

RESEARCH

Open Access



Deep learning-enabled ECG system for detecting left ventricular hypertrophy and predicting cardiovascular prognoses

Zhe-Yu Yang^{1,8}, Shi-Chue Hsing², Dung-Jang Tsai³, Chin Lin^{3,4,5,6}, Chin-Sheng Lin², Chih-Hung Wang^{7,8} and Wen-Hui Fang^{1,4,6*}

*Correspondence:
Wen-Hui Fang
rumaf.fang@gmail.com

Full list of author information is available at the end of the article

Abstract

Left ventricular hypertrophy (LVH) is a common condition with a prevalence of 15%–20% in general population. Prior studies have suggested that deep learning model (DLM)–enabled electrocardiogram (ECG) systems can aid LVH detection and cardiovascular risk assessment; however, conventional manual ECG criteria have limited sensitivity and their prognostic utility remains suboptimal. Therefore, this study aimed to develop a DLM-enabled ECG system to detect LVH and evaluate its prognostic associations with incident cardiovascular outcomes. A total of 40,736 patients from hospital A were used for model development (training and tuning) and internal validation (29,595/5,935/5,206 patients, respectively), and 6,271 patients from hospital B were used for external validation. LVH was defined by left ventricular mass index (LVMI) derived from echocardiography. Prognostic outcomes included new-onset acute myocardial infarction (AMI), heart failure (HF), and atrial fibrillation (AFib). In the external validation set, our AI-ECG-LVH model achieved area under the receiver operating characteristic curve (AUC) values of 0.82 in males and 0.77 in females. Furthermore, the hazard ratios for incident AMI, HF, and AFib were 2.67, 3.15, and 2.23 for AI-ECG-LVH, compared with 2.76, 3.78, and 2.25 for echocardiography-defined LVH (ECHO-LVH). Our AI-ECG-LVH model may provide a straightforward, affordable, and noninvasive approach for LVH screening and first-contact risk stratification.

Keywords Artificial intelligence, Electrocardiogram, Deep learning, Left ventricular hypertrophy, Previvor, Cardiovascular disease

Introduction

Left ventricle hypertrophy (LVH) has a prevalence of 15% to 20% in the general population [1]. The risk of LVH escalates with the severity of hypertension (HTN), obesity, and advancing age. HTN stands out as one of the most prevalent noncommunicable diseases globally [2]. LVH can precipitate various cardiovascular diseases (CVDs) including atrial fibrillation (AFib), ventricular tachycardia (VT)/ventricular fibrillation (VF), and sudden cardiac death (SCD), consequently elevating mortality rates [3]. Moreover, an increase in the thickness of the left ventricle diminishes blood supply to its muscles, heightening



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

the risk of ischemia, acute myocardial infarction (AMI), and heart failure (HF) [4]. Therefore, early detection of LVH is crucial for preventing serious CVDs and potentially reducing mortality rates, thereby enhancing healthcare outcomes [5]. Through early intervention with medication to prevent abnormal LV thickness growth, it may be possible to avert the onset of serious CVDs such as HF, AMI, and SCDs.

The most common diagnostic tool for LVH is currently echocardiography [6]; by measuring the diameter of the ventricle to discover the abnormalities. While echocardiogram is a routine and reasonably priced procedure, its widespread application for population-level screening is occasionally constrained by its high dependency on the skill level of echocardiographers and technical challenges such as poor acoustic windows in certain patient populations [7]. Therefore, there is a need for an alternative, more affordable approach to ensure better healthcare outcome. Electrocardiogram (ECG) is a common diagnostic tool that had been widely used to screen left ventricle activities with advantages of being non-invasive, rapid and inexpensive [8]. There are numbers of studies have already demonstrated that LVH can be diagnosed using ECGs. For instance, criteria such as the Sokolow-Lyon criteria and Cornell voltage criteria offer reasonably high accuracy, reaching around 90%. However, their sensitivity remains relatively low, typically ranging from 20% to 40% [9].

As rapid growth of deep learning (DL) technology application in medicines recently [5], DL-enabled ECG models have demonstrated encouraging performance for LVH detection. Several studies in artificial intelligence (AI)-based ECG classification have demonstrated high sensitivity in identifying patients with various diseases. For example, previous research that diagnose asymptomatic left ventricular dysfunction (ALVD) using 12-lead ECG AI model had area under receiver operating characteristic curves (AUC) of 0.93 and sensitivity of 86.3%, respectively [10]. Previous studies also used AI-ECG to detect LVH and gained high accuracy, with sufficiently high sensitivity. In specific, the AUC of AI enhanced ECG model can reach about 0.78, with sensitivity of around 85% to 90%, which is much higher compared to cardiologists based on conventional ECG criteria (around AUC of 0.63 and 34.2% of sensitivity) [11]. Therefore, it is suggested that deep learning model (DLM)-enhanced model using ECGs is a promising solution to the insufficient sensitivity problem. However, achieving high diagnostic sensitivity for structurally defined LVH should not be viewed as the ultimate clinical objective. A critical unresolved question remains whether LVH detected by AI conveys prognostic significance comparable to—or potentially exceeding—that of echocardiography-defined LVH (ECHO-LVH). Most existing models focus on replicating anatomical findings and reporting discrimination metrics, few have investigated whether these AI-driven patterns can effectively predict long-term cardiovascular outcomes as accurately as the clinical gold standard.

The objective of this study is to bridge the gap between AI-driven detection and clinical risk stratification. Specifically, we developed an AI-ECG-detected LVH (AI-ECG-LVH) model not only to detect LVH but also to provide long-term prognostic insights for AMI, HF, and AFib. We compared the prognostic performance of our AI-ECG-LVH model with ECHO-LVH. By demonstrating that an AI-enabled ECG can provide prognostic value equivalent to echocardiography, we aim to offer a more accessible, cost-effective, and scalable solution for cardiovascular risk management.

Methods

Data source and population

The criteria for pairing ECGs with echocardiographic measurements were strictly defined to ensure the reliability of the reference standard. For the development and tuning sets, we utilized a bi-directional 7-day window; any ECG recorded within 7 days before or after an echocardiogram was paired with that measurement to maximize the diversity of the training data. For the internal and external validation sets, a more stringent one-directional 7-day window was applied. Specifically, we selected the single ECG performed within the 7 days prior to the echocardiogram that was in closest chronological proximity to the ultrasound examination, thereby reflecting a real-world clinical screening scenario. Echocardiographic quality was ensured through a standardized institutional protocol, with all studies performed in accordance with the American Society of Echocardiography clinical recommendations [12]. All examinations were acquired by certified sonographers and subsequently reviewed and finalized by board-certified cardiologists with expertise in echocardiography. Patients with missing height or weight required for left ventricular mass index (LVMI) calculation were excluded. All ECGs were recorded in a standardized digital format (12-lead, 10-second duration, 500 Hz sampling rate) and annotated with the corresponding LVMI values and baseline history of CVDs.

These data were collected from two facilities within the Tri-Service General Hospital System in Taipei, Taiwan. Hospital A is a tertiary academic medical center located in the Neihu District, with approximately 1,800 inpatient beds. Hospital B is a community-oriented hospital situated in the Zhongzheng District, focused on primary and secondary care with a capacity of approximately 100 beds. A total of 47,007 patients were involved in this study, including 40,736 from Hospital A and 6,271 from Hospital B (Fig. 1). The development set comprised 58,025 ECGs from 29,595 patients who presented after January 2017, while the tuning set included 17,181 ECGs from 5,935 patients seen between January and December 2016. In these two sets, all available longitudinal ECGs per patient were utilized to expose the model to natural intra-individual variability—such

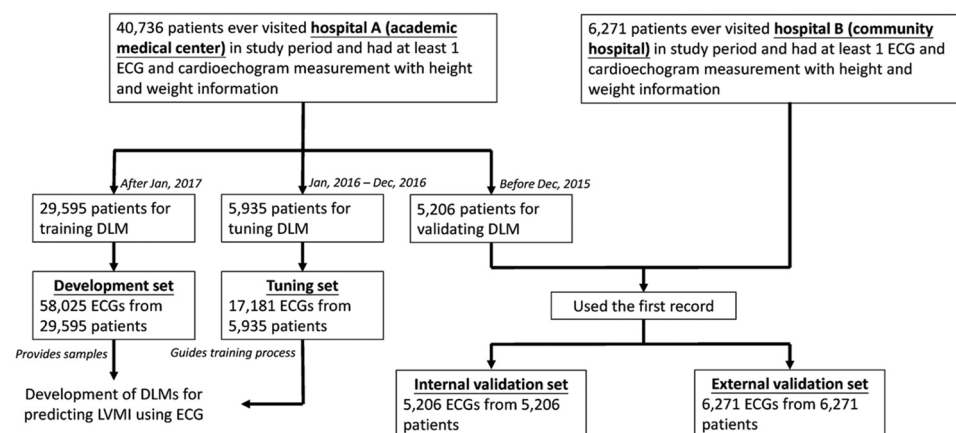


Fig. 1 Development, tuning, internal validation, and external validation sets generation and ECG labeling of left ventricular mass index. Schematic of the data set creation and analysis strategy, which was devised to assure a robust and reliable data set for training, validating, and testing of the network. Once a patient's data were placed in one of the data sets, that individual's data were used only in that set, avoiding 'cross-contamination' among the training, validation, and test data sets. The details of the flow chart and how each of the data sets was used are described in the methods

as fluctuations in electrode placement or physiological states—thereby enhancing the robustness of feature extraction.

In contrast, the internal validation set consisted of only the first ECG from 5,206 patients who visited Hospital A before December 2015. Similarly, for the external validation, only the first ECG from each of the 6,271 patients at Hospital B was used. This restriction to the initial encounter serves a dual purpose: it ensures statistical independence for performance evaluation and, crucially, simulates the clinical scenario of initial screening at the first point of medical contact. This design prioritizes clinical utility and minimizes the risk of performance inflation that might arise from repeated measurements within the validation cohorts. The specific chronological partitioning scheme was designed to facilitate the evaluation of long-term cardiovascular prognoses. Utilizing the earliest available cohort (pre-December 2015) for validation provided a sufficient observation window to capture longitudinal clinical events (e.g., AMI, HF, and AFib), which would not have been possible with more recent data. This retrospective study was approved by the institutional review board of Tri-Service General Hospital, Taipei, Taiwan (IRB NO. B202405084).

Baseline characteristics and follow-up information

Patient clinical status and cardiovascular outcomes were ascertained through the International Classification of Diseases, Ninth Revision and Tenth Revision (ICD-9 and ICD-10, respectively) and hospital's Electronic Health Record (EHR) system, which provides integrated data on clinical diagnoses, laboratory tests, and medication prescriptions. To maintain a comprehensive longitudinal profile, diagnoses from both inpatient and outpatient encounters were included. The detailed list of all ICD-9 and ICD-10 codes used for baseline comorbidities and follow-up outcomes is provided in Supplemental Table 1.

Definition of LVH

The diagnose of LVH was based on the LVMI, a quantified severity of LVH, which is the gold standard and has been widely used today. The equation for determining LVMI is LVM/BSA , where LVM is left ventricular mass and BSA is body surface area [13]. From previous study, LVM can be calculated through the following formula:

$$LVM(g) = 0.8\{1.04[(PWT+LVID + SWT)^3-LVID^3]\} + 0.6$$

Where PWT is posterior wall thickness, LVID is LV internal diameter, and SWT is septal wall thickness [14]. On the other hand, BSA is determined using the Mosteller formula below:

$$BSA = \frac{\sqrt{h \times w}}{60}$$

Where h is height in centimeters and w is weight in kilograms [15]. According to the guidelines, LVH can be confirmed morphologically when LVMI exceeds 115 g/m² in males and 95 g/m² in females. Severe LVH is indicated when the LVMI is greater than 149 g/m² in males and 122 g/m² in females [13].

The echocardiography was performed with a Philips (EPIQ 5, EPIQ 7, CX50 and IE33) ultrasound system.

Deep learning model

The proposed DLM for LVH detection is summarized as follows. Each ECG was acquired as a standard 12-lead recording with 5,000 samples per lead (10 s at 500 Hz), and the model input was constructed as a multichannel waveform array of size $12 \times 5,000$. Using these raw ECG signals, we trained separate convolutional neural network–based DLMs for LVH classification and LVMI regression under an identical hyperparameter setting. Model training used a batch size of 32 and the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with an initial learning rate of 1×10^{-3} . When the validation loss showed no further improvement, the learning rate was reduced tenfold to facilitate continued convergence [16, 17]. To mitigate overfitting, we employed early stopping by checkpointing model weights after each epoch and retaining the checkpoint with the lowest validation loss. L2 weight decay was applied as the primary regularization strategy, with a coefficient of 1×10^{-4} . The automated LVH interpretation from the Philips ECG system (PageWriter TC30 and TC50; Philips, Amsterdam, the Netherlands) was also included as a comparator.

Statistical analysis

Baseline characteristics of the patients were reported as using means with standard deviations, numbers of patients, or percentages as appropriately. All statistical analyses were done using R version 3.4.4 with significance level of $p < 0.05$.

The comparison of predicted ECG-LVMI and actual LVMI was analyzed using scatter plots of the mean difference and standard deviation, along with the Pearson correlation coefficient (r), and mean absolute error (MAE). Furthermore, to evaluate the diagnostic value for mild and severe LVH in both internal and external validation sets, AUC, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated with operating point of maximum Youden's index from the training set. Lastly, multivariable Cox proportional hazard model were used to examine the relationship between the LVH, included ECG and echocardiography-based diagnoses, and the interested follow-up new onset CVDs. The index date was the date of the initial ECG. Follow-up duration was calculated from the index date to the first event occurrence or censoring (death, loss to follow-up, or study termination, April 2021). The proportional hazards assumption was assessed and confirmed using Schoenfeld residuals (all $p > 0.05$). Primary prognostic models were adjusted for age and sex to approximate a realistic first-contact screening context. In particular, the hazard ratios (HRs) and 95% confidence intervals (95% CIs) were determined and used for the comparisons.

