Canadian Journal of Cardiology ■ (2022) 1–11

Clinical Research

Detection of Left Ventricular Systolic Dysfunction Using an Artificial Intelligence–Enabled Chest X-Ray

Chih-Weim Hsiang, MD,[a] Chin Lin, PhD,[b,c,d] Wen-Cheng Liu, MD,[e] Chin-Sheng Lin, MD, PhD,[e,f] Wei-Chou Chang, MD, PhD,[a,f] Hsian-He Hsu, MD,[a] Guo-Shu Huang, MD,[a] Yu-Sheng Lou, MS,[b,c] Chia-Cheng Lee, MD,[g] Chih-Hung Wang, MD, PhD,[f,h] and Wen-Hui Fang, MD[i]

[a] *Department of Radiology, Tri-Service General Hospital, National Defense Medical Centre, Taipei, Taiwan*

[b] *Graduate Institute of Life Sciences, National Defense Medical Centre, Taipei, Taiwan*

[c] *School of Public Health, National Defense Medical Centre, Taipei, Taiwan*

[d] *Medical Technology Education Center, School of Medicine, National Defense Medical Centre, Taipei, Taiwan*

[e] *Division of Cardiology, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Centre, Taipei, Taiwan*

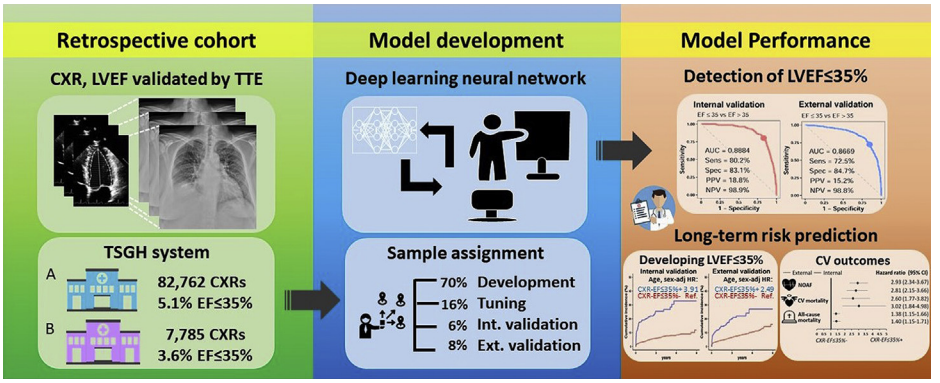
[f] *Graduate Institute of Medical Sciences, National Defense Medical Centre, Taipei, Taiwan*

[g] *Department of Medical Informatics, Tri-Service General Hospital, National Defense Medical Centre, Taipei, Taiwan*

[h] *Department of Otolaryngology–Head and Neck Surgery, Tri-Service General Hospital, National Defense Medical Centre, Taipei, Taiwan*

[i] *Department of Family and Community Medicine, Tri-Service General Hospital, National Defense Medical Centre, Taipei, Taiwan*

See editorial by Lauzier and Chow, pages xxx-xxx of this issue.



ABSTRACT

Background: Assessment of left ventricular systolic dysfunction provides essential information related to the prognosis and management

RÉSUMÉ

Contexte : L'évaluation de la dysfonction systolique ventriculaire gauche fournit des informations essentielles liées au pronostic et à la

Received for publication October 16, 2021. Accepted December 18, 2021.

Corresponding author: Dr Wen-Hui Fang, Department of Family and Community Medicine, Tri-Service General Hospital, National Defense Medical Centre, Taipei, Taiwan. Tel.: +886-2-87923100, ext 12322; fax: +886-2-87923147.

E-mail: rumaf.fang@gmail.com

See page 10 for disclosure information.

Heart failure (HF) is a growing epidemic attributed to the aging population worldwide.[1,2] The main terminology used to describe HF is based on the measurement of left ventricular ejection fraction (LVEF).[3,4] HF with reduced LVEF (HFrEF) is associated with poor quality of life, and increased morbidity and mortality. Recent studies have

https://doi.org/10.1016/j.cjca.2021.12.019

of cardiovascular diseases. The aim of this study was to develop a deep-learning model to identify left ventricular ejection fraction (LVEF) $\leq$ 35% via chest X-ray (CXR [*CXR-EF*$\leq$*35%*]) features and investigate the performance and clinical implications.

Methods: This study collected 90,547 CXRs with the corresponding LVEF according to transthoracic echocardiography from the outpatient department in an academic medical center. Among these, 77,227 CXRs were used to develop the identification of *CXR-EF*$\leq$*35%*. Another 13,320 CXRs were used to validate the performance, which was evaluated by area under the receiver operating characteristic curve (AUC). Furthermore, *CXR-EF*$\leq$*35%* was tested to assess the long-term risks of developing LVEF $\leq$ 35% and cardiovascular outcomes, which were evaluated by Kaplan-Meier survival analysis and the Cox proportional hazards model.

Results: The AUCs of *CXR-EF*$\leq$*35%* for the detection of LVEF $\leq$ 35% were 0.888 and 0.867 in the internal and external validation cohorts, respectively. Patients with baseline LVEF $>$ 50% but detected as *CXR-EF*$\leq$*35%* were at higher risk of long-term development of LVEF $\leq$ 35% (hazard ratio, internal validation cohort [HR_i] 3.91, 95% CI 2.98-5.14; hazard ratio, external validation cohort [HR_e] 2.49, 95% CI 1.89-3.27). Furthermore, patients detected as LVEF $\leq$ 35% by *CXR-EF*$\leq$*35%* had significantly higher future risks of all-cause mortality (HR_i 1.40, 95% CI 1.15-1.71; HR_e 1.38, 95% CI 1.15-1.66), cardiovascular mortality (HR_i 3.02, 95% CI 1.84-4.98; HR_e 2.60, 95% CI 1.77-3.82), and new-onset atrial fibrillation (HR_i 2.81, 95% CI 2.15-3.66; HR_e 2.93, 95% CI 2.34-3.67) compared with those detected as no LVEF $\leq$ 35%.

Conclusions: *CXR-EF*$\leq$*35%* may serve as a screening tool for early detection of LVEF $\leq$ 35% and could independently contribute to predictions of long-term development of LVEF $\leq$ 35% and cardiovascular outcomes. Further prospective studies are needed to confirm the model performance.

prise en charge des maladies cardiovasculaires. L'objectif de cette étude était de développer un modèle d'apprentissage profond pour identifier une fraction d'éjection ventriculaire gauche (FEVG) $\leq$ 35 % via les données de radiographie thoracique (RT [*RT-FE*$\leq$*35%*]) et d'en étudier les performances et les implications cliniques.

