ORIGINAL PAPER

# A Deep-Learning-Enabled Electrocardiogram and Chest X-Ray for Detecting Pulmonary Arterial Hypertension

Pang-Yen Liu[1,2] · Shi-Chue Hsing[1] · Dung-Jang Tsai[3,5] · Chin Lin[3,4,5] · Chin-Sheng Lin[1] · Chih-Hung Wang[6,7] · Wen-Hui Fang[3,8]

## Abstract

The diagnosis and treatment of pulmonary hypertension have changed dramatically through the re-defined diagnostic criteria and advanced drug development in the past decade. The application of Artificial Intelligence for the detection of elevated pulmonary arterial pressure (ePAP) was reported recently. Artificial Intelligence (AI) has demonstrated the capability to identify ePAP and its association with hospitalization due to heart failure when analyzing chest X-rays (CXR). An AI model based on electrocardiograms (ECG) has shown promise in not only detecting ePAP but also in predicting future risks related to cardiovascular mortality. We aimed to develop an AI model integrating ECG and CXR to detect ePAP and evaluate their performance. We developed a deep-learning model (DLM) using paired ECG and CXR to detect ePAP (systolic pulmonary artery pressure > 50 mmHg in transthoracic echocardiography). This model was further validated in a community hospital. Additionally, our DLM was evaluated for its ability to predict future occurrences of left ventricular dysfunction (LVD, ejection fraction < 35%) and cardiovascular mortality. The AUCs for detecting ePAP were as follows: 0.8261 with ECG (sensitivity 76.6%, specificity 74.5%), 0.8525 with CXR (sensitivity 82.8%, specificity 72.7%), and 0.8644 with a combination of both (sensitivity 78.6%, specificity 79.2%) in the internal dataset. In the external validation dataset, the AUCs for ePAP detection were 0.8348 with ECG, 0.8605 with CXR, and 0.8734 with the combination. Furthermore, using the combination of ECGs and CXR, the negative predictive value (NPV) was 98% in the internal dataset and 98.1% in the external dataset. Patients with ePAP detected by the DLM using combination had a higher risk of new-onset LVD with a hazard ratio (HR) of 4.51 (95% CI: 3.54–5.76) in the internal dataset and cardiovascular mortality with a HR of 6.08 (95% CI: 4.66–7.95). Similar results were seen in the external validation dataset. The DLM, integrating ECG and CXR, effectively detected ePAP with a strong NPV and forecasted future risks of developing LVD and cardiovascular mortality. This model has the potential to expedite the early identification of pulmonary hypertension in patients, prompting further evaluation through echocardiography and, when necessary, right heart catheterization (RHC), potentially resulting in enhanced cardiovascular outcomes.

**Keywords** Artificial Intelligence · Electrocardiogram · Chest X-ray · Deep learning · Pulmonary arterial hypertension

## Introduction

Pulmonary hypertension (PH), a condition characterized by elevated pulmonary artery pressure (ePAP), affects 1% of the global population and rises to 10% among individuals aged 65 and older [1]. Left untreated, pulmonary hypertension can lead to premature disability and death, primarily due to heart failure (HF). Over the past two decades, the field of pharmacological treatment for PH has undergone significant transformation with the development of novel medications, including prostacyclin analogs, endothelin receptor

antagonists, and phosphodiesterase 5 inhibitors [2]. These advancements have expanded the therapeutic options available to individuals with PH, significantly enhancing their quality of life. The condition is so insidious that there can be a delay of up to 2.25 years between the onset of symptoms and the diagnosis. At the time of presentation, 75% of patients already have advanced disease, classified under New York Heart Association (NYHA) functional class III or IV [3]. The current guidelines recommend that early detection of populations at risk is a crucial objective for improving outcomes. In cases where the prevalence rate is low and asymptomatic screening is not supported, population-based systemic strategies are not recommended. However, it is

---

Extended author information available on the last page of the article

Springer

suggested to apply early detection approaches in asymptomatic at-risk individuals or symptomatic patients in dyspnea clinics to facilitate early diagnosis [4, 5].

Chest X-ray radiography (CXR) is a cost-effective and readily available diagnostic tool for screening patients with ePAP [6, 7]. Recent advancements in deep learning models (DLM) have enabled the use of CXR to identify elevated mean pulmonary arterial pressure (PAP) based on data from right heart catheter (RHC) procedures and assess the risk of hospitalization due to HF in PH cases [8]. A recent study has demonstrated that a model based on electrocardiograms (ECG) can effectively identify individuals with ePAP (PAP estimated to be > 50 mm Hg by transthoracic echocardiography) and predict their future risk of cardiovascular mortality [9].

However, previous studies employing AI algorithms assessed pulmonary hypertension using a single diagnostic tool. Our study aimed to utilize an integrated AI model, incorporating both ECG and CXR data, to identify patients with elevated pulmonary arterial pressure (ePAP) and to provide prognostic predictions regarding their outcomes.

## Methods

### Data Source

This study, with the institutional ethics committee approval from the Tri-Service General Hospital (C202105049), conducted a retrospective investigation to develop a DLM and assess its performance through both internal and external validation. Patient consent was waived due to the retrospective nature of data collection. ECGs and CXRs were sourced from two hospitals, specifically an academic medical center (Hospital A) and a community hospital (Hospital B), covering the period from January 1, 2010, to April 30, 2021. Each ECG or CXR was annotated by the echocardiography within 7 days. Patients below 20 years of age were excluded from the study.

The ECG signal was recorded in a digital format with a sampling frequency of 500 Hz and 10 s for each lead. The CXR image was recorded in DICOM format with a resolution of more than $3000 \times 3000$ pixels.

