Deep learning-based long-term risk evaluation of incident type 2 diabetes using electrocardiogram in a non-diabetic population: a retrospective, multicentre study

Junmo Kim,[a] Hyun-Lim Yang,[b,c] Su Hwan Kim,[b,k] Siun Kim,[b] Jisoo Lee,[a] Jiwon Ryu,[d,e] Kwangsoo Kim,[f,g] Zio Kim,[a] Gun Ahn,[a] Doyun Kwon,[h] and Hyung-Jin Yoon[a,i,j,*]

[a]Interdisciplinary Program in Bioengineering, Seoul National University, Seoul, Republic of Korea
[b]Biomedical Research Institute, Seoul National University Hospital, Seoul, Republic of Korea
[c]Department of Anesthesiology and Pain Medicine, Seoul National University Hospital, Seoul, Republic of Korea
[d]Division of General Internal Medicine, Department of Internal Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Republic of Korea
[e]Hospital Medicine Center, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
[f]Department of Transdisciplinary Medicine, Institute of Convergence Medicine with Innovative Technology, Seoul National University Hospital, Seoul, Republic of Korea
[g]Department of Medicine, College of Medicine, Seoul National University, Seoul, Republic of Korea
[h]Interdisciplinary Program of Medical Informatics, Seoul National University College of Medicine, Seoul, Republic of Korea
[i]Medical Bigdata Research Center, Seoul National University College of Medicine, Seoul, Republic of Korea
[j]Department of Biomedical Engineering, Seoul National University College of Medicine, Seoul, Republic of Korea
[k]Department of Information Statistics, Gyeongsang National University, Jinju, Gyeongsangnam-do, Republic of Korea

eClinicalMedicine
2024;68: 102445
Published Online xxx
https://doi.org/10.1016/j.eclinm.2024.102445

Summary

Background Diabetes is a major public health concern. We aimed to evaluate the long-term risk of incident type 2 diabetes in a non-diabetic population using a deep learning model (DLM) detecting prevalent type 2 diabetes using electrocardiogram (ECG).

Methods In this retrospective study, participants who underwent health checkups at two tertiary hospitals in Seoul, South Korea, between Jan 1, 2001 and Dec 31, 2022 were included. Type 2 diabetes was defined as glucose ≥126 mg/dL or glycated haemoglobin (HbA1c) ≥ 6.5%. For survival analysis on incident type 2 diabetes, we introduced an additional variable, diabetic ECG, which is determined by the DLM trained on ECG and corresponding prevalent diabetes. It was assumed that non-diabetic individuals with diabetic ECG had a higher risk of incident type 2 diabetes than those with non-diabetic ECG. The one-dimensional ResNet-based model was adopted for the DLM, and the Guided Grad-CAM was used to localise important regions of ECG. We divided the non-diabetic group into the diabetic ECG group (false positive) and the non-diabetic ECG (true negative) group according to the DLM decision, and performed a Cox proportional hazard model, considering the occurrence of type 2 diabetes more than six months after the visit.

Findings 190,581 individuals were included in the study with a median follow-up period of 11.84 years. The areas under the receiver operating characteristic curve for prevalent type 2 diabetes detection were 0.816 (0.807–0.825) and 0.762 (0.754–0.770) for the internal and external validations, respectively. The model primarily focused on the QRS duration and, occasionally, P or T waves. The diabetic ECG group exhibited an increased risk of incident type 2 diabetes compared with the non-diabetic ECG group, with hazard ratios of 2.15 (1.82–2.53) and 1.92 (1.74–2.11) for internal and external validation, respectively.

Interpretation In the non-diabetic group, those whose ECG was classified as diabetes by the DLM were at a higher risk of incident type 2 diabetes than those whose ECG was not. Additional clinical research on the relationship between the phenotype of ECG and diabetes to support the results and further investigation with tracked data and various ECG recording systems are suggested for future works.

Funding National Research Foundation of Korea.

*Corresponding author. Medical Bigdata Research Center, Seoul National University College of Medicine, Seoul, Republic of Korea.
E-mail address: hjyoon@snu.ac.kr (H.-J. Yoon).

Keywords: Deep learning; Type 2 diabetes; Electrocardiogram; Risk evaluation

Research in context

Evidence before this study

We searched PubMed, MEDLINE, medRxiv, and arXiv with the terms 'diabetes', 'electrocardiogram', 'machine learning', 'deep learning', and 'survival analysis' for relevant peer-reviewed publications up to 12 August 2023. We included and reviewed reports written in English with the full text available and found several studies that detected type 2 diabetes using electrocardiogram (ECG) with machine learning methods. None of the studies restricted the study population to the general public, and the exclusion criteria for cohort definition did not include information from the health questionnaire. In addition, previous studies were not externally validated with a multicentre population and did not verify the decision of their model by survival analysis of incident type 2 diabetes.

Added value of this study

To the best of our knowledge, this is the first study to determine the long-term risk of incident type 2 diabetes in a non-diabetic population using ECG from two locally separated tertiary hospitals. 190,581 participants who underwent health checkups at two separate hospitals in the Republic of Korea were included in this study. As we strictly excluded participants with previous diagnoses or prescriptions for diabetes using electronic medical records and a health questionnaire, the population of this study represents the general public without any concerns about diabetes. We used residual convolutional neural networks for model development and a Cox proportional hazard model for long-term risk evaluation of incident type 2 diabetes. Individuals without diabetes whose ECG was diagnosed with type 2 diabetes by the model exhibited an increased risk of type 2 diabetes compared with those with a normal ECG.

Implications of all the available evidence

This study demonstrated that a deep learning model trained using ECG data is able to detect type 2 diabetes in the general population and that the risk of incident type 2 diabetes differs according to the decision of the model. Based on this, in theory, individuals without previous diabetes concerns can perceive their potential type 2 diabetes risk through an ECG-based diagnosis and can be offered additional tests or regular checkups. This approach could be used to contribute to the early prevention of type 2 diabetes and the reduction in the associated medical costs. More work is needed. Clinical research on the relationship between the phenotype of ECG and diabetes to explain the model and further investigation with tracked data and various ECG recording systems are suggested as future works.

Introduction

Diabetes is one of the most critical worldwide public health issues, reducing life expectancy and causing various complications—such as coronary heart disease, stroke, peripheral arterial disease, kidney disease, and retinopathy.[1,2]

It is estimated that 578 million people will have diabetes in 2030, and among all types of diabetes, type 2 diabetes accounts for an overwhelming majority of adult diabetics.[3,4] Type 2 diabetes encompasses a range of metabolic conditions associated with hyperglycaemia and is caused by insulin resistance.[5] Because type 2 diabetes usually develops in adulthood and is affected by lifestyle, the potential benefits of its early detection and treatment have been discussed. The U.S. Preventive Service Task Force indicated that changing lifestyle can help people with prediabetes prevent progression to type 2 diabetes.[6] In addition, the incidence of 10-year cardiovascular disease and all-cause mortality increased as the diagnosis and treatment of type 2 diabetes were delayed.[7]

Cardiovascular disease is one of the most common complications of type 2 diabetes, as over one-third of patients with diabetes also go on to establish cardiovascular disease.[8] Electrocardiogram (ECG) is an essential test performed non-invasively during a medical check and contains considerable information about cardiac electrical activities and potential cardiovascular diseases. According to the close relation between type 2 diabetes and cardiovascular disease, ECG and machine learning methods have been utilised to detect prevalent type 2 diabetes in several studies. Lin et al. proposed ECG-based glycated haemoglobin (HbA1c) using a one-dimensional residual module, detecting prevalent diabetes with an area under the receiver operating characteristic curve (AUROC) of 0.826.[9] There were other single-centre studies sharing the same task with AUROCs over 0.9, but the number of participants was fewer than 1200.[10,11]

Although detecting prevalent type 2 diabetes using ECG and machine learning algorithms is effective, a multicentre general population study has never been conducted. Furthermore, because most studies performed to date have included all types of patients (such as inpatients, outpatients, emergency patients, and people undergoing medical checkups), it was difficult to guarantee that the results from these previous studies will apply to the general public. Most of all, no studies

have addressed the risk of incident type 2 diabetes, only prevalent type 2 diabetes.

In this study, we aimed to assess the long-term risk of incident type 2 diabetes in a non-diabetic general population using ECG independently of glucose and HbA1c from two locally separated tertiary hospitals over a median 11-year follow-up. Individuals who visited hospitals for medical checkups were included in our study population. For survival analysis on incident type 2 diabetes in a non-diabetic population, we introduced an additional variable, diabetic ECG, which is determined by a deep learning model (DLM) trained on ECG and corresponding prevalent diabetes. In addition, we adjusted the survival model by patient information, comorbidities, and laboratory values, including fasting serum glucose and HbA1c, for the further demonstration of the diabetic ECG. To the best of our knowledge, this is the first study to evaluate incident type 2 diabetes using ECG and DLM and to externally validate the model in a multicentre general population with more than 190,000 participants.

Methods

Study design, participants, and main outcomes

This study used data from the Seoul National University Hospital Healthcare System Gangnam Center (SNUH-HSGC) and Seoul National University Hospital Health Promotion Center (SNUH-HPC). Accordingly, individuals who visited hospitals for general health checkups were included. In the Republic of Korea, the national health screening program, which is organised by the National Health Insurance Service, is offered free of charge and biennially for citizens. Thus, visiting hospitals for health checkups is commonplace in the Republic of Korea, and the participants might have visited hospitals routinely. Although ECG and HbA1c are not on the list of mandatory tests, many people include several additional tests, such as ECG. We identified individuals with at least one 10-s, 500-Hz, and 12-lead ECG and blood tests, including glucose and HbA1c, during the first health checkup at SNUH-HSGC between 1 January 2004 and 31 December 2022 and at SNUH-HPC between 1 January 2001 and 31 December 2022. All participants had ECG, glucose, and HbA1c tests on the same day; for each person, we used only the first record for both training and validation.

We excluded individuals under 18 years of age and those with a previous diagnosis of diabetes or antidiabetic drug prescriptions using a health questionnaire. Any history of diabetes can cause individual regulation in eating or exercise habits and can intervene the model training as a potential confounding factor. Questionnaire data were computerised for each question and answer. Questions regarding diabetes were identified to examine previous diagnoses or prescriptions.

We collected data, including ECG, age, sex, glucose and HbA1c levels, and comorbidity records. ECGs were collected using the GE MUSE ECG system (GE Healthcare, Chicago, IL, USA), and the remaining items were obtained from the clinical data warehouse (CDW). Comorbidity records were obtained using the Charlson Comorbidity Index and the International Classification of Disease codes (ICD-10).[12]

Data from SNUH-HSGC were used for model development and internal validation, and data from SNUH-HPC were used for external validation. Note that even though SNUH-HSGC and SNUH-HPC are in the same hospital group, they are approximately seven miles apart. In addition, most visitors of SNUH-HSGC are working people near Gangnam-gu, Seoul, whereas those of SNUH-HPC are everyone in the Republic of Korea. Accordingly, SNUH-HSGC has a relatively high proportion of males, a low age group, and a low rate of comorbidities compared with SNUH-HPC.

Type 2 diabetes was defined as cases with fasting serum glucose values of 126 mg/dL or above or HbA1c values of 6.5% or above.[13] We introduced an additional variable, diabetic ECG, which is determined by a DLM trained on ECG and corresponding prevalent type 2 diabetes, for survival analysis on incident type 2 diabetes in a non-diabetic population. A DLM was trained to classify prevalent type 2 diabetes using ECG signals, age, and sex as inputs. Using the trained model, we divided the non-diabetic group into the diabetic ECG group (false positive) and the non-diabetic ECG group (true negative) according to the DLM decision.

The Institutional Review Board (IRB) of Seoul National University Hospital (IRB approval No. 2204-001-1310) approved the study with a waiver of informed consent, considering that our study used retrospective and observational electronic medical records and ECG data. The approval aligns with the principles outlined in the Declaration of Helsinki, the Korean Bioethics and Safety Act (Law No. 16372), and the Human Research Protection Program–Standard Operating Procedure of Seoul National University Hospital.

Data preprocessing

For all leads of ECG, we applied a fifth-order Butterworth filter with a frequency range between 0.05 and 150 Hz according to the American Heart Association standards and scaled the filtered signal to range between −1 and 1. The dimensions of the ECG were 12×5000, and the leads were arranged in the following order: I, II, V1–V6, III, aVR, aVL, and aVF. Considering a scenario using smartwatch, we conducted experiments with 1- and 2-lead ECGs in addition to 12-lead ECG. For 2-lead ECG, we used leads I and II because they can be measured by smartwatch users themselves; lead I can be measured by wearing a smartwatch on the left hand and placing the right finger on it, and lead II can be measured by wearing a smartwatch on the left ankle and

the right finger on it. For 1-lead ECG, we used lead II which is commonly used for basic cardiac monitoring.[14,15] For glucose and HbA1c values, we removed outliers outside the range [Q1—5 IQR, Q3 + 5 IQR], where Q1 and Q3 are the first and third quartiles, respectively, and the IQR is Q3—Q1. Data from SNUH-HSGC were randomly split into development (60% for training and 20% for validation) and internal validation (20%) datasets. After removing outliers, the means and standard deviations of glucose and HbA1c were calculated from development datasets, and those values were used for standard normalisation of both development and internal validation datasets. For external validation datasets, the outlier removal and the normalisation were performed in the same way independently.

Long-term risk evaluation

This study aimed to evaluate the long-term risk of incident type 2 diabetes in a non-diabetic population. For survival analysis on incident type 2 diabetes, we introduced an additional variable, diabetic ECG, determined by a DLM trained on ECG and corresponding prevalent diabetes. The washout period of six months was applied to prevent potential data contamination caused by a mixture of prevalent and incident cases. The excluded cases were assumed to be offered limited benefits because they were considered to have already had diabetes risk from the visit. In addition, those prevalent cases may cause contamination to risk evaluation of incident cases. Individuals without diabetes were divided into two groups and compared: the diabetic ECG group and the non-diabetic ECG group. We fitted a Cox proportional hazard (CPH) model, regressing the occurrence of type 2 diabetes on the DLM decision. Three CPH models were created, the first of which was adjusted for age and sex. Subsequently, comorbidities and laboratory values for glucose and HbA1c were cumulatively added for additional adjustments in the second and last models. Comorbidities included malignant tumours, peptic ulcer disease, cerebrovascular disease, connective tissue disease, myocardial infarction, pulmonary disease, and liver disease. The rest are excluded because of low incidence ratios. To fit the CPH model using more normal individuals, we redefined the normal ranges of glucose and HbA1c levels (glucose < 110 mg/dL and HbA1c < 5.6%) to exclude borderline cases.

Deep learning training

Our DLM consists of three layers: an ECG waveform processing layer based on a one-dimensional ResNet, a patient information processing layer that uses a multilayer perceptron (MLP), and a type 2 diabetes detection layer that uses another MLP to return the probability of type 2 diabetes based on the concatenated outputs of the previous two layers. The type 2 diabetes detection layer ends with double nodes that reflect the two exclusive classes (normal and type 2 diabetes). The model architecture is shown in Supplementary Fig. S1, and its detailed configuration is described in Supplementary Table S1. We set the batch size to 2048 and used the cross-entropy loss and Adam optimizer[16] with a learning rate of 0.0003. The model was trained for a maximum of 100 epochs, and early stopping was set with a patience of 20 on the performance measured using the AUROC. Experiments were conducted on 1-, 2-, and 12-lead ECGs with and without patient information and on patient information only. In the last case, only the MLP was used. We used Guided Grad-CAM[17] to determine where the model focused on when detecting type 2 diabetes. PyTorch (version 1.12.0) in Python was used for deep learning training.

Statistical analysis

The characteristics such as age, sex, ECG features, and comorbidities, between diabetic and non-diabetic groups and between hospitals were compared by calculating P-values using the Student's t-test for continuous variables and the χ^2 and Fisher's exact tests for categorical variables. To measure and compare the performance of the models, we used the AUROC and area under the precision–recall curve (AUPRC). The confidence intervals (CIs) of the AUROC and AUPRC were calculated using DeLong's method,[18] and those of sensitivity, specificity, precision, and F1-score were calculated using Wilson's method.[19] For survival analysis, all CPH models were fitted by Breslow's method with a penaliser of 0.01, considering the potential collinearity of the datasets.[20,21] The proportional hazard (PH) assumption was verified using scaled Schoenfeld residuals and a log (–log (survival)) plot. The Nelson–Aalen cumulative hazard curve was used to assess the risk of incident type 2 diabetes. Hazard ratios (HRs) with a 95% CI were reported for all variables. As a sensitivity analysis, we screened more normal individuals with lower glucose and HbA1c levels and implemented the same procedures. Statistical significance was set at $\alpha = 0.05$. All statistical analyses were performed using Python (version 3.8.10), and Lifelines (version 0.27.7) in Python was used for survival analysis.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

This study included 187,524 individuals who visited SNUH-HSGC between 1 January 2004 and 31 December 2022 and 84,449 who visited SNUH-HPC between 1 January 2001 and 31 December 2022. 43,292 individuals from SNUH-HSGC and 38,100 individuals from SNUH-HPC were excluded based on our

cohort criteria, and 144,232 individuals from SNUH-HSGC and 46,349 individuals from SNUH-HPC were finally included in our study population, as shown in Fig. 1. The baseline characteristics of the hospitals are presented in Table 1. The overall age and rate of comorbidities were higher for SNUH-HPC, and the proportion of males was higher for SNUH-HSGC. For both hospitals, the diabetic group was older, and the male group had a high prevalence of type 2 diabetes. The proportion of patients with type 2 diabetes was approximately 5.35% (7717/144232) for SNUH-HSGC and 5.66% (2622/46349) for SNUH-HPC. Out of 144,232 participants in SNUH-HSGC, 1540 were diagnosed using only glucose, 3085 using only HbA1c, and 3092 using both glucose and HbA1c. Among the 46,349 participants in SNUH-HPC, 477 were diagnosed using only glucose, 1332 using only HbA1c, and 813 using both glucose and HbA1c. The distributions of the ECG features in both hospitals were generally similar. For both hospitals, the PP interval, QT interval, RR interval, ST-T segment, and TP interval were shorter in the diabetic group, and the remaining ECG features were similar between the diabetic and non-diabetic groups. All the ECG features were derived from lead II.

The performance of the DLM in detecting type 2 diabetes is shown in Table 2. The model trained by 12-lead ECG and patient information (age and sex) exhibited the best detection performance, with an AUROC of 0.816 (0.807–0.825) for internal validation and 0.762 (0.754–0.770) for external validation. This model also exhibited the best AUPRC with 0.179 (0.170–0.188) in the internal validation and 0.145 (0.136–0.153) in the external validation. The model trained using 2-lead ECG and patient information exhibited a fair performance, with an AUROC of 0.800 (0.791–0.810) in internal validation and 0.744 (0.736–0.753) in external validation. The use of 1-lead ECG was ineffective. The detection performance significantly improved for all cases except when using a 1-lead ECG when the ECG was added to the patient information. The receiver operating characteristic (ROC) and precision–recall curves of all the experiments are shown in Fig. 2, and the results of Guided Grad-CAM are shown in Fig. 3. For both hospitals, we randomly selected 10 true-positive ECG samples from the DLM trained by 12-lead ECG and patient information, and provided the results of lead II. For most samples, the model focused on the QRS complex to detect type 2 diabetes; however, some exceptions focused on P or T waves.

The median follow-up from the first visit for a health checkup to incident type 2 diabetes or the end of the study was 11.84 years (average 11.08 years; standard deviation 5.14 years; range 0–18.65 years). Note that we used the decision of DLM trained by 12-lead ECG and patient information for all survival analyses. The median diabetes assessment intervals of SNUH-HSGC and SNUH-HPC were 301 days and 119 days, respectively. The Nelson–Aalen cumulative hazard curves for incident type 2 diabetes are shown in Fig. 4, and the HRs of all variables are listed in Table 3. The gap between the curves of the diabetic ECG group and the non-diabetic ECG group was clearly proportional to time in both the internal and external validation datasets. The risk of incident type 2 diabetes was significantly higher in the diabetic ECG group than the non-diabetic ECG group. For all CPH models, the diabetic ECG group exhibited an increased risk of incident type 2 diabetes compared with the well-classified group with an HR of 2.15 (1.83–2.53) for internal validation and 1.92 (1.75–2.12) for external validation in CPH model 1. For CPH model 2, the HR was almost the same. In CPH model 3, glucose and HbA1c values were added for adjustment, and the HR was still significant at 1.67 (1.41–1.96) for internal validation and 1.44 (1.31–1.59) for external validation. The average glucose and HbA1c levels of the

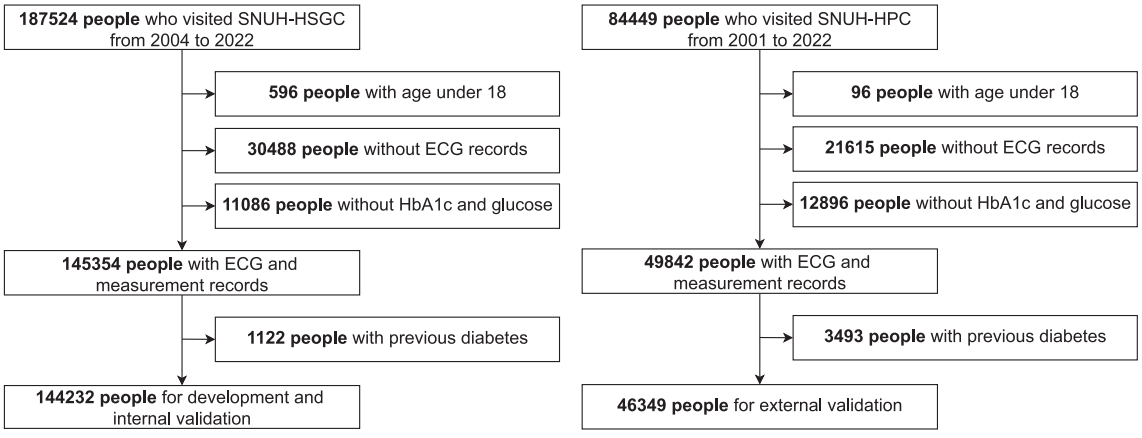

Fig. 1: Population flowchart.

	SNUH-HSGC: Development and internal validation set			SNUH-HPC: External validation set			Between hospitals
	Diabetic (n = 7717)	Non-diabetic (n = 136,515)	P-value	Diabetic (n = 2622)	Non-diabetic (n = 43,727)	P-value	P-value
Age	56.36 ± 10.20	45.69 ± 11.92	<0.0001	59.63 ± 10.51	50.96 ± 12.94	<0.0001	<0.0001
Male sex	5276 (68.37)	67,918 (49.75)	<0.0001	1485 (56.64)	20,782 (47.53)	<0.0001	<0.0001
ECG features (sec)							
P wave duration	0.07 ± 0.02	0.07 ± 0.02	0.0836	0.07 ± 0.02	0.07 ± 0.02	<0.0001	<0.0001
PR interval	0.15 ± 0.03	0.15 ± 0.03	<0.0001	0.15 ± 0.03	0.15 ± 0.03	<0.0001	<0.0001
PP interval	0.88 ± 0.14	0.92 ± 0.14	<0.0001	0.89 ± 0.15	0.93 ± 0.14	<0.0001	<0.0001
PR segment	0.08 ± 0.03	0.08 ± 0.04	<0.0001	0.09 ± 0.03	0.09 ± 0.04	0.3328	<0.0001
QRS duration	0.11 ± 0.05	0.11 ± 0.05	<0.0001	0.11 ± 0.05	0.11 ± 0.05	0.0007	<0.0001
QT interval	0.37 ± 0.05	0.38 ± 0.05	<0.0001	0.37 ± 0.06	0.39 ± 0.05	<0.0001	<0.0001
RR interval	0.88 ± 0.14	0.92 ± 0.14	<0.0001	0.89 ± 0.15	0.93 ± 0.14	<0.0001	<0.0001
ST segment	0.13 ± 0.06	0.13 ± 0.06	0.0014	0.14 ± 0.06	0.14 ± 0.06	0.0484	<0.0001
ST-T segment	0.26 ± 0.05	0.27 ± 0.05	<0.0001	0.27 ± 0.06	0.28 ± 0.06	<0.0001	<0.0001
TP interval	0.36 ± 0.11	0.39 ± 0.10	<0.0001	0.37 ± 0.11	0.39 ± 0.10	<0.0001	<0.0001
Comorbidities							
Liver disease	14 (0.18)	248 (0.18)	1.0000	17 (0.65)	338 (0.77)	0.5514	<0.0001
Myocardial infarction	10 (0.13)	56 (0.04)	0.0011	15 (0.57)	134 (0.31)	0.0311	<0.0001
Cerebrovascular disease	21 (0.27)	186 (0.14)	0.0036	36 (1.37)	358 (0.82)	0.0038	<0.0001
Peripheral vascular disease	4 (0.05)	36 (0.03)	0.1643	3 (0.11)	93 (0.21)	0.3774	<0.0001
Peptic ulcer disease	25 (0.32)	193 (0.14)	0.0001	3 (0.11)	78 (0.18)	0.6297	0.2937
Malignant tumour	30 (0.39)	351 (0.26)	0.0377	30 (1.14)	538 (1.23)	0.7655	<0.0001
Metastatic carcinoma	0 (0.00)	13 (0.01)	1.0000	3 (0.11)	14 (0.03)	0.0679	0.0001
Pulmonary disease	10 (0.13)	179 (0.13)	1.0000	10 (0.38)	210 (0.48)	0.5693	<0.0001
Renal disease	2 (0.03)	11 (0.01)	0.1512	9 (0.34)	40 (0.09)	0.0004	<0.0001
Paraplegia and hemiplegia	0 (0.00)	3 (0.00)	1.0000	1 (0.04)	6 (0.01)	0.3348	0.0027
Connective tissue disease	1 (0.01)	53 (0.04)	0.3694	0 (0.00)	73 (0.17)	0.0359	<0.0001
Heart failure	1 (0.01)	4 (0.00)	0.2404	1 (0.04)	9 (0.02)	0.4414	0.0004
Dementia	1 (0.01)	6 (0.00)	0.3195	2 (0.08)	15 (0.03)	0.2496	<0.0001

[a]All data are shown as mean ± SD and n (%).

Table 1: **Baseline characteristics of the included participants.**[a]

diabetic ECG group were higher than the non-diabetic ECG group as shown in Supplementary Table S2. The scaled Schoenfeld residuals and a log–log plot for the verification of the PH assumption are shown in Supplementary Figs. S2 and S3. To fit the CPH model using the more normal group, we screened the individuals with glucose < 110 mg/dL and HbA1c < 5.6%, and the HR of model classification adjusted by all variables was 1.81 (1.32–2.48) and 1.48 (1.20–1.82) for internal and external validation datasets, respectively. The Nelson–Aalen cumulative hazard curves and HRs of all CPH models for the more normal group are shown in Supplementary Fig. S4 and Table S3.

Discussion

This study included 190,581 individuals without any history of diabetes who visited two separate hospitals for health checkups. We divided non-diabetic individuals into the diabetic ECG group and the non-diabetic ECG group, and compared the risk of incident type 2 diabetes by the CPH model with appropriate adjustments. The diabetic ECG group was observed to be more susceptible to incident type 2 diabetes than the non-diabetic ECG group, with a significant HR of 2.15 for internal validation and 1.92 for external validation. It can be an opportunity for those with normal glucose and HbA1c but diabetic ECG to learn what lifestyle they should make and which interventions are needed to prevent the development of diabetes.[22–27]

The correlation between ECG and type 2 diabetes was ascertained by training the model using ECG exclusively. The model trained by the 12-lead ECG exhibited fair performance, achieving an AUROC of 0.787. While the performance declined with a reduced number of leads, it still provided sufficient evidence for the association between ECG and prevalent type 2 diabetes. We then compared the performance of the models trained with both ECG and patient information (age and sex) with that of the models trained with

	AUROC	P-value[b]	AUPRC	Sensitivity	Specificity	Precision	F1-score
Internal validation							
Patient information (Age & Sex)	0.770 (0.760–0.780)	..	0.135 (0.125–0.145)	0.787 (0.783–0.792)	0.631 (0.626–0.637)	0.109 (0.105–0.112)	0.191 (0.186–0.195)
12-lead ECG only	0.787 (0.777–0.797)	0.1179	0.165 (0.154–0.175)	0.826 (0.822–0.831)	0.611 (0.605–0.616)	0.108 (0.104–0.112)	0.191 (0.187–0.196)
12-lead ECG + Patient information	**0.816 (0.807–0.825)**	<0.0001	**0.179 (0.170–0.188)**	**0.846 (0.842–0.850)**	0.651 (0.646–0.657)	0.122 (0.118–0.125)	0.213 (0.208–0.217)
2-lead ECG only	0.734 (0.722–0.746)	0.0037	0.121 (0.109–0.133)	0.774 (0.769–0.779)	0.592 (0.586–0.598)	0.098 (0.094–0.101)	0.173 (0.169–0.178)
2-lead ECG + Patient information	0.800 (0.791–0.810)	<0.0001	0.158 (0.149–0.167)	0.803 (0.799–0.808)	**0.680 (0.674–0.685)**	**0.125 (0.121–0.129)**	**0.217 (0.212–0.221)**
1-lead ECG only	0.640 (0.627–0.654)	<0.0001	0.083 (0.069–0.096)	0.632 (0.627–0.638)	0.589 (0.583–0.595)	0.081 (0.078–0.084)	0.143 (0.139–0.147)
1-lead ECG + Patient information	0.771 (0.761–0.781)	0.4529	0.135 (0.125–0.145)	0.771 (0.766–0.776)	0.650 (0.644–0.655)	0.112 (0.108–0.115)	0.195 (0.190–0.200)
External validation							
Patient information (Age & Sex)	0.705 (0.696–0.714)	..	0.109 (0.099–0.118)	**0.812 (0.808–0.815)**	0.491 (0.487–0.496)	0.087 (0.085–0.090)	0.158 (0.154–0.161)
12-lead ECG only	0.748 (0.740–0.757)	<0.0001	0.139 (0.130–0.147)	0.788 (0.784–0.792)	0.586 (0.581–0.590)	0.102 (0.100–0.105)	0.181 (0.178–0.185)
12-lead ECG + Patient information	**0.762 (0.754–0.770)**	<0.0001	**0.145 (0.136–0.153)**	0.801 (0.798–0.805)	0.598 (0.593–0.602)	**0.107 (0.104–0.110)**	**0.188 (0.185–0.192)**
2-lead ECG only	0.701 (0.691–0.710)	0.9326	0.109 (0.099–0.118)	0.694 (0.690–0.698)	**0.607 (0.602–0.611)**	0.096 (0.093–0.098)	0.168 (0.165–0.172)
2-lead ECG + Patient information	0.744 (0.736–0.753)	<0.0001	0.131 (0.123–0.140)	0.791 (0.788–0.795)	0.584 (0.580–0.589)	0.102 (0.100–0.105)	0.181 (0.178–0.185)
1-lead ECG only	0.625 (0.614–0.636)	<0.0001	0.082 (0.072–0.093)	0.603 (0.598–0.607)	0.591 (0.587–0.596)	0.081 (0.079–0.084)	0.143 (0.140–0.146)
1-lead ECG + Patient information	0.704 (0.695–0.713)	0.8140	0.108 (0.099–0.117)	0.796 (0.793–0.800)	0.508 (0.504–0.513)	0.089 (0.086–0.091)	0.159 (0.156–0.163)

[a]Bold is the best and underlined is the second best. [b]P-values were calculated using the DeLong method.

Table 2: **Performance of the DLM for detecting type 2 diabetes.**[a]

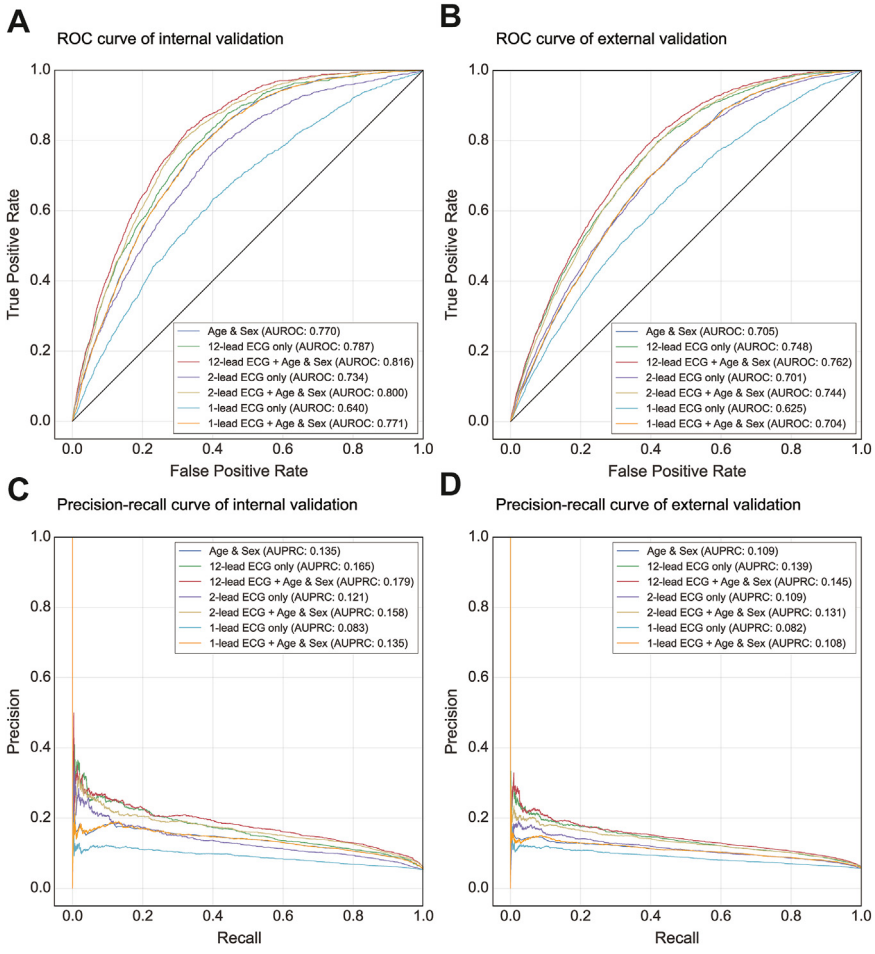

Fig. 2: Receiver operating characteristic and precision–recall curves of the deep learning models for detecting type 2 diabetes. AUROC; area under the receiver operating characteristic curve, AUPRC; area under the precision precision–recall curve. ECG; electrocardiogram.

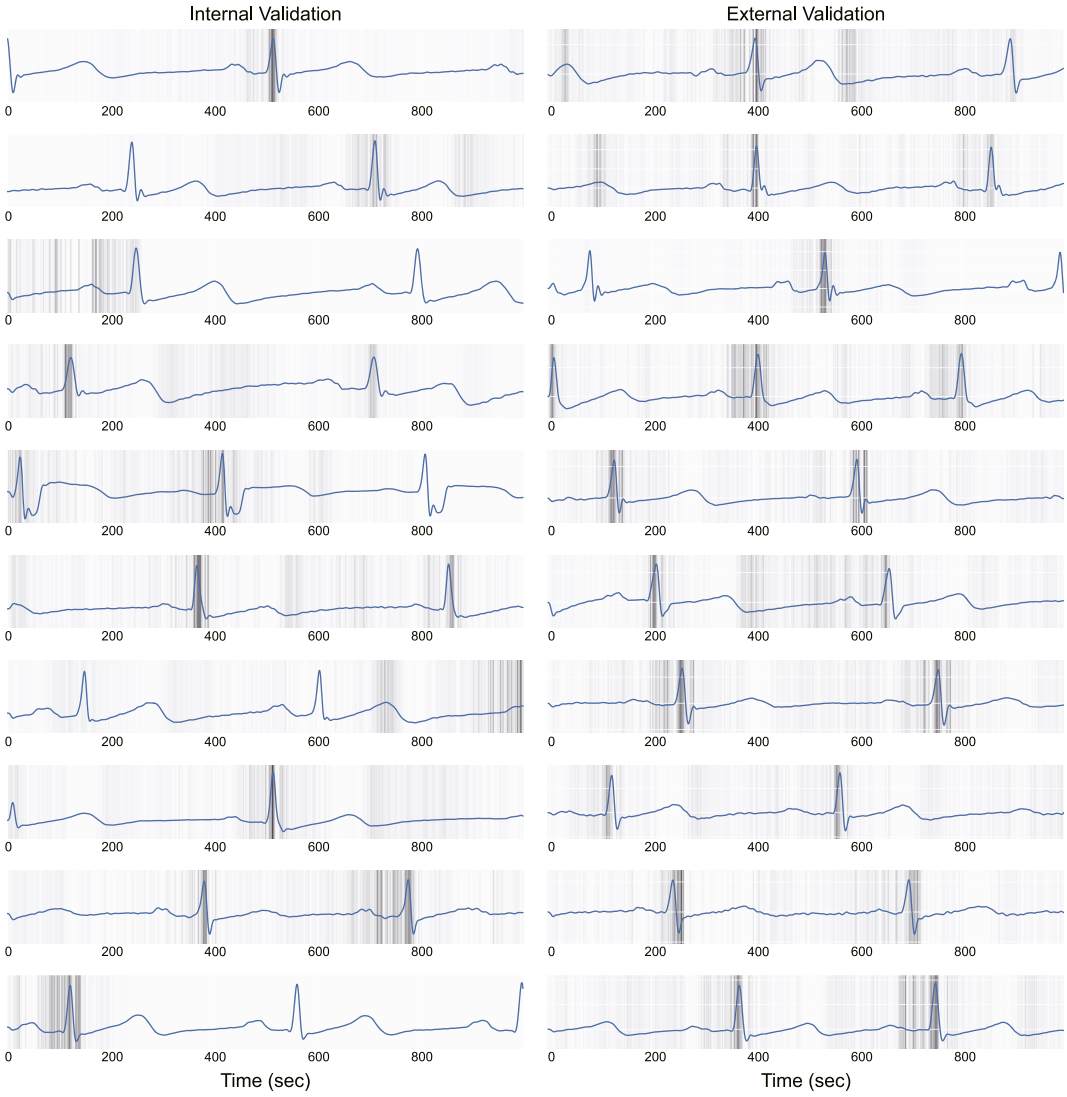

Fig. 3: Samples of Guided Grad-CAM applied ECG. All cases were observed with lead II.

patient information only. We observed that the inclusion of ECG data alongside patient information significantly enhanced the detection performance. Our model is highly likely to be utilised in a wearable system; the model trained by 2-lead ECG and patient information showed fair performance (AUROC of 0.800), and the size of the model was less than 8 MB, which is sufficiently small to be embedded in a device. Because smartwatches can easily measure leads I and II, users can check the risk of incident type 2 diabetes in real-time. In addition to the CNN-based model, we applied a representative recurrent neural network (RNN) model, long short-term memory (LSTM).[28] However, the performance of the LSTM trained with the 12-lead ECG and patient information was almost the same as that of the model trained with patient information, with an AUROC of 0.771. The CNN filters can appear to extract proper information by compressing the phenotype of the ECG itself, whereas the RNN receives numerical data individually without any consideration of the phenotype.

Diabetes and cardiovascular disease are closely associated.[29,30] However, there is no standard report that explains the characteristics of ECG in patients with diabetes compared with those without diabetes. Although some ECG features (PP interval, QT interval, RR interval, ST-T segment, and TP interval) with significant differences were observed between the diabetic and non-diabetic groups (Table 1), we could not find any literature supporting this finding. Considering the differences in ECG features and assuming that those features can classify prevalent type 2 diabetes, we trained a

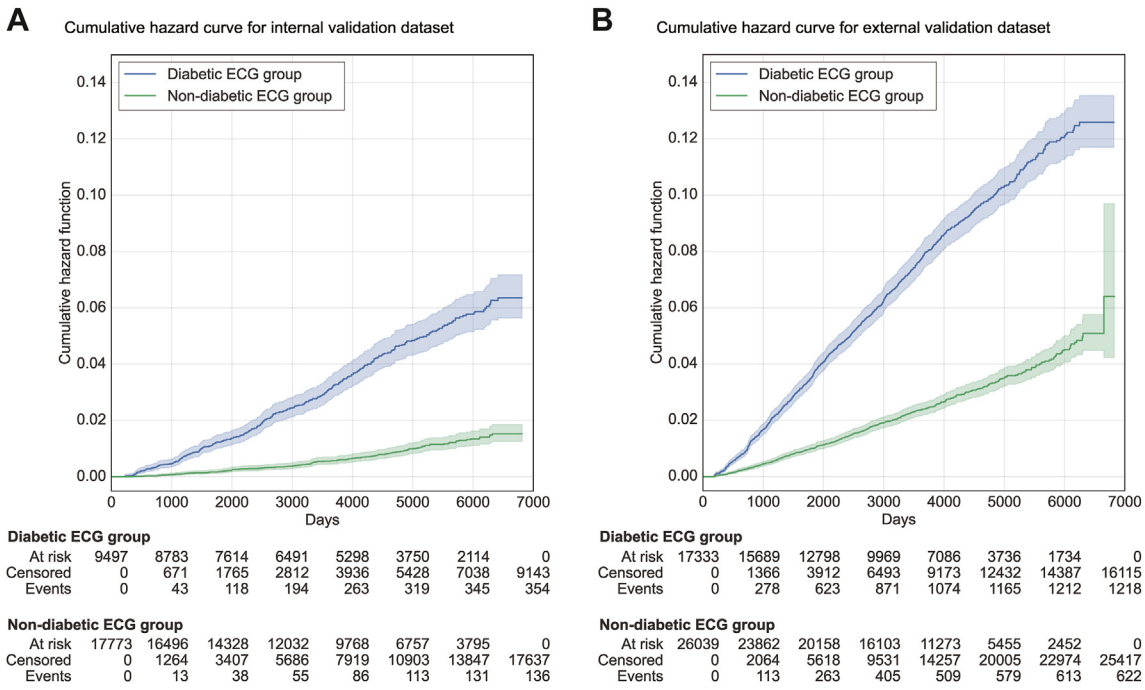

Fig. 4: Nelson–Aalen cumulative hazard curves on incident type 2 diabetes for individuals without diabetes. ECG; electrocardiogram.

random forest and LightGBM[31] using a total of 10 ECG features. However, the performance was poor, with an AUROC of less than 0.63. Accordingly, we confirmed the suitability of our CNN-based model. However, concerns remain regarding the black box of neural networks. To determine where our DLM focused on detecting prevalent type 2 diabetes, we adapted Guided Grad-CAM[17] to localise the important regions. The QRS complex was primarily focused on, and T and P waves were also occasionally noted; however, no clinical research supports these results. No additional patterns were observed.

This study primarily focused on showing a higher long-term risk of incident type 2 diabetes in non-diabetic individuals with diabetic ECG than in those with non-diabetic ECG. Initially, we adjusted the CPH model using age and sex and then added comorbidity variables to the CPH model. While the HR of diabetic ECG was expected to decline after adjustment because comorbidity variables, such as liver disease and myocardial infarction, are known to be related to type 2 diabetes,[32,33] the HR changed slightly and was still statistically significant. To further verify the effectiveness of the diabetic ECG, we adjusted the CPH model with values of glucose and HbA1c, which were the variables most related to type 2 diabetes in this study, and the HR was still significant even after adjustment. As a sensitivity analysis, we reduced the upper limits of glucose (from 126 to 110 mg/dL) and HbA1c (from 6.5 to 5.6%) to exclude borderline cases and validate the diabetic ECG on more normal people. We fit CPH models again, and the diabetic ECG group still exhibited an increased risk of incident type 2 diabetes compared with the non-diabetic ECG group, with an HR of 1.84 (1.34–2.53) for internal validation and 1.58 (1.28–1.94) for external validation. This implies that the risk of incident type 2 diabetes can be assessed using ECG in individuals with a relatively low risk of developing type 2 diabetes. Consequently, non-diabetic individuals with diabetic ECG can perceive a potential risk of type 2 diabetes and can be exposed to proper lifestyle and interventions to prevent the development of diabetes. Regarding diabetes assessment interval, the main cause of the difference between the two hospitals is as follows: SNUH-HSGC is specialised in health checkups and is locally separated from SNUH, while SNUH-HPC is locally a part of SNUH, which is one of the biggest tertiary hospitals in the Republic of Korea. Thus, people visiting SNUH-HPC might have had other health concerns rather than diabetes, and those people might have undergone blood tests as an inpatient or outpatient.

The limitations of our study are as follows: First, the CPH model would benefit from additional adjustments using variables such as body mass index. However, we could not access height and weight data owing to internal circumstances, despite their significance in influencing type 2 diabetes.[34] Second, although we externally validated our model, it should be further investigated if it is valid for ECG from another system because this study used a single ECG recording system.

	CPH model 1	P-value	CPH model 2	P-value	CPH model 3	P-value
	Hazard ratio (CI 95%)		Hazard ratio (CI 95%)		Hazard ratio (CI 95%)	
Internal validation						
Baseline						
Diabetic ECG	2.15 (1.83–2.53)	<0.0001	2.15 (1.82–2.53)	<0.0001	1.67 (1.41–1.96)	<0.0001
Age	1.02 (1.01–1.03)	<0.0001	1.02 (1.01–1.03)	<0.0001	1.01 (1.01–1.02)	0.0003
Sex	1.27 (1.09–1.47)	0.0016	1.27 (1.10–1.47)	0.0015	1.17 (1.01–1.36)	0.0349
Comorbidities						
Malignant tumour			2.40 (0.78–7.39)	0.1264	2.53 (0.81–7.84)	0.1090
Peptic ulcer disease			1.08 (0.20–5.70)	0.9304	1.02 (0.20–5.34)	0.9786
Cerebrovascular disease			0.51 (0.03–7.49)	0.6244	0.52 (0.04–7.71)	0.6345
Connective tissue disease			0.67 (0.01–67.54)	0.8670	0.76 (0.01–92.24)	0.9089
Myocardial infarction			3.10 (0.43–22.52)	0.2638	3.00 (0.42–21.65)	0.2757
Pulmonary disease			1.38 (0.25–7.61)	0.7103	1.41 (0.25–7.88)	0.6925
Liver disease			1.13 (0.22–5.70)	0.8814	1.22 (0.24–6.27)	0.8130
Laboratory values						
Glucose					1.04 (1.03–1.05)	<0.0001
HbA1c					4.31 (3.41–5.45)	<0.0001
External validation						
Baseline						
Diabetic ECG	1.92 (1.75–2.12)	<0.0001	1.92 (1.74–2.11)	<0.0001	1.44 (1.31–1.59)	<0.0001
Age	1.02 (1.01–1.02)	<0.0001	1.02 (1.01–1.02)	<0.0001	1.01 (1.00–1.01)	0.0012
Sex	1.20 (1.10–1.31)	<0.0001	1.19 (1.10–1.30)	<0.0001	1.19 (1.09–1.29)	0.0001
Comorbidities						
Malignant tumour			1.45 (1.01–2.08)	0.0420	1.30 (0.91–1.85)	0.1471
Peptic ulcer disease			0.39 (0.11–1.36)	0.1412	0.48 (0.13–1.74)	0.2646
Cerebrovascular disease			1.29 (0.86–1.92)	0.2158	1.17 (0.79–1.74)	0.4409
Connective tissue disease			1.42 (0.50–4.09)	0.5109	1.85 (0.63–5.48)	0.2647
Myocardial infarction			3.52 (2.26–5.47)	<0.0001	2.98 (1.92–4.62)	<0.0001
Pulmonary disease			1.41 (0.83–2.38)	0.2054	1.30 (0.77–2.20)	0.3237
Liver disease			2.37 (1.64–3.43)	<0.0001	2.34 (1.62–3.38)	<0.0001
Laboratory values						
Glucose					1.05 (1.04–1.05)	<0.0001
HbA1c					4.97 (4.32–5.71)	<0.0001

Table 3: **Adjusted HRs for incident type 2 diabetes for individuals without diabetes.**

Detailed ECG phenotypes may differ depending on the equipment used. Third, the specificities of the models were relatively low and this may cause many false alarms of wearable devices even though our model can be built into wearable devices such as smartwatches. Accordingly, improving specificity and precision through developing better models and proposing appropriate protocols for diabetic risk monitoring using wearable devices is suggested for future works. Fourth, the results of Guided Grad-CAM should be analysed qualitatively using relevant clinical domain knowledge. The Guided Grad-CAM exhibited that our model mainly focused on the QRS complex and occasionally on P or T waves to detect type 2 diabetes. Unfortunately, there was no previous clinical research investigating the relationship between the phenotype of ECG and diabetes to support our result. If obvious regional patterns are discovered and supported by proper clinical reasons, our model would be more likely to be used in the medical field with improved explainability. Finally, because this study constructed a censored dataset using the database of each hospital, the diagnoses of incident type 2 diabetes in other hospitals were not considered. The reliability of the study will be improved if individual diabetes records can be tracked using the diagnostic records of all hospitals in the Republic of Korea through institutions such as the National Health Insurance Service.

In conclusion, this study demonstrated that the DLM can effectively evaluate the long-term risk of incident type 2 diabetes in a non-diabetic population using ECG. Individuals without prior diabetes concerns can perceive their potential risk of type 2 diabetes through an ECG-based diagnosis. This study had limitations in data

acquisition, a single ECG recording system, low specificity, finite explanations of the model without supporting clinical background, and participant tracking. For future works, discovering better models with improved specificity, clinical research on the relationship between the phenotype of ECG and diabetes to support our result, and further investigation with tracked data and various ECG recording systems are suggested.

Contributors

JK and HJY contributed to the conceptualisation and design of the study. JK handled and mainly analysed the research data, and all authors interpreted the results. JK constructed the deep learning model, conducted statistical analysis, and wrote the original draft of the paper. HLY, SHK, SK, HJY revised the paper. SHK reviewed the validity of statistical analysis. HLY and SK reviewed the deep learning training process. JR and HJY reviewed the clinical evidence of the study. KSK and GA provided the data and JK and JL verified the quality of the data. JK, JL, KSK, ZK, GA, DK, and HJY had full access to the raw data. All authors had the final responsibility to submit for publication.

Data sharing statement

The raw data used in this study are not publicly available to preserve participant privacy. The data generated during and/or analysed during the study are available from the corresponding author upon reasonable request. The codes for deep learning training and statistical analysis are available at our GitHub repository (https://github.com/kicarussays/ecg-diabetes).

Declaration of interests

All authors declare no competing interests.

Acknowledgements

This work was supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Korean government (Ministry of Science and ICT) (No. 2018R1A5A1060031) and the AI Institute at Seoul National University. HJY receives funding from the National Research Foundation of Korea, and JK is supported by the fellowship program of the AI Institute at Seoul National University.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2024.102445.

References

1 Marshall SM, Flyvbjerg A. Prevention and early detection of vascular complications of diabetes. *BMJ*. 2006;333:475–480. https://doi.org/10.1136/bmj.38922.650521.80.
2 Fowler MJ. Microvascular and macrovascular complications of diabetes. *Clin Diabetes*. 2008;26:77–82. https://doi.org/10.2337/diaclin.26.2.77.
3 Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the international diabetes federation diabetes atlas, 9th ed. *Diabetes Res Clin Pract*. 2019;157:107843. https://doi.org/10.1016/j.diabres.2019.107843.
4 Sun H, Saeedi P, Karuranga S, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res Clin Pract*. 2022;183:109119. https://doi.org/10.1016/j.diabres.2021.109119.
5 Roglic G. WHO Global report on diabetes: a summary. *Int J Noncommun Dis*. 2016;1:3.
6 Siu AL. Screening for abnormal blood glucose and type 2 diabetes mellitus: U.S. Preventive services task Force recommendation statement. *Ann Intern Med*. 2015;163:861–868. https://doi.org/10.7326/M15-2345.
7 Herman WH, Ye W, Griffin SJ, et al. Early detection and treatment of type 2 diabetes reduce cardiovascular morbidity and mortality: a simulation of the results of the anglo-Danish-Dutch study of intensive treatment in people with screen-detected diabetes in primary care (ADDITION-Europe). *Diabetes Care*. 2015;38:1449–1455. https://doi.org/10.2337/dc14-2459.
8 Mosenzon O, Alguwaihes A, Leon JLA, et al. CAPTURE: a multinational, cross-sectional study of cardiovascular disease prevalence in adults with type 2 diabetes across 13 countries. *Cardiovasc Diabetol*. 2021;20:154. https://doi.org/10.1186/s12933-021-01344-0.
9 Lin C-S, Lee Y-T, Fang W-H, et al. Deep learning algorithm for management of diabetes mellitus via electrocardiogram-based glycated hemoglobin (ECG-HbA1c): a retrospective cohort study. *J Personalized Med*. 2021;11:725. https://doi.org/10.3390/jpm11080725.
10 Cordeiro R, Karimian N, Park Y. Hyperglycemia identification using ECG in deep learning era. *Sensors*. 2021;21:6263. https://doi.org/10.3390/s21186263.
11 Swapna G, Vinayakumar R, Soman K. Diabetes detection using deep learning algorithms. *ICT Express*. 2018;4:243–246. https://doi.org/10.1016/j.icte.2018.10.005.
12 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chron Dis*. 1987;40:373–383. https://doi.org/10.1016/0021-9681(87)90171-8.
13 Jeon JY, Ko S-H, Kwon H-S, et al. Prevalence of diabetes and prediabetes according to fasting plasma glucose and HbA1c. *DMJ*. 2013;37:349–357. https://doi.org/10.4093/dmj.2013.37.5.349.
14 Acharya UR, Fujita H, Oh SL, Hagiwara Y, Tan JH, Adam M. Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Inf Sci*. 2017;415–416:190–198. https://doi.org/10.1016/j.ins.2017.06.027.
15 Kwon S, Lee S-R, Choi E-K, et al. Comparison between the 24-hour holter test and 72-hour single-lead electrocardiogram monitoring with an adhesive patch-type device for atrial fibrillation detection: prospective cohort study. *J Med Internet Res*. 2022;24:e37970. https://doi.org/10.2196/37970.
16 Kingma DP, Ba J. Adam: a method for stochastic optimization. preprint arXiv:14126980 *arXiv*. 2014. https://doi.org/10.48550/arXiv.1412.6980.
17 Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE international conference on computer vision*. 2017:618–626.
18 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–845. https://doi.org/10.2307/2531595.
19 Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc*. 1927;22:209–212. https://doi.org/10.1080/01621459.1927.10502953.
20 Breslow NE. Analysis of survival data under the proportional hazards model. *Int Stat Rev/Rev Int Stat*. 1975;43:45–57. https://doi.org/10.2307/1402659.
21 Wang B, Wang Z, Poundarik AA, et al. Unmasking fracture risk in type 2 diabetes: the association of longitudinal glycemic hemoglobin level and medications. *J Clin Endocrinol Metab*. 2022;107:e1390–e1401. https://doi.org/10.1210/clinem/dgab882.
22 Ekoe J-M, Goldenberg R, Katz P. Screening for diabetes in adults. *Can J Diabetes*. 2018;42:S16–S19. https://doi.org/10.1016/j.jcjd.2017.10.004.
23 Selph S, Dana T, Blazina I, Bougatsos C, Patel H, Chou R. Screening for type 2 diabetes mellitus: a systematic Review for the U.S. Preventive services task Force. *Ann Intern Med*. 2015;162:765–776. https://doi.org/10.7326/M14-2221.
24 Li R, Zhang P, Barker LE, Chowdhury FM, Zhang X. Cost-effectiveness of interventions to prevent and control diabetes mellitus: a systematic Review. *Diabetes Care*. 2010;33:1872–1894. https://doi.org/10.2337/dc10-0843.
25 Millard AV, Graham MA, Mier N, et al. Diabetes screening and prevention in a high-risk, medically isolated border community. *Front Public Health*. 2017;5:135. https://doi.org/10.3389/fpubh.2017.00135. Community Case Study.
26 Chatterjee R, Narayan KMV, Lipscomb J, et al. Screening for diabetes and prediabetes should Be cost-saving in patients at high risk. *Diabetes Care*. 2013;36:1981–1987. https://doi.org/10.2337/dc12-1752.
27 Chatterjee R, Venkat Narayan KM, Lipscomb J, Phillips LS. Screening adults for pre-diabetes and diabetes may Be cost-saving. *Diabetes Care*. 2010;33:1484–1490. https://doi.org/10.2337/dc10-0054.

28 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9:1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

29 Strain WD, Paldánius PM. Diabetes, cardiovascular disease and the microcirculation. *Cardiovasc Diabetol*. 2018;17:57. https://doi.org/10.1186/s12933-018-0703-2.

30 Resnick HE, Howard BV. Diabetes and cardiovascular disease. *Annu Rev Med*. 2002;53:245–267. https://doi.org/10.1146/annurev.med.53.082901.103904.

31 Ke G, Meng Q, Finley T, et al. Lightgbm: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst*. 2017;30:3149–3157.

32 Stefan N, Cusi K. A global view of the interplay between non-alcoholic fatty liver disease and diabetes. *Lancet Diabetes Endocrinol*. 2022;10:284–296. https://doi.org/10.1016/S2213-8587(22)00003-1.

33 Cui J, Liu Y, Li Y, Xu F, Liu Y. Type 2 diabetes and myocardial infarction: recent clinical evidence and perspective. *Front Cardiovas Med*. 2021;8:644189. https://doi.org/10.3389/fcvm.2021.644189. Mini Review.

34 Rubino DM, Greenway FL, Khalid U, et al. Effect of weekly subcutaneous semaglutide vs daily liraglutide on body weight in adults with overweight or obesity without diabetes: the STEP 8 randomized clinical trial. *JAMA*. 2022;327:138–150. https://doi.org/10.1001/jama.2021.23619.