1. Data sources and labeling 2.7. Multinational retrospective cohort analyses
	5c	D;V	Give details of treatments received, if relevant.	2.1. Data sources and labeling 2.7. Multinational retrospective cohort analyses

Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	2.5. Outcomes 2.6. Performance evaluation
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	2.5. Outcomes 2.6. Performance evaluation
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	2.3. AI-enabled ECG framework development 2.4. Outcomes
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	2.3. AI-enabled ECG model development 2.5. Outcomes
Sample size	8	D;V	Explain how the study size was arrived at.	2.1. Data sources and labeling 2.7. Multinational retrospective cohort analyses
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	2.1. Data sources and labeling 2.7. Multinational retrospective cohort analyses
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	2.2. Data preprocessing
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	2.3. AI-enabled ECG model development 2.5. Outcomes
	10c	V	For validation, describe how the predictions were calculated.	2.3. AI-enabled ECG model development 2.5. Outcomes
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	2.6. Performance evaluation
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	N/A (not done)
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	2.7. Multinational retrospective cohort analyses
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	2.1. Data sources and labeling 2.7. Multinational retrospective cohort analyses
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of	3.1. Dataset characteristics

			participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	3.1. Dataset characteristics
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	3.1. Dataset characteristics
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	3.1. Dataset characteristics
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	3.1. Dataset characteristics
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	3.2. Model performance 3.3. Multinational retrospective cohort analyses
	15b	D	Explain how to the use the prediction model.	2.3. AI-enabled ECG model development 2.5. Outcomes
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	3.2. Model performance 3.3. Multinational retrospective cohort analyses
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	N/A (not done)
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	4.4. Limitations
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	4.2. Implications 4.3. Strengths
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	4.2. Implications 4.3. Strengths
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	4.2. Implications 4.3. Strengths
Other information				

Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	All information is provided in the main manuscript.
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	Not relevant to this manuscript.

TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis

4. DISCUSSION

4.1. Summary of key findings

In this study, we developed an AI-ECG model to predict coronary artery calcification and validated its potential for opportunistic screening using a health screening dataset. Our AI-ECG model showed strong performance in predicting CAC, achieving an AUROC of 0.841 for CACS ≥ 400 and an AUROC of 0.720 for CAC > 0 in the health screening test dataset. Our AI-ECG model maintained robustness in external validation health screening datasets, underscoring its generalizability to various environments. In the PCE low-risk category, 24.9% had CAC > 0; among these individuals, the proportion increased to 45.7% when selecting those identified as high-risk by the AI-ECG model. In the PCE moderate-risk category, 35.4% had CACS = 0; among these individuals, the proportion increased to 60.9% when selecting those identified as low-risk by the AI-ECG model. Among PCE low-risk individuals who were reclassified as high-risk by AI-ECG, the IR of MACE was higher compared to those in the PCE moderate-risk category who were reclassified as low-risk by AI-ECG (SH cohort analysis dataset MACE IR per 1000 PY: 6.0 vs. 3.3, UKB cohort analysis dataset MACE IR per 1000 PY: 8.3 vs. 7.0). AI-ECG was an independent risk factor for MACE (adjusted HR [95% CI]: 1.087 [1.053–1.123] in the SH cohort analysis dataset and 1.117 [1.061–1.175] in the UKB cohort analysis dataset). AI-ECG provided additional predictive value beyond the PCE, with the combined PCE plus AI-ECG score outperforming the PCE alone in terms of C-index. The association between AI-ECG and MACE remained consistent across all demographic and PCE-based subgroups. We provided visual morphological interpretations of ECG factors associated with increased predicted risk, identifying potential changes such as upward shift of the ST segment in the anteroseptal leads with reciprocal downward shift in the inferolateral leads, downward shift of the ST segment in all leads, longer PR interval, and others, to be associated with CAC.

4.2. Implications

We demonstrated the potential integration of our AI-ECG model into the clinical workflow by assessing its effectiveness for CAC screening and its ability to reclassify individuals for initiating or withholding primary prevention decisions. Specifically, we showed that the AI-ECG model could screen individuals in the PCE low-risk group with the highest likelihood of having CAC, and those

in the PCE moderate-risk group with the lowest likelihood of having CAC. In the PCE low-risk category, where detecting CAC could recommend initiating primary prevention strategies, 24.9% of individuals had a CACS > 0. Of these, 3.2%–5.2% were classified as high risk by the AI-ECG model, with the proportion of individuals with CAC in this subgroup rising to 45.7%. Similarly, in the PCE moderate-risk category, where ruling out CAC could recommend withholding primary prevention strategies, 35.4% had a CACS = 0. Among these, 4.6%–5.7% were classified as low risk by the AI-ECG model, with the proportion of CAC-free individuals in this subgroup reaching 60.9%. This suggests that the AI-ECG model could be used to screen and prioritize a small subset of individuals for whom CACS measurement would be most useful, especially given that routine CACS measurement is impractical due to cost, radiation exposure, or lack of insurance coverage[21,22]. Particularly in routine health screenings where ECGs are universally performed, AI-ECG-based CAC prediction could enable opportunistic CAC screening in the general population, paving the way for earlier detection and timely implementation of primary prevention strategies.

Additionally, decisions regarding initiating or withholding statin therapy could be directly guided by AI-ECG risk reclassification. Across two cohort datasets, individuals classified as moderate risk by the PCE but as low risk by the AI-ECG model had a lower MACE IR than those classified as low risk by the PCE but as high risk by the AI-ECG model (MACE IR per 1000 PY: 3.3 vs. 6.0 in the SH cohort; 7.0 vs. 8.3 in the UKB cohort). Thus, it would be more reasonable to withhold statin therapy (down-risk) in individuals classified as PCE moderate risk but AI-ECG low risk, and to initiate statin therapy (up-risk) in those classified as PCE low risk but AI-ECG high risk. This approach is further supported by the NRI for up-risking or down-risking based on AI-ECG within the PCE low- or moderate-risk categories, which was 2.8% (95% CI: 1.4%–4.1%) and 1.5% (95% CI: -0.6%–3.7%) in the two cohort datasets, respectively.

We demonstrated that AI-ECG prediction serves as an independent risk factor for MACE and enhances the predictive value of conventional risk stratification tools, such as the PCE, with the combined PCE and AI-ECG scores outperforming the PCE alone in terms of the C-index. Given that ECG is a ubiquitous and cost-effective tool, these findings suggest that incorporating our AI-ECG model as a novel risk factor into existing CVD risk prediction tools—or developing new CVD risk prediction tools that integrate AI-ECG—could provide significant advantages in predictive ability. Furthermore, the association between AI-ECG and MACE remained robust across all demographic and PCE-based subgroups, indicating that the AI-ECG model is effective across a diverse range of individuals and cardiovascular risk levels. Moreover, regarding the model's generalizability and reproducibility, similar trends were observed when the AI-ECG model, primarily developed using data from an Asian population, was applied to a United Kingdom dataset for survival analysis.

Previous studies have developed AI-ECG models capable of predicting CAC using only ECGs and have validated these models on a more general patient spectrum[38,39]. However, a notable

limitation of these studies is the lack of exclusion of individuals with clinical ASCVD. According to guidelines[5-7], CAC scoring is not recommended for individuals with clinical ASCVD, as they are already candidates for primary prevention or more advanced treatments and evaluations. Consequently, AI-ECG-based CAC prediction holds no utility for this group. Without excluding individuals with clinical ASCVD, it is unclear whether the model's performance is biased toward those with an existing diagnosis or symptoms. In our current study, we sought to address this limitation by validating the AI-ECG model on individuals undergoing health screenings. Health screenings embody the principles of preventive medicine, primarily targeting asymptomatic individuals to proactively detect early-stage diseases and implement primary prevention when necessary. Although it is not possible to confirm that the health screening cohort was entirely subclinical, these individuals represented a healthier spectrum than patients undergoing outpatient CAC scoring, as evidenced by their younger age and lower CAC scores compared to the model development dataset. Furthermore, our retrospective cohort analyses excluded individuals with prior cardiovascular events, ensuring that our findings are clinically relevant for individuals without overt disease.

While the association between CAC and structural changes in the heart has not been extensively studied, some recent studies indicates that higher CACS is linked to adverse cardiac remodeling, including increased left ventricular mass and larger aortic root diameter[60,61]. These findings suggest that higher CACS are linked to structural changes in the heart. Such structural changes may manifest as subtle alterations in the ECG that were previously undetectable to the human eye but can now be identified through the application of AI techniques[30]. In our results, the observation that higher CACS correlated with higher outputs from the AI-ECG model suggests that elevated CACS may be associated with more pronounced structural changes in the heart that the model was able to detect.

We integrated a VAE model (pre-trained using 5.6 million ECGs) to provide visual morphological explainability, as described in previous studies[35,36]. This approach allowed us to mitigate the “black box” issue prevalent in traditional end-to-end deep learning techniques, enabling us to provide quantifiable visual interpretations of the morphological ECG changes linked to our prediction task. We found potential ECG changes such as upward shift of the ST segment in the anteroseptal leads with reciprocal downward shift of the ST segment in the inferolateral leads, downward shift of the ST segment in all leads, longer PR interval, increased QRS amplitude in the anterolateral leads with decreased QRS amplitude in the inferoseptal leads, and shorter TP interval that might be associated with CAC. Our literature review revealed that studies investigating ECG abnormalities associated with CAC are limited, and the findings are often conflicting: While some studies have identified associations between ST-T or Q wave abnormalities and CAC[62], others have reported no such links[63]. Although further research is necessary to establish the associations between the possible morphological ECG changes identified in our study and CAC, novel data-driven AI-based approaches, such as the one we employed, can open up new opportunities for exploration.

4.3. Strengths

Our study has numerous strengths. Firstly, our AI-ECG model uniquely requires only ECGs for analysis, without the need for any additional clinical data, enhancing its usability in various clinical settings. It includes all ECGs regardless of medical anomalies, such as arrhythmias or ischemia, thereby increasing its applicability; the only exclusions are cases of lead misplacements, unwanted artifacts, and the presence of artificial pacemakers. This inclusive approach ensures that the model can be widely applicable across a diverse range of ECG readings. Second, our model has consistently demonstrated strong performance across multiple external validation cohorts, underscoring its reliability and adaptability in various clinical environments. Third, while there have been concerns about applying AI-ECG models developed on one racial or ethnic group to others[64], our findings showed that similar trends were observed when our AI-ECG model, initially developed using data primarily from an Asian population, was applied to the United Kingdom dataset for survival analysis. This highlights the model's robust performance across different racial and ethnic groups, reinforcing its potential for broader clinical application. Fourth, the training processes for both the AI-ECG and VAE models were conducted using exceptionally large datasets, which is a significant advantage. Specifically, nearly 200,000 ECGs were employed for the training and validation of the AI-ECG model, while more than 5 million ECGs were utilized to train the VAE model. Utilizing such large datasets helps to minimize the risk of overfitting, allows the models to capture a wider range of variabilities, and ultimately increases the confidence and generalizability of the models in clinical applications[65,66].

4.4. Limitations

This study's findings should be interpreted in light of the following limitations. First, the retrospective design introduces some limitations. Although we validated our AI-ECG model using health screening data, potential selection bias may exist depending on who decided or was recommended to undergo CAC scoring during health screenings, a factor that cannot be precisely known due to the retrospective nature of the study. Additionally, because of its retrospective design, data not originally recorded, such as the presence or absence of symptoms, could not be extracted. Future prospective studies are necessary to confirm the findings of the study. Second, although the ACC/AHA PCE is a widely used CVD risk prediction model, applying it to the Korean and UK cohorts may introduce inaccuracies. Specifically, studies have shown that the PCE, developed based on American cohorts, tends to overestimate the risk in Asian and European populations[67-70]. This misalignment may affect the validity of the AI-ECG based risk reclassification of individuals

compared to the PCE in Korean and UK populations. Third, while mortality data linked using KOSTAT mortality records and resident registration numbers ensures 100% accuracy up to the end of follow-up date, the occurrence of cardiovascular events, extracted from EMRs, may be prone to incomplete capture: Events diagnosed at other hospitals may not be captured, and it is not possible to ascertain whether the diagnosis codes extracted from EMRs represent the patient's first-ever diagnosis. Future studies should utilize longitudinal data with comprehensive diagnosis records that include diagnoses made at any institution. Fourth, in the UKB, clinical data were collected at the first assessment date, while resting ECGs were performed during a subsequent imaging visit, creating an unavoidable time gap that could have influenced the results. Fifth, the current model's explainability with the pre-trained VAE model is limited due to the ambiguous nature of some factors, which capture multiple ECG alterations simultaneously. Advancements in AI-based feature extraction methods might enhance our understanding of these ECG factors.

5. CONCLUSIONS

In conclusion, our AI-ECG model proves to be an effective tool for predicting coronary artery calcification. We demonstrated the potential integration of our AI-ECG model into clinical workflow by showing its dual utility: it can either screen individuals who would benefit most from CACS measurement, or directly guide decisions regarding statin therapy initiation or withholding through patient reclassification. The ubiquitous availability of ECGs, combined with our finding that the AI-ECG model serves as an independent risk factor for cardiovascular events, suggests its potential for incorporation into CVD risk prediction tools. Particularly in routine health screenings where ECGs are universally performed, AI-ECG-based CAC prediction could enable opportunistic CAC screening in the general population, paving the way for earlier detection and timely implementation of primary prevention strategies.

REFERENCES

1. Fryar CD, Chen T-C, Li X. Prevalence of uncontrolled risk factors for cardiovascular disease: United States, 1999-2010. NCHS Data Brief 2012 Aug;(103):1–8. PMID:23101933

2. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Delling FN, Djousse L, Elkind MSV, Ferguson JF, Fornage M, Khan SS, Kissela BM, Knutson KL, Kwan TW, Lackland DT, Lewis TT, Lichtman JH, Longenecker CT, Loop MS, Lutsey PL, Martin SS, Matsushita K, Moran AE, Mussolino ME, Perak AM, Rosamond WD, Roth GA, Sampson UKA, Satou GM, Schroeder EB, Shah SH, Shay CM, Spartano NL, Stokes A, Tirschwell DL, VanWagner LB, Tsao CW, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association. Circulation 2020 Mar 3;141(9):e139–e596. PMID:31992061

3. Birger M, Kaldjian AS, Roth GA, Moran AE, Dieleman JL, Bellows BK. Spending on Cardiovascular Disease and Cardiovascular Risk Factors in the United States: 1996 to 2016. Circulation 2021 Jul 27;144(4):271–282. PMID:33926203

4. Stone NJ, Robinson JG, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, Goldberg AC, Gordon D, Levy D, Lloyd-Jones DM, McBride P, Schwartz JS, Shero ST, Smith SC Jr, Watson K, Wilson PWF, American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2014 Jul 1;63(25 Pt B):2889–2934. PMID:24239923

5. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, Braun LT, de Ferranti S, Faiella-Tommasino J, Forman DE, Goldberg R, Heidenreich PA, Hlatky MA, Jones DW, Lloyd-Jones D, Lopez-Pajares N, Ndumele CE, Orringer CE, Peralta CA, Saseen JJ, Smith SC Jr, Sperling L, Virani SS, Yeboah J. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 2019 Jun 18;139(25):e1082–e1143. PMID:30586774

6. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, Himmelfarb CD, Khera A, Lloyd-Jones D, McEvoy JW, Michos ED, Miedema MD, Muñoz D, Smith SC Jr, Virani SS, Williams KA Sr, Yeboah J, Ziaeian B. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of

Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 2019 Sep 10;140(11):e596–e646. PMID:30879355

7. Orringer CE, Blaha MJ, Blankstein R, Budoff MJ, Goldberg RB, Gill EA, Maki KC, Mehta L, Jacobson TA. The National Lipid Association scientific statement on coronary artery calcium scoring to guide preventive strategies for ASCVD risk reduction. J Clin Lipidol 2021 Jan-Feb;15(1):33–60. PMID:33419719

8. Alexopoulos N, Raggi P. Calcification in atherosclerosis. Nat Rev Cardiol 2009 Nov;6(11):681–688. PMID:19786983

9. Onnis C, Virmani R, Kawai K, Nardi V, Lerman A, Cademartiri F, Scicolone R, Boi A, Congiu T, Faa G, Libby P, Saba L. Coronary Artery Calcification: Current Concepts and Clinical Implications. Circulation 2024 Jan 16;149(3):251–266. PMID:38227718

10. Raff GL, Gallagher MJ, O'Neill WW, Goldstein JA. Diagnostic accuracy of noninvasive coronary angiography using 64-slice spiral computed tomography. J Am Coll Cardiol 2005 Aug 2;46(3):552–557. PMID:16053973

11. Clouse ME. How useful is computed tomography for screening for coronary artery disease? Noninvasive screening for coronary artery disease with computed tomography is useful. Circulation 2006 Jan 3;113(1):125–46; discussion 125–46. PMID:16391167

12. Bergström G, Persson M, Adiels M, Björnson E, Bonander C, Ahlström H, Alfredsson J, Angerås O, Berglund G, Blomberg A, Brandberg J, Börjesson M, Cederlund K, de Faire U, Duvernoy O, Ekblom Ö, Engström G, Engvall JE, Fagman E, Eriksson M, Erlinge D, Fagerberg B, Flinck A, Gonçalves I, Hagström E, Hjelmgren O, Lind L, Lindberg E, Lindqvist P, Ljungberg J, Magnusson M, Mannila M, Markstad H, Mohammad MA, Nystrom FH, Ostenfeld E, Persson A, Rosengren A, Sandström A, Själander A, Sköld MC, Sundström J, Swahn E, Söderberg S, Torén K, Östgren CJ, Jernberg T. Prevalence of Subclinical Coronary Artery Atherosclerosis in the General Population. Circulation 2021 Sep 21;144(12):916–929. PMID:34543072

13. Gatto L, Prati F. Subclinical atherosclerosis: how and when to treat it? Eur Heart J Suppl 2020 Jun;22(Suppl E):E87–E90. PMID:32523447

14. Rumberger JA, Simons DB, Fitzpatrick LA, Sheedy PF, Schwartz RS. Coronary artery calcium area by electron-beam computed tomography and coronary atherosclerotic plaque area. A histopathologic correlative study. Circulation 1995 Oct 15;92(8):2157–2162. PMID:7554196

15. Hecht HS. Coronary artery calcium scanning: past, present, and future. JACC Cardiovasc Imaging 2015 May;8(5):579–596. PMID:25937196

16. Detrano R, Guerci AD, Carr JJ, Bild DE, Burke G, Folsom AR, Liu K, Shea S, Szklo M, Bluemke DA, O'Leary DH, Tracy R, Watson K, Wong ND, Kronmal RA. Coronary calcium as a predictor of coronary events in four racial or ethnic groups. N Engl J Med 2008 Mar 27;358(13):1336–1345. PMID:18367736

17. Blaha M, Budoff MJ, Shaw LJ, Khosa F, Rumberger JA, Berman D, Callister T, Raggi P, Blumenthal RS, Nasir K. Absence of coronary artery calcification and all-cause mortality. JACC Cardiovasc Imaging 2009 Jun;2(6):692–700. PMID:19520338

18. Erbel R, Möhlenkamp S, Moebus S, Schmermund A, Lehmann N, Stang A, Dragano N, Grönemeyer D, Seibel R, Kälsch H, Bröcker-Preuss M, Mann K, Siegrist J, Jöckel K-H, Heinz Nixdorf Recall Study Investigative Group. Coronary risk stratification, discrimination, and reclassification improvement based on quantification of subclinical coronary atherosclerosis: the Heinz Nixdorf Recall study. J Am Coll Cardiol 2010 Oct 19;56(17):1397–1406. PMID:20946997

19. Becker A, Leber AW, Becker C, von Ziegler F, Tittus J, Schroeder I, Steinbeck G, Knez A. Predictive value of coronary calcifications for future cardiac events in asymptomatic patients with diabetes mellitus: a prospective study in 716 patients over 8 years. BMC Cardiovasc Disord 2008 Oct 10;8:27. PMID:18847481

20. Choi E-K, Choi SI, Rivera JJ, Nasir K, Chang S-A, Chun EJ, Kim H-K, Choi D-J, Blumenthal RS, Chang H-J. Coronary computed tomography angiography as a screening tool for the detection of occult coronary artery disease in asymptomatic individuals. J Am Coll Cardiol 2008 Jul 29;52(5):357–365. PMID:18652943

21. Peng AW, Mirbolouk M, Orimoloye OA, Osei AD, Dardari Z, Dzaye O, Budoff MJ, Shaw L, Miedema MD, Rumberger J, Berman DS, Rozanski A, Al-Mallah MH, Nasir K, Blaha MJ. Long-Term All-Cause and Cause-Specific Mortality in Asymptomatic Patients With CAC ≥1,000: Results From the CAC Consortium. JACC Cardiovasc Imaging 2020 Jan;13(1 Pt 1):83–93. PMID:31005541

22. Greenland P, Maron DJ, Budoff MJ. Insurance Payers Should Cover Selective Coronary Artery Calcium Testing in Intermediate Risk Primary Prevention Patients. Circulation 2022 Aug 23;146(8):585–586. PMID:35994564

23. Naghavi M, Maron DJ, Kloner RA, Berman DS, Budoff M, Superko HR, Shah PK. Coronary artery calcium testing: A call for universal coverage. Prev Med Rep Elsevier BV; 2019 Sep;15(100879):100879. PMID:31193256

24. Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, Ferreira MPS, Andersson CR, Macfarlane PW, Meira W Jr, Schön TB, Ribeiro ALP. Automatic diagnosis of

the 12-lead ECG using a deep neural network. Nat Commun 2020 Apr 9;11(1):1760. PMID:32273514

25. Galloway CD, Valys AV, Shreibati JB, Treiman DL, Petterson FL, Gundotra VP, Albert DE, Attia ZI, Carter RE, Asirvatham SJ, Ackerman MJ, Noseworthy PA, Dillon JJ, Friedman PA. Development and Validation of a Deep-Learning Model to Screen for Hyperkalemia From the Electrocardiogram. JAMA Cardiol 2019 May 1;4(5):428–436. PMID:30942845

26. Han C, Song Y, Lim H-S, Tae Y, Jang J-H, Lee BT, Lee Y, Bae W, Yoon D. Automated Detection of Acute Myocardial Infarction Using Asynchronous Electrocardiogram Signals-Preview of Implementing Artificial Intelligence With Multichannel Electrocardiographs Obtained From Smartwatches: Retrospective Study. J Med Internet Res 2021 Sep 10;23(9):e31129. PMID:34505839

27. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nat Rev Cardiol 2021 Jul;18(7):465–478. PMID:33526938

28. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 2019 Jan;25(1):65–69. PMID:30617320

29. Yoon D, Jang J-H, Choi BJ, Kim TY, Han CH. Discovering hidden information in biosignals from patients using artificial intelligence. Korean J Anesthesiol 2020 Aug;73(4):275–284. PMID:31955546

30. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, Carter RE, Yao X, Rabinstein AA, Erickson BJ, Kapa S, Friedman PA. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet 2019 Sep 7;394(10201):861–867. PMID:31378392

31. Noseworthy PA, Attia ZI, Behnken EM, Giblon RE, Bews KA, Liu S, Gosse TA, Linn ZD, Deng Y, Yin J, Gersh BJ, Graff-Radford J, Rabinstein AA, Siontis KC, Friedman PA, Yao X. Artificial intelligence-guided screening for atrial fibrillation using electrocardiogram during sinus rhythm: a prospective non-randomised interventional trial. Lancet 2022 Oct 8;400(10359):1206–1212. PMID:36179758

32. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, Pellikka PA, Enriquez-Sarano M, Noseworthy PA, Munger TM, Asirvatham SJ, Scott CG, Carter RE, Friedman PA. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. Nat Med 2019 Jan;25(1):70–74. PMID:30617318

33. Yao X, Rushlow DR, Inselman JW, McCoy RG, Thacher TD, Behnken EM, Bernard ME, Rosas SL, Akfaly A, Misra A, Molling PE, Krien JS, Foss RM, Barry BA, Siontis KC, Kapa S, Pellikka PA, Lopez-Jimenez F, Attia ZI, Shah ND, Friedman PA, Noseworthy PA. Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. Nat Med 2021 May;27(5):815–819. PMID:33958795

34. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. 2017 IEEE International Conference on Computer Vision (ICCV) IEEE; 2017. doi: 10.1109/iccv.2017.74

35. van de Leur RR, Bos MN, Taha K, Sammani A, Yeung MW, van Duijvenboden S, Lambiase PD, Hassink RJ, van der Harst P, Doevendans PA, Gupta DK, van Es R. Improving explainability of deep neural network-based electrocardiogram interpretation using variational auto-encoders. Eur Heart J Digit Health 2022 Sep;3(3):390–404. PMID:36712164

36. Wouters PC, van de Leur RR, Vessies MB, van Stipdonk AMW, Ghossein MA, Hassink RJ, Doevendans PA, van der Harst P, Maass AH, Prinzen FW, Vernooy K, Meine M, van Es R. Electrocardiogram-based deep learning improves outcome prediction following cardiac resynchronization therapy. Eur Heart J 2023 Feb 21;44(8):680–692. PMID:36342291

37. Farjo PD, Yanamala N, Kagiyama N, Patel HB, Casaclang-Verzosa G, Nezarat N, Budoff MJ, Sengupta PP. Prediction of coronary artery calcium scoring from surface electrocardiogram in atherosclerotic cardiovascular disease: a pilot study. Eur Heart J Digit Health 2020 Nov;1(1):51–61. PMID:37056293

38. Han C, Kang K-W, Kim TY, Uhm J-S, Park J-W, Jung IH, Kim M, Bae S, Lim H-S, Yoon D. Artificial Intelligence-Enabled ECG Algorithm for the Prediction of Coronary Artery Calcification. Front Cardiovasc Med 2022 Apr 6;9:849223. PMID:35463761

39. Awasthi S, Sachdeva N, Gupta Y, Anto AG, Asfahan S, Abbou R, Bade S, Sood S, Hegstrom L, Vellanki N, Alger HM, Babu M, Medina-Inojosa JR, McCully RB, Lerman A, Stampehl M, Barve R, Attia ZI, Friedman PA, Soundararajan V, Lopez-Jimenez F. Identification and risk stratification of coronary disease by artificial intelligence-enabled ECG. EClinicalMedicine 2023 Nov;65:102259. PMID:38106563

40. Shen Y-W, Wu Y-J, Hung Y-C, Hsiao C-C, Chan S-H, Mar G-Y, Wu M-T, Wu F-Z. Natural course of coronary artery calcium progression in Asian population with an initial score of zero. BMC Cardiovasc Disord 2020 May 6;20(1):212. PMID:32375648

41. McEvoy JW, Blaha MJ, Defilippis AP, Budoff MJ, Nasir K, Blumenthal RS, Jones SR. Coronary artery calcium progression: an important clinical measurement? A review of published reports. J Am Coll Cardiol 2010 Nov 9;56(20):1613–1622. PMID:21050970

42. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015 Mar;12(3):e1001779. PMID:25826379

43. Goldberger E. A simple, indifferent, electrocardiographic electrode of zero potential and a technique of obtaining augmented, unipolar, extremity leads. Am Heart J Elsevier BV; 1942 Apr;23(4):483–492.

44. Machado DB, Crow RS, Boland LL, Hannan PJ, Taylor HA Jr, Folsom AR. Electrocardiographic findings and incident coronary heart disease among participants in the Atherosclerosis Risk in Communities (ARIC) study. Am J Cardiol 2006 Apr 15;97(8):1176–1181. PMID:16616022

45. Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional Neural Networks. arXiv; 2019; doi: 10.48550/ARXIV.1905.11946

46. Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw 2018 Oct;106:249–259. PMID:30092410

47. Mohammed A, Kora R. A comprehensive review on ensemble deep learning: Opportunities and challenges. J King Saud Univ - Comput Inf Sci Elsevier BV; 2023 Feb;35(2):757–774.

48. Ganaie MA, Hu M, Malik AK, Tanveer M, Suganthan PN. Ensemble deep learning: A review. Eng Appl Artif Intell Elsevier BV; 2022 Oct;115(105151):105151.

49. Kingma DP, Welling M. Auto-Encoding Variational Bayes. arXiv; 2013; doi: 10.48550/ARXIV.1312.6114

50. Budoff MJ, Shaw LJ, Liu ST, Weinstein SR, Mosler TP, Tseng PH, Flores FR, Callister TQ, Raggi P, Berman DS. Long-term prognosis associated with coronary calcification: observations from a registry of 25,253 patients. J Am Coll Cardiol 2007 May 8;49(18):1860–1870. PMID:17481445

51. Budoff MJ, Mayrhofer T, Ferencik M, Bittner D, Lee KL, Lu MT, Coles A, Jang J, Krishnam M, Douglas PS, Hoffmann U, PROMISE Investigators. Prognostic Value of Coronary Artery Calcium in the PROMISE Study (Prospective Multicenter Imaging Study for Evaluation of Chest Pain). Circulation 2017 Nov 21;136(21):1993–2005. PMID:28847895

52. Ho JS, Fitzgerald SJ, Stolfus LL, Wade WA, Reinhardt DB, Barlow CE, Cannaday JJ. Relation of a coronary artery calcium score higher than 400 to coronary stenoses detected using multidetector computed tomography and to traditional cardiovascular risk factors. Am J Cardiol

2008 May 15;101(10):1444–1447. PMID:18471456

53. Kawai K, Finn AV, Virmani R, Subclinical Atherosclerosis Collaborative. Subclinical Atherosclerosis: Part 1: What Is it? Can it Be Defined at the Histological Level? Arterioscler Thromb Vasc Biol 2024 Jan;44(1):12–23. PMID:38150517

54. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. arXiv; 2017; doi: 10.48550/ARXIV.1705.07874

55. Bosco E, Hsueh L, McConeghy KW, Gravenstein S, Saade E. Major adverse cardiovascular event definitions used in observational analysis of administrative databases: a systematic review. BMC Med Res Methodol 2021 Nov 6;21(1):241. PMID:34742250

56. 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 Jan 30;27(2):157–72; discussion 207–12. PMID:17569110

57. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics JSTOR; 1988 Sep;44(3):837–845. PMID:3203132

58. DiCiccio TJ, Efron B. Bootstrap confidence intervals. Stat Sci Institute of Mathematical Statistics; 1996 Sep 1;11(3):189–228.

59. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med 2015 Jan 6;13:1. PMID:25563062

60. Castro-Diehl C, Song RJ, Mitchell GF, McManus D, Cheng S, Vasan RS, Xanthakis V. Association of subclinical atherosclerosis with echocardiographic indices of cardiac remodeling: The Framingham Study. PLoS One 2020 May 15;15(5):e0233321. PMID:32413074

61. Yared GS, Moreira HT, Ambale-Venkatesh B, Vasconcellos HD, Nwabuo CC, Ostovaneh MR, Reis JP, Lloyd-Jones DM, Schreiner PJ, Lewis CE, Sidney S, Carr JJ, Gidding SS, Lima JAC. Coronary Artery Calcium From Early Adulthood to Middle Age and Left Ventricular Structure and Function. Circ Cardiovasc Imaging 2019 Jun;12(6):e009228. PMID:31195818

62. Desai CS, Ning H, Soliman EZ, Burke GL, Shea S, Nazarian S, Lloyd-Jones DM, Greenland P. Electrocardiographic abnormalities and coronary artery calcium for coronary heart disease prediction and reclassification: the Multi-Ethnic Study of Atherosclerosis (MESA). Am Heart J 2014 Sep;168(3):391–397. PMID:25173552

63. Lloyd-Jones DM, Walsh JA, Prineas RJ, Ning H, Liu K, Daviglus ML, Shea S, Detrano RC, Tandri H, Greenland P. Association of electrocardiographic abnormalities with coronary artery

calcium and carotid artery intima-media thickness in individuals without clinical coronary heart disease (from the Multi-Ethnic Study of Atherosclerosis [MESA]). Am J Cardiol 2009 Oct 15;104(8):1086–1091. PMID:19801030

64. Tat E, Bhatt DL, Rabbat MG. Addressing bias: artificial intelligence in cardiovascular medicine. Lancet Digit Health 2020 Dec;2(12):e635–e636. PMID:33328028

65. Zhang Q, Yang LT, Chen Z, Li P. A survey on deep learning for big data. Inf Fusion Elsevier BV; 2018 Jul;42:146–157.

66. Ying X. An Overview of Overfitting and its Solutions. J Phys Conf Ser IOP Publishing; 2019 Feb;1168:022022.

67. Cook NR, Ridker PM. Calibration of the Pooled Cohort Equations for Atherosclerotic Cardiovascular Disease: An Update. Ann Intern Med 2016 Dec 6;165(11):786–794. PMID:27723890

68. Damen JA, Pajouheshnia R, Heus P, Moons KGM, Reitsma JB, Scholten RJPM, Hooft L, Debray TPA. Performance of the Framingham risk models and pooled cohort equations for predicting 10-year risk of cardiovascular disease: a systematic review and meta-analysis. BMC Med 2019 Jun 13;17(1):109. PMID:31189462

69. Jung KJ, Jang Y, Oh DJ, Oh B-H, Lee SH, Park S-W, Seung K-B, Kim H-K, Yun YD, Choi SH, Sung J, Lee T-Y, Kim SH, Koh SB, Kim MC, Chang Kim H, Kimm H, Nam C, Park S, Jee SH. The ACC/AHA 2013 pooled cohort equations compared to a Korean Risk Prediction Model for atherosclerotic cardiovascular disease. Atherosclerosis 2015 Sep;242(1):367–375. PMID:26255683

70. Cho YG. Cardiovascular Risk Prediction in Korean Adults. Korean J Fam Med 2018 May;39(3):135–136. PMID:29788700

ABSTRACT IN KOREAN

관상동맥 석회화 예측을 위한 설명 가능한 인공지능 기반 심전도 모델

관상동맥 석회화(coronary artery calcium, CAC) 점수는 관상동맥에 칼슘이 침착된 정도를 양적으로 평가하기 위해 컴퓨터 단층촬영(computed tomograph, CT)으로 측정되며, 죽상동맥경화증의 지표이자 관상동맥 사건의 강력한 예측 인자이다. 현재의 심혈관 질환 위험 예측 모델, 예를 들어 ACC/AHA Pooled Cohort Equations (PCE) 등은 일차 예방을 위한 지침을 제공하지만, 종종 경계선 위험 분류를 초래하여 의사 결정이 불확실해지는 경우가 많다. 이러한 경우, CAC 점수는 의사 결정을 안내하기 위한 추가 도구로 지침에서 권장된다. CAC가 존재할 경우 일차 예방 조치(예: 스타틴 요법) 시작이 권장되는 반면, CAC가 없을 경우 스타틴 사용이 보류될 수 있다.

그러나 CAC 점수 측정은 비용, 방사선 노출, 보험 적용 부족으로 인해 일상적인 사용이 제한된다. 반면 심전도(electrocardiogram, ECG)는 널리 사용되며, 비침습적이고, 비용 효과적이며 방사선 노출이 없다. 딥 컨볼루션 신경망의 발전은 심전도를 통해 이전에는 탐지할 수 없었던 상태를 탐지할 수 있는 인공지능(artificial intelligence, AI) 모델을 가능하게 했다. CAC를 예측할 수 있는 AI-ECG는 심혈관 질환 위험에 대한 통찰력을 제공할 수 있다. 특히 심전도가 널리 수행되는 건강 검진에서 이러한 접근법은 일반 인구에서 기회적인 CAC 탐지를 가능하게 하여 관상동맥 석회화를 조기에 발견하고 일차 예방을 적시에 실행하게 할 수 있다.

본 연구에서는 CAC를 예측하기 위한 AI-ECG 모델을 개발하고, 건강 검진 환경에서의 기회적 선별 가능성을 검증하고자 하였다. 더 넓은 적용 가능성을 보장하기 위해 두 개의 별도 기관에서 건강 검진 데이터를 이용하여 외부 검증을 수행하고자 했다. 또한, 본 연구는 두 개국에 걸친 다국적 후향적 코호트 분석을 통해 AI-ECG 모델의 임상적 함의와 잠재적 영향을 평가하고자 하였다. 마지막으로, 모델 예측에 영향을 미치는 심전도 특징의 시각적 형태적 설명을 제공하고자 하였다.

AI-ECG 모델은 연세대학교 세브란스병원에서 CAC 점수로 라벨링된 194,000개 이상의 심전도를 사용해 훈련되었다. 이 모델은 동일 방문에서 심전도와 CT 기반 CAC 측정을 모두 수행한 14,242개의 세브란스병원 건강 검진 데이터를 통해 테스트되었다. 외부 검증은 용인세브란스병원(729개의 심전도)과 아주대학교병원(2,056개의 심전도) 데이터를 사용하여 수행되었다. 다국적 후향적 코호트 분석에서는 세브란스병원 건강 검진에서 52,400개의 심전도와 United Kingdom Biobank (UKB)에서 수집된 30,623개의 심전도를 활용하였다. 해석 가능성을 높이기 위해 500만 개 이상의 심전도로 사전 훈련된 변분 오토인코더를 사용하여 예측에

영향을 미치는 심전도 특징을 시각적으로 설명하였다.

AI-ECG 모델은 CAC를 예측하는 데 있어 강력한 성능을 보여주었으며, 건강 검진 테스트 데이터셋에서 CAC 점수 ≥ 400의 경우 AUROC 0.841, CAC 점수 > 0의 경우 AUROC 0.720을 기록하였다. 외부 검증에서도 용인세브란스병원과 아주대학교병원 데이터셋에서 각각 CAC 점수 ≥ 400의 AUROC 0.784와 0.814, CAC 점수 > 0의 AUROC 0.691과 0.701로 강력한 성능을 입증하였다.

AI-ECG 모델은 PCE 저위험군에서 CAC 가능성이 가장 높은 개인과 PCE 중간 위험군에서 CAC 가능성이 가장 낮은 개인을 선별할 수 있었다: PCE 저위험군에서 24.9%가 CAC > 0을 나타냈으며, AI-ECG 모델에 의해 고위험으로 식별된 경우 이 비율은 45.7%로 증가했다. PCE 중간 위험군에서 35.4%가 CACS = 0을 나타냈으며, AI-ECG 모델에 의해 저위험으로 식별된 경우 이 비율은 60.9%로 증가했다; PCE 저위험군에서 AI-ECG에 의해 고위험으로 재분류된 그룹은 PCE 중간 위험군에서 AI-ECG에 의해 저위험으로 재분류된 그룹보다 주요 심혈관 사건 발생률이 더 높았다(세브란스병원 코호트 분석 데이터셋에서 1000 인년당 주요 심혈관 사건 발생률: 6.0 vs. 3.3, P = 0.007, UKB 코호트 분석 데이터셋에서 1000 인년당 주요 심혈관 사건 발생률: 8.3 vs. 7.0, P = 0.360). 따라서, PCE 중간 위험군으로 분류되었지만 AI-ECG에서 저위험으로 분류된 개인에게는 스타틴을 보류하는 것이 더 합리적이며, PCE 저위험군으로 분류되었지만 AI-ECG에서 고위험으로 분류된 개인에게는 스타틴을 시작하는 것이 더 합리적일 것이다. AI-ECG는 주요 심혈관 사건의 독립적 위험 요인으로 확인되었다(조정된 위험비 [95% 신뢰구간]: 세브란스병원 코호트 분석 데이터셋에서 1.087 [1.053–1.123], UKB 코호트 분석 데이터셋에서 1.117 [1.061–1.175]). PCE와 AI-ECG 점수를 결합한 경우, 단독 PCE보다 높은 C-index를 보였다. 증가된 예측 위험과 관련된 심전도 요인을 시각적으로 해석했으며, upward shift of the ST segment in the anteroseptal leads with reciprocal downward shift in the inferolateral leads, downward shift of the ST segment in all leads, longer PR interval 등과 같은 변화를 CAC와 연관지었다.

AI-ECG 모델은 관상동맥 석회화를 예측하는 데 효과적인 도구임을 입증하였다. 본 연구는 AI-ECG 모델이 이중적 유틸리티를 통해 임상 워크플로에 통합될 가능성을 보여주었다. 즉, CAC 점수 측정이 가장 필요한 개인을 선별하거나, 환자 재분류를 통해 스타틴 치료 시작 또는 보류 결정을 직접 안내할 수 있음을 입증하였다. 심전도가 널리 사용된다는 점과 AI-ECG 모델이 심혈관 사건의 독립적 위험 요인으로 확인된 결과를 바탕으로, 본 모델은 심혈관 질환 위험 예측 도구로 통합될 잠재력을 가지고 있다. 특히 심전도가 일반적으로 수행되는 정기 건강 검진에서, AI-ECG 기반 CAC 예측은 일반 인구에서 기회적 CAC 검사를 가능하게 하여 조기 발견과 초기 예방 전략의 적시 실행으로 이어질 수 있다.

핵심되는 말: 관상동맥 석회화, 관상동맥 석회화 점수, 인공지능, 심전도, 인공지능 기반 심전도, 건강검진, 일차 예방

PUBLICATIONS

*First author(s)

1. Kim, D. W., Seo, J., Kwon, S., Park, C. M., Han, C., Kim, Y., ... & Kim, K. M. (accepted, 2024). Predicting In-Hospital Fall Risk Using Machine Learning with Real-Time Location System and Electronic Medical Records. *Journal of Cachexia, Sarcopenia and Muscle*.

2. Kim, S., Kim, D., Shin, H. J., Lee, S. H., Kang, Y., ... Han, C., Yoon, D. (accepted, 2024). Large-Scale Validation of the Feasibility of GPT-4 as a Proofreading Tool for Head CT Reports. *Radiology*.

3. **Han, C.***, Jung, Y. J.*, Park, J. E., Chung, W. Y., & Yoon, D. (accepted, 2024). Artificial Intelligence-Based Early Prediction of Acute Respiratory Failure in the Emergency Department Using Biosignal and Clinical Data. *Yonsei Medical Journal*.

4. **Han, C.***, Kim, H. I.*, Soh, S., Choi, J. W., Song, J. W., & Yoon, D. (2024). Machine learning with clinical and intraoperative biosignal data for predicting postoperative delirium after cardiac surgery. *Iscience*, 27(6).

5. **Han, C.***, Soh, S., Kim, H. I., Song, J. W., & Yoon, D. (2024). Machine Learning with Clinical and Intraoperative Biosignal Data for Predicting Cardiac Surgery-Associated Acute Kidney Injury. *Studies in health technology and informatics*, 316, 286-290.

6. Yoon, D., Han, C., Kim, D. W., Kim, S., Bae, S., Ryu, J. A., & Choi, Y. (2024). Redefining Health Care Data Interoperability: Empirical Exploration of Large Language Models in Information Exchange. *Journal of Medical Internet Research*, 26, e56614.

7. **Han, C.***, Kim, D. W.*, Kim, S.*, You, S. C., Park, J. Y., Bae, S., & Yoon, D. (2024). Evaluation of GPT-4 for 10-year cardiovascular risk prediction: Insights from the UK Biobank and KoGES data. *Iscience*, 27(2).

8. **Han, C.***, & Yoon, D. (2024). An Explainable Artificial Intelligence-enabled ECG Framework for the Prediction of Subclinical Coronary Atherosclerosis. *AMIA Summits on Translational Science Proceedings*, 2024, 535.

9. Lee, D.*, **Han, C.***, Kim, H., Uhm, J. S., Yoon, D., & Park, J. Y. (2023). Changes in the Circadian Rhythm of High-Frequency Heart Rate Variability Associated With Depression. *Journal of Korean Medical Science*, 38(19).

10. **Han, C.***, Park, C. M., Kim, Y., Kang, S., Park, T. J., & Yoon, D. (2022). Approach for Electronic Medical Record Data Analysis. *Journal of Health Informatics and Statistics*, 47(Suppl 1), S1-S8.

11. **Han, C.***, Kwon, O.*, Chang, M., Joo, S., Lee, Y., Lee, J. S., ... & Yoon, D. (2022). Evaluating the risk of paroxysmal atrial fibrillation in noncardioembolic ischemic stroke using artificial intelligence-enabled ECG algorithm. *Frontiers in Cardiovascular Medicine*, 9, 865852.

12. **Han, C.***, Kang, K. W.*, Kim, T. Y., Uhm, J. S., Park, J. W., Jung, I. H., ... & Yoon, D. (2022). Artificial intelligence-enabled ECG algorithm for the prediction of coronary artery calcification. *Frontiers in Cardiovascular Medicine*, 9, 849223.

13. **Han, C.***, Song, Y.*, Lim, H. S.*, Tae, Y., Jang, J. H., Lee, B. T., ... & Yoon, D. (2021). Automated detection of acute myocardial infarction using asynchronous electrocardiogram signals—preview of implementing artificial intelligence with multichannel electrocardiographs obtained from Smartwatches: retrospective study. *Journal of Medical Internet Research*, 23(9), e31129.

14. Park, J. E., Kim, T. Y., Jung, Y. J., Han, C., Park, C. M., Park, J. H., ... & Chung, W. Y. (2021). Biosignal-based digital biomarkers for prediction of ventilator weaning success. *International Journal of Environmental Research and Public Health*, 18(17), 9229.

15. Yoon, D., Jang, J. H., Choi, B. J., Kim, T. Y., & Han, C. H. (2020). Discovering hidden information in biosignals from patients using artificial intelligence. *Korean journal of anesthesiology*, 73(4), 275.

CONFERENCE PRESENTATIONS

1. **Han, C. (oral presenter)**, “AI-ECG-Afib as a Predictor of New-Onset Postoperative Atrial Fibrillation Following Cardiac Surgery”, Fall Conference of the Korean Society of Medical Informatics (KOSMI) 2024, November 2024.
2. **Han, C. (oral presenter)**, “Machine Learning with Clinical and Intraoperative Biosignal Data for Predicting Cardiac Surgery-Associated Acute Kidney Injury”, Medical Informatics Europe Conference 2024 (MIE2024), August 2024.
3. **Han, C. (oral presenter)**, “Redefining Clinical Prediction Models with LLMs: Data-free Clinical Prediction Model Development Using LLM Knowledge Extraction”, Spring Conference of the Korean Society of Medical Informatics (KOSMI) 2024, June 2024.
4. **Han, C. (oral presenter)**, “Predicting cardiac surgery-associated acute kidney injury using machine learning with clinical and intraoperative biosignal data”, Korean Society of Cardiothoracic and Vascular Anesthesiologists (KSCTVA) Conference 2024, May 2024.
5. **Han, C. (oral presenter)**, “An Explainable Artificial Intelligence-enabled ECG Framework for the Prediction of Subclinical Coronary Atherosclerosis”, American Medical Informatics Association (AMIA) Informatics Summit 2024, March 2024.
6. **Han, C. (oral presenter)**, “Evaluating risk of paroxysmal atrial fibrillation presence in non-cardioembolic ischemic stroke using artificial intelligence-enabled ECG algorithm”, European Society of Cardiology (ESC) Heart & Stroke Conference 2021, June 2021.
7. **Han, C. (oral presenter)**, “Development and Validation of Machine Learning Models for the Prediction of Delirium after Cardiac Surgery”, Fall Conference of the Korean Society of Medical Informatics (KOSMI) 2021, November 2021.
8. **Han, C. (oral presenter)**, “Artificial Intelligence-enabled ECG algorithm for Early Detection of Coronary Artery Atherosclerosis”, Fall Conference of the Korean Society of Medical Informatics (KOSMI) 2020, November 2020.
9. **Han, C. (oral presenter)**, “Biosignal Analysis for the Development of an Artificial Neural Network Model to Predict Successful Extubation in Intensive Care Units”, Fall Conference of the Korean Society of Medical Informatics (KOSMI) 2019, November 2019.

AWARDS AND HONORS

1. Best paper and presenter, Fall conference, Korean Society of Medical Informatics (KOSMI), November 2024.
2. Best researcher, Research fair, Department of Biomedical Systems Informatics, Yonsei University College of Medicine, December 2022.
3. Best paper and presenter, Fall conference, Korean Society of Medical Informatics (KOSMI), November 2021.
4. Best paper and presenter, Fall conference, Korean Society of Medical Informatics (KOSMI), November 2020.

PATENTS

1. Dukyong Yoon, Songsoo Kim, Jee An Ryu, Yujin Choi, **Changho Han**, “APPARATUS AND METHOD FOR EVALUATING THE QUALITY OF MEDICAL DATA BASED ON FIDELITY”, KR patent, application number 10-2024-0094441, filed Jul. 17, 2024, awaiting registration decision.
2. Dukyong Yoon, **Changho Han**, SungA Bae, Songsoo Kim, Dong Won Kim, “METHOD FOR PROVIDING INFORMATION ON THE RISK OF DEVELOPING CARDIOVASCULAR DISEASES AND DEVICE USING THE SAME”, KR patent, application number 10-2024-0091928, filed Jul. 11, 2024, awaiting registration decision.
3. Dukyong Yoon, **Changho Han**, Songsoo Kim, Jee An Ryu, Yujin Choi, “APPARATUS AND METHOD FOR EVALUATING THE QUALITY OF MEDICAL DATA BASED ON RELIABILTY”, KR patent, application number 10-2024-0089258, filed Jul. 5, 2024, awaiting registration decision.
4. Dukyong Yoon, Yujin Choi, Jee An Ryu, **Changho Han**, Songsoo Kim, “APPARATUS AND METHOD FOR EVALUATING THE QUALITY OF MEDICAL DATA BASED ON DATA CONSISTENCY”, KR patent, application number 10-2024-0050332, filed Apr. 15, 2024, awaiting registration decision.
5. Dukyong Yoon, Jee An Ryu, **Changho Han**, Songsoo Kim, Yujin Choi, “APPARATUS AND METHOD FOR EVALUATING THE QUALITY OF MEDICAL DATA BASED ON PRECEDENCE AND PRECEDENCE RELATIONSHIPS”, KR patent, application number 10-2024-0046975, filed Apr. 5, 2024, awaiting registration decision.
6. Dukyong Yoon, Jihoon Seo, Chan Min Park, **Changho Han**, “METHOD FOR PROVIDING INFORMATION HYPOTENSION AND DEVICE USING THE SAME”, KR patent, application number 10-2024-0012559, filed Jan, 26, 2024, awaiting registration decision.
7. Dukyong Yoon, Chan Min Park, **Changho Han**, Soo Kyeong Jang, Shin Beom Hong, “COMPUTING DEVICES AND METHOD THAT ANALYZE THE LIKELIHOOD OF DELIRIUM”, KR patent, application number 10-2022-0140183, filed Oct. 27, 2022, published May. 7, 2024.