Méthodes : Cette étude a recueilli 90 547 RT, chacune associée à une FEVG selon les critères de l'échocardiographie transthoracique issue du service ambulatoire d'un centre médical universitaire. Parmi celles-ci, 77 227 RT ont été utilisées pour développer l'identification de *RT-FE*$\leq$*35%*. Les 13 320 RT restantes ont été utilisées pour en valider la performance, qui a été évaluée via l'aire sous la courbe caractéristique d'exploitation du récepteur (ASC). En outre, la *RT-FE*$\leq$*35%* a été testée pour évaluer les risques à long terme de développer une FEVG $\leq$ 35 % et ses conséquences cardiovasculaires, qui ont été évalués par une analyse de survie Kaplan-Meier et le modèle des risques proportionnels de Cox.

Résultats : Les ASC de la *RT-FE*$\leq$*35%* pour la détection d'une FEVG $\leq$ 35 % étaient de 0,888 et 0,867 dans les cohortes de validation interne et externe, respectivement. Les patients présentant une FEVG initiale $>$ 50 % mais détectées comme *RT-FE*$\leq$*35%* présentaient un risque plus élevé, à long terme, de développer une FEVG $\leq$ 35 % (rapport de risque, cohorte de validation interne [RRi] 3,91, intervalle de confiance [IC] à 95 % 2,98-5,14; rapport de risque, cohorte de validation externe [RRe] 2,49, IC à 95 % 1,89-3,27). En outre, les patients détectés comme ayant une FEVG $\leq$ 35 % par *RT-FE*$\leq$*35%* présentaient des risques de mortalité future significativement plus élevés, toutes causes confondues (RRi 1,40, IC à 95 % 1,15-1,71; RRe 1,38, IC à 95 % 1,15-1,66), de mortalité cardiovasculaire (RRi 3,02, IC à 95 % 1,84-4,98; RRe 2,60, IC à 95 % 1,77-3,82) et de fibrillation auriculaire d'apparition récente (RRi 2,81, IC à 95 % 2,15-3,66; RRe 2,93, IC à 95 % 2,34-3,67) par rapport aux personnes détectées comme n'ayant pas de FEVG $\leq$ 35 %.

Conclusions : La *RT-FE*$\leq$*35%* peut servir d'outil de dépistage pour la détection précoce de la FEVG $\leq$ 35 % et pourrait indépendamment contribuer aux prédictions de l'évolution à long terme de la FEVG $\leq$ 35 % et des effets cardiovasculaires. D'autres études prospectives sont nécessaires pour confirmer la performance du modèle.

demonstrated that all-cause mortality is generally higher in HFrEF than in HF with preserved LVEF.[5,6] Guideline-directed management and therapies have improved survival and reduced rehospitalisation in patients with HFrEF.[7,8]

Strategies for the early identification of left ventricular systolic dysfunction (LVSD) may prevent progression to symptomatic HF. Cardiac imaging plays an essential role in diagnosis and guiding treatment. Among several imaging modalities, transthoracic echocardiography (TTE) is the primary and standard method owing to its optimal accuracy, portability, and safety. Nevertheless, TTE is an expensive and time-consuming tool and, importantly, requires specialists for operation and interpretation.

Chest X-ray (CXR) is a widespread, noninvasive, and preliminary radiologic screen examination, providing primary information about patients' pulmonary and heart conditions. Although it has a limited role in the diagnostic work-up of patients with suspected HF, it is still helpful in the acute setting, rather than in the nonacute setting, in HF patients.[9] It is difficult, however, for general practitioners to make a precise evaluation of LVSD with the use of CXR alone.

Artificial intelligence (AI) using deep-learning models (DLMs) has been applied to the sophisticated interpretation of subtle patterns of digital images in multiple applications to facilitate physicians' diagnostic support or prompt treatment.[10,11] Many studies have focused on cardiac computed tomography (CT) or magnetic resonance imaging (MRI), which are expensive and time consuming, radiation exposing, and of concern to renal function and require contrast injection and expert interpretation. Recent studies have shown a significant association between AI-enabled CXR and the diagnosis of HF.[12,13] Nevertheless, research on the use of DLMs to detect LVSD based on CXR is scarce. Therefore, we hypothesised that AI-enabled CXR may be a potential tool for large-scale screening and detecting of LVSD.

In this cohort study, we developed a DLM to detect LVEF $\leq$ 35% based on CXR, with the corresponding LVEF

validated by TTE. We verified the performance of the DLM to evaluate LVEF $\leq$ 35% using CXR (*CXR-EF$\leq$35%*) and explored the correlation between *CXR-EF$\leq$35%* and TTE parameters. Furthermore, this study set out to demonstrate the capability of *CXR-EF$\leq$35%* to predict long-term development of LVEF $\leq$ 35% and cardiovascular outcomes to enhance the potential value in future clinical applications.

Methods

Data source

This study was approved by the Institutional Review Board of Tri-Service General Hospital (TSGH), Taipei, Taiwan (IRB no. C202105150). We performed a retrospective 2-site study in the TSGH system from January 1, 2010, to December 31, 2020. Information on study sites is provided in Suppemental Appendix S1. We identified CXRs in the posterior-anterior view with at least 1 TTE obtained within 7 days of the index CXR. A total of 45,188 patients in the study period at hospital A with more than 1 CXR-TTE pair within 7 days were included. Figure 1 shows the generation of each dataset. There were 5535 patients before January 1, 2015, in the validation set and 5437 patients before January 1, 2016, in the tuning set for guiding the training process. The remaining 34,216 patients, who were used for training a DLM, provided 62,821 CXRs with the corresponding LVEF to construct a development set. Details of sample assignment are provided in the Supplemental Appendix S1.

To conduct the accuracy test of the DLM, an internal validation set including 5535 independent patients was enrolled and the other 7785 nonoverlapping patients at hospital B meeting the same criteria were included in the external validation set. To avoid overrepresenting sicker patients who undergo more CXRs, we selected a single CXR per patient in both validation sets. Rather than selecting the most recent or earliest CXR, we randomly sampled 1 CXR from all CXRs available for a given patient. This strategy was considered to be most representative of deploying the model on CXRs from new patients, which in each case would be at a random time point in that patient's life. This strategy has bee previously described.[14]

Quantitative echocardiographic data collection

Comprehensive 2-dimensional TTE was available for all patients. Quantitative data were recorded at the time of acquisition with a Philips image system. LVEF is routinely acquired by experienced technicians or cardiologists using standardised methods. LVEF was determined with the use of the M-mode, modified Simpson method in 2-dimensional imaging, and the reported visually estimated LVEF. We traced the endocardial border in both apical 4-chamber and 2-chamber views in end-systole and end-diastole. After dividing the LV cavity into predetermined numbers of slices, LV volume and LVEF were calculated.

Implementation of the DLM

The CXR image was recorded in DICOM format with a resolution of more than 3000 $\times$ 3000 pixels. All CXRs were labelled as a binary classification according to LVEF $\leq$ 35%. The training details of the DLM for CXR were revised from a previous study.[10] The major feature extraction architecture is based on a 50-layer SE-ResNeXt, which won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2017.[15] This SE-ResNeXt is pretrained by ImageNet, and the last feature map is saved for further application. The output features of SE-ResNeXt are downsampled by 32-fold compared with the original images, which were passed through a subsequent global pooling layer and a fully connected layer with 1 output with a sigmoid function. The output of our revised DLM was a probability to describe the likelihood of LVEF $\leq$ 35%. We resized our CXR to let the short side be 256 pixels without changing the aspect ratio. In the training stage, we randomly cropped 224 $\times$ 224 pixels as input and applied a random lateral inversion with 50% probability. In the inference stage, the 10-crop evaluation based on a previous study was used to generate 10 probabilities for each CXR.[16] An average of a total of 10 probabilities was the final output of our AI network. Detailed information on the DLM implantation is provided in Supplemental Appendix S1.

Baseline information and data collection

The electronic medical record (EMR) of each hospital provided baseline information. The disease histories were based on a new diagnosis according to the corresponding International Classification of Diseases (ICD), Ninth and Tenth Revisions, or laboratory tests. The details of enrolled ICD codes are provided in the *Data Collection* section of Supplemental Appendix S1. Patients with at least 2 records of $\geq$ 7 mmol/L fasting glucose or $\geq$ 48 mmol/mol glycated hemoglobin for 6 months were considered to have diabetes. We also defined at least 2 records of estimated glomerular filtration rate $<$ 1 mL/s as chronic kidney disease (CKD).

Outcomes

We followed several outcomes of interest in both the internal and external validation sets. The primary outcome was the ability of *CXR-EF$\leq$35%* to identify individuals with LVEF $>$ 50% at the time of screening, who had an increasing risk of subsequent developing LVEF $\leq$ 35% during follow-up.[17] The follow-up time was calculated with reference to the date of CXR. Patient status was defined through continuous LVEF changes, which were updated by each TTE. Moreover, data for unchanged patients were censored at the patient's last known TTE to limit bias from incomplete records. The secondary outcomes including subsequent all-cause mortality, cardiovascular (CV) mortality, acute myocardial infarction (AMI), and new-onset atrial fibrillation (AF) were evaluated. Patients with a history of AMI or AF were excluded from the analysis of the same follow-up outcomes of interest. These statuses were defined as the corresponding events or ICD records from our EMR, which were updated by each hospital activity. The data for at-risk patients were also censored at the

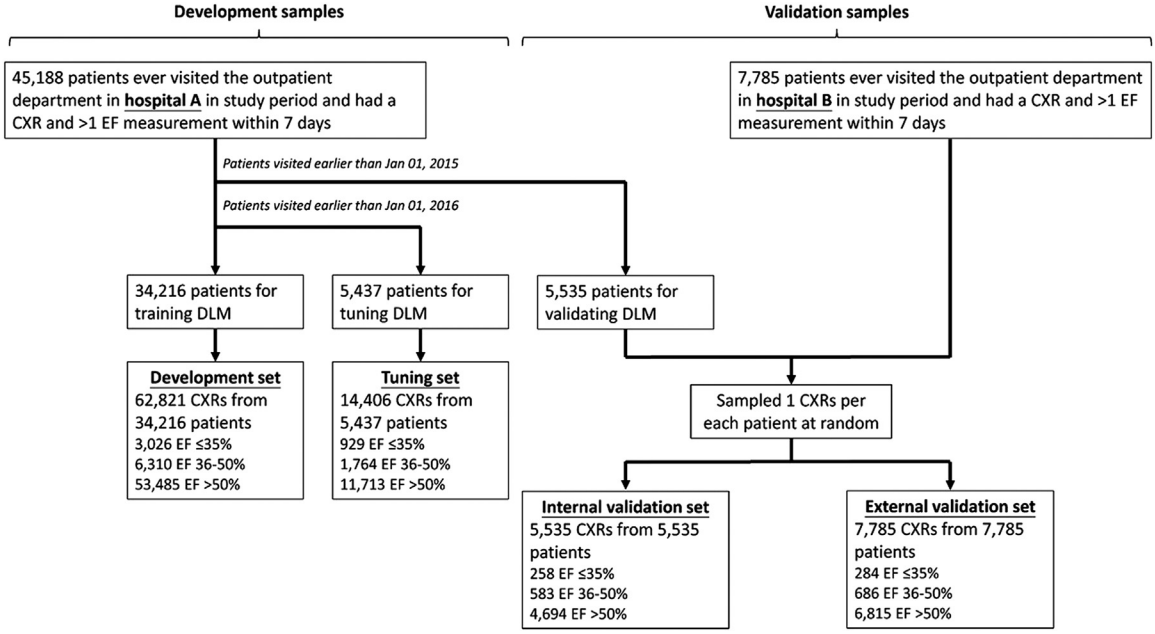


Figure 1. Development, tuning, internal validation, and external validation set generation and chest X-ray (CXR) labelling of ejection fraction (EF). The dataset creation and analysis strategy was devised to assure a robust and reliable dataset for training, validating, and testing of the network. Once a patient's data were placed in one of the datasets, those data were used only in that set, avoiding "cross-contamination" among the development, tuning, and validation sets. DLM, deep-learning model.

patient's last known hospital encounter to limit bias from incomplete records.

Statistical analysis

We presented the characteristics of different datasets as mean and standard deviation, number of patients, or percentage as appropriate. They were compared by means of either analysis of variance or chi-square test as appropriate. The performance of *CXR-EF≤35%* was determined by receiver operating characteristic (ROC) curve analysis to detect an LVEF $\leq$ 35%, and the area under the curve (AUC), sensitivity, and specificity were used to demonstrate the performance. The operating point was selected based on the maximum Youden index in the tuning set, which was used for both validation sets as the same value. We also used baseline information for stratifying patients to explore the performance of *CXR-EF≤35%* in each population. All statistical analyses were completed in R software version 3.4.4 (R Foundation for Statistical Computing, Vienna, Austria). The significance level was set as $P < 0.05$.

For the estimation error analysis, we explored the difference in patient characteristics between *CXR-EF≤35%+* and *CXR-EF≤35%−* patients. Linear regression or logistic regression was used for statistical testing where appropriate. We also performed Kaplan-Meier survival analysis with the available follow-up data stratified by *CXR-EF≤35%* prediction on each outcome of interest. A Cox proportional hazards model was used to calculate the hazard ratios (HRs) with 95% confidence intervals (95% CIs), which were reported for all data. According to the potential competing risk between all-cause mortality and other outcomes of interest, we used the R package "cmprsk" to calculate cumulative incidence and conduct competing risks regression.[18]

Results

Baseline characteristics

Patient characteristics of the development, tuning, internal and external validation cohorts were shown in Table 1. Patients in the development cohort were younger and more male, had a lower BMI and fewer comorbidities than patients in the validation cohorts. In the development cohort, 3026 patients (4.8%) had an LVEF $\leq$ 35%, 6310 patients (10.1%) had an LVEF $>$ 35% to 50%, and 53,485 patients (85.1%) had an LVEF $>$ 50%. In the internal validation cohort, 258 patients (4.7%) had an LVEF $\leq$ 35%, 583 patients (10.5%) had an LVEF $>$ 35% to 50%, and 4694 patients (84.8%) had an LVEF $>$ 50%, while in the external validation cohort, 284 patients (3.7%) had an LVEF $\leq$ 35%, 686 patients (8.8%) had an LVEF $>$ 35% to 50%, and 6815 patients (87.5%) had an LVEF $>$ 50%.

Performance of *CXR-EF≤35%* to identify LVEF ≤ 35%

The algorithm provided discrimination between LVEF $\leq$ 35% and LVEF $>$ 35% with an AUC of 0.888 and corresponding sensitivity of 80.2% and specificity of 83.1% in the internal validation cohort and an AUC of 0.867 with corresponding sensitivity of 72.5% and specificity of 84.7% in the external validation cohort, as shown in Figure 2. The model performance was further adjusted for age and sex, but did not improve as shown in Supplemental Figure S1 and described in the *Results* section of Supplemental Appendix S1. In addition, the performance of *CXR-EF≤35%* to distinguish between

Table 1. Corresponding characteristics in development, tuning, internal validation, and external validation sets

	Development (n = 62,821)	Tuning (n = 14,406)	Internal validation (n = 5535)	External validation (n = 7785)	*P* value
Demographics					
Male	33,650 (54.4%)	7452 (51.7%)	2902 (52.4%)	3931 (50.5%)	< 0.001
Age, years	67.1 ± 16.7	70.3 ± 15.6	67.8 ± 15.6	70.5 ± 16.6	< 0.001
BMI, kg/m^2	24.2 ± 4.4	24.0 ± 4.5	24.6 ± 4.4	24.3 ± 4.3	< 0.001
Disease history					
DM	17,596 (28.4%)	5701 (39.6%)	2114 (38.2%)	2943 (37.8%)	< 0.001
HTN	27,624 (44.6%)	8914 (61.9%)	3381 (61.1%)	5007 (64.3%)	< 0.001
HLP	19,363 (31.3%)	6593 (45.8%)	2698 (48.7%)	3972 (51.0%)	< 0.001
CKD	20,415 (33.0%)	7543 (52.4%)	2218 (40.1%)	3053 (39.2%)	< 0.001
AMI	2402 (3.9%)	982 (6.8%)	311 (5.6%)	293 (3.8%)	< 0.001
CAD	14,772 (23.9%)	5294 (36.7%)	2118 (38.3%)	2926 (37.6%)	< 0.001
Stroke	10,567 (17.1%)	3,668 (25.5%)	1285 (23.2%)	1977 (25.4%)	< 0.001
AF	4464 (7.2%)	1868 (13.0%)	571 (10.3%)	786 (10.1%)	< 0.001
COPD	10,546 (17.0%)	3805 (26.4%)	1423 (25.7%)	2480 (31.9%)	< 0.001
Echocardiographic data					
LVEF, %	63.2 ± 12.5	61.6 ± 13.5	63.3 ± 12.3	64.1 ± 11.5	< 0.001
LV-D, mm	47.5 ± 7.3	47.9 ± 7.9	47.6 ± 7.4	47.3 ± 7.1	< 0.001
LV-S, mm	30.5 ± 7.1	31.1 ± 7.7	30.5 ± 7.1	30.1 ± 6.7	< 0.001
IVS, mm	11.4 ± 2.6	11.7 ± 2.6	11.5 ± 2.6	11.4 ± 2.6	< 0.001
LVPW, mm	9.4 ± 1.8	9.6 ± 1.8	9.5 ± 1.7	9.3 ± 1.7	< 0.001
LA, mm	39.1 ± 7.9	40.1 ± 8.4	39.5 ± 8.0	39.5 ± 7.8	< 0.001
AO, mm	33.1 ± 4.3	33.3 ± 4.4	33.3 ± 4.4	33.1 ± 4.3	< 0.001
RV, mm	24.0 ± 5.1	24.3 ± 5.2	24.4 ± 5.1	24.1 ± 5.0	< 0.001
PASP, mm Hg	34.4 ± 12.1	35.6 ± 12.9	33.8 ± 11.7	34.6 ± 12.0	< 0.001

Values are n (%) or mean ± SD.

AF, atrial fibrillation; AMI, acute myocardial infarction; AO, aortic root diameter; BMI, body mass index; CAD, coronary artery disease; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; DM, diabetes mellitus; HLP, hyperlipidemia; HTN, hypertension; IVS, interventricular septum thickness; LA, left atrial diameter; LV-D, left ventricular end-diastolic diameter; LV-S, left ventricular end-systolic diameter; LVEF, ventricular ejection fraction; LVPW, left ventricular posterior wall thickness; PASP, pulmonary artery systolic pressure; RV, right ventricular outflow tract proximal diameter in parasternal long axis view.

LVEF ≤ 50% and > 50% was investigated. These results are shown in Supplemental Figure S2 and described in the *Results* section of Supplemental Appendix S1. *CXR-EF≤35%* appeared to have relative better performance in detecting LVEF ≤ 35% in patients without CKD and AMI in both validation cohorts, as shown in Figure 3.

Echocardiographic characteristics and risk analysis stratified by *CXR-EF≤35%* in different LVEF groups

Patients with a baseline LVEF > 50%, detected as LVEF ≤ 35% by *CXR-EF≤35%* (*CXR-EF≤35%+*) were more physically incompetent, showing male predominance, larger LV, right ventricular (RV), and LA size, and thicker LV posterior wall and interventricular septum, compared with those who were detected as no LVEF ≤ 35% by *CXR-EF≤35%* (*CXR-EF≤35%−*) in both validation cohorts, as shown in Figure 4. This implied that *CXR-EF≤35%* might provide the capability for evaluating the long-term outcome according to worsening of underlying TTE parameters in those with a baseline LVEF > 50%.

Prediction of long-term risk of developing LVEF ≤ 35%

The prediction of long-term development of LVEF ≤ 35% in patients with a baseline LVEF > 50% stratified by the *CXR-EF≤35%* after the adjustment for age and sex was showed in Figure 5. The cumulative incidence of developing LVEF ≤ 35% was 25.4% and 32.9% for 3 and 6 years, respectively, in those stratified as *CXR-EF≤35%+*, compared with 6.0% and 8.5% for 3 and 6 years in those stratified as *CXR-EF≤35%−* in the internal validation cohort, and the cumulative incidence of developing LVEF ≤ 35% was 20.1% and 27.0% for 3 and 6 years in those stratified as *CXR-EF≤35%+*, compared with 6.6% and 10.6% for 3 and 6 years in those stratified as *CXR-EF≤35%−* in the external validation cohort. This indicated a higher risk of future LVEF ≤ 35% when patients with LVEF > 50% detected as LVEF ≤ 35% compared with those detected as LVEF ≤ 35% by *CXR-EF≤35%* (internal validation cohort HR [HR$_i$] 3.91, 95% CI 2.98-5.14; $P < 0.01$; external validation cohort HR [HR$_i$] 2.49, 95% CI 1.89-3.27; $P < 0.01$). It demonstrated the capability of *CXR-EF≤35%* to predict the long-term development of LVEF ≤ 35% in patients with LVEF > 50%. Furthermore, a sensitivity analysis was evaluated to consider potential competing risk of death and long-term development of LVEF ≤ 35%. We conducted a competing risk analysis in Supplemental Fig. S3. The results were similar to those in Figure 5, which implied independence between the 2 events and no potential competing risk.

Prediction of long-term risk of CV outcomes

We analysed the HR comparison of each outcome, which was followed along with developing LVEF ≤ 35%, as shown in Figure 6. Stratified by LVEF, lower LVEF predicted higher long-term risks of CV outcomes, including all-cause mortality (HR$_i$ 1.20, 95% CI 1.03-1.40 [$P = 0.017$]; HR$_e$ 1.17, 95% CI 1.01-1.37 [$P = 0.043$]), CV mortality (HR$_i$ 2.06, 95% CI

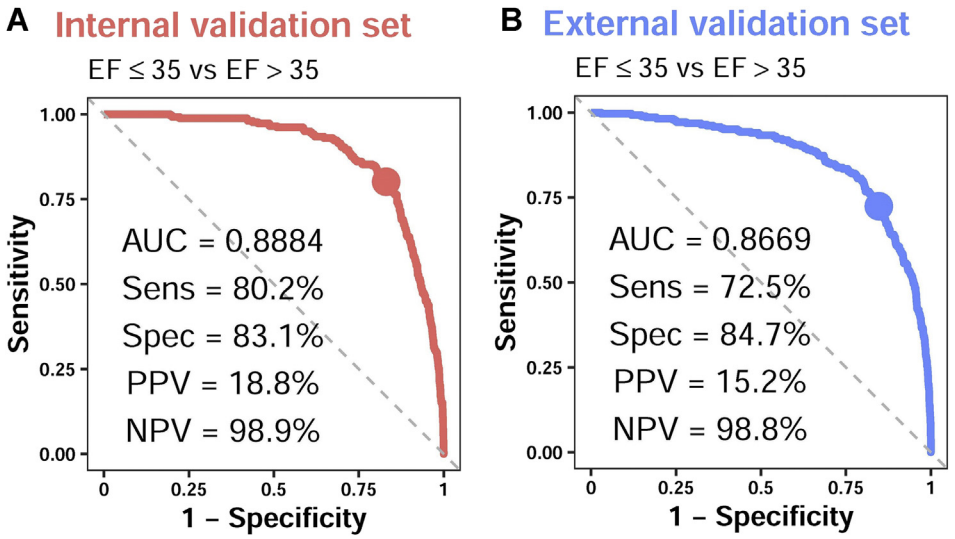


Figure 2. The area under the receiver operating characteristic curve (AUC) of *CXR-EF≤35%* to identify left ventricular ejection fraction (LVEF) ≤ 35%. The AUC (*x*-axis = 1-specificity; *y*-axis = sensitivity) were calculated in the internal and external validation sets. The operating point was selected based on the maximum of Youden's index in tuning set, which was used for calculating the corresponding sensitivities and specificities in both validation sets. *CXR-EF≤35%,* deep-learning model to identify LVEF ≤ 35% via chest X-ray; NPV, negative predictive value; PPV, positive predictive value; Sens, sensitivity; Spec, specificity.

1.54-2.76 [$P < 0.01$]; HR_e 1.68, 95% CI 1.28-2.21 [$P < 0.01$]), AMI (HR_i 2.50, 95% CI 1.92-3.27 [$P < 0.01$]; HR_e 2.80, 95% CI 2.22-3.54 [$P < 0.01$]), and new-onset AF (HR_i 1.36, 95% CI 1.12-1.65 [$P < 0.01$]; HR_e 1.42, 95% CI 1.18-1.69 [$P < 0.01$]), compared to normal LVEF in both validation cohorts. Importantly, patients with *CXR-EF≤35%+* significantly contributed to higher risks of all-cause mortality (HR_i 1.40, 95% CI 1.15-1.71; HR_e 1.38, 95% CI 1.15-1.66), CV mortality (HR_i3.02, 95% CI 1.84-4.98; HR_e 2.60, 95% CI 1.77-3.82), and new-onset AF (HR_i 2.81, 95% CI 2.15-3.66; HR_e 2.93, 95% CI 2.34-3.67) compared with those with *CXR-EF≤35%−*. These features were noticeably demonstrated in patients with baseline LVEF > 50%. The competing risk analysis is shown in Supplemental Figure S4. The results were similar to those in Figure 6, demonstrating a robust finding in extensive outcomes.

Discussion

In this study, we developed a DLM to detect LVEF ≤ 35% based on CXR (*CXR-EF≤35%*), which achieved an AUC of 0.888 with a sensitivity of 80.2% and a specificity of 83.1% in the internal validation cohort. Our model may be useful for early detection of LVEF ≤ 35% via CXR, especially in patients without a history of CKD and AMI, for general practitioners in primary care. Moreover, *CXR-EF≤35%* may serve as a screening tool in patients with an initial LVEF > 50% for the long-term development of LVEF ≤ 35% and CV outcomes.

The present study was the first to explore the detection of LVEF ≤ 35% by means of a DLM using CXR. The application of AI-enabled CXR to detect pulmonary diseases or anomalies, including nodule, tumour, pneumonia, or tuberculosis, with regard to diagnosis, staging, exacerbations, and survival is widespread.[19] Automated CXR reading based on DLMs is currently an intense field of research. Nevertheless, most studies are limited to the lungs.[10] Recently, a growing number of studies have reported the use of AI-enabled CXR to detect heart structures and conditions.[20,21] CXR is a 2-dimensional (2D) presentation of the 3D heart structure. In contrast, there are other advanced diagnostic methods, such as electrocardiography (ECG), echocardiography, and cardiac CT or MRI, that can provide extra information better than CXR about how efficiently the heart is pumping and which chambers of the heart are enlarged.[19] Nevertheless, CXR offers the fundamental evaluation in integrated cardiopulmonary conditions and serves as a common, feasible, affordable clinical examination than the above tools for screening and detecting heart functions.

Asymptomatic LVSD is present in 3% of the general population and confers a 4-fold increased risk of clinical HF and a 2.0-fold increased risk of all-cause mortality.[22-24] Effective population-wide screening for LVSD is lacking. Previous studies using ECG to identify patients with an LVEF ≤ 35% revealed an AUC of 0.933 with a sensitivity of 86.3% and a specificity of 85.7%. In contrast, *CXR-EF≤35%* in the internal validation set showed an AUC of 0.888 with a sensitivity of 80.2% and a specificity of 83.1%. Basically, both CXR and ECG are modalities for the diagnostic work-up in patients with HF. With the application of DLM, the performances of the 2 modalities are strengthened for detecting LVEF ≤ 35%. Based on the different characteristics of CXR and ECG, incorporation of both modalities was considered to enhance the capability to detect LVEF ≤ 35%.

B-Type natriuretic peptide (BNP) and N-terminal proBNP (NT-proBNP) have been previously proposed for the screening or detection of LVSD. The Dallas Heart Study screened LVSD with the use of both markers and found an AUC < 0.700, and performance did not improve more

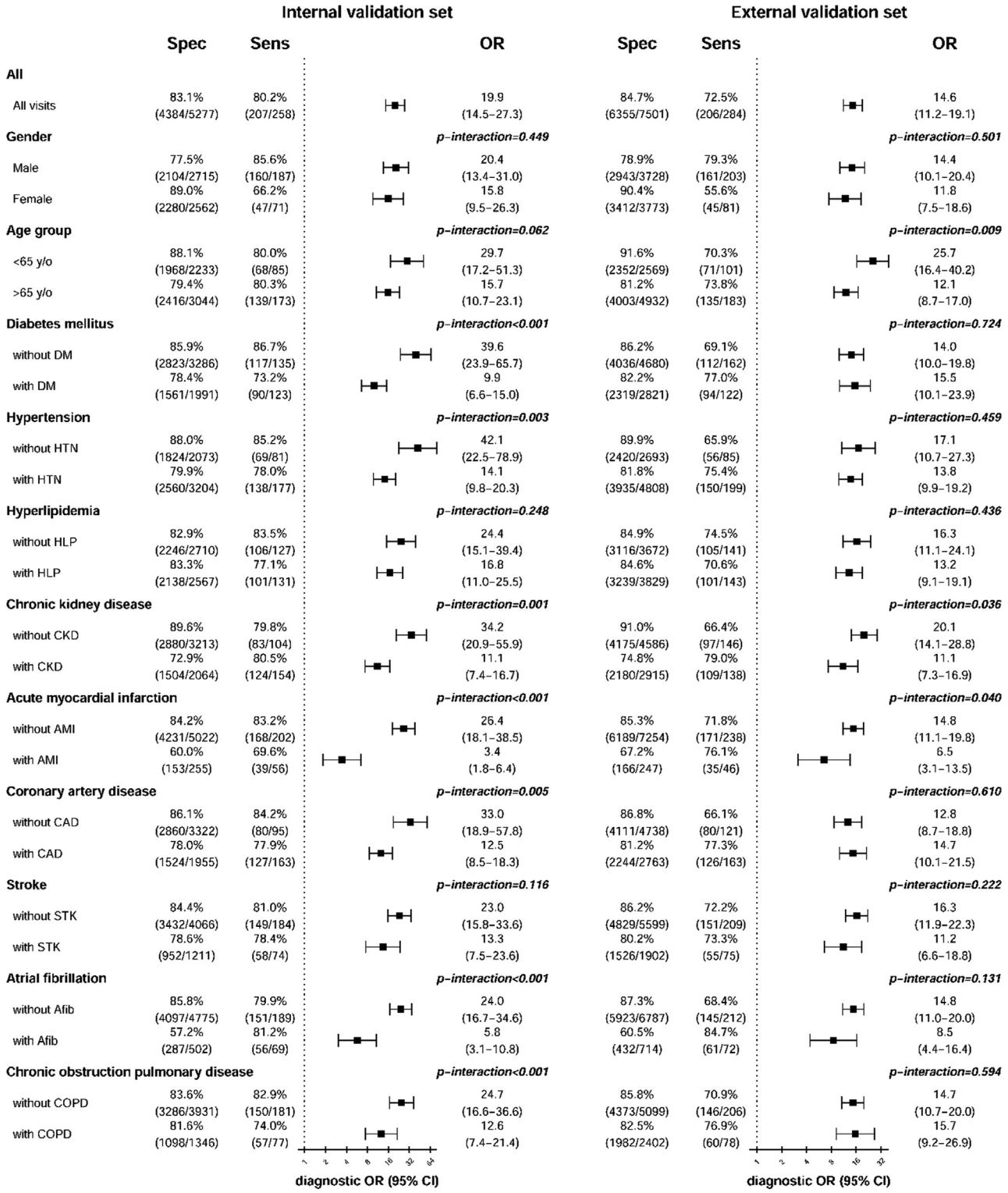


Figure 3. The sensitivity and specificity of *CXR-EF≤35%* to detect left ventricular ejection fraction (LVEF) ≤ 35% tabulated across a series of stratified analyses. The diagnostic odds ratio (OR), which is the ratio of the positive likelihood ratio (sensitivity/[1−specificity]) to the negative likelihood ratio ([1−sensitivity]/specificity), as well as the associated 95% confidence interval (CI), is shown for each feature. *CXR-EF≤35%*, deep-learning model to identify LVEF ≤ 35% via chest X-ray; Sens, sensitivity; Spec, specificity.

among subgroups aged ≥ 50 or with hypertension.[25] Leong et al. appraised community screening for LVSD with the use of plasma and urinary natriuretic peptides and found AUCs for urinary and plasma NT-proBNP of 0.831 and 0.840, respectively. The plasma-urinary NT-proBNP product could improve the performance with an AUC of 0.923.[26] A Mayo

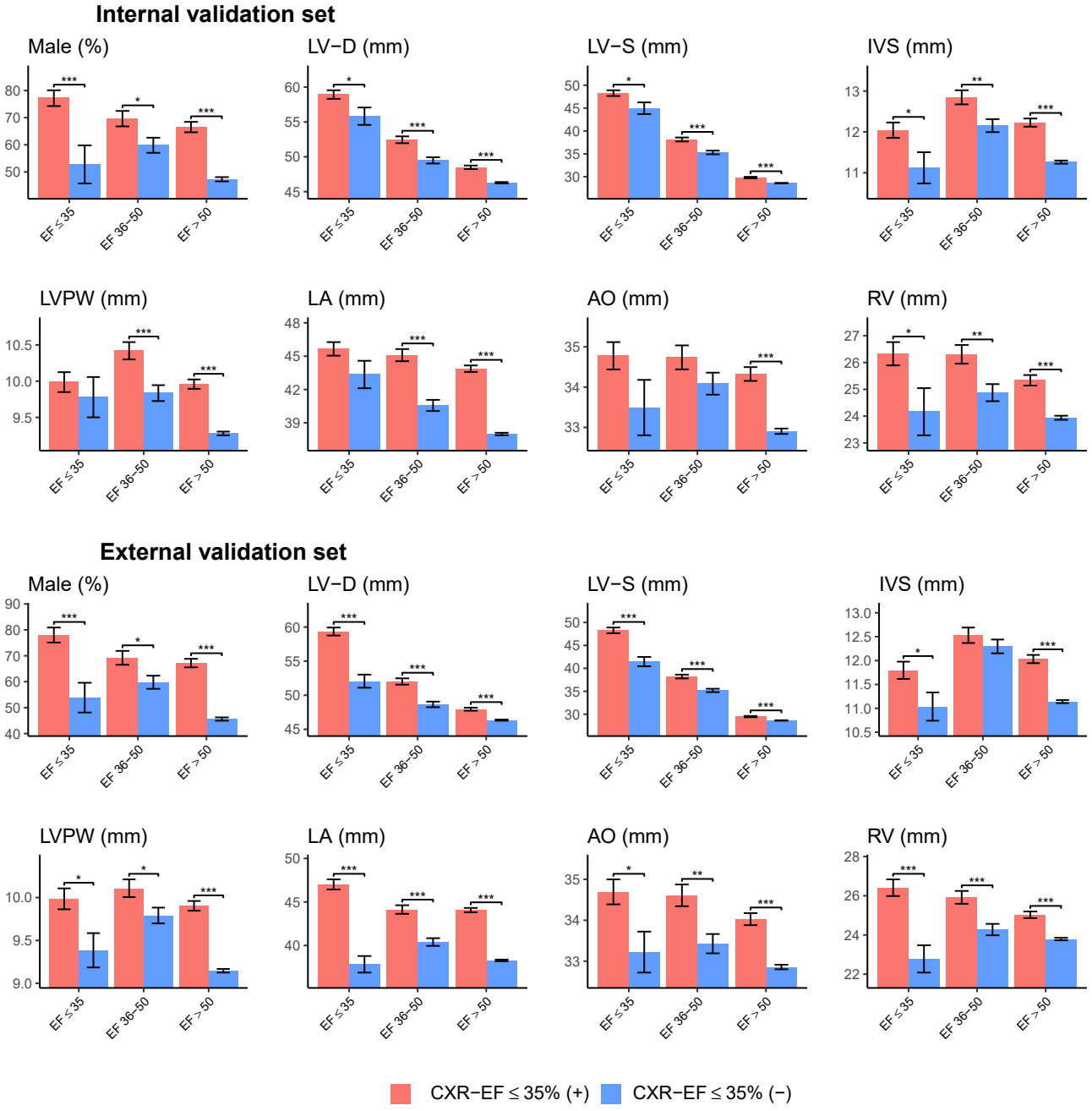


Figure 4. Echocardiographic characteristics stratified by *CXR-EF≤35%* and baseline left ventricular ejection fraction (LVEF). **Bars** represent the mean or proportion as appropriate and corresponding 95% conference intervals, which were adjusted by baseline LVEF via linear or logistic regression (**P* for trend < 0.05; ***P* for trend < 0.01; ****P* for trend < 0.001). AO, aortic root diameter; *CXR-EF≤35%,* deep-learning model to identify LVEF ≤ 35% via chest X-ray; IVS, interventricular septum thickness; LA, left atrial diameter; LV-D, left ventricular end-diastolic diameter; LV-S, left ventricular end-systolic diameter; LVPW, left ventricular posterior wall thickness; RV, right ventricular outflow tract proximal diameter in parasternal long axis view.

Clinic study screening a population aged ≥ 45 with the use of BNP and NT-proBNP demonstrated that the AUCs were greater for individuals with more severe systolic dysfunction, (0.89 and 0.94, respectively) than for those with any systolic dysfunction (0.72 and 0.78).[27] The screening method with biomarkers was time consuming and required blood draws. In addition, optimal discriminatory levels for both natriuretic peptide biomarkers varied with age, sex, and baseline comorbidities.[8] Overall, BNP or NT-proBNP was a suboptimal screening test for LVSD in the general population. In contrast, *CXR-EF≤35%* presented with equivalent performance across age and sex strata. This capability appears to be unique for the screening of LVSD.

In addition to validly identifying patients with LVEF ≤ 35%, intriguingly, *CXR-EF≤35%* may predict long-term development of LVEF ≤ 35%. *CXR-EF≤35%* identified patients with an initial LVEF > 50% who were at risk of subsequently developing LVEF ≤ 35%. The long-term

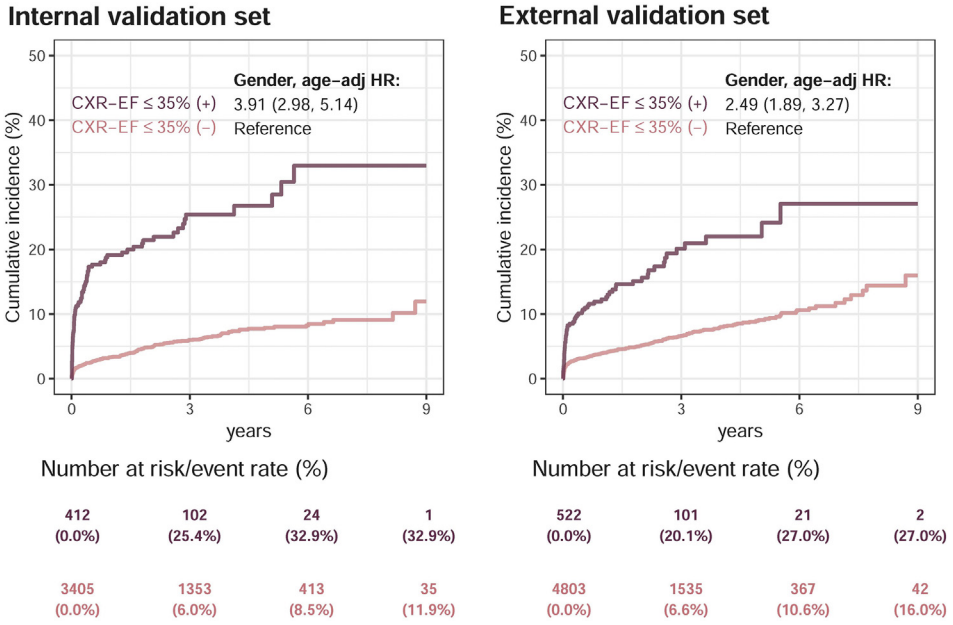


Figure 5. Long-term incidence of developing left ventricular ejection fraction (LVEF) ≤ 35% in patients with an initial LVEF > 50% stratified by *CXR-EF≤35%*. The *y*-axis indicates the cumulative incidence of developing LVEF ≤ 35%, and the *x*-axis indicates years from the time of the first chest X-ray (CXR). Sex and age were adjusted in this model. *CXR-EF≤35%*, deep-learning model to identify LVEF ≤ 35% via CXR; HR, hazard ratio.

incidence of developing LVEF ≤ 35% in patients with an initial LVEF > 50% stratified by *CXR-EF≤35%+* was significantly increased by a 2- to 4-fold compared with those stratified by *CXR-EF≤35%−* in both validation cohorts. These results indicated that *CXR-EF≤35%* may have the capability to identify CXR manifestations to predict long-term development of LVEF ≤ 35% before overt LVEF ≤ 35% manifests. Nevertheless, further prospective studies are needed to validate this prediction model.

CXR-EF≤35% may have several potential clinical applications. First, with the incorporation of *CXR-EF≤35%*, CXR may be used not only to detect cardiopulmonary abnormalities, but also preliminarily to identify LVEF ≤ 35% with subsequent TTE validation. Second, it may serve as a

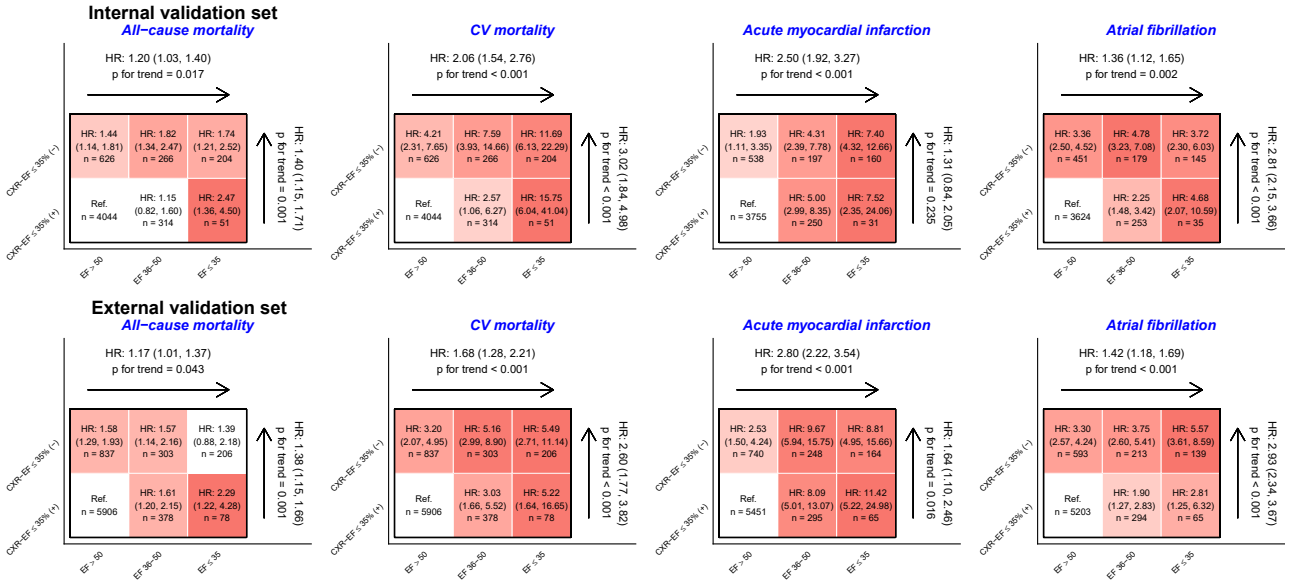


Figure 6. The risk matrixes of long-term cardiovascular (CV) outcomes stratified by *CXR-EF≤35%* and left ventricular ejection fraction (LVEF). The hazard ratios were based on Cox proportional hazard model adjusting for sex and age. The **color gradient** represents the risk of the corresponding group, and nonsignificant results are shown as **white**. *CXR-EF≤35%*, deep-learning model to identify LVEF ≤ 35% via chest X-ray; HR, hazard ratio (with 95% confidence limits).

screening tool to early identify potentially high-risk LVSD patientsm such as asymptomatic American College of Cardiology Foundation/American Heart Association HF stage A/B patients or HF with preserved LVEF with potential progression to reduced LVEF, probably enabling these individuals to early adopt further CV image examination, aggressive risk modification, or guideline-directed medical therapies. Third, this algorithm may be translated into preventive medicine or primary health care to identify individuals with long-term risk of developing LVEF $\leq$ 35% and CV outcomes. Fourth, to strengthen the ability of *CXR-EF$\leq$35%* to identify patients with LVEF $\leq$ 35%, it could be linked to the patient's medical history, risk factor assessment, biomarkers, or 12-lead ECG, and finally provide recommendations for clinicians to arrange further examination or imaging modalities to recognise patients with LVEF $\leq$ 35% and prevent future worse outcomes.

Limitations

Several limitations need to be addressed in this work. First, this was a retrospective study. Although CXRs were collected in outpatient departments, further community-based prospective studies are necessary to validate the accuracy and application of *CXR-EF$\leq$35%*. Second, despite the methods applied to reduce class imbalance and overfitting of the DLM, the generalisation of the DLM should be carefully evaluated. Further studies are warranted to confirm the model performance. Third, the CXR characteristics found by the DLM cannot be ascertained. A set of methods is used to create the model with raw data for automatic identification of features and relationships. Fourth, the CXR-TTE pairs were not simultaneously acquired. However, we collected all echocardiograms within 7 days, and 80% of TTEs were performed within 3 days. The inaccuracy related to temporal delay is small. Fifth, the detection of LVEF by *CXR-EF$\leq$35%* was divided into only $\leq$ 35% and $>$ 35%. An LVEF cutoff of 35% was selected owing to the well-established outcome and therapeutic implications of this value.[17] However, the identification of LVEF $<$ 50% is still clinically significant, reflecting an abnormal LVEF. Further detailed stratification of LVEF may be needed, especially in the range of 35%-50%. Finally, *CXR-EF$\leq$35%* was trained to identify LVEF $\leq$ 35% via CXR, serving as a screening tool, but it was unable to recognise patient's clinical condition, including acute decompensation, New York Heart Association functional class, or asymptomatic status.

Conclusion

We developed a DLM from a large set of CXRs validated by TTE, to identify LVEF $\leq$ 35%. The novel strategy provides a common, feasible, and affordable method to assist physicians in early identifying individuals with current or subsequent LVEF $\leq$ 35%. AI-enabled CXR may permit the addition of significant prognostic implication for LVEF $\leq$ 35% and serve as a screening tool to improve the quality of care in the CV field in the near future.

Funding Sources

This study was supported by the Smart Healthcare Project from the Medical Affairs Bureau, Ministry of National Defense, Taiwan. The source of funding was supported by the Medical Affairs Bureau, Ministry of National Defense (MND-MAB-D-111143 to Dr Fang), the Ministry of Science and Technology, Taiwan (MOST TSGH-B-110009 to Dr Wang and MOST109-2314-B-016-021 to Dr Fang), and the Tri-Service General Hospital, Taipei, Taiwan (TSGH-E-110263 to Dr Liu).

Disclosures

The authors have no conflicts of interest to disclose.

References

1. Bleumink GS, Knetsch AM, Sturkenboom MC, et al. Quantifying the heart failure epidemic: prevalence, incidence rate, lifetime risk and prognosis of heart failure: the Rotterdam Study. Eur Heart J 2004;25:1614-9.

2. Ambrosy AP, Fonarow GC, Butler J, et al. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. J Am Coll Cardiol 2014;63:1123-33.

3. Garbi M, McDonagh T, Cosyns B, