BMC Medical Informatics and Decision Making

RESEARCH

Open Access

ECG-based deep learning for chronic kidney disease detection and cardiovascular risk prediction

Ping-Huang Tsai[1], Shang-Yang Lee[2,3], Chia-Ling Helen Wei[1,4], Yu-Juei Hsu[1,5] and Chin Lin[2,3,6,7*]

Abstract

Background Chronic kidney disease (CKD) is a global health burden with low awareness among both patients and healthcare providers. Deep learning models (DLMs) have shown promise in interpreting electrocardiograms (ECGs) for various disease and may offer new opportunities for early CKD detection.

Methods We enrolled 66,587 outpatients with estimated glomerular filtration rate (eGFR) data from January 2010 to October 2020. A total of 72,618 ECGs from 49,632 patients were used to develop DLMs. Internal validation was performed on 16,955 nonoverlapping patients, and external validation involved 10,476 patients from a community hospital. The primary outcome was the detection of CKD, defined as eGFR < 60 mL/min/1.73 m^2. Secondary outcomes included all-cause mortality and major cardiovascular events.

Results The DLM achieved an AUC of 0.885 and 0.861 in the internal and external validation sets, respectively. Patients flagged by the DLM as having CKD showed more clinical risk factors for CKD progression and cardiovascular disease. Among patients without baseline CKD, those with a positive DLM screen had a significantly higher risk of incident CKD (hazard ratios 2.14 and 1.38; 95% CIs: 1.76–2.60 and 1.09–1.74). DLM stratification also predicted adverse outcomes such as stroke, heart failure, and atrial fibrillation more effectively than eGFR classification alone.

Conclusion An ECG-based deep learning model can help identify individuals at risk for CKD and its complications, even before laboratory abnormalities emerge. This approach may support early detection and risk stratification in clinical practice.

Clinical trial number Not applicable.

Keywords Chronic kidney disease, Deep learning model, Electrocardiograms, Estimated glomerular filtration rate

*Correspondence:
Chin Lin
xup6fup0629@gmail.com
[1]Division of Nephrology, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical University, Taipei, Taiwan, R.O.C.
[2]School of Public Health, College of Public Health, National Defense Medical University, No. 161 Min-Chun E. Rd., Sec. 6, Neihu, Taipei 114, Taiwan, R.O.C.
[3]Military Digital Medical Center, Tri-Service General Hospital, National Defense Medical University, Taipei, Taiwan
[4]Department of Medicine, Cardinal Tien Hospital, New Taipei City, Taiwan, R.O.C.
[5]Department of Biochemistry, College of Biomedical Sciences, National Defense Medical University, Taipei, Taiwan, R.O.C.
[6]Graduate Institute of Life Sciences, College of Biomedical Sciences, National Defense Medical University, Taipei, Taiwan, R.O.C.
[7]Medical Technology Education Center, School of Medicine, College of Medicine, National Defense Medical University, Taipei, Taiwan, R.O.C.

Background

Chronic kidney disease (CKD) is a global health issue linked to kidney failure, cardiovascular complications, and increased mortality [1, 2]. Current CKD assessments—such as estimated glomerular filtration rate (eGFR) and albuminuria—are based on blood and urine tests that are routinely available but may not fully capture future risk. While several prediction models based on demographic and clinical variables have been developed [3–5], they often require multiple inputs and show reduced performance in external validation cohorts [6]. Moreover, most existing models are not designed to predict longitudinal renal outcomes, cardiovascular disease, or mortality.

Electrocardiogram (ECG)-based screening may provide a scalable, low-cost alternative for opportunistic CKD risk assessment. A standard 12-lead ECG reflects both electrical and structural changes in the heart, and subtle ECG alterations—such as prolonged PR intervals, QRS widening, and increased P wave dispersion—have been observed in individuals with impaired kidney function [7, 8].

Deep learning models (DLMs) have shown potential in extracting disease-relevant features from ECGs, enabling prediction of conditions such as vulnerable patients [9], dyskalemia [10–12], anemia [13], arrhythmia [14], atrial fibrillation [15, 16], long QT syndrome [17], heart dysfunction [18–20], and even all-cause mortality [21]. Recent studies have also demonstrated that machine learning models can detect moderate to severe CKD using ECGs [22–24]. However, ECG remains underexplored as a prognostic tool for predicting long-term renal outcomes, cardiovascular risk, or mortality. The main challenge lies in collecting sufficient clinical and outcome data at scale.

In this study, we propose a novel application of a DLM (ECG12Net) to both detect CKD and predict long-term risks—including cardiovascular events—in a large outpatient population. Our approach leverages ECG signals alone, without additional clinical variables, to explore whether cardiac electrophysiological signatures can serve as early, accessible markers of CKD-related risk.

Methods

Population and study characteristics

This prospective, multisite study was conducted at two hospitals within the same healthcare system: Hospital A (Tri-Service General Hospital, Neihu) and Hospital B (Tingzhou Branch Hospital, Zhongzheng). The study received ethical approval from the Institutional Review Board (IRB No. C202105049) and was conducted in accordance with the Declaration of Helsinki.

Between January 1, 2010, and October 31, 2020, we collected data from 66,587 patients at Hospital A for model development, tuning, and internal validation. An external validation cohort was drawn from 10,476 patients at Hospital B. These two hospitals function independently, each operating its own outpatient and emergency services with separate medical equipment.

Only ECGs recorded during outpatient visits were included. ECGs from emergency department visits were excluded to minimize potential confounding from acute conditions. Across all cohorts, 62,139 ECG–eGFR pairs were included. All data were de-identified, and informed consent was waived by the IRB. Figure 1 illustrates the cohort construction process.

Data source and measurements

Standard 12-lead ECGs were acquired using a Philips machine (PH080A) at a sampling frequency of 500 Hz for 10 s. We extracted eight continuous ECG measurements and 31 diagnostic pattern classes from final clinical reports. The eight measurements included heart rate, PR interval, QRS duration, QT interval, corrected QT interval, and the axes of P wave, RS wave, and T wave. Data for these variables were 93–100% complete, and missing values were imputed using multiple imputation [25]. The patterns included abnormal T wave, atrial fibrillation, atrial flutter, atrial premature complex, complete atrioventricular block, complete left bundle branch block, complete right bundle branch block, first-degree atrioventricular block, incomplete left bundle branch block, incomplete right bundle branch block, ischemia/infarction, junctional rhythm, left anterior fascicular block, left atrial enlargement, left axis deviation, left posterior fascicular block, left ventricular hypertrophy, low QRS voltage, pacemaker rhythm, prolonged QT interval, right atrial enlargement, right ventricular hypertrophy, second-degree AV block, sinus bradycardia, sinus pause, sinus rhythm, sinus tachycardia, supraventricular tachycardia, ventricular premature complex, ventricular tachycardia, and Wolff–Parkinson–White syndrome. The 31 clinical diagnosis patterns were parsed from the structured findings statements based on key phrases standardized within the Philips system.

Baseline kidney function was assessed using eGFR and urinary albumin-to-creatinine ratio (UACR), both obtained from electronic medical records. eGFR was calculated using the CKD-EPI equation, and albuminuria was defined as UACR ≥ 30 mg/g. CKD stages were defined based on a single measurement of eGFR and UACR at baseline as follows: (1) without CKD [eGFR ≥ 60 mL/min/1.73 m^2 without albuminuria] (2), stage 1 [eGFR ≥ 90 mL/min/1.73 m^2 with micro/macroalbuminuria] (3), stage 2 [60 ≤ eGFR < 90 mL/min/1.73 m^2 with micro/macroalbuminuria] (4), stage 3 [30 ≤ eGFR < 60 mL/min/1.73 m^2] (5), stage 4 [15 ≤ eGFR < 30 mL/min/1.73 m^2], and (6) stage 5 [eGFR < 15 mL/min/1.73

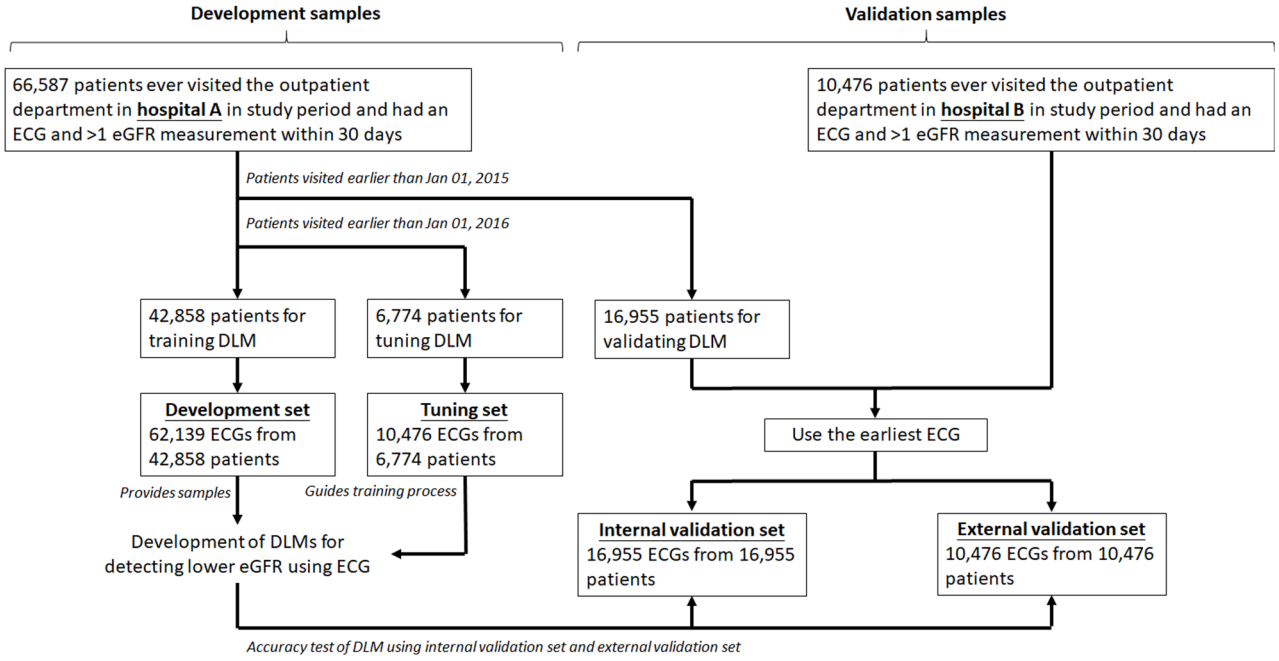

Fig. 1 Development, tuning, internal validation, and external validation set generation and ECG labeling of eGFR. Schematic of the dataset creation and analysis strategy that was devised to ensure a robust and reliable dataset for training, validating, and testing the network. Once a patient's data were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

m^2]. We acknowledge that this approach does not satisfy the persistence criterion in the KDIGO 2023 guideline [26], which defines CKD as eGFR < 60 or evidence of kidney damage persisting for at least 3 months. Due to the limitations of our dataset, longitudinal laboratory values were not available. As such, CKD prevalence—particularly in early-stage groups such as G1A2, G1A3, G2A2, and G2A3—may have been underestimated. This limitation is further discussed in the Discussion section.

Quantitative echocardiography data were recorded at the time of the ECG within 30 days using a Philips imaging system®. The disease histories were based on a new diagnosis according to the corresponding International Classification of Diseases, Ninth Revision and Tenth Revision (ICD-9 and ICD-10, respectively) or laboratory tests. They included diabetes mellitus (DM, ICD-9 codes 250.x and ICD-10 codes E11.x), hypertension (HTN, ICD-9 codes 401.x to 404.x, and ICD-10 codes I10.x to I16.x), hyperlipidemia (HLP, ICD-9 codes 272.x and ICD-10 codes E78.x), acute myocardial infarction (AMI, ICD-9 codes 410.x and ICD-10 codes I21.x), coronary artery disease (CAD, ICD-9 codes 410.x to 414.x, and 429.2, and ICD-10 codes I20.x to I25.x), stroke (STK, ICD-9 codes 430.x to 438.x and ICD-10 codes I60.x to I63.x), heart failure (HF, ICD-9 codes 428.x, 398.91, and 402.x1, and ICD-10 codes I50.x), and atrial fibrillation (Afib, ICD-9 codes 427.31 and ICD-10 codes I48.x). DM was defined as ≥ 2 fasting glucose results ≥ 126 mg/dL or HbA1c ≥ 6.5% within six months. HF was defined by ejection fraction ≤ 35%. Baseline labs were collected within 3 days of enrollment.

The primary outcome was CKD progression, as defined above. Secondary outcomes included all-cause mortality, as well as AMI, CAD, STK, HF, and Afib, assessed among patients without a prior history of these conditions. Data for living patients were censored at their last known hospital visit to limit bias from incomplete records.

Deep learning and machine learning models

The ECG12Net architecture with 82 convolutional layers and an attention mechanism was developed in our previous study [11]. We trained a new DLM for incident CKD detection using the ECG12Net model, which requires a 1,024-sequence input. Each 12-lead ECG signal originally contained 5,000 data points per lead recorded over 10 s at a sampling rate of 500 Hz. To accommodate the input length of the network while preserving computational efficiency, we randomly cropped 1,024-sequence segments from the raw ECG signal during training. This strategy enables the model to focus on localized temporal features associated with CKD. During inference, we extracted nine overlapping 1,024-sequence segments per ECG, and the final prediction was computed as the average of these outputs, thereby improving robustness and signal representativeness [27, 28]. To ensure adequate recognition of rare CKD cases, we applied an oversampling process, weighting each class based on CKD stage prevalence in the development cohort. We trained a

series of DLMs to identify stages 4 and 5 CKD and estimate eGFR, aiming to explore the relationship between eGFR and ECG patterns. These models were trained with a batch size of 32 and an initial learning rate of 0.001 using the Adam optimizer with standard parameters (β1 = 0.9, β2 = 0.999). The learning rate was reduced by a factor of 10 whenever the loss on the tuning set plateaued after an epoch. To prevent overfitting, early stopping was implemented by saving the model after each epoch and selecting the DLMs with the lowest validation loss. L2 regularization, with a coefficient of 10^{-4}, was the sole regularization method used in this study. The dataset was split randomly at the patient level to ensure no data leakage between the development, tuning, and validation sets. Hyperparameter tuning for the DLM was performed through a systematic manual search, guided by the model's performance (validation loss) on the tuning set. For reproducibility, a fixed random seed was used for all data splitting and model training processes.

We trained an eXtreme Gradient Boosting (XGB) model and an elastic net model using 31 diagnostic pattern classes and 8 ECG measurements to estimate CKD stage in the development set, enabling a comparison between ECG voltage-time traces and clinically reported ECG measures. The XGBoost model was implemented using the xgboost package in R (version 0.71.2). Based on performance on the tuning set, the final model was trained with the following key hyperparameters: nrounds = 300, max.depth = 3, and the objective was set to reg: squarederror for regression. Both models provided variable importance rankings, helping to explore relationships between specific features and eGFR. Since eGFR calculation incorporates sex and age—both validated as related to ECG—these factors were considered in model training [29]. To benchmark our model's performance, we compared the DLM trained on raw ECG voltage-time traces against these two machine learning models (XGB and elastic net), as well as a baseline model using only sex and age. We also trained two DLMs to predict sex and age using the same architecture and details in the development set. Moreover, due to the strong relationship between hyperkalemia and ECG [10, 11], a new DLM was trained to estimate the K + level. We also conducted a linear regression analysis to integrate the original estimated eGFR, sex, age, and potassium level and to further improve the accuracy.

DLM-augmented ECG to predict future incident CKD development

Based on DLM identification, patients in the internal and external validation sets were divided into "ECG-CKD" and "ECG-normal" subgroups. For estimation error analysis, we categorized patients by actual eGFR (< 60 mL/min/1.73 m^2 or ≥ 60 mL/min/1.73 m^2) and examined differences in clinical characteristics and ECG traits between the ECG-CKD and ECG-normal groups. We employed linear or logistic regression analysis for statistical testing as appropriate. We hypothesized that ECG signals would exhibit subtle abnormal patterns prior to the apparent onset of CKD. Our developed DLM might classify some of these cases as abnormal, leading to initial false positives (individuals classified as ECG-CKD despite actual eGFR ≥ 60 mL/min/1.73 m^2), which could later be confirmed as true positives. To test this hypothesis, we conducted a subgroup analysis of patients from the internal and external validation datasets meeting these criteria: (i) eGFR ≥ 60 mL/min/1.73 m^2; (ii) an ECG performed within 30 days predicted by the DLM to indicate CKD (DLM-predicted eGFR < 60 mL/min/1.73 m^2), representing a false positive; and (iii) availability of an additional eGFR measurement (not used for training or testing) at a later date. A control group was established using true negative cases where both CKD-EPI determinations and DLM-predicted eGFR indicated eGFR ≥ 60 mL/min/1.73 m^2. We conducted a Kaplan–Meier survival analysis to assess CKD incidence over time for both false positives and true negatives, using the models' optimal operating point. Data were censored based on the most recent encounter. Following this, we performed Cox proportional hazard regression analysis to estimate the hazard of incident CKD, adjusting for sex, age, and eGFR.

Statistical analysis

All statistical analyses were performed using R version 3.4.4, with a significance level set at $p < 0.05$. Patient characteristics were presented as means and standard deviations, counts, or percentages, and compared using analysis of variance (ANOVA) or the chi-square test as appropriate. Our primary focus was on identifying incident CKD in both internal and external validation sets. DLM probabilities generated receiver operating characteristic (ROC) curves and areas under the curve (AUCs) to evaluate diagnostic performance, with the operating point selected based on the maximum Yuden's index in the tuning set. In secondary analysis, we calculated Pearson's correlation coefficients (COR) and mean absolute error (MAE) to assess the relationship between DLM-predicted eGFR and actual eGFR computed using the CKD-EPI formula. We analyzed a series of models under various conditions to assess their predictive ability. This included examining the interactions between DLM-based ECG stratification (ECG-CKD or ECG-normal) and actual eGFR regarding the risk of outcomes in both internal and external validation datasets. The ECG-normal category combined with actual eGFR > 60 mL/min/1.73 m^2 served as the reference group. Models were adjusted for hospital, sex, age, and eGFR, with hazard

ratios (HRs) and 95% confidence intervals (CIs) reported for all data.

Results

The prevalence of CKD was 7.9% (1,343 out of 16,955) in the internal validation cohort and 13.6% (1,421 out of 10,476) in the external validation cohort. Table 1 summarizes patient characteristics across development, tuning, internal, and external cohorts. All ECGs in both validation datasets were evaluated using the DLM. Variances in kidney function, demographics, medical history, ECG data, and biochemistry among these cohorts may limit the DLM's generalizability by introducing spurious relationships. The DLM achieved strong performance in detecting CKD. As shown in Fig. 2a, the AUC was 0.885 for internal and 0.861 for external validation. At the optimal cut-off (model-predicted eGFR < 84.76 mL/min/1.73 m^2), sensitivity and specificity were 81.0% and 80.8% internally, and 85.8% and 68.2% externally. In the internal validation set, we compared the DLM with a baseline model using only sex and age, as well as XGB and elastic net models. The baseline model achieved an AUC of 0.881 (95% CI: 0.873–0.890). For XGB, the AUC was 0.820 (95% CI: 0.808–0.832) and improved to 0.896 (95% CI: 0.888–0.905) with age and sex included. Similarly, the elastic net model's AUC increased from 0.791 (95% CI: 0.778–0.804) to 0.892 (95% CI: 0.883–0.900) after adding these covariates. In the external validation set, the DLM consistently outperformed the XGB and elastic net models, regardless of whether sex and age were included. In both validation sets, it also showed superior performance in detecting severe CKD (eGFR < 30 and < 15 mL/min/1.73 m^2) using ECG voltage–time traces alone (Fig. 2b). Since eGFR is a continuous variable, we evaluated prediction accuracy using mean absolute error (MAE) and Pearson's correlation. As shown in Fig. 3, the mean absolute errors (MAEs) for the internal and external validations were 13.615 and 13.649, respectively, with Pearson's correlation coefficients of 0.642 and 0.667. After adjusting for sex and age, correlations improved to 0.727 in the internal validation set and 0.725 in the external validation set. Figure 4 shows that the DLM outperformed both XGB and elastic net models in eGFR prediction, regardless of adjustment for sex and age. Among all features, age contributed most to model performance. Figure 5 ranks important ECG traits in XGB and elastic net models, with prolonged PR/QT intervals, altered T waves, and RS wave axis among the top predictors.

Using combined validation samples, we analyzed clinical differences between ECG-CKD and ECG-normal groups. As shown in Fig. 6, patients classified as ECG-CKD exhibited more high-risk features—such as older age, higher systolic blood pressure, more comorbidities, worse echocardiographic findings, and abnormal lab values—regardless of actual eGFR. Figure 7 further shows that ECG abnormalities (e.g., prolonged PR and QRS intervals) were also more prevalent in the ECG-CKD group. The ECG-CKD group also had more frequent ECG abnormalities—such as prolonged QTc, lower RS wave axis, and higher T wave axis—based on XGB and elastic net models. We then performed a subgroup analysis in patients with baseline eGFR ≥ 60 mL/min/1.73 m^2 to evaluate long-term CKD development. Patients misclassified by the DLM as having eGFR < 60 (false positives) but with CKD-EPI-confirmed eGFR ≥ 60 had a significantly higher risk of developing CKD than true negatives, after adjusting for sex, age, and baseline eGFR (HR 2.14 and 1.38 in internal and external sets; Fig. 8). A stratified analysis of CKD stage progression (Fig. 9) showed that in patients with normal function or early-stage CKD (stages 1–2), the ECG-CKD subgroup had a higher risk of progression than ECG-normal. In advanced CKD, a significant difference was observed only in stage 3 in the internal validation set (HR 3.47, 95% CI: 1.61–7.44). No significance was seen in stages 4–5, likely due to small sample sizes and short follow-up.

We performed prognostic analyses using combined validation data stratified by both actual eGFR and DLM-based ECG predictions (Fig. 10). Among patients with actual eGFR ≥ 60, those classified as ECG-CKD (false positives) had significantly higher risks of all-cause mortality, AMI, STK, CAD, HF, and Afib (HRs ranging from 1.18 to 3.83) compared to ECG-normal (true negatives). In contrast, ECG-normal patients with true CKD (false negatives) showed increased risks of mortality and HF only. As expected, the highest risks across all outcomes were seen in true positives (ECG-CKD with eGFR < 60). Notably, DLM-based ECG stratification showed stronger prognostic value than actual eGFR, especially for stroke (HR 1.57 vs. 1.36), heart failure (HR 3.46 vs. 2.77), and Afib (HR 3.13 vs. 1.25).

Discussion

Our study demonstrates that a DLM, trained solely on 12-lead ECG voltage–time data, can accurately identify CKD without relying on additional clinical variables. The model achieved an AUC of 0.88 for CKD detection, outperforming both XGBoost and elastic net models. Importantly, the DLM identified patients with normal baseline eGFR who later developed CKD, highlighting its potential for early detection. Furthermore, the ECG-based CKD classification remained predictive of all-cause mortality and cardiovascular events, independent of traditional risk factors.

Prior research has shown only weak associations between CKD and ECG features like PR or QT intervals, especially after adjusting for confounders [30]. These inconsistencies likely stem from the nonlinear, complex

Table 1 Corresponding characteristics of patients in the development, tuning, internal validation, and external validation sets

	Development	Tuning	Internal validation	External validation	*p* value
Kidney function					
eGFR (mL/min/1.73 m^2)	89.6 ± 24.6	90.7 ± 25.8	93.6 ± 23.1	84.9 ± 23.9	< 0.001
CKD stage					< 0.001
Without	52,739 (84.9%)	9287 (85.1%)	14,779 (87.2%)	8568 (81.8%)	
Stage 1	1258 (2.0%)	264 (2.4%)	539 (3.2%)	243 (2.3%)	
Stage 2	1083 (1.7%)	177 (1.6%)	294 (1.7%)	244 (2.3%)	
Stage 3	5181 (8.3%)	775 (7.1%)	946 (5.6%)	1065 (10.2%)	
Stage 4	863 (1.4%)	127 (1.2%)	171 (1.0%)	180 (1.7%)	
Stage 5	1015 (1.6%)	289 (2.6%)	226 (1.3%)	176 (1.7%)	
Albuminuria					< 0.001
Without	55,527 (89.4%)	9719 (89.0%)	15,340 (90.5%)	9160 (87.4%)	
Micro/macro	6612 (10.6%)	1200 (11.0%)	1615 (9.5%)	1316 (12.6%)	
BUN (mg/dl)	16.3 ± 9.8	16.0 ± 10.7	15.1 ± 9.4	17.1 ± 10.6	< 0.001
Demographics					
Sex (male)	34,263 (55.1%)	5852 (53.6%)	9637 (56.8%)	5427 (51.8%)	< 0.001
Age (years)	54.2 ± 16.9	51.8 ± 17.1	50.1 ± 17.0	59.3 ± 16.4	< 0.001
BMI (kg/m^2)	24.8 ± 4.9	24.5 ± 4.1	24.6 ± 4.1	25.0 ± 4.6	< 0.001
SBP (mmHg)	127.4 ± 21.9	125.6 ± 22.2	125.3 ± 21.0	130.3 ± 22.7	< 0.001
DBP (mmHg)	78.9 ± 12.9	78.2 ± 12.9	78.3 ± 12.8	79.6 ± 13.5	< 0.001
Smoking/Ever	16,393 (26.4%)	2836 (26.0%)	4614 (27.2%)	2829 (27.0%)	0.053
Disease history					
DM	15,694 (25.3%)	1936 (17.7%)	2148 (12.7%)	2926 (27.9%)	< 0.001
HTN	19,084 (30.7%)	3576 (32.8%)	5311 (31.3%)	5258 (50.2%)	< 0.001
HLP	18,816 (30.3%)	2669 (24.4%)	2985 (17.6%)	4145 (39.6%)	< 0.001
AMI	1152 (1.9%)	185 (1.7%)	135 (0.8%)	97 (0.9%)	< 0.001
CAD	11,663 (18.8%)	2059 (18.9%)	2614 (15.4%)	2567 (24.5%)	< 0.001
STK	3517 (5.7%)	665 (6.1%)	771 (4.5%)	833 (8.0%)	< 0.001
HF	2499 (4.0%)	627 (5.7%)	609 (3.6%)	620 (5.9%)	< 0.001
Afib	2160 (3.5%)	388 (3.6%)	295 (1.7%)	400 (3.8%)	< 0.001
Echocardiography data					
LV-D (mm)	46.9 ± 6.5	46.8 ± 6.4	46.9 ± 6.3	47.1 ± 6.3	0.016
LV-S (mm)	29.1 ± 6.0	29.0 ± 6.0	29.0 ± 5.9	29.2 ± 5.9	0.047
IVS (mm)	10.6 ± 2.5	10.5 ± 2.5	10.5 ± 2.4	10.9 ± 2.5	< 0.001
LVPW (mm)	9.0 ± 1.6	8.9 ± 1.6	8.9 ± 1.6	9.1 ± 1.6	< 0.001
LA (mm)	37.2 ± 6.6	36.9 ± 6.7	36.7 ± 6.5	38.1 ± 6.7	< 0.001
AO (mm)	32.1 ± 4.3	31.8 ± 4.4	31.8 ± 4.2	32.6 ± 4.3	< 0.001
RV (mm)	23.7 ± 4.6	23.5 ± 4.7	23.5 ± 4.6	24.0 ± 4.7	< 0.001
PASP (mmHg)	31.0 ± 9.1	30.8 ± 9.6	30.2 ± 8.7	31.6 ± 9.5	< 0.001
EF (%)	67.0 ± 9.9	67.0 ± 9.9	67.2 ± 9.8	67.0 ± 9.6	0.195
Laboratory tests					
GLU (g/d)	105.7 ± 33.3	101.1 ± 29.4	98.9 ± 28.3	108.9 ± 33.4	< 0.001
Na^+ (mEq/L)	139.3 ± 3.1	139.3 ± 3.0	139.4 ± 2.9	139.1 ± 3.4	< 0.001
K^+ (mEq/L)	4.1 ± 0.4	4.1 ± 0.4	4.1 ± 0.4	4.1 ± 0.4	< 0.001
Alb (g/dl)	4.3 ± 0.4	4.3 ± 0.4	4.4 ± 0.4	4.3 ± 0.4	< 0.001
Hb (g/dL)	13.7 ± 1.9	13.7 ± 1.9	13.9 ± 1.8	13.6 ± 1.9	< 0.001
AST (U/L)	22.6 ± 20.9	22.1 ± 14.7	22.7 ± 32.0	23.4 ± 19.9	0.001
ALT (U/L)	23.2 ± 25.0	22.5 ± 21.8	23.7 ± 23.9	23.8 ± 29.2	< 0.001
TG (g/dl)	130.9 ± 100.7	126.0 ± 88.2	129.8 ± 101.8	135.8 ± 99.1	< 0.001
TC (g/dl)	182.5 ± 40.7	183.4 ± 38.6	186.1 ± 38.9	183.5 ± 40.6	< 0.001

Table 1 (continued)

	Development	Tuning	Internal validation	External validation	*p* value
LDL (g/dl)	110.2 ± 34.1	111.7 ± 33.1	114.6 ± 33.9	110.2 ± 34.2	< 0.001
HDL (g/dl)	51.2 ± 13.9	51.2 ± 13.7	51.0 ± 14.0	50.1 ± 13.9	< 0.001

Abbreviations: eGFR, estimated glomerular filtration rate; CKD, chronic kidney disease; BUN, blood urea nitrogen; BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; DM, diabetes mellitus; HTN, hypertension; HLP, hyperlipidemia; AMI, acute myocardial infarction; CAD, coronary artery disease; STK, stroke; HF, heart failure; Afib, atrial fibrillation; LV-D, left ventricle (end-diastole); LV-S, left ventricle (end-systole); IVS, interventricular septum; LVPW, left ventricular posterior wall; LA, left atrium; AO, aortic root; RV, right ventricle; PASP, pulmonary artery systolic pressure; EF, ejection fraction; GLU, fasting glucose; Na^+, sodium; K^+, potassium; Alb, albumin; Hb: hemoglobin; AST, aspartate aminotransferase; ALT, alanine aminotransferase; TG, triglyceride; TC, total cholesterol; LDL, low density lipoprotein

interactions between ECG signals and systemic diseases like CKD—interactions that conventional statistical methods may overlook. Our study focused on the ECG patterns and measurements integrated into the DLM. However, a key limitation is the model's relative opacity. Unlike conditions with more localized ECG markers, CKD's systemic nature means its effects on the ECG are often diffuse and non-specific. Consequently, standard visualization techniques like saliency maps may not yield clinically intuitive insights. While a deep mechanistic investigation was beyond the scope of this validation-focused study, it remains a critical direction for future research to fully understand the features driving the model's predictions. Prolonged corrected QT interval, prolonged PR interval, prolonged QRS wave duration, and lower RS wave axis were significantly more common in the ECG-CKD group, regardless of actual eGFR. This finding indicates that ECG signals may display subtle abnormalities before a significant decline in renal function occurs. While the XGB and elastic net models relied on similar ECG features, as shown in the correlation heatmap (Fig. 4), the DLM outperformed both models in CKD detection, even after adjusting for age and sex. We analyzed patients' demographics, comorbidities, abnormal cardiac structures, and laboratory variables using a combined sample from the internal and external validation sets. As shown in Fig. 6, risk factors for CKD, including advanced age, high systolic blood pressure, diabetes, hypertension, and increased albuminuria, were more common in the ECG-CKD group, regardless of actual eGFR. This supports the idea that the DLM can extract additional information from ECG signals.

The DLM not only identified individuals with CKD but also flagged patients with initially normal renal function (eGFR ≥ 60 mL/min/1.73 m^2) who were at risk of developing CKD. Although classified as "false positives," these patients had a significantly higher risk of CKD development over time compared to those with a negative DLM screen(Fig. 8). While some nonsignificant findings emerged from the stratified analysis (Fig. 9), the overall trend indicates that the DLM captures prognostic correlations of subclinical metabolic or structural abnormalities before renal function declines. Further research is needed to assess whether ongoing screening or treatment could benefit the false-positive group. While our study focused on ECG-based prediction without laboratory data, future studies could compare or integrate our model with established biomarker-based tools such as the Kidney Failure Risk Equation [31], particularly for predicting disease progression in patients with known CKD. Such hybrid models may offer complementary strengths, balancing accessibility and biochemical specificity.

CKD is a well-established independent risk factor for cardiovascular disease [32, 33] and related mortality [34, 35]. eGFR and albuminuria are standard metrics for CKD staging and cardiovascular risk assessment [34, 35], though their evaluation methods and the inclusion of cardiovascular outcomes can vary across studies. We evaluated all-cause mortality and cardiovascular outcomes using both actual eGFR and DLM-based ECG stratification. Incorporating ECG-CKD significantly enhanced risk stratification beyond traditional methods. We also found that "false-positive" patients—those predicted as ECG-CKD by the DLM but with actual eGFR ≥ 60 mL/min/1.73 m^2—had significantly higher risks of all-cause mortality and cardiovascular events. Although these individuals were not classified as having CKD by traditional biomarkers, their elevated risk suggests that further evaluation and closer management of cardiovascular risk factors may be warranted. In this context, the potential benefits of early intervention likely outweigh the minimal risks associated with false-positive predictions. Overall, DLM-based ECG stratification may serve as a complementary risk assessment tool, helping to identify high-risk individuals and estimate their cardiovascular risk more accurately. Holmstrom et al. developed a CNN-based model for CKD detection from ECGs, reporting accuracies of 0.767 and 0.709 in internal and external cohorts, with better performance in advanced CKD [23]. In contrast, our model demonstrated consistent performance across cohorts and identified higher CKD progression risk in ECG-CKD patients with normal or early-stage renal function, underscoring its potential for early detection.

Several limitations should be acknowledged. First, CKD status was defined based on a single measurement of eGFR and/or albuminuria, without confirmation of persistence over 3 months as recommended by the KDIGO

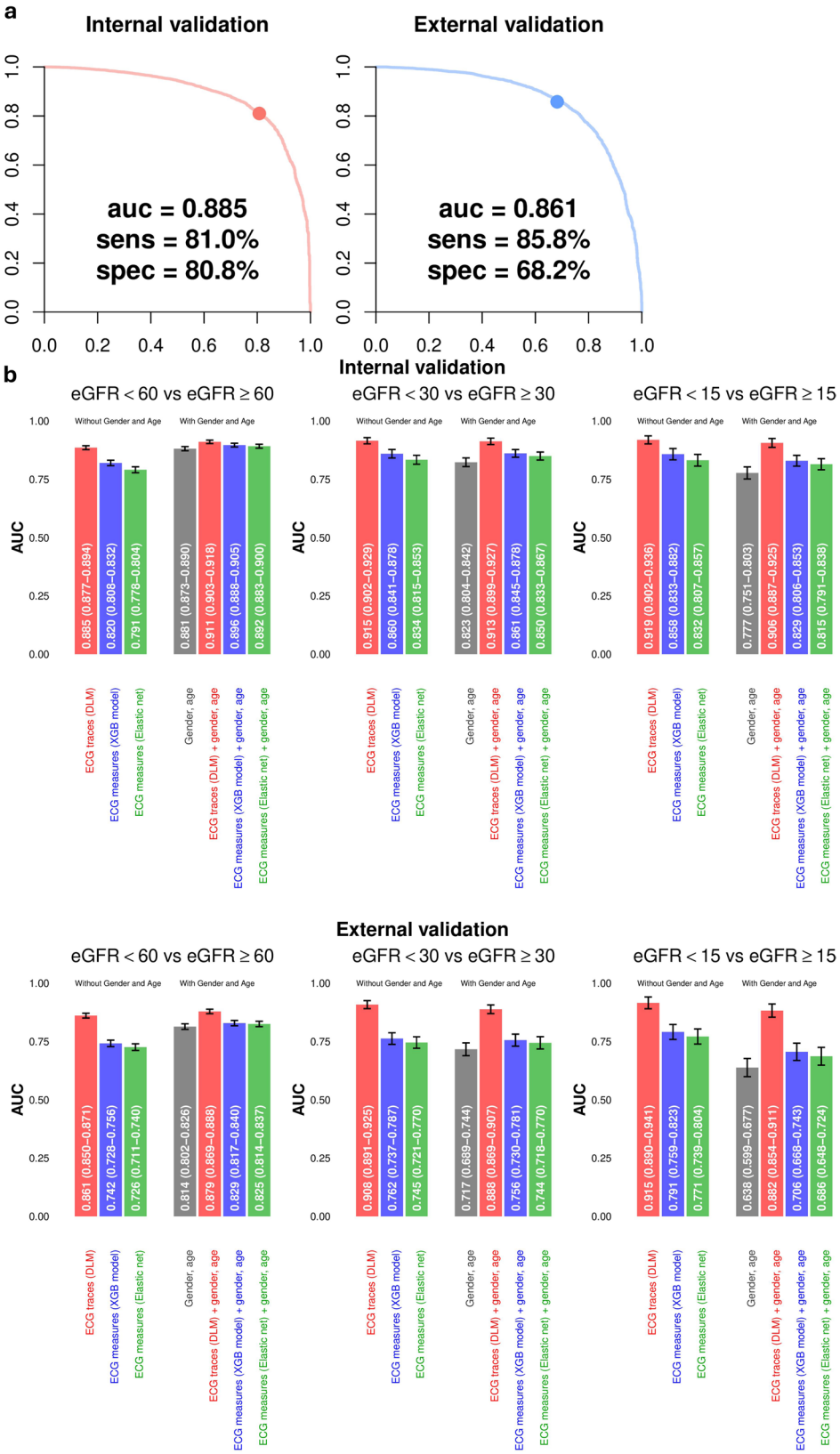

Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 **a**. The ROC of the deep learning model used to identify patients with an eGFR < 60 mL/min/1.73 m^2. The ROC curve (x-axis = specificity and y-axis = sensitivity) and AUC were calculated using the internal validation set and external validation set. The operating point was selected based on the maximum Youden's index in the tuning set, which was used to calculate the corresponding sensitivities and specificities in the two validation sets. **b**. Results obtained from the predictive algorithms as areas under the receiver operating characteristic curves for identifying CKD. The AUCs for the identification of CKD based on the indicated input data, including (i) using the DLM for analyzing ECG voltage–time traces, (ii) using the XGB model for analyzing clinically acquired ECG measures (8 numerical values and 31 diagnostic labels), (iii) using Elastic net for analyzing ECG measures, (iv) sex and age alone, (v) DLM with sex and age, (vi) XGB model with sex and age, (vii) Elastic net with sex and age. 'eGFR < 60 mL/min/1.73 m^2 vs. eGFR $\geq$ 60 mL/min/1.73 m^2', 'eGFR ≤ 30 mL/min/1.73 m^2 vs. eGFR >30 mL/min/1.73 m^2', and 'eGFR ≤ 15 mL/min/1.73 m^2 vs. eGFR >15 mL/min/1.73 m^2' refer to the patients in the test set separated by the original CKD-EPI equation at the time of ECG acquisition with an eGFR = 60, 30, and 15 mL/min/1.73 m^2, respectively

2023 guidelines. Previous studies have shown that reliance on single-timepoint data can result in substantial variability in CKD prevalence estimates and misclassification of transient kidney dysfunction, especially in early-stage or borderline cases [36, 37]. Second, laboratory data were allowed within 30 days of ECG acquisition to ensure data completeness, which may have introduced temporal mismatches between cardiac and renal function, further contributing to potential misclassification. Third, the study cohort consisted entirely of Taiwanese individuals. The eGFR values were calculated using the original CKD-EPI equation, which included race as a parameter prior to its 2021 revision. These factors may limit generalizability to racially diverse populations or cohorts using race-neutral eGFR equations. Fourth, the data used for model development spanned a 10-year period, during which changes in CKD definitions, laboratory assay methods, ECG device calibrations, and clinical practices may have occurred. This temporal drift and potential domain shift could affect model robustness. Although we attempted to mitigate this by using temporal validation, future studies using contemporary data and periodic re-training will be essential to maintain performance. Fifth, CKD prevalence in our development cohort was higher than in the general population, which may affect applicability in lower-risk groups. Lastly, as with most neural networks, interpretability remains limited. Although we explored model behavior using feature importance and correlation heatmaps, further work is needed to clarify the physiological basis of predictions and enhance clinical trust.

Conclusions

This study demonstrates that the DLM trained on 12-lead ECG data significantly enhances the detection of CKD and cardiovascular outcomes, as well as risk stratification. These findings support the DLM-based ECG analysis as a valuable tool for early CKD detection and cardiovascular risk assessment. Additionally, the 12-lead ECG combined with the DLM offers essential predictive insights, making it a powerful resource for identifying CKD and cardiovascular event risk.

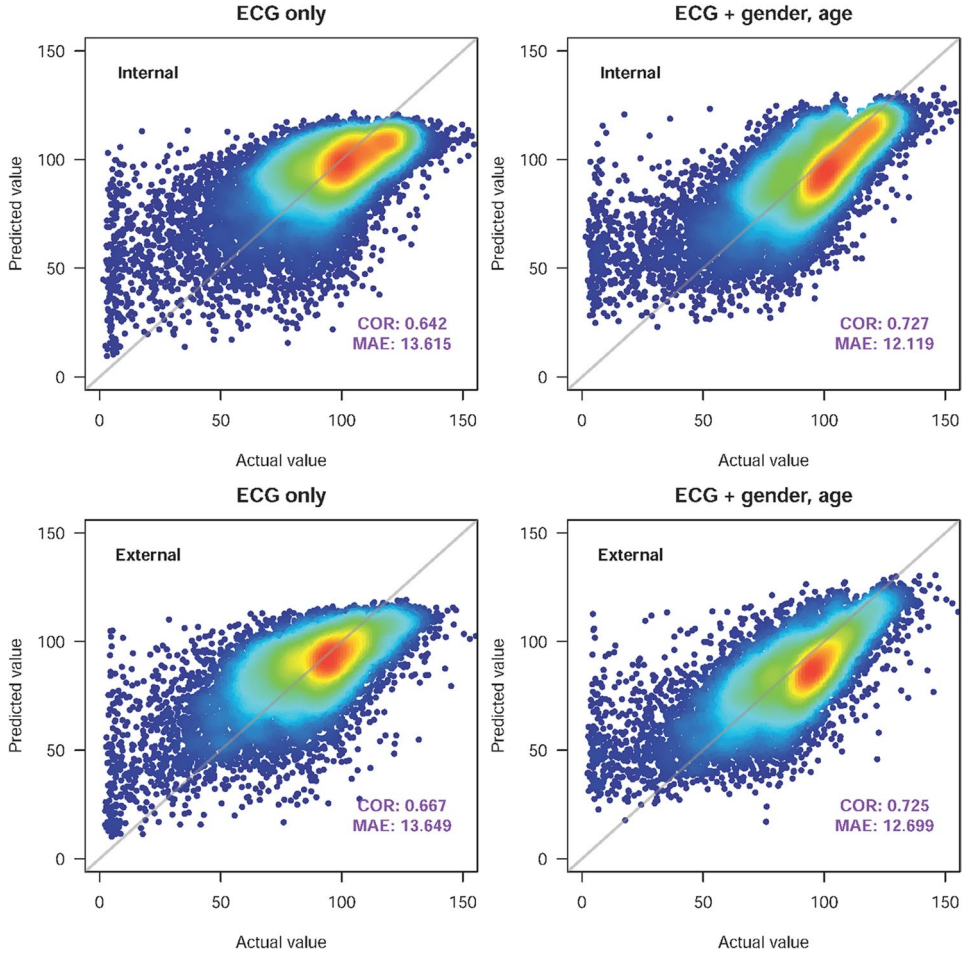

Fig. 3 Scatter plots of predicted and actual eGFR values. The x-axis indicates the actual eGFR, and the y-axis presents the DLM predictions. Red points represent the highest density, followed by yellow, green light blue, and dark blue. Perfect model performance would fall only along the diagonal line. We presented the Pearson correlation coefficients (COR) and mean absolute errors (MAE) to show the accuracy of the DLM in the two validation sets

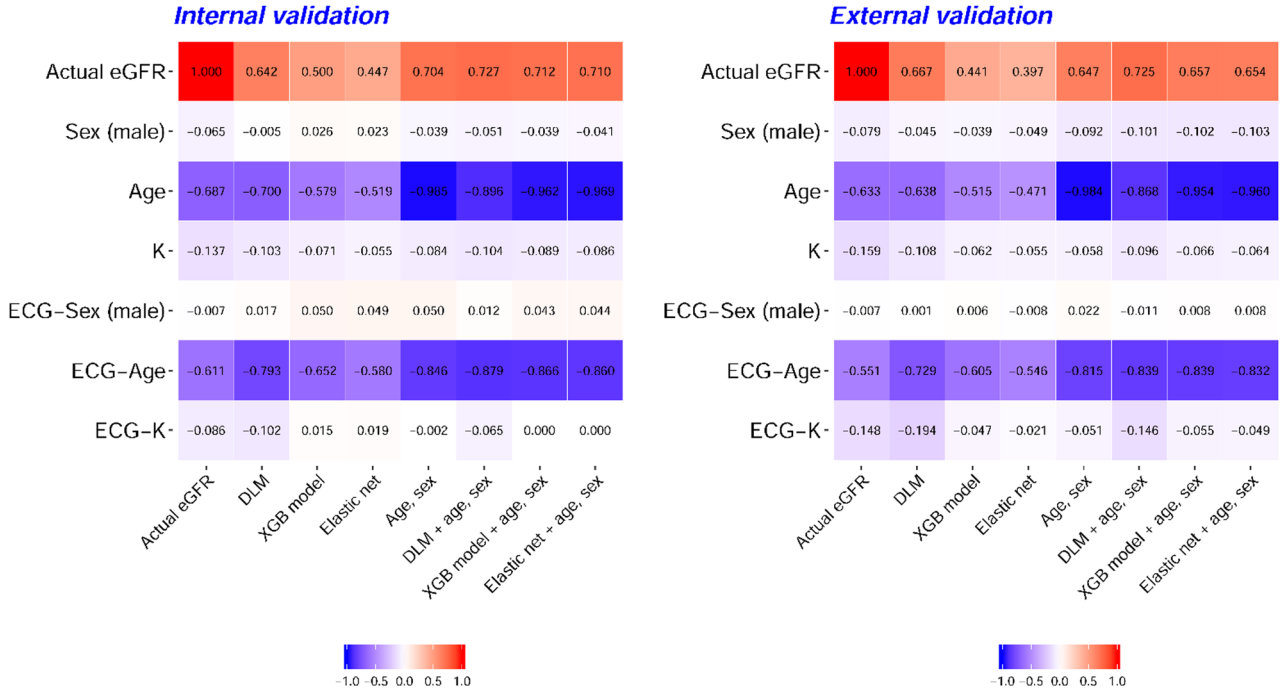

Fig. 4 The correlation heatmap of ECG-based eGFRs generated using different models in the two validation sets. Pearson's correlation coefficients for eGFR prediction and ECG-predictable features, including sex, age, and potassium level. The DLM consistently shows superior predictive performance compared to the XGB and elastic net models, regardless of sex and age adjustments. Age is the primary predictive factor for these ECG-based models

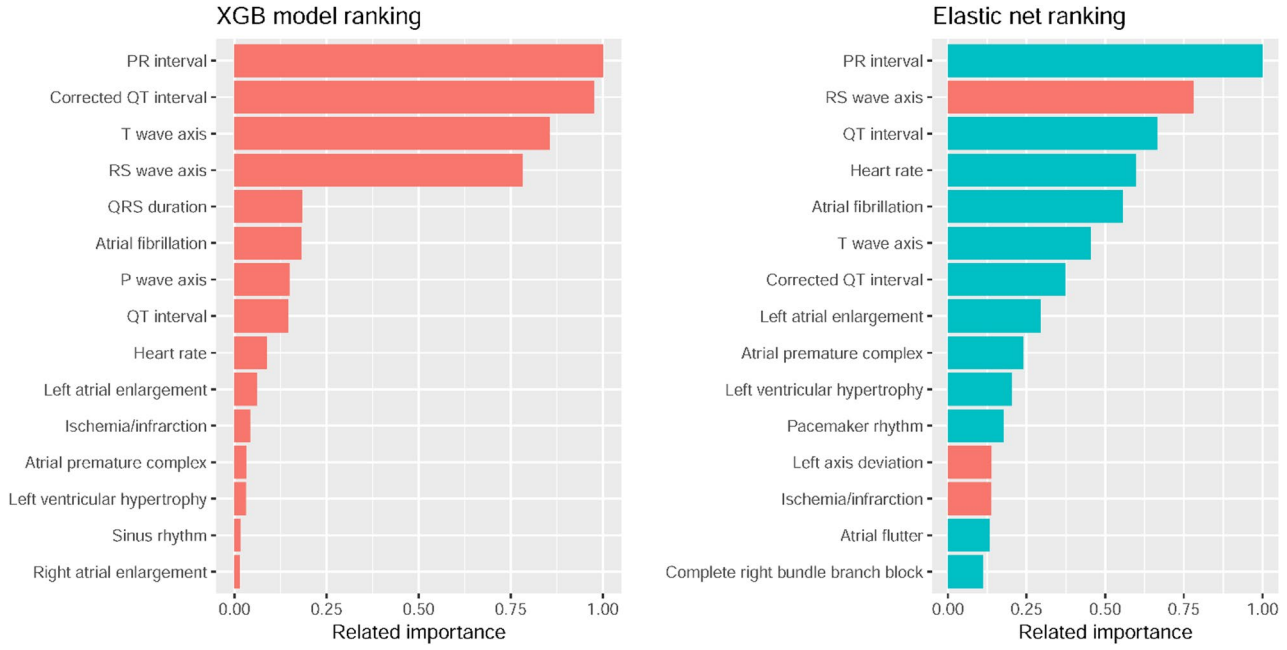

Fig. 5 Related feature importance ranking in the XGB model (information gain) and elastic net (standard coefficient). Only the top 15 important variables in each model are shown. The blue color indicates a negative relationship between these variables and CKD, while the red color indicates a positive relationship

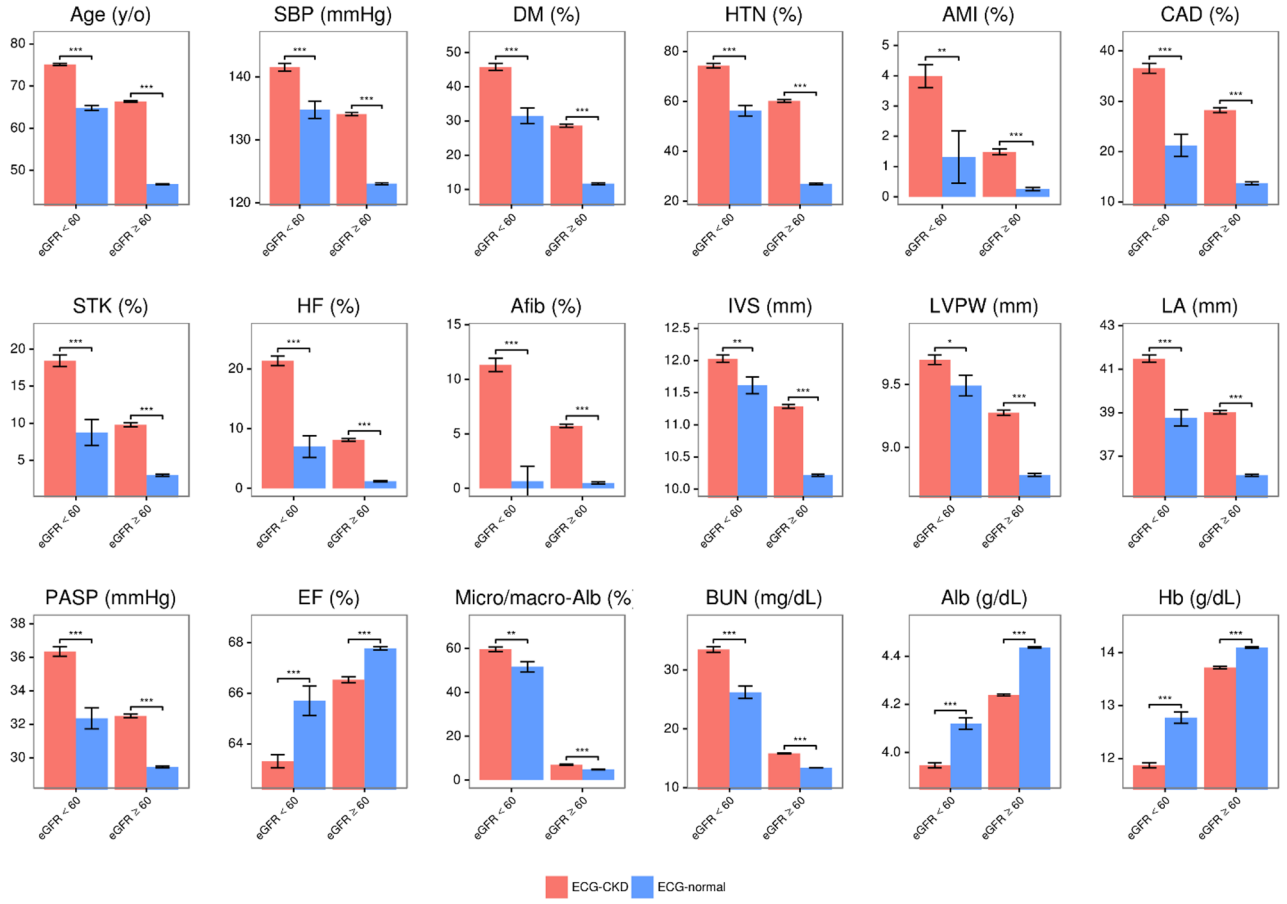

Fig. 6 Selected patient characteristics used in the DLM stratification of different eGFR groups. Bars represent the means or proportions, as appropriate, and corresponding 95% conference intervals, which were adjusted for the hospital, sex, age, and eGFR via a linear or logistic regression analysis (*: p for trend < 0.05; **: p for trend < 0.01; and ***: p for trend < 0.001). Abbreviations: eGFR, estimated glomerular filtration rate; CKD, chronic kidney disease; BUN, blood urea nitrogen; SBP, systolic blood pressure; DM, diabetes mellitus; HTN, hypertension; AMI, acute myocardial infarction; CAD, coronary artery disease; HF, heart failure; Afib, atrial fibrillation; IVS, interventricular septum; LVPW, left ventricular posterior wall; LA, left atrium; PASP, pulmonary artery systolic pressure; EF, ejection fraction; Alb, albumin; Hb: hemoglobin

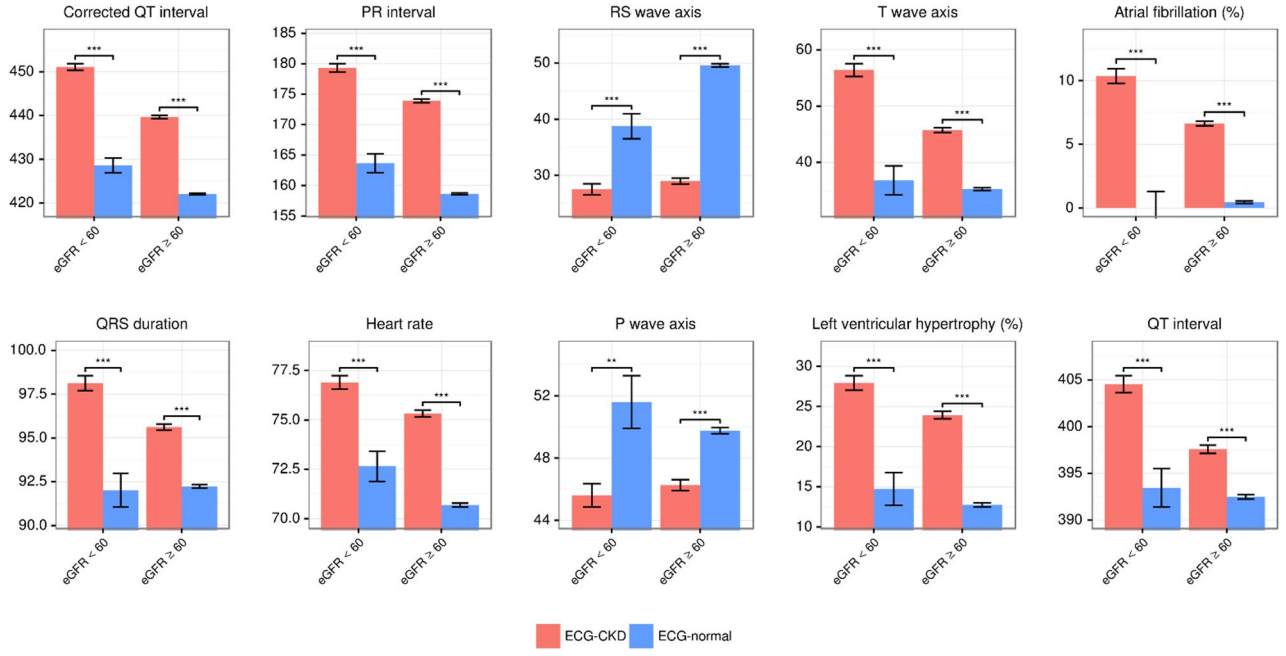

Fig. 7 Selected ECG measurements and ECG patterns in the patients with ECG-based CKD and normal controls in different eGFR groups. Bars represent the means or proportions, as appropriate, and corresponding 95% conference intervals, which were adjusted for the hospital, sex, age, and eGFR via a linear or logistic regression analysis (*: p for trend < 0.05; **: p for trend < 0.01; and ***: p for trend < 0.001)

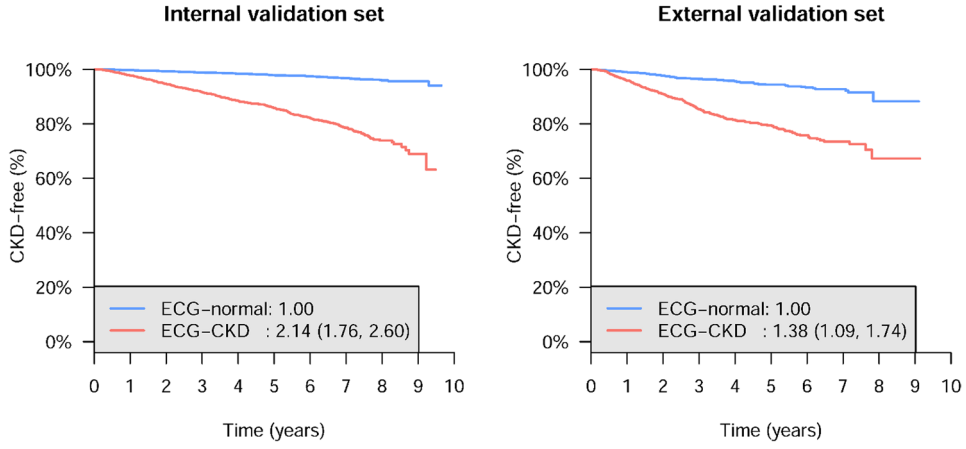

Fig. 8 Long-term incidence of developing incident CKD in patients with an initially normal eGFR stratified by the DLM classification. The ordinate shows the cumulative incidence of developing incident CKD (eGFR < 60 mL/min/1.73 m^2), and the abscissa indicates years from the time of the first ECG. The values of 2.14 (95% CI: 1.76–2.60) and 1.38 (95% CI: 1.09–1.74) for risks of future incident CKD were obtained when the AI algorithm defined the ECG as abnormal (adjusting for sex, age, and eGFR) in internal and external validation sets, respectively, compared with patients with a normal ECG classified by the DLM

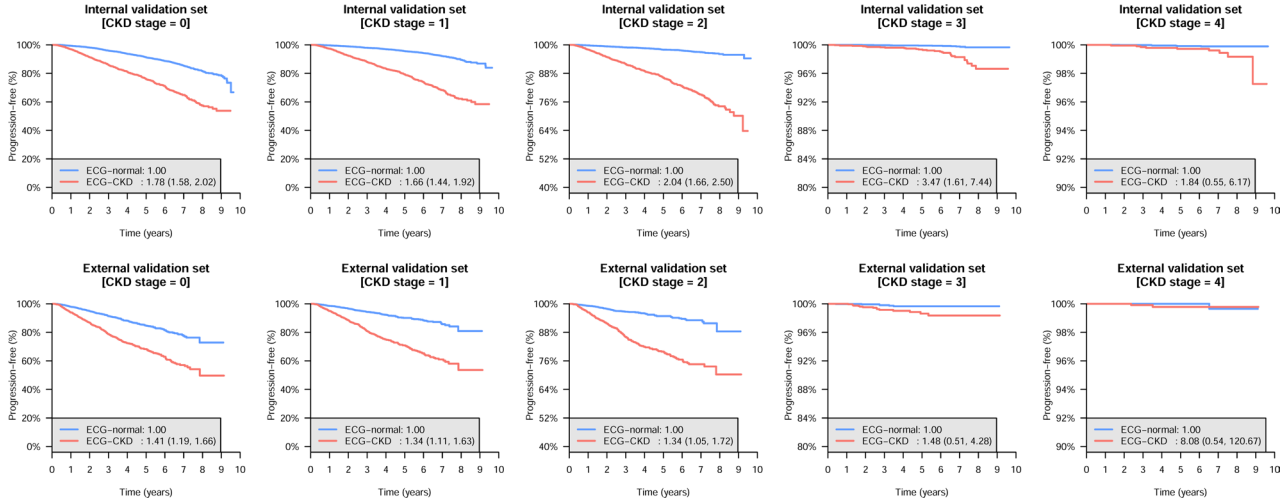

Fig. 9 Long-term incidence of stage progression in patients with different CKD stages stratified by AI classification. Long-term outcomes of patients with stage 0–4 CKD at the time of the initial classification, stratified by the initial DLM classification. The definitions of stage progression were the development of stage 1 in patients with stage 0 CKD, stage 2 in patients with stage 1 CKD, to stage 3 in patients with stage 2 CKD, stage 4 in patients with stage 3 CKD, and stage 5 in patients with stage 4 CKD

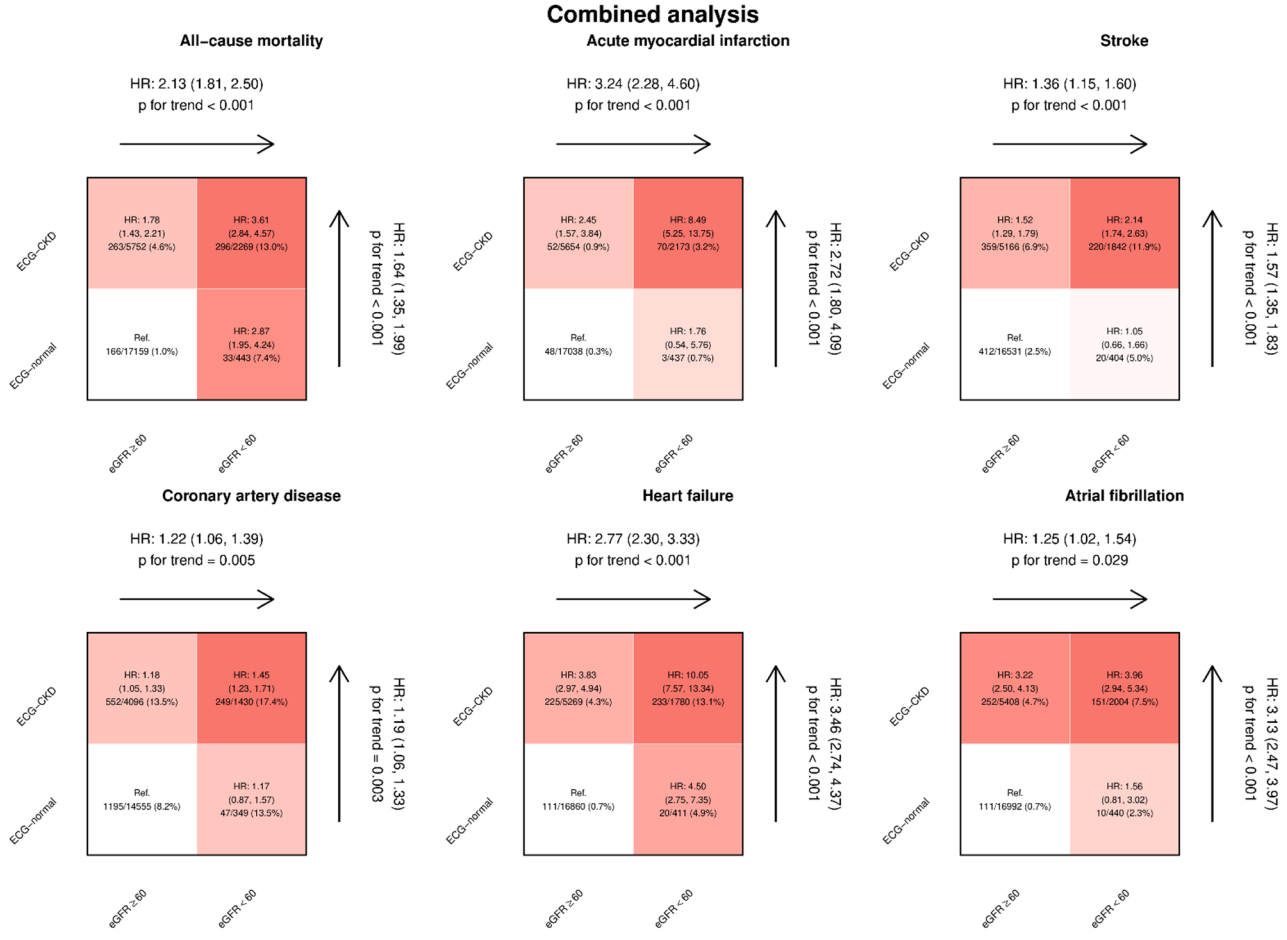

Fig. 10 Risk matrices of different DLM-based ECG stratification and eGFR groups for adverse events in the combined analysis. The hazard ratios (HRs) were based on the Cox proportional hazard model adjusted for hospital, sex, age, and eGFR. Nonsignificant results are depicted in white, and the color gradient represents the risk of the corresponding group. ECG-CKD significantly contributes to risk stratification for all-cause mortality and adverse cardiovascular outcomes

Acknowledgements
Not applicable.

Author contributions
YJH and CL conceptualised the study. CL and SYL prepared the data and figures and performed statistical analyses. PHT and CL interpreted the data and analyses. All authors have verified the underlying data, analyses, and interpretations. PHT wrote the first draft of the manuscript with input from CLHW. Critical revisions of the Article were provided by YJH and CL. All authors reviewed, contributed to, and approved the manuscript. All authors had final responsibility for the decision to submit for publication.

Funding
This study was supported by funding from the Ministry of Science and Technology, Taiwan (MOST 108-2314-B-016-017-MY3 to Y.J. Hsu, MOST 108-2314-B-016-001 to C. Lin, and MOST 109-2314-B-016-026 to C. Lin), the Tri-Service General Hospital, Taiwan (TSGH-A-111005 to Y.J. Hsu), the National Science and Technology Development Fund Management Association, Taiwan (MOST 108-3111-Y-016-009 and MOST 109-3111-Y-016-002 to C. Lin), and the Cheng Hsin General Hospital, Taiwan (CHNDMC114-11205 to C. Lin).

Data availability
The data analyzed in this study is not publicly available due to privacy and security concerns. The data may be shared with a third party upon execution of data sharing agreement for reasonable requests, such requests should be addressed to corresponding author C.L.

Declarations

Ethics approval and consent to participate
This study was approved by the Institutional Review Board of Tri-Service General Hospital (IRB No. C202105049) and conducted in accordance with the Declaration of Helsinki. The data were derived from de-identified medical records, and the institutional review board waived the requirement for signed informed consent.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 11 June 2025 / Accepted: 7 November 2025
Published online: 03 December 2025

References

1. Jha V, Garcia-Garcia G, Iseki K, Li Z, Naicker S, Plattner B, et al. Chronic kidney disease: global dimension and perspectives. Lancet. 2013;382(9888):260–72.
2. Global regional, national burden of chronic kidney disease. 1990–2017: a systematic analysis for the global burden of disease study 2017. Lancet. 2020;395(10225):709–33.
3. Kshirsagar AV, Bang H, Bomback AS, Vupputuri S, Shoham DA, Kern LM, et al. A simple algorithm to predict incident kidney disease. Arch Intern Med. 2008;168(22):2466–73.
4. Obermayr RP, Temml C, Knechtelsdorfer M, Gutjahr G, Kletzmayr J, Heiss S, et al. Predictors of new-onset decline in kidney function in a general middle-european population. Nephrol Dial Transpl. 2008;23(4):1265–73.
5. Nelson RG, Grams ME, Ballew SH, Sang Y, Azizi F, Chadban SJ, et al. Development of risk prediction equations for incident chronic kidney disease. JAMA. 2019;322(21):2104–14.
6. Fraccaro P, van der Veer S, Brown B, Prosperi M, O'Donoghue D, Collins GS, et al. An external validation of models to predict the onset of chronic kidney disease using population-based electronic health records from Salford, UK. BMC Med. 2016;14:104.
7. Majima S, Tanaka M, Okada H, Senmaru T, Asano M, Yamazaki M, et al. The PR interval and QRS duration could be predictors of renal function decline. Atherosclerosis. 2015;240(1):105–9.
8. Huang JC, Wei SY, Chen SC, Chang JM, Hung CC, Su HM, et al. P wave dispersion and maximum P wave duration are associated with renal outcomes in chronic kidney disease. PLoS ONE. 2014;9(7):e101962.
9. Lin CS, Liu WT, Tsai DJ, Lou YS, Chang CH, Lee CC, et al. AI-enabled electrocardiography alert intervention and all-cause mortality: a pragmatic randomized clinical trial. Nat Med. 2024.
10. Galloway CD, Valys AV, Shreibati JB, Treiman DL, Petterson FL, Gundotra VP, et al. Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram. JAMA Cardiol. 2019;4(5):428–36.
11. 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.
12. Lin C, Chau T, Lin CS, Shang HS, Fang WH, Lee DJ, 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.
13. Kwon J-m, Cho Y, Jeon K-H, Cho S, Kim K-H, Baek SD, et al. A deep learning algorithm to detect anaemia with ecgs: a retrospective, multicentre study. Lancet Digit Health. 2020;2(7):e358–67.
14. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med. 2019;25(1):65–9.
15. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet. 2019;394(10201):861–7.
16. Christopoulos G, Graff-Radford J, Lopez CL, Yao X, Attia ZI, Rabinstein AA, et al. Artificial Intelligence-Electrocardiography to predict incident atrial fibrillation: A Population-Based study. Circ Arrhythm Electrophysiol. 2020;13(12):e009355.
17. Bos JM, Attia ZI, Albert DE, Noseworthy PA, Friedman PA, Ackerman MJ. Use of artificial intelligence and deep neural networks in evaluation of patients with electrocardiographically concealed long QT syndrome from the surface 12-lead electrocardiogram. JAMA Cardiol. 2021.
18. 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.
19. Kwon J-m, Kim K-H, Eisen HJ, Cho Y, Jeon K-H, Lee SY, et al. Artificial intelligence assessment for early detection of heart failure with preserved ejection fraction based on electrocardiographic features. Eur Heart J - Digit Health. 2020.
20. Ko WY, Siontis KC, Attia ZI, Carter RE, Kapa S, Ommen SR, et al. Detection of hypertrophic cardiomyopathy using a convolutional neural Network-Enabled electrocardiogram. J Am Coll Cardiol. 2020;75(7):722–33.
21. Raghunath S, Ulloa Cerna AE, Jing L, vanMaanen DP, Stough J, Hartzel DN, et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. Nat Med. 2020;26(6):886–91.
22. Kwon JM, Kim KH, Jo YY, Jung MS, Cho YH, Shin JH, et al. Artificial intelligence assessment for early detection and prediction of renal impairment using electrocardiography. Int Urol Nephrol. 2022;54(10):2733–44.
23. Holmstrom L, Christensen M, Yuan N, Weston Hughes J, Theurer J, Jujjavarapu M, et al. Deep learning-based electrocardiographic screening for chronic kidney disease. Commun Med (Lond). 2023;3(1):73.
24. Binsawad M. Enhancing kidney disease prediction with optimized forest and ECG signals data. Heliyon. 2024;10(10):e30792.
25. van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. J Stat Softw. 2011;45(3):1–67.
26. KDIGO 2024 Clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int. 2024;105(4s):S117-s314.
27. Chang DW, Lin CS, Tsao TP, Lee CC, Chen JT, Tsai CS, et al. Detecting digoxin toxicity by artificial intelligence-assisted electrocardiography. Int J Environ Res Public Health. 2021;18(7).
28. Wen-Cheng L, Chin-Sheng L, Chien-Sung T, Tien-Ping T, Cheng-Chung C, Jun-Ting L, et al. A deep-learning algorithm for detecting acute myocardial infarction. EuroIntervention. 2021.
29. Attia ZI, Friedman PA, Noseworthy PA, Lopez-Jimenez F, Ladewig DJ, Satam G, et al. Age and sex Estimation using artificial intelligence from standard 12-Lead ECGs. Circ Arrhythm Electrophysiol. 2019;12(9):e007284.
30. Kestenbaum B, Rudser KD, Shlipak MG, Fried LF, Newman AB, Katz R, et al. Kidney function, electrocardiographic findings, and cardiovascular events among older adults. Clin J Am Soc Nephrol. 2007;2(3):501–8.
31. Chu CD, McCulloch CE, Hsu RK, Powe NR, Bieber B, Robinson BM, et al. Utility of the kidney failure risk equation and estimated GFR for estimating time to kidney failure in advanced CKD. Am J Kidney Dis. 2023;82(4):386–e941.

32. K/DOQI clinical. Practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Am J Kidney Dis. 2002;39(2 Suppl 1):S1–266.
33. Sarnak MJ, Levey AS, Schoolwerth AC, Coresh J, Culleton B, Hamm LL, et al. Kidney disease as a risk factor for development of cardiovascular disease: a statement from the American heart association councils on kidney in cardiovascular Disease, high blood pressure Research, clinical Cardiology, and epidemiology and prevention. Circulation. 2003;108(17):2154–69.
34. Association of estimated. Glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. Lancet. 2010;375(9731):2073–81.
35. Matsushita K, Coresh J, Sang Y, Chalmers J, Fox C, Guallar E, et al. Estimated glomerular filtration rate and albuminuria for prediction of cardiovascular outcomes: a collaborative meta-analysis of individual participant data. Lancet Diabetes Endocrinol. 2015;3(7):514–25.
36. Fernández-Llaneza D, Hilbrands LB, Vogt L, Engberink R, Klopotowska JE. Identifying a cohort of hospitalized chronic kidney disease patients using electronic health records: lessons learnt and implications for future research and clinical practice guidelines. Clin Kidney J. 2025;18(4):sfaf073.
37. Vestergaard SV, Christiansen CF, Thomsen RW, Birn H, Heide-Jørgensen U. Identification of patients with CKD in medical databases: A comparison of different algorithms. Clin J Am Soc Nephrol. 2021;16(4):543–51.

Publisher's note