The term "ePAP" was defined as systolic pulmonary artery pressure greater than 50 mmHg (indicated by a peak tricuspid regurgitation velocity greater than 3.4 m/s) through transthoracic echocardiography. This criterion indicates a high probability of PH in accordance with established studies and guidelines [4, 10, 11]. As illustrated in Fig. 1, we employed the following methods for developing and validating the DLM. We conducted a retrospective analysis involving patients who had at least one ECG or CXR within 7 days of an echocardiography exam. In Hospital A, a total of 85,193 patients were included in the study. Out of these, 42,499 patients were randomly assigned to the development set, which contributed 72,948 ECG records from 36,851 patients and 110,396 CXR records from 33,493 patients for training the DLM for detecting ePAP. Furthermore, 17,182 patients were randomly assigned to the tuning set, which provided 11,080 pairs of ECG and CXR to find hyperparameter optimization. The best combination of ECG and CXR models was determined by predictive value in each
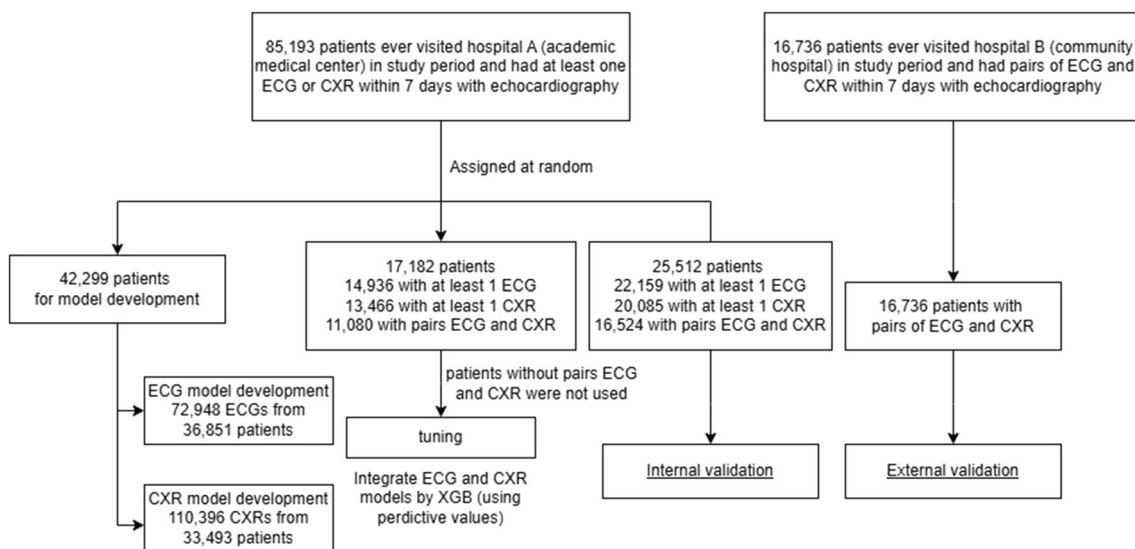


**Fig. 1** Development, tuning, internal validation, and external validation sets generation and labeling of echocardiography. We designed a schematic for the creation and analysis of the data set to ensure its robustness and reliability during network development, tuning, and validation. Each patient's data were exclusively assigned to one of the designated data sets, preventing any cross-contamination between sets

models. Finally, 25,512 patients were randomly assigned to the internal validation set for conducting accuracy tests and follow-up analyses. In order to assess the generalizability of the DLM, we also gathered data from Hospital B, involving 16,736 patients who met the same inclusion criteria as those from Hospital A, for external validation.

## Data Annotation and Variables

In this study, the medical histories of the patients were determined based on the International Classification of Diseases, Ninth Revision and Tenth Revision (ICD-9 and ICD-10). Here is how each medical condition was classified:

- Diabetes mellitus (DM, ICD-9 codes 250.x and ICD-10 codes E08.x to E13.x)
- Hypertension (HTN, ICD-9 codes 401.x to 404.x and ICD-10 codes I10.x to I16.x)
- Hyperlipidemia (ICD-9 codes 272.x and ICD-10 codes E78.x)
- Chronic kidney disease (CKD, ICD-9 codes 585.x and ICD-10 codes N18.x)
- Acute myocardial infarction (AMI, ICD-9 codes 410.x and ICD-10 codes I21.x)
- Stroke (ICD-9 codes 430.x to 438.x and ICD-10 codes I60.x to I63.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)
- Heart failure (HF, ICD-9 codes 428.x and ICD-10 codes I50.x)
- Atrial fibrillation (Af, ICD-9 codes 427.31 and ICD-10 codes I48.x)
- Chronic obstructive pulmonary disease (COPD, ICD-9 codes 490.x to 496.x and ICD-10 codes J44.9)

The primary outcomes of interest in this study were new-onset left ventricular dysfunction (LVD, EF ≤ 35%) and cardiovascular (CV) death. New-onset LVD was defined as the presence of at least one recorded estimated ejection fraction (EF) of ≤ 35%. For CV mortality, the survival time was calculated with reference to the date of the patient's ECG/CXR. The patients' status (whether deceased or alive) was determined through electronic medical records, which were regularly updated by each hospital. Data for patients who were still alive were censored at their last known hospital encounter to mitigate any potential bias arising from incomplete records.

## Model Development and Statistical Analysis

We employed the developed DLM using ECG data to predict ePAP values, which was based on a convolutional neural network with 82 trainable layers as described in our previous study [12]. The training details of the DLM for CXR were adapted from a previous study, which utilized a 121-layer DenseNet model [13]. To integrate the information obtained from the DLM for ECG and CXR in predicting ePAP, we utilized an eXtreme gradient boosting (XGB) model by predictive value in each models. These two DLMs were trained with a 32 batch size and used an initial learning rate of 0.001 using the Adam optimizer with standard parameters ($\beta1 = 0.9$ and $\beta2 = 0.999$). An oversampling process was implemented to ensure that the patients were adequately recognized. We sampled 16 cases and 16 controls in the development set for each batch. The learning rate was decayed by a factor of 10 each time the loss on the tuning set plateaued after an epoch. Early stopping was performed to prevent the networks from overfitting by saving the network after every epoch and choosing the saved DLMs with the lowest loss on the tuning set. L2 regularization was also applied to avoid overfitting.

We utilized the receiver operating characteristic (ROC) curve and the area under the curve (AUC) to assess the performance of the model. The operating point was selected based on the maximum Youden's index for the detection of ePAP in the tuning set, and this same operating point was applied for both internal and external validation to calculate sensitivity, specificity, positive predictive value, and negative predictive value. Furthermore, we employed multivariable Cox proportional hazard models to examine the relationship between AI predictions and the occurrence of new-onset left ventricular dysfunction (LVD, EF ≤ 35%) and cardiovascular (CV) outcomes. Sex and age-adjusted hazard ratios (HRs) and their corresponding 95% confidence intervals (95% CIs) were used for comparison, and Kaplan−Meier curve analysis was employed for visualization. All statistical analyses were conducted using the R software environment, version 3.4.4, with a significance level set at $p < 0.05$.

## Result

### Diagnostic Performance of the AI Model for ePAP

Table 1 displays patient characteristics across the development, tuning, internal validation, and external validation cohorts. The patients in the internal validation set were, on average, 65.2 ± 16.3 years old. 7.1% of the patients had ePAP (PAP > 50 mmHg), and 50.8% of the patients were male. In the external validation set, the average age of patients was 66.3 ± 17.2 years. 7.7% of the patients had ePAP, and 49.9% of the patients were male. The average pulmonary arterial systolic pressure in the "Internal validation" set is slightly lower at 32.9 ± 10.9 mmHg, while in the "External validation" set, it is slightly higher at 33.3 ± 11.1 mmHg.

**Table 1** Patient characteristics across the development, tuning, internal validation, and external validation cohorts

| | Development | | Tuning | Internal validation | External validation |
|---|---|---|---|---|---|
| | ECG subset | CXR subset | | | |
| PASP (mmHg) | 33.8 ± 11.9 | 35.5 ± 13.0 | 33.0 ± 11.0 | 32.9 ± 10.9 | 33.3 ± 11.1 |
| < 30 mmHg | 29,374 (40.3%) | 40,024 (36.3%) | 4660 (42.1%) | 7009 (42.4%) | 7001 (41.8%) |
| 30–39 mmHg | 27,365 (37.5%) | 38,208 (34.6%) | 4241 (38.3%) | 6333 (38.3%) | 6265 (37.4%) |
| 40–49 mmHg | 9721 (13.3%) | 18,513 (16.8%) | 1386 (12.5%) | 2000 (12.1%) | 2178 (13.0%) |
| 50–59 mmHg | 3755 (5.1%) | 7758 (7.0%) | 468 (4.2%) | 697 (4.2%) | 761 (4.5%) |
| ≥ 60 mmHg | 2733 (3.7%) | 5893 (5.3%) | 325 (2.9%) | 485 (2.9%) | 531 (3.2%) |
| Demography | | | | | |
| Gender (male) | 38,798 (54.9%) | 62,515 (56.6%) | 5814 (52.5%) | 8399 (50.8%) | 8357 (49.9%) |
| Age (years) | 65.7 ± 17.0 | 68.4 ± 16.1 | 65.2 ± 16.3 | 65.2 ± 16.2 | 66.3 ± 17.2 |
| Height (cm) | 162.2 ± 9.1 | 161.9 ± 9.0 | 161.9 ± 9.0 | 161.7 ± 9.0 | 162.0 ± 9.0 |
| Weight (kg) | 64.7 ± 14.2 | 63.2 ± 14.1 | 64.2 ± 14.0 | 64.3 ± 14.1 | 64.6 ± 14.1 |
| BMI (kg/m$^2$) | 24.5 ± 4.4 | 24.0 ± 4.5 | 24.5 ± 4.4 | 24.5 ± 4.4 | 24.5 ± 4.4 |
| Disease history | | | | | |
| DM | 19,959 (28.2%) | 34,486 (31.2%) | 2512 (22.7%) | 3688 (22.3%) | 5072 (30.3%) |
| HTN | 6702 (9.5%) | 9737 (8.8%) | 609 (5.5%) | 919 (5.6%) | 1565 (9.4%) |
| HLP | 24,267 (34.3%) | 33,148 (30.0%) | 2907 (26.2%) | 4384 (26.5%) | 7160 (42.8%) |
| CKD | 18,947 (26.8%) | 36,295 (32.9%) | 1903 (17.2%) | 2804 (17.0%) | 3530 (21.1%) |
| AMI | 5771 (8.2%) | 6864 (6.2%) | 425 (3.8%) | 625 (3.8%) | 522 (3.1%) |
| Stroke | 11,941 (16.9%) | 22,792 (20.6%) | 1664 (15.0%) | 2398 (14.5%) | 3129 (18.7%) |
| CAD | 23,244 (32.9%) | 30,747 (27.9%) | 2424 (21.9%) | 3595 (21.8%) | 5168 (30.9%) |
| HF | 12,285 (17.4%) | 24,126 (21.9%) | 1112 (10.0%) | 1580 (9.6%) | 2408 (14.4%) |
| Afib | 6601 (9.3%) | 10,298 (9.3%) | 590 (5.3%) | 852 (5.2%) | 1274 (7.6%) |
| COPD | 10,854 (15.4%) | 19,422 (17.6%) | 1349 (12.2%) | 2027 (12.3%) | 3796 (22.7%) |

Abbreviations: *PASP* pulmonary artery systolic pressure, *BMI* body mass index, *DM* diabetes mellitus, *HTN* hypertension, *HLP* hyperlipidemia, *CKD* chronic kidney disease, *AMI* acute myocardial infarction, *CAD* coronary artery disease, *HF* heart failure, *Afib* atrial fibrillation, *COPD* chronic obstructive pulmonary disease

However, these differences are not statistically significant. In the "External validation" set, there are significantly higher prevalences of diabetes mellitus (DM), hyperlipidemia, hypertension, coronary artery disease (CAD), and chronic obstructive pulmonary disease (COPD) compared to the "Internal validation" set. These differences suggest variations in the prevalence of certain medical conditions between the internal and external validation sets.

Figure 2 illustrates DLM predictions based on ECG, CXR, and combination to detect ePAP in both the internal validation and external validation sets. ROC curves were generated, presenting sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The AUCs for detecting ePAP were 0.8261 by ECG, 0.8525 by CXR, and 0.8644 by combination, respectively, in the internal validation set. In the external validation set, the AUCs for detecting ePAP were 0.8348 by ECG, 0.8605 by CXR, and 0.8734 by combination, respectively. The DLM exhibited ePAP detection through ECGs with a sensitivity of 76.6%, specificity of 74.5%, positive predictive value of 18.8%, and high negative predictive value of 97.6% in the internal validation cohort. The DLM for detecting ePAP by CXR achieved a sensitivity of 82.8%, a specificity of 72.7%, and a notably high negative predictive value of 98.2% in the internal validation cohort. The combination of ECGs and CXR achieved a sensitivity of 78.6%, a specificity of 79.2%, and a negative predictive value of 98% in the internal validation cohort for detecting ePAP. In the external validation cohort, the Diagnostic Learning Model (DLM) achieved the following results for detecting elevated pulmonary arterial pressures (ePAPs): ECG exhibited a sensitivity of 78.9%, specificity of 73.7%, PPV of 20%, and a high NPV of 97.7%. CXR achieved a sensitivity of 85.4%, a specificity of 71.8%, and a NPV of 98.3%. The combination of ECG and CXR yielded a sensitivity of 81.6%, a specificity of 78.1%, a PPV of 23.8%, and a NPV of 98.1%.

In Fig. 3, the AI diagnostic performance in the combined use of ECG and CXR was strong in the internal validation set across different genders and various comorbidities. However, in specific subgroups such as the elderly (≥ 65-year-old), those with chronic kidney disease (CKD), HF, and Af, the diagnostic performance (measured by the AUC) in the

**Fig. 2** The ROC curve of DLM predictions based on ECG, CXR, and combination to detect ePAP. The cut-off point was chosen based on the maximum value of Yunden's index in the tuning set and indicated by a circle mark. Subsequently, the area under the ROC curve (AUC), sensitivity (Sens.), specificity (Spec.), positive predictive value (PPV), and negative predictive value (NPV) were calculated based on this selected cut-off point
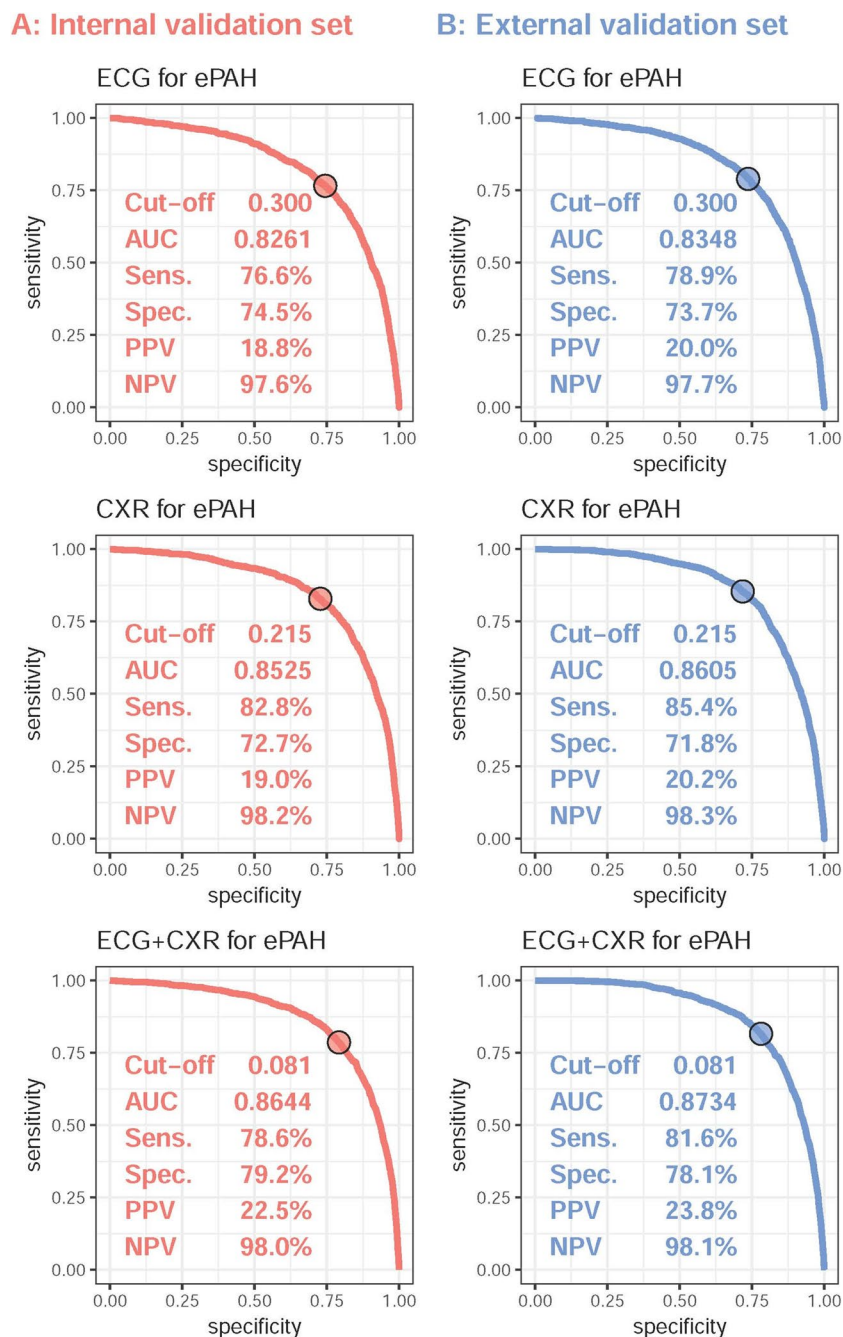


DLM predictions based on ECG, CXR, and their combination was significantly reduced in the internal validation. Additionally, in cases of chronic obstructive pulmonary disease (COPD), the diagnostic performance in CXR-based DLM predictions was decreased. Elderly ($\geq$ 65-year-old), CKD, DM, HTN, stroke, HF, Af, and COPD were significantly association with a reduction of the diagnostic performance in the DLM predictions based on ECG, CXR, and their combination in the external validation. In cases of CAD, the diagnostic performance in DLM predictions by ECG and combination was decreased.

We conducted stratified analyses to compare the AUC in the detection of ePAH using DLM predictions based on ECG, CXR, and their combination. These analyses were stratified by demographic characteristics and the medical history of the populations. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

## Long-Term Cardiovascular Outcome Predicted by Multimodal AI in Patients Initially Without ePAP

New onset of left ventricular dysfunction (LVD, EF $\leq$ 35%) was analyzed for patients in 3 subsets (AI-ECG, AI-CXR,
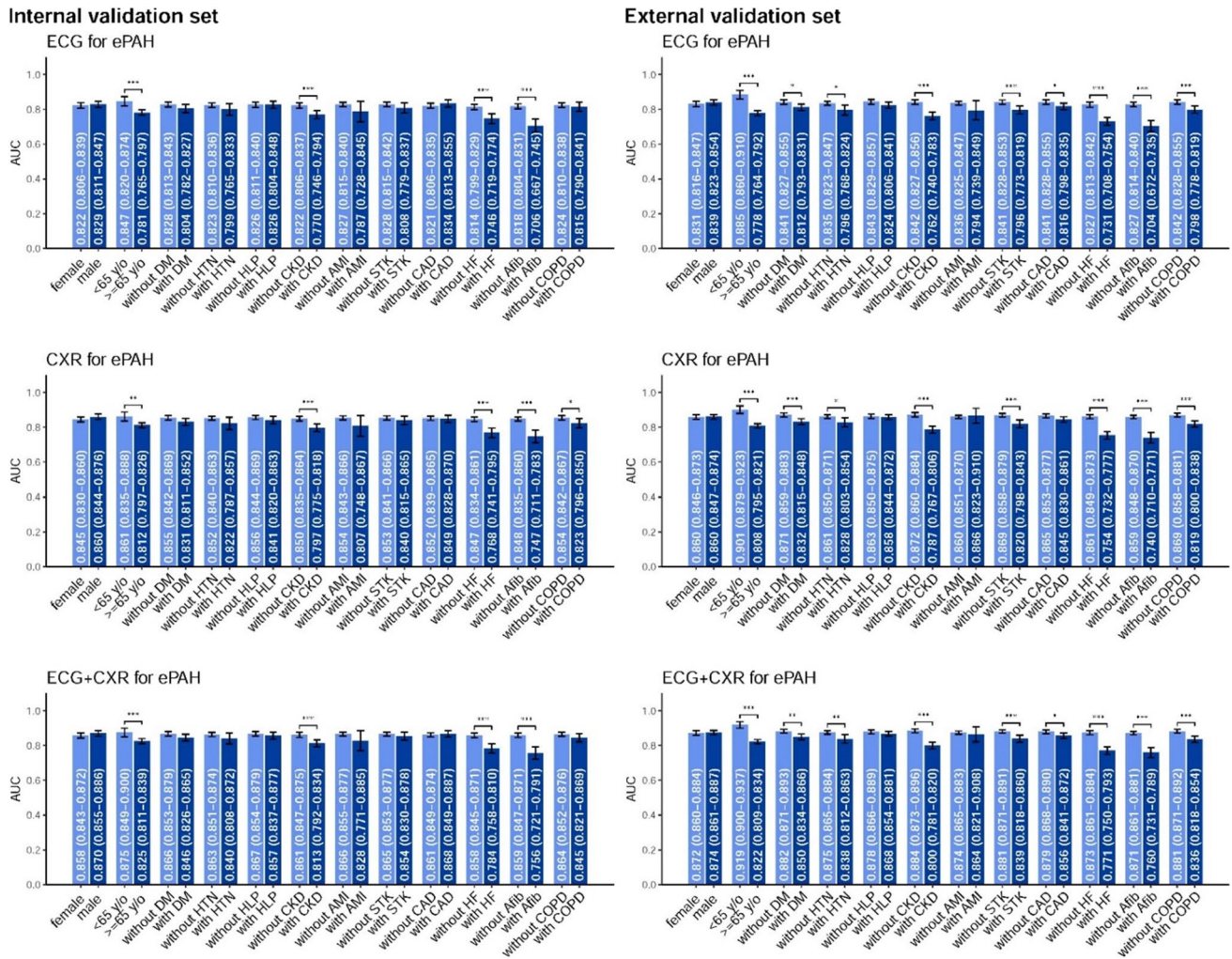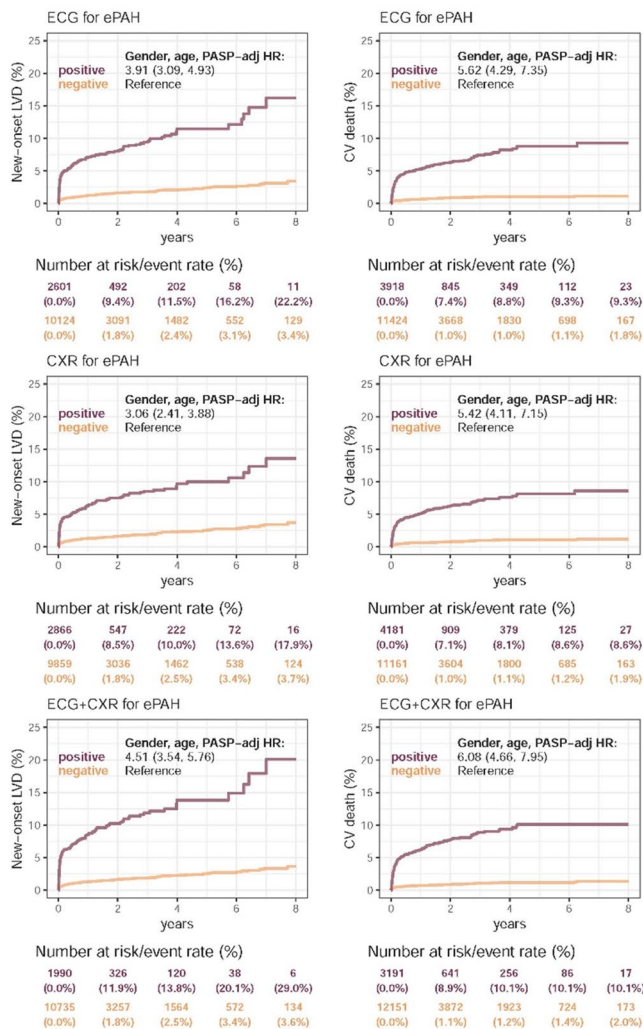
**Fig. 3** The association of subgroups and the DLM diagnostic performance

and AI-ECG + CXR) in Fig. 4. Of the 10,124 patients labeled by the AI-ECG model as having a normal PAPs who had a documented normal PAPs and normal EF by TTE within a 1-week window (Fig. 4, left upper, the cure labeled negative, AI-ECG-predicted non-ePAPs), the cumulative incidence of new onset of LVD at each time period were 1.8% in 2 years, 2.4% in 4 years, 3.1% in 6 years, and 3.4% in 8 years. By contrast, for the 2601 patients labeled by the AI-ECG as having ePAP but with normal PAPs and normal EF by TTE (the curve labeled positive, AI-ECG-predicted ePAP), the cumulative incidence of LVD were 9.4% in 2 years, 11.5% in 4 years, 16.2% in 6 years, and 22.2% in 8 years. This marked difference was observed with an adjusted hazard ratio (HR) of 3.91 (95% CI: 3.09–4.93). Of the 9859 patients labeled by the AI-CXR model as having normal PAPs who had documented normal PAPs and normal EF by TTE (AI-CXR-predicted non-ePAP), the cumulative incidence of LVD were 1.8% in 2 years, 2.5% in 4 years, 3.4% in 6 years, and 3.7% in 8 years. By contrast, for the

2866 patients labeled by the AI-CXR as having ePAP but with normal PAPs and normal EF by TTE (AI-CXR-predicted ePAP), the cumulative incidence of LVD were 8.5% in 2 years, 10% in 4 years, 13.6% in 6 years, and 17.9% in 8 years. This marked difference was observed with an adjusted hazard ratio (HR) of 3.06 (95% CI: 2.41–3.88). Of the 10,735 patients labeled by the combination of ECG and CXR as having normal PAPs who had documented normal PAPs and normal EF by TTE (AI-ECG + CXR-predicted non-ePAPs), the cumulative incidence of LVD were 1.8% in 2 years, 2.5% in 4 years, 3.4% in 6 years, and 3.6% in 8 years. By contrast, for the 1990 patients labeled by the AI-ECG + CXR as having ePAP but with normal PAP and normal EF by TTE (AI-CXR-predicted ePAP), the cumulative incidence of LVD were 11.9% in 2 years, 13.8% in 4 years, 20.1% in 6 years, and 29% in 8 years (Fig. 4, left lower). This marked difference was observed with an adjusted hazard ratio (HR) of 4.51 (95% CI: 3.54–5.76). The hazard ratio was the largest in the AI-ECG + CXR subset.
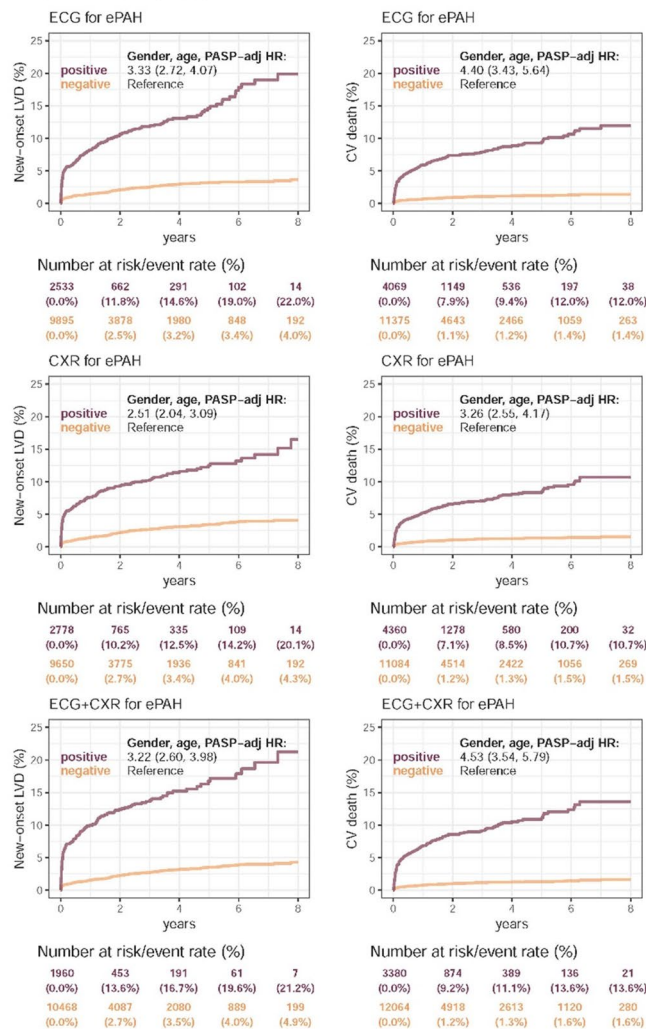
**Fig. 4** Long-term incidence of developing new-onset left ventricular dysfunction (LVD, EF≤35%) and cardiovascular (CV) death, stratifying patients based on the presence of ePAP as determined by DLM. These analyses were performed in both the internal and external validation sets. It is important to note that the analysis for new-onset LVD only included patients with an initial EF>50%. The table presents information on the at-risk population and cumulative risk for specific time intervals within each risk stratification

Cardiovascular mortality was analyzed for patients in 3 subsets (AI-ECG, AI-CXR, and AI-ECG+CXR). During the 8-year follow-up, cumulative incidence analysis showed that in comparison to patients labeled as non-ePAP by the AI model, those stratified as ePAP were associated with higher cardiovascular mortality. The hazard ratios with 95% confidence intervals were 5.62 (4.29–7.35), 5.42 (4.11–7.15), and 6.08 (4.66–7.95) in AI-ECG, AI-CXR, and AI-ECG+CXR subset, respectively. The hazard ratio was the largest in the AI-ECG+CXR subset, and the prediction of 2-year cardiovascular mortality from multimodal AI was a significant discriminator of long-term cardiovascular mortality for up to 8 years after the initial ECG and CXR (Fig. 4; the second row from the left). Similar prediction and performance for LVD and cardiovascular mortality were replicated in the external cohort. The hazard ratios for new onset of left ventricular dysfunction (LVD, EF≤35%) with 95% confidence intervals were 3.33 (2.72–4.07) in the AI-ECG subset, 2.51 (2.04–3.09) in the AI-CXR subset, and 3.22 (2.60–3.98) in the AI-ECG+CXR subset, respectively. The hazard ratios for subsequent cardiovascular mortality, with 95% confidence intervals, were 4.03 (3.43–5.64) in the AI-ECG subset, 3.26 (2.55–4.17) in the AI-CXR subset, and 4.53 (3.54–5.79) in the AI-ECG+CXR subset, respectively. Notably, the largest hazard ratio was observed in the AI-ECG+CXR subset.

## Discussion

In the present study, the multimodal AI model recaptured patients with ePAP and predicted their future risk for developing new-onset LVD and cardiovascular mortality. Combination of ECG and CXR performed more powerfully than either in diagnosis, but not overt. In the internal validation set, the AUCs for detecting ePAP were 0.8261 for ECG, 0.8525 for CXR, and 0.8644 for the combination of both. In the external validation set, the AUCs for detecting ePAP were 0.8348 for ECG, 0.8605 for CXR, and 0.8734 for the combination. Furthermore, the combination method demonstrated a high NPV, reaching 98% in the internal validation set and 98.1% in the external validation set. Due to the high negative predictive value, the utilization of DLM may prompt further evaluation with echocardiography and lead to a reduction in the number of unnecessary RH procedures. On the other hand, the diagnostic performance was adversely affected by certain comorbidities such as advanced age, CKD, HF, and Af. In patients who were initially found to have normal LV systolic function, those predicted as positive for ePAP by the combination method were independently associated with the development of future LVD (EF $\leq$ 35%) with a HR of 4.51 (95% CI: 3.54–5.76) in the internal dataset. Additionally, they were also independently associated with a higher risk of cardiovascular mortality (HR: 6.08, CI: 4.66–7.95) at an 8-year follow-up. The AI model identified ECG/CXR abnormalities and thus predicted LVD and cardiovascular mortality. The AI-ECG + CXR was more powerful than AI-ECG or AI-CXR, but not overt.

This model has the potential to be valuable in clinical settings for screening patients with pulmonary hypertension due to its high NPV, allowing for early intervention and improved long-term cardiovascular outcomes.

Detecting PH in its early stages and initiating timely treatment can significantly enhance survival rates [14, 15]. However, traditional detection methods such as laboratory data, ECG, CXR parameters (such as widening of pulmonary hilum, projection of the right side of the heart border, the ratio of these parameters to the chest diameter), and physical examinations have shown limited effectiveness. The AUC for the parameters used to detect PH in a CXR study did not show significant differences [7]. Traditional ECG findings used to identify ePAP were reported as unreliable screening tools due to their low sensitivity [16, 17], a finding echoed in studies of pulmonary function and laboratory data [18, 19]. Therefore, screening algorithms incorporating clinical characteristics and a variety of tests were employed to enhance the likelihood of detecting PH, guided by reliable expert opinions [19, 20].

In comparison to these tests, new approaches with AI models demonstrated a significantly higher likelihood of detecting PH. An AI study showed that standard CXR effectively identifies ePAP and predicts future HF [8]. AUC by the AI algorithm was lower than our DLM (0.71 vs 0.8525 in internal validation and 0.8605 in external validation). But different pulmonary hypertension definition was used (PAP > 20 mmHg by RHC). Those predicted as AI-positive for ePAP in that study had a two-fold higher risk of future HF admission than those predicted as AI-negative. Our results are consistent with this, affirming the role of AI in predicting heart failure. Besides CXR, ECG is a cost-effective and commonly used diagnostic tool. An AI ECG-based cohort study demonstrated high accuracy in ePAP through echocardiography, with an AUC of 0.859 in internal and 0.902 in external validation [21]. Our study achieved an AUC closely matching these results. And AI defined higher risk patient had a significantly higher chance of developing PH than those in the low-risk group (31.5% vs 5.9%) during the follow-up period [21].

Moreover, a recent study utilizing an AI algorithm with ECG data, which yielded an AUC of 0.88 for ePAP through echocardiography, revealed that patients identified by the AI as having ePAP experienced a higher cardiovascular mortality rate during a 6-year follow-up (HR: 3.69, 95% CI: 3.27–4.17) [9]. Our study not only demonstrated the predictive value of cardiovascular mortality using DLM applied to ECG (HR: 5.62, 95% CI: 4.29–7.35) but also using CXR (HR: 5.42, 95% CI: 4.11–7.15), as well as a combination of CXR and ECG (HR: 6.08, 95% CI: 4.66–7.95) for DLM-predicted positive ePAP. The presence of comorbidities, such as advanced age, CKD, HF, and Af, was associated with reduced diagnostic performance in the DLM when using ECG for ePAP detection. This suggests a potential interference of ECG changes from these conditions in the neural network's deep learning process. Similarly, the presence of COPD was linked to reduced diagnostic performance in DLM when using CXR for ePAP detection, likely due to the pathologic effects of COPD on CXR-based diagnosis.

In summary, traditional detection methods are inadequate for effectively predicting PH. However, with the advent of AI, both CXR and ECG can be used as training materials for models to achieve high accuracy. Our study, combining CXR and ECG to train AI, has yielded better results and demonstrated superior performance in predicting heart failure. Our model demonstrated a high NPV of approximately 98%. This indicates that the diagnostic performance across modalities is clinically similar and reliable as a screening tool. Notably, prior to our study, there was limited data on the prediction of cardiovascular mortality using AI with CXR for ePAP assessment [8]. In our study, we provide evidence that AI applied to CXR reliably predicted cardiovascular outcomes (HR: 5.42, 95% CI: 4.11–7.15).

## Study Limitations

Firstly, the retrospective design of our study emphasizes the importance of conducting further prospective studies to assess the clinical feasibility of this method in real-world settings.

Secondly, while ePAP > 50 mmHg by transthoracic echocardiography indicates a high probability of pulmonary hypertension (PH) according to established guidelines [5, 10], it is important to note that right heart catheterization remains the definitive diagnostic tool for PH and was not utilized in this study. The reliance on transthoracic echocardiography as an ePAP marker, without invasive right heart catheterization for a definitive PH diagnosis, may limit the clinical feasibility of this method. And our study did not distinguish between specific types of pulmonary hypertension. The ability to classify these types is crucial for determining appropriate therapeutic strategies. However, our study's ability to yield a high NPV suggests that it could potentially serve as a cost-effective screening tool to detect PH, which can then prompt further evaluation with echocardiography and RHC when necessary.

Thirdly, our patient was derived from Taiwan. Given that ECG characteristics can vary across different racial groups [22], it is essential to validate these results in diverse racial and geographical contexts.

Lastly, the opacity of current DLMs presents methodological limitations. Human experts currently lack the capacity to diagnose ePAP through ECG and CXR. To enhance the transparency of DLMs, future studies should investigate the correlation and interpretability of ECG features in relation.

## Conclusion

In this study, we demonstrated that DLM applied to different modalities could effectively identify patients with elevated ePAP, achieving high AUC and NPV. This innovative approach provides a practical and cost-effective means to assist physicians to recognize individuals at risk of PH. Additionally, it has the potential to predict their future risk of developing HF and experiencing cardiovascular mortality. Lastly, the diagnostic performance and predictive capabilities of DLM are clinically similar across modalities, including ECG, CXR, and their combination.

**Data Availability** The data supporting the findings of this study are available from the corresponding author upon reasonable request. Due to privacy and ethical considerations, the data are not publicly accessible.

## Declarations

**Conflict of Interest** The authors declare no competing interests.

**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.

## References

1. Hassoun, P.M. Pulmonary arterial hypertension. *New England Journal of Medicine* **385**, 2361-2376 (2021).
2. Elliott, C.G., *et al.* Worldwide physician education and training in pulmonary hypertension: pulmonary vascular disease: the global perspective. *Chest* **137**, 85S-94S (2010).
3. Humbert, M., *et al.* Pulmonary arterial hypertension in France: results from a national registry. *Am J Respir Crit Care Med* **173**, 1023-1030 (2006).
4. Galiè, N., *et al.* 2015 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension: the joint task force for the diagnosis and treatment of pulmonary hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). *European heart journal* **37**, 67-119 (2016).
5. Humbert, M., *et al.* 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur Heart J* **43**, 3618-3731 (2022).
6. Matthay, R.A., Niederman, M.S. & Wiedemann, H.P. Cardiovascular-pulmonary interaction in chronic obstructive pulmonary disease with special reference to the pathogenesis and management of cor pulmonale. *Med Clin North Am* **74**, 571-618 (1990).
7. Mirsadraee, M., Nazemi, S., Hamedanchi, A. & Naghibi, S. Simple screening of pulmonary artery hypertension using standard chest x ray: an old technique, new landmark. *Tanaffos* **12**, 17-22 (2013).
8. Kusunose, K., Hirata, Y., Tsuji, T., Kotoku, J. & Sata, M. Deep learning to predict elevated pulmonary artery pressure in patients with suspected pulmonary hypertension using standard chest X ray. *Sci Rep* **10**, 19311 (2020).
9. Liu, C.M., *et al.* Artificial Intelligence-Enabled Electrocardiogram Improves the Diagnosis and Prediction of Mortality in Patients With Pulmonary Hypertension. *JACC Asia* **2**, 258-270 (2022).
10. McLaughlin, V.V., *et al.* ACCF/AHA 2009 expert consensus document on pulmonary hypertension a report of the American College of Cardiology Foundation Task Force on Expert Consensus Documents and the American Heart Association developed in collaboration with the American College of Chest Physicians; American Thoracic Society, Inc.; and the Pulmonary Hypertension Association. *J Am Coll Cardiol* **53**, 1573-1619 (2009).
11. Rudski, L.G., *et al.* Guidelines for the echocardiographic assessment of the right heart in adults: a report from the American Society of Echocardiography endorsed by the European Association of Echocardiography, a registered branch of the European Society of Cardiology, and the Canadian Society of Echocardiography. *J Am Soc Echocardiogr* **23**, 685-713; quiz 786-688 (2010).

12. Lin, C.-S., *et al.* A deep-learning algorithm (ECG12Net) for detecting hypokalemia and hyperkalemia by electrocardiography: algorithm development. *JMIR medical informatics* **8**, e15931 (2020).

13. Liu, W.-T., *et al.* A deep-learning algorithm-enhanced system integrating electrocardiograms and chest X-rays for diagnosing aortic dissection. *Canadian Journal of Cardiology* **38**, 160-168 (2022).

14. Humbert, M., *et al.* Screening for pulmonary arterial hypertension in patients with systemic sclerosis: Clinical characteristics at diagnosis and long-term survival. *Arthritis & Rheumatism* **63**, 3522-3530 (2011).

15. Taichman, D.B., *et al.* Pharmacologic Therapy for Pulmonary Arterial Hypertension in Adults: CHEST Guideline and Expert Panel Report. *Chest* **146**, 449-475 (2014).

16. Kiely, D.G., Lawrie, A. & Humbert, M. Screening strategies for pulmonary arterial hypertension. *European Heart Journal Supplements* **21**, K9-K20 (2019).

17. Speich, R. Diagnosing pulmonary hypertension: is there a revival of the electrocardiogram? *European Respiratory Journal* **37**, 994-996 (2011).

18. Gladue, H., *et al.* Combination of Echocardiographic and Pulmonary Function Test Measures Improves Sensitivity for Diagnosis of Systemic Sclerosis-associated Pulmonary Arterial Hypertension: Analysis of 2 Cohorts. *The Journal of Rheumatology* **40**, 1706-1711 (2013).

19. Thakkar, V., *et al.* The inclusion of N-terminal pro-brain natriuretic peptide in a sensitive screening strategy for systemic sclerosis-related pulmonary arterial hypertension: a cohort study. *Arthritis Research & Therapy* **15**, R193 (2013).

20. Coghlan, J.G., *et al.* Evidence-based detection of pulmonary arterial hypertension in systemic sclerosis: the DETECT study. *Annals of the Rheumatic Diseases* **73**, 1340-1349 (2014).

21. Kwon, J.M., *et al.* Artificial intelligence for early prediction of pulmonary hypertension using electrocardiography. *J Heart Lung Transplant* **39**, 805-814 (2020).

22. Noseworthy, P.A., *et al.* Assessing and mitigating bias in medical artificial intelligence: the effects of race and ethnicity on a deep learning model for ECG analysis. *Circulation: Arrhythmia and Electrophysiology* **13**, e007988 (2020).

## Authors and Affiliations

**Pang-Yen Liu[1,2] · Shi-Chue Hsing[1] · Dung-Jang Tsai[3,5] · Chin Lin[3,4,5] · Chin-Sheng Lin[1] · Chih-Hung Wang[6,7] · Wen-Hui Fang[3,8]**

✉ Wen-Hui Fang
rumaf.fang@gmail.com

Pang-Yen Liu
liupydr@gmail.com

Shi-Chue Hsing
lars0121@gmail.com

Dung-Jang Tsai
oo800217@gmail.com

Chin Lin
xup6fup0629@gmail.com

Chin-Sheng Lin
littlelincs@gmail.com

Chih-Hung Wang
chw@ms3.hinet.net

[1] Division of Cardiology, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center Taipei, Taipei, Taiwan R.O.C.

[2] Department of Cardiovascular Medicine, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[3] Artificial Intelligence of Things Center, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan R.O.C.

[4] School of Public Health, National Defense Medical Center, Taipei, Taiwan R.O.C.

[5] Medical Technology Education Center, School of Medicine, National Defense Medical Center, Taipei, Taiwan R.O.C.

[6] Department of Otolaryngology-Head and Neck Surgery, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan R.O.C.

[7] Graduate Institute of Medical Sciences, National Defense Medical Center, Taipei, Taiwan R.O.C.

[8] Department of Family and Community Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan R.O.C.