Results

Characteristics of the study population

Table 1 presents the demographic characteristics of patients, diseases history, echocardiography data and follow-up new onset diseases in each data set which includes development set, tuning set, internal validation set, and external validation set. Among these sets, development set and tuning set have much more males than females while both internal and external validation have the same percentage of 50.4% of males instead. In addition, the mean age of development set and both internal and external validation sets were around 64 to 65 in average, while 68 only in the tuning set. And all the data sets have an average body mass index around 24.

Table 1 Baseline characteristics

	Development set	Tuning set	Internal validation set	External validation set
Demography				
Sex (male)	32,290(55.6%)	9136(53.2%)	2624(50.4%)	3160(50.4%)
Age (years)	64.4±16.3	68.4±15.8	64.7±15.8	65.4±16.8
BMI (kg/m ²)	24.6±4.4	24.3±4.4	24.6±4.3	24.5±4.3
Disease history				
DM	15,498(26.7%)	6483(37.7%)	1720(33.0%)	2129(33.9%)
HTN	24,243(41.8%)	10,299(59.9%)	2826(54.3%)	3490(55.7%)
HLP	19,303(33.3%)	8093(47.1%)	2321(44.6%)	2972(47.4%)
CKD	16,456(28.4%)	8055(46.9%)	1464(28.1%)	1702(27.1%)
CAD	17,431(30.0%)	7370(42.9%)	1685(32.4%)	1993(31.8%)
COPD	7564(13.0%)	3872(22.5%)	1133(21.8%)	1584(25.3%)
Echocardiography data				
LVMI (g/m ²)	106.5±35.8	112.5±38.3	104.6±34.8	102.0±33.0
IVS (mm)	11.3±2.6	11.6±2.6	11.3±2.6	11.1±2.5
LVPW (mm)	9.4±1.7	9.6±1.8	9.3±1.7	9.2±1.6
LV-D (mm)	47.8±7.3	48.1±7.8	47.3±7.1	47.3±6.8
LV-S (mm)	30.8±7.3	31.4±7.8	29.8±6.7	29.6±6.2
LA (mm)	38.9±7.6	39.7±8.0	38.6±7.5	38.8±7.2
AO (mm)	33.0±4.4	33.1±4.4	33.0±4.4	32.9±4.3
RV (mm)	23.8±4.9	24.2±5.1	24.2±5.0	23.9±4.9
PASP (mmHg)	34.0±11.5	35.0±12.5	32.4±10.2	32.6±10.4
PE (mm)	0.7±2.3	0.6±2.2	0.4±2.0	0.4±1.8
EF (%)	62.5±13.2	60.8±14.3	65.1±11.4	65.6±10.6
Follow up data				
Present LVH			2141(41.1%)	2445(40.0%)
Follow-up (years), median (IQR)			2.3(1.0–4.0)	2.2(1.0–3.9)
New-onset LVH			613(43.1%)	715(43.7%)
Present AMI			188(3.6%)	160(2.6%)
Follow-up (years), median (IQR)			3.5(1.5–5.5)	2.8(1.1–5.0)
New-onset AMI			194(3.9%)	176(2.9%)
Present HF			654(12.6%)	721(11.5%)
Follow-up (years), median (IQR)			3.3(1.3–5.2)	2.6(0.9–4.7)
New-onset HF			477(10.5%)	545(9.9%)
Present Afib			336(6.5%)	359(5.7%)
Follow-up (years), median (IQR)			3.4(1.3–5.4)	2.7(1.0–4.9)
New-onset Afib			402(8.3%)	448(7.6%)

Abbreviations: BMI, body mass index; DM, diabetes mellitus; HTN, hypertension; HLP, hyperlipidemia; CKD, chronic kidney disease; CAD, coronary artery disease; COPD, chronic obstructive pulmonary disease; LVMI, left ventricular mass index; IVS, Inter-ventricular septum; LVPW, left ventricular posterior wall; LV-D, left ventricle (end-diastole); LV-S, left ventricle (end-systole); LA, left atrium; AO, aortic root; RV, right ventricle; PASP, pulmonary artery systolic pressure; PE, pericardial effusion; EF, ejection fraction; LVH, left ventricular hypertrophy; AMI, acute myocardial infarction; HF, heart failure; Afib, atrial fibrillation

Diagnostic performance and interpretability of AI-ECG-LVH

Figure 2 is a scatter plot of predicted ECG-LVMI compared to the actual LVMI. In the internal validation set, the MAE was 24.12 with a correlation coefficient of 0.57 (mean difference: -0.60 ± 31.84). In the external validation set, performance was slightly lower, with an MAE of 23.84 g/m² and a correlation coefficient of 0.55 (mean difference: -1.94 ± 31.34). Figure 3 presents sex-stratified receiver operating characteristic (ROC) analyses for AI-ECG-LVH in detecting echocardiography-confirmed LVH. Overall, the model showed good discrimination across subgroups (AUC, 0.75–0.82; 95% CIs in Supplemental Table 2), and calibration plots are shown in Supplemental Fig. 1, with

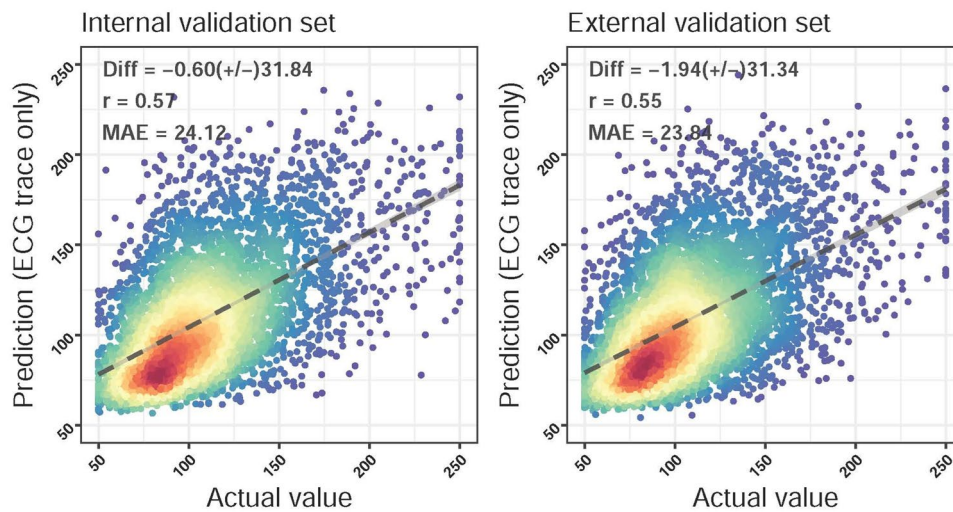


Fig. 2 Scatter plots of predicted left ventricular mass index (ECG-LVMI) via ECG only compared to the actual LVMI. The x-axis indicates the actual LVMI, and the y-axis presents the ECG-LVMI. Red points represent the highest density, followed by yellow, green light blue, and dark blue. We presented the mean difference, Pearson correlation coefficients (COR), and mean absolute errors (MAE) to demonstrate the accuracy of DLM. The black lines with 95% confidence intervals are fitted via simple linear regression

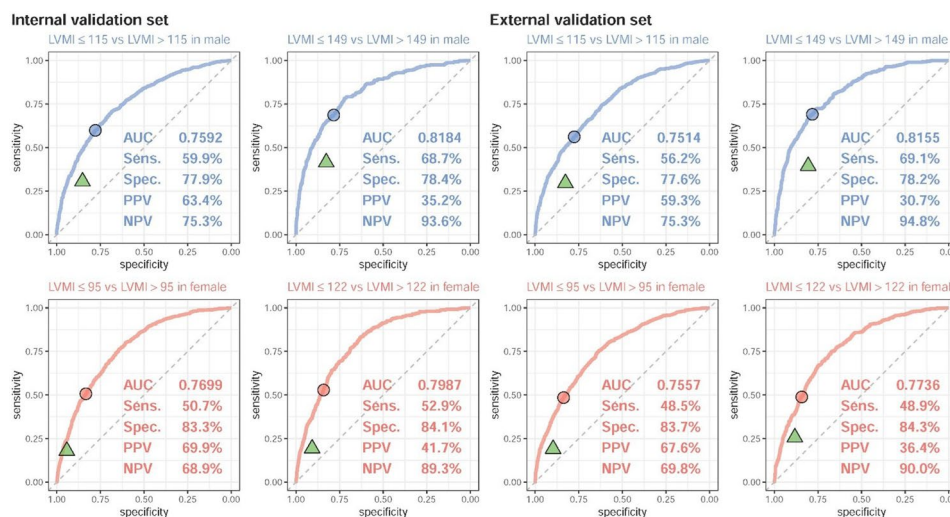


Fig. 3 The ROC curve of DLM predictions based on ECG to detect minor and severe left ventricular hypertrophy (LVH). The minor LVH cases are defined as a left ventricular mass index (LVMI) of > 115 in male and > 95 in female, respectively. The severe LVH cases are defined as a LVMI of > 149 in male and > 122 in female, respectively. The operating point was selected based on the maximum of Youden’s index in tuning set and presented using a circle mark, and the area under ROC curve (AUC), sensitivity (Sens.), specificity (Spec.), positive predictive value (PPV), and negative predictive value (NPV) were calculated based on it. The green triangle was the diagnosis based on Philips ECG automatic analysis system

higher sensitivity observed for severe LVH in men at the expense of lower precision. Specifically, for severe LVH in males, sensitivity was 68.7% (internal) and 69.1% (external), whereas the PPV was modest (35.2% internal; 30.7% external). For severe LVH in females, sensitivity was 52.9% and 48.9% with PPV of 41.7% and 36.4% in internal and external sets, respectively. Compared to the traditional rule-based criteria from the Philips ECG system (indicated by green triangles), our model achieved improved discriminatory performance and sensitivity.

To elucidate the decision-making process of the DL model, we utilized saliency maps to visualize the key ECG features contributing to the prediction of LVH (Fig. 4). As shown in the representative heatmap of an LVH patient, the model's attention was primarily localized to the QRS complexes and ST-T segments, particularly in the lateral precordial leads (V5 and V6) and lead II. This focus aligns with the established electrophysiological criteria for LVH.

Risk stratification for incident structural LVH

To evaluate the potential of AI-ECG-LVH in identifying individuals at high risk for future LVH, we conducted an analysis on patients who did not initially manifest structural LVH on echocardiography. Figure 5 illustrates the risk stratification of AI-ECG-LVH in these “baseline-normal” individuals ($\text{LVMI} \leq 115$ for males and ≤ 95 for females). Regarding the prediction of new-onset LVH, the AI-ECG-LVH demonstrated a clear risk gradient in the internal validation set. The HRs for the minor and severe categories were 1.28 and 1.55 in males, and 1.34 and 1.67 in females, respectively. In the external validation set, the predictive value was primarily driven by the severe category, which maintained a robust association with new-onset LVH (HR: 1.52 in males and 1.22 in females).

The predictive power of the AI model was further amplified when evaluating the risk of new-onset severe LVH. In the internal validation set, severe AI-ECG-LVH was associated with HRs of 1.87 in males and 2.05 in females. This high-risk signal remained highly consistent in the external validation set, with HRs of 2.29 in males and 2.19 in females. The clinical significance of these findings is further illustrated by the 6-year cumulative incidence rates. Consistent with these associations, in the external validation set, the absolute risk of developing new-onset LVH for individuals with a severe AI-ECG-LVH signature reached 83.3% in males and 77.7% in females.

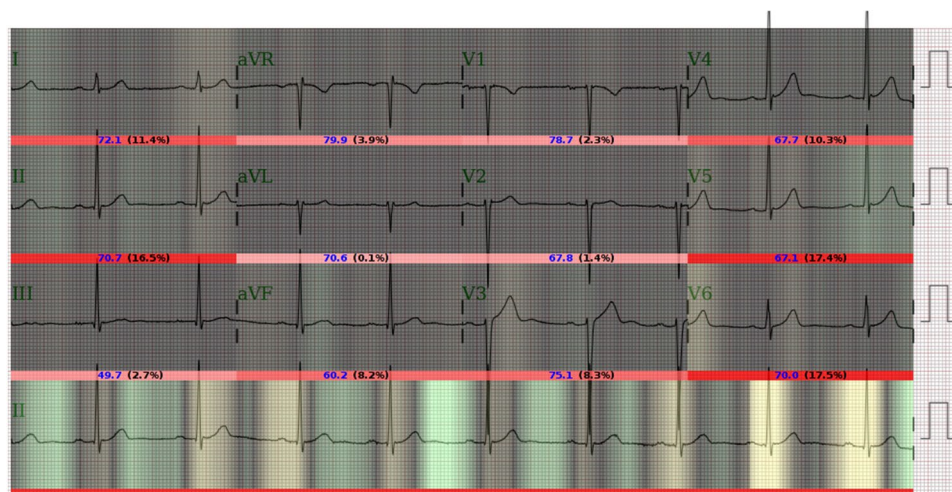


Fig. 4 Interpretability and feature attribution of the DL model for LVH prediction. Using the class activation mapping and attention mechanism to explain the AI-ECG prediction, we used white-to-red gradient bars to indicate the importance of each lead (with contribution percentages), and the darker-to-light gradient to indicate the contribution of each temporal position in the prediction of ECG-LVH. In this representative case, the model's attention is primarily focused on the QRS complexes and ST-T segments, particularly within the lateral precordial leads (V5 and V6) and lead II. This localization aligns with the established clinical criteria for left ventricular hypertrophy, emphasizing the biological plausibility of the model's diagnostic performance

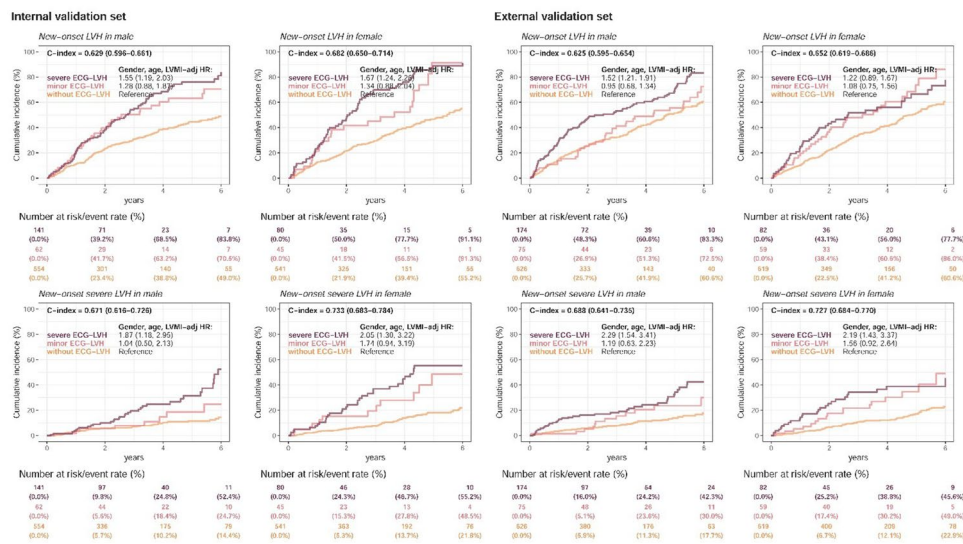


Fig. 5 Long-term incidence of developing a new-onset minor and severe left ventricular hypertrophy (LVH) in patients with an initially normal left ventricular mass index (LVMI) stratified by AI-ECG prediction. Patients in this analysis had an initial LVMI of ≤ 115 in male and ≤ 95 in female, respectively. The AI-ECG predictions were classified as without ECG-LVH, minor ECG-LVH, and severe ECG-LVH based on the operating points as the same with previous ROC curve analysis. The outcomes included new-onset LVH (> 115 in male and > 95 in female) and new-onset severe LVH (> 149 in male and > 122 in female), respectively. The C-index is calculated based on the continuous value combined with sex and age. The analyses are conducted both in internal and external validation sets. The table shows the at-risk population and cumulative risk for the given time intervals in each risk stratification

Prognostic equivalence of AI-ECG-LVH versus ECHO-LVH for incident cardiovascular outcomes

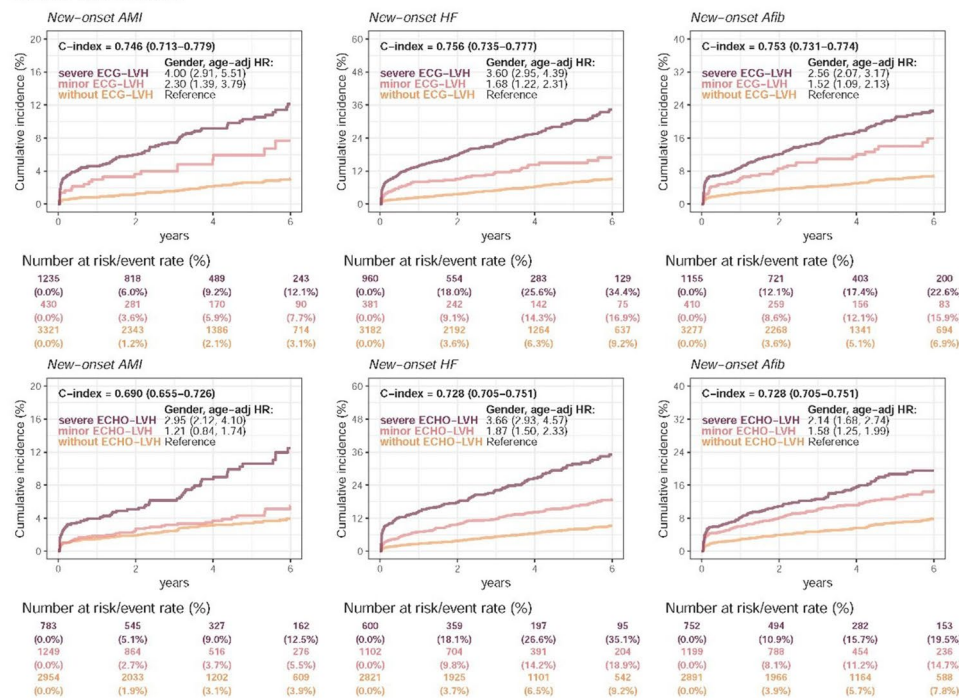
Figure 6 compares the prognostic associations of AI-ECG-LVH and ECHO-LVH with incident AMI, HF, and AFib in both validation sets. Individuals with a prior history of the corresponding outcome were excluded; analyses pooled sexes and were adjusted for age and sex. In the internal validation set, severe AI-ECG-LVH was associated with increased risks of AMI, HF, and AFib (HRs: 4.00, 3.60, and 2.56), with comparable estimates observed for severe ECHO-LVH (HRs: 2.95, 3.66, and 2.14). Similar concordance between severe AI-ECG-LVH and severe ECHO-LVH was observed in the external validation set.

6-year cumulative incidence estimates were consistent with the HR patterns. In the internal validation set, the 6-year risk among individuals with severe AI-ECG-LVH was 12.1% for AMI, 34.4% for HF, and 22.6% for AFib, which was comparable to the corresponding risks for severe ECHO-LVH (12.5%, 35.1%, and 19.5%, respectively). Similar concordance between severe AI-ECG-LVH and severe ECHO-LVH was observed in the external validation set (AMI: 7.2% vs. 6.7%; HF: 30.9% vs. 34.7%; AFib: 22.4% vs. 23.3%).

Discussion

This study developed a DLM to detect LVH from raw 12-lead ECG recordings and evaluated the clinical significance of AI-ECG-LVH in relation to incident cardiovascular outcomes. Our AI-ECG-LVH model achieved diagnostic performance with AUCs of 0.75–0.81 in validation sets. Furthermore, the model showed prognostic associations with incident AMI, HF, and AFib that were comparable to ECHO-LVH, supporting the potential of AI-ECG-LVH as a pragmatic first-contact risk stratification tool.

Internal validation set



External validation set

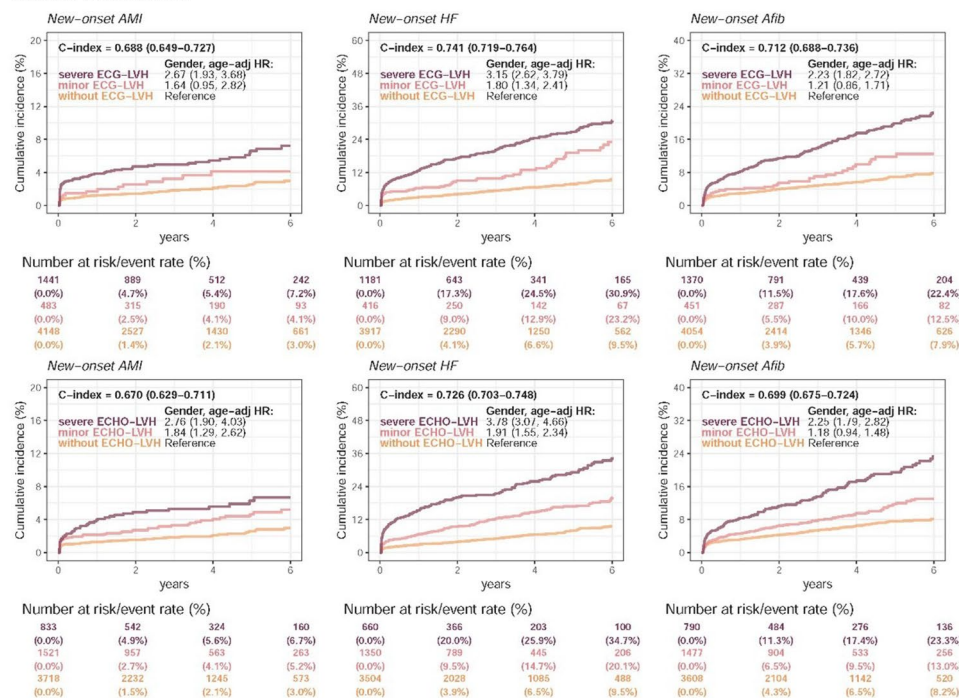


Fig. 6 Kaplan–Meier curves for each severity of electrocardiogram based left ventricular hypertrophy (ECG-LVH) and cardioechogram based LVH (ECHO-LVH) on new-onset complications. This analysis used the same operating points in male and female described previously, and combined male and female into one analysis. We excluded patients with present corresponding disease histories to analyze the new-onset events. The C-index is calculated based on the continuous value combined with sex and age. The analyses are conducted both in internal and external validation sets. The table shows the at-risk population and cumulative risk for the given time intervals in each risk stratification

Compared with conventional rule-based ECG criteria, the diagnostic performance of AI-ECG-LVH was improved. Prior studies have reported that the best AUCs of conventional ECG criteria for LVH are approximately 0.68 for the Sokolow–Lyon criterion and 0.60 for the Cornell criterion [18]. While prior machine learning approaches have shown promise, they often faced limitations regarding data generalizability. For instance, Lim DY et al. achieved an AUC of approximately 0.873 using GLMNet, but their dataset was biased toward young military conscripts with a lack of actual LVH cases [19]. Similarly, Liu CW et al. exhibited an even higher AUC of 0.96 using a back propagation neural network (BPN), although the study was limited by a small sample size [20]. A prior model combining ECG with transthoracic echocardiography (TTE) was developed to detect LVH and predict cardiovascular mortality [11]. In that study, the combined AI-ECG–TTE model achieved an AUC of 0.89 in individuals aged 20–60 years and was associated with a HR of 1.91 for cardiovascular mortality. While such multimodal approaches may yield higher discrimination, our results indicate that an AI model based solely on raw ECG signals can still provide clinically meaningful LVH detection and subsequent risk stratification, offering a simpler and less resource-intensive strategy that may be more scalable for first-line assessment. Regarding the quantification of LVMI, the model demonstrated correlation coefficients of 0.55–0.57 and an MAE of approximately 24 g/m². This level of precision indicates that the model should not be interpreted as a direct substitute for echocardiography-derived LVMI at the individual level. Rather, the LVMI estimation may be useful for screening-oriented risk stratification and for prioritizing confirmatory echocardiography in higher-risk individuals.

Our analyses further indicated that AI-ECG-LVH positivity in individuals without echocardiographic LVH at baseline was not necessarily benign. Individuals classified as severe AI-ECG-LVH despite being echocardiography-normal exhibited a higher risk of incident minor and severe LVH over the subsequent 6 years in both internal and external validation sets. Similar “false-positive” patterns have been described in prior AI-ECG studies for left ventricular systolic dysfunction [21], AFib, dyskalemia [22] and low ejection fraction (EF) [23], giving rise to the concept of “previvors” [24]. Consistently, the landmark AI-ECG screening study for cardiac contractile dysfunction reported that individuals with positive AI-ECG findings but no overt dysfunction at the time of testing had an increased risk of subsequently developing reduced EF [10]. Together, these observations support the interpretation that AI-ECG may capture latent or subclinical disease signatures that precede overt abnormalities on standard testing. This interpretation is further supported by emerging evidence that deep learning models can infer structural heart disease directly from ECG signals, indicating that ECG contains recoverable signatures of underlying structural pathology [25].

LVH is a well-established marker associated with CVDs, including AMI, HF, and AFib [2]. Earlier outcome studies based on conventional ECG-LVH criteria generally concluded that ECHO-LVH provides stronger long-term risk stratification than ECG-LVH. For example, Pedersen et al. indicated that LVH diagnosed by ECHO provided superior HRs for cardiovascular events compared to conventional ECG criteria [26]. However, recent studies suggest that AI-enhanced ECG analysis can achieve prognostic accuracy that rivals conventional imaging. Huang et al. demonstrated that AI-identified LVH was significantly associated with all-cause mortality over long-term follow-up and showed stronger mortality association than echocardiography-confirmed LVH [27]. Similarly,

Jentzer et al. highlighted the prognostic value of AI-ECG in predicting survival outcomes related to diastolic dysfunction in cardiac intensive care units [28]. AI-ECG models have also been shown to predict incident cardiovascular outcomes. Hong et al. developed an AI-ECG model to identify HF with preserved EF, and reported that AI-ECG positivity was associated with subsequent risk stratification, including elevated risks of cardiac death and HF hospitalization [29]. Kim et al. reported that an ECG-based AI approach could also be used to identify HF with mildly reduced EF and to stratify the subsequent clinical course [30]. Consistent with recent evidence, our study extends prognostic validation by evaluating AI-ECG-LVH against ECHO-LVH for incident AMI, HF, and AFib. Concordant prognostic associations were observed across internal and external validation, supporting AI-ECG-LVH as a pragmatic first-contact risk stratification signal rather than a purely anatomical surrogate.

The PPV of the AI-ECG-LVH model, particularly in the external validation cohort, was modest, underscoring the inherent trade-off between sensitivity and precision in screening-oriented deployment. Importantly, indiscriminate referral of all AI-positive individuals to echocardiography could increase downstream testing burden. Therefore, AI-ECG-LVH should be interpreted as a risk-stratification signal in conjunction with clinical context and baseline cardiovascular risk factors (e.g., HTN, DM, or metabolic syndrome), rather than as an automatic mandate for echocardiography. This perspective is supported by recent work highlighting that AI-enhanced ECG signatures can exhibit phenotypic selectivity and carry actionable risk information beyond single-label disease detection [31]. In practice, a positive AI-ECG-LVH result may be used to prioritize echocardiographic evaluation for those with higher pre-test probability and to prompt more intensive risk-factor management and early preventive interventions, positioning the model as a decision-support tool rather than a binary diagnostic test.

This study has several limitations. First, the analysis was retrospective and observational; therefore, causal inferences cannot be made, and prospective research is warranted. Second, we did not apply etiology-specific exclusions for conditions that may influence LV remodeling or ECG morphology (e.g., hyperthyroidism or HF), which may have introduced clinical heterogeneity in both the reference standard and outcome associations. Third, echocardiography-derived LVMI—used as the reference standard—may be affected by image quality, loading conditions, and inter-operator variability despite standardized protocols; such measurement error could contribute to misclassification of LVH and attenuate diagnostic performance. Fourth, ECG–echocardiography pairing within a predefined time window may introduce temporal mismatch, as physiological status can change between acquisitions, potentially impacting both LVMI and ECG features. In addition, physician-adjudicated labels for ECG rhythm/conduction patterns (e.g., paced rhythm, conduction abnormalities) were not systematically available in our database; thus, these factors were not explicitly modeled and may act as unmeasured confounders. Finally, heterogeneity in follow-up duration may affect the comparability of HRs across Cox models and should be considered when interpreting prognostic estimates. Although comparisons between AI-ECG-LVH and ECHO-LVH were performed within the validation set under identical adjustment and follow-up settings, HRs across different sets (internal vs. external) or across outcomes should not be over-interpreted as directly comparable, as follow-up distributions and event timing may differ. Residual

confounding also remains possible given the pragmatic adjustment strategy designed to reflect a first-contact screening scenario.

Conclusion

In conclusion, we developed an AI-enabled ECG model for LVH detection and prognostic risk stratification. The AI-ECG-LVH model demonstrated improved diagnostic performance over conventional criteria-based approaches and automated ECG interpretations, with notably higher sensitivity. Importantly, individuals with AI-ECG-LVH positivity despite no echocardiographic LVH at baseline exhibited an increased risk of subsequent LVH, supporting the concept that AI-ECG-LVH may capture subclinical signatures that precede overt structural changes. Moreover, the prognostic associations of AI-ECG-LVH with incident AMI, HF, and AFib were broadly comparable to those of ECHO-LVH. Collectively, these findings suggest that AI-enabled ECG may offer a practical, affordable, and noninvasive strategy for opportunistic LVH screening and first-contact cardiovascular risk stratification.

Abbreviations

LVH	Left ventricular hypertrophy
HTN	Hypertension
CVDs	Cardiovascular diseases
AFib	Atrial fibrillation
VT	Ventricular tachycardia
VF	Ventricular fibrillation
SCD	Sudden cardiac death
AMI	Acute myocardial infarction
HF	Heart failure
ECG	Electrocardiogram
DL	Deep learning
AI	Artificial intelligence
ALVD	Asymptomatic left ventricular dysfunction
AUC	Area under receiver operating characteristic curve
DLM	Deep learning model
ECHO-LVH	Echocardiography-defined LVH
AI-ECG-LVH	AI-ECG-detected LVH
LVMi	Left ventricular mass index
HER	Electronic Health Record
BSA	Body surface area
PWT	Posterior wall thickness
LVID	LV internal diameter
SWT	Septal wall thickness
MAE	Mean absolute error
PPV	Positive predictive value
NPV	Negative predictive value
HRs	Hazard ratios
95% CIs	95% confidence intervals
ROC	Receiver operating characteristic
BPN	Back propagation neural network
TTE	Transthoracic echocardiography
EF	Ejection fraction
MACEs	Major adverse cardiovascular events

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13040-026-00536-2>.

Supplementary Material 1

Author contributions

All authors participated in designing the study, generating hypotheses, interpreting the data, and critically reviewing the paper. ZYY and SCH wrote the first draft, and CSL, CHW, and WHF contributed substantially to writing subsequent versions. ZYY designed and conducted statistical analyses with support from CL and DJT. All authors had full access to all the data in the study and accepted responsibility for the decision to submit for publication. ZYY and WHF verified all the

data used in this study. The corresponding author (WHF) attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding

This study was supported by funding from the Ministry of Science and Technology, Taiwan (MOST110-2314-B-016-010-MY3 to C. Lin and MOST110-2321-B-016-002 to C.H. Wang).

Data availability

The datasets generated and/or analysed during the current study are not publicly available due to patient privacy and confidentiality requirements but are available from the corresponding author on reasonable request and with institutional review board approval.

Declarations

Ethics approval and consent to participate

This retrospective study was approved by the institutional review board of Tri-Service General Hospital, Taipei, Taiwan (IRB NO. B202405084).

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author used chatgpt to enhance language. After using this tool, the author thoroughly reviewed and edited the content as necessary and takes full responsibility for the final publication.

Competing interests

The authors declare no competing interests.

Author details

¹Division of Family Medicine, Department of Family and Community Medicine, Tri-Service General Hospital, and School of Medicine, National Defense Medical University, Taipei, Taiwan, Republic of China

²Division of Cardiology, Department of Internal Medicine, Tri-Service General Hospital, and School of Medicine, National Defense Medical University, Taipei, Taiwan, Republic of China

³Medical Technology Education Center, School of Medicine, National Defense Medical University, Taipei, Taiwan, Republic of China

⁴Artificial Intelligence of Things Center, Tri-Service General Hospital, National Defense Medical University, Taipei, Taiwan, Republic of China

⁵School of Public Health, National Defense Medical University, Taipei, Taiwan, Republic of China

⁶Department of Education and Research, Taipei City Hospital, Taipei, Taiwan, Republic of China

⁷Department of Otolaryngology-Head and Neck Surgery, Tri-Service General Hospital, National Defense Medical University, Taipei, Taiwan, Republic of China

⁸Graduate Institute of Medical Sciences, National Defense Medical University, Taipei, Taiwan, Republic of China

Received: 26 December 2025 / Accepted: 26 February 2026

Published online: 04 March 2026

References

1. Bornstein AB, Rao SS, Marwaha K. Left Ventricular Hypertrophy. StatPearls. Treasure Island (FL) ineligible companies. Disclosure: Suman Rao declares no relevant financial relationships with ineligible companies. Disclosure: Komal Marwaha declares no relevant financial relationships with ineligible companies.: StatPearls Publishing Copyright ©. 2024, StatPearls Publishing LLC.; 2024.
2. Alkema M, Spitzer E, Soliman OI, Loewe C. Multimodality imaging for left ventricular hypertrophy severity grading: a methodological review. *J Cardiovasc Ultrasound*. 2016;24(4):257–67. <https://doi.org/10.4250/jcu.2016.24.4.257>.
3. Shenasa M, Shenasa H. Hypertension, left ventricular hypertrophy, and sudden cardiac death. *Int J Cardiol*. 2017;237:60–3. <https://doi.org/10.1016/j.ijcard.2017.03.002>.
4. Torpy JM, Glass TJ, Glass RM. Left ventricular hypertrophy. *JAMA*. 2004;292(19):2430. <https://doi.org/10.1001/jama.292.19.430>.
5. Duffy G, Cheng PP, Yuan N, He B, Kwan AC, Shun-Shin MJ, et al. High-throughput precision phenotyping of left ventricular hypertrophy with cardiovascular deep learning. *JAMA Cardiol*. 2022;7(4):386–95. <https://doi.org/10.1001/jamacardio.2021.6059>.
6. Movahed MR, Ramaraj R, Manrique C, Hashemzadeh M. Left ventricular hypertrophy is independently associated with all-cause mortality. *Am J Cardiovasc Dis*. 2022;12(1):38–41.
7. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J Am Soc Echocardiogr*. 2015;28(1):1–e3914. <https://doi.org/10.1016/j.echo.2014.10.003>.
8. Kokubo T, Kodera S, Sawano S, Katsushika S, Nakamoto M, Takeuchi H, et al. Automatic detection of left ventricular dilatation and hypertrophy from electrocardiograms using deep learning. *Int Heart J*. 2022;63(5):939–47. <https://doi.org/10.1536/ihj.22-132>.
9. Peguero JG, Lo Presti S, Perez J, Issa O, Brenes JC, Tolentino A. Electrocardiographic criteria for the diagnosis of left ventricular hypertrophy. *J Am Coll Cardiol*. 2017;69(13):1694–703. <https://doi.org/10.1016/j.jacc.2017.01.037>.
10. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med*. 2019;25(1):70–4. <https://doi.org/10.1038/s41591-018-0240-2>.

11. Liu CM, Hsieh ME, Hu YF, Wei TY, Wu IC, Chen PF, et al. Artificial intelligence-enabled model for early detection of left ventricular hypertrophy and mortality prediction in young to middle-aged adults. *Circ Cardiovasc Qual Outcomes*. 2022;15(8):e008360. <https://doi.org/10.1161/circoutcomes.121.008360>.
12. Mitchell C, Rahko PS, Blauwet LA, Canaday B, Finstuen JA, Foster MC, et al. Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: recommendations from the american society of echocardiography. *J Am Soc Echocardiogr*. 2019;32(1):1–64. <https://doi.org/10.1016/j.echo.2018.06.004>.
13. Zipes DPL, Bonow P, Braunwald RO. E. Braunwald's heart disease: a textbook of cardiovascular medicine. Eleventh edition. Philadelphia: Elsevier Saunders; 2019.
14. Myerson SG, Montgomery HE, World MJ, Pennell DJ. Left ventricular mass: reliability of M-mode and 2-dimensional echocardiographic formulas. *Hypertension*. 2002;40(5):673–8. <https://doi.org/10.1161/01.hyp.0000036401.99908.db>.
15. Mosteller RD. Simplified calculation of body-surface area. *N Engl J Med*. 1987;317(17):1098. <https://doi.org/10.1056/nejm198710223171717>.
16. Hung Y, Lin C, Lin C-S, Lee C-C, Fang W-H, Lee C-C, et al. Artificial intelligence-enabled electrocardiography predicts future pacemaker implantation and adverse cardiovascular events. *J Med Syst*. 2024;48(1):67. <https://doi.org/10.1007/s10916-024-02088-6>.
17. Lin CS, Lin C, Fang WH, Hsu CJ, Chen SJ, Huang KH, et al. A deep-learning algorithm (ECG12Net) for detecting hypokalemia and hyperkalemia by electrocardiography: algorithm development. *JMIR Med Inf*. 2020;8(3):e15931. <https://doi.org/10.2196/15931>.
18. Siranart N, Deepan N, Techasatian W, Phutinart S, Sowalerat W, Kaewkanha P, et al. Diagnostic accuracy of artificial intelligence in detecting left ventricular hypertrophy by electrocardiograph: a systematic review and meta-analysis. *Sci Rep*. 2024;14(1):15882. <https://doi.org/10.1038/s41598-024-66247-y>.
19. Lim DY, Sng G, Ho WH, Hankun W, Sia CH, Lee JS, et al. Machine learning versus classical electrocardiographic criteria for echocardiographic left ventricular hypertrophy in a pre-participation cohort. *Kardiol Pol*. 2021;79(6):654–61. <https://doi.org/10.33963/kp.15955>.
20. Liu C-W, Wu F-H, Hu Y-L, Pan R-H, Lin C-H, Chen Y-F, et al. Left ventricular hypertrophy detection using electrocardiographic signal. *Sci Rep*. 2023;13(1):2556. <https://doi.org/10.1038/s41598-023-28325-5>.
21. Jentzer JC, Kashou AH, Attia ZI, Lopez-Jimenez F, Kapa S, Friedman PA, et al. Left ventricular systolic dysfunction identification using artificial intelligence-augmented electrocardiogram in cardiac intensive care unit patients. *Int J Cardiol*. 2021;326:114–23. <https://doi.org/10.1016/j.ijcard.2020.10.074>.
22. Lin C, Chau T, Lin C-S, Shang H-S, Fang W-H, Lee D-J, et al. Point-of-care artificial intelligence-enabled ECG for dyskalemia: a retrospective cohort analysis for accuracy and outcome prediction. *npj Digit Med*. 2022;5(1):8. <https://doi.org/10.1038/s41746-021-00550-0>.
23. Lee CH, Liu WT, Lou YS, Lin CS, Fang WH, Lee CC, et al. Artificial intelligence-enabled electrocardiogram screens low left ventricular ejection fraction with a degree of confidence. *Digit Health*. 2022;8:20552076221143249. <https://doi.org/10.1177/20552076221143249>.
24. Attia ZI, Harmon DM, Behr ER, Friedman PA. Application of artificial intelligence to the electrocardiogram. *Eur Heart J*. 2021;42(46):4717–30. <https://doi.org/10.1093/eurheartj/ehab649>.
25. Poterucha TJ, Jing L, Ricart RP, Adjei-Mosi M, Finer J, Hartzel D, et al. Detecting structural heart disease from electrocardiograms using AI. *Nature*. 2025;644(8075):221–30. <https://doi.org/10.1038/s41586-025-09227-0>.
26. Pedersen LR, Kristensen AMD, Petersen SS, Vaduganathan M, Bhatt DL, Juel J, et al. Prognostic implications of left ventricular hypertrophy diagnosed on electrocardiogram vs echocardiography. *J Clin Hypertens (Greenwich)*. 2020;22(9):1647–58. <https://doi.org/10.1111/jch.13991>.
27. Huang JT, Tseng CH, Huang WM, Yu WC, Cheng HM, Chao HL, et al. Comparison of machine learning and conventional criteria in detecting left ventricular hypertrophy and prognosis with electrocardiography. *Eur Heart J Digit Health*. 2025;6(2):252–60. <https://doi.org/10.1093/ehjdh/ztaf003>.
28. Jentzer JC, Lee E, Attia Z, Hillerson D, Kane GC, Lopez-Jimenez F, et al. Artificial intelligence ECG diastolic dysfunction and survival in cardiac intensive care unit patients. *J Am Heart Association*. 2025;14(5):e037839. <https://doi.org/10.1161/JAHA.124.037839>.
29. Hong D, Song SH, Shin H, Bak M, Kim J, Kim D, et al. Artificial intelligence-enabled electrocardiogram model for predicting heart failure with preserved ejection fraction: a single-center study. *Eur Heart J Digit Health*. 2025;6(5):959–68. <https://doi.org/10.1093/ehjdh/ztaf080>.
30. Kim D-Y, Lee S-W, Lee D-H, Lee S-C, Jang J-H, Shin S-H, et al. Electrocardiography-based artificial intelligence predicts the upcoming future of heart failure with mildly reduced ejection fraction. *Front Cardiovasc Med*. 2025;12–2025. <https://doi.org/10.3389/fcvm.2025.1418914>.
31. Croon PM, Dhingra LS, Biswas D, Oikonomou EK, Khera R. Phenotypic selectivity of artificial intelligence-enhanced electrocardiography in cardiovascular diagnosis and risk prediction. *Circulation*. 2025;152(18):1282–94. <https://doi.org/10.1161/circulationaha.125.076279>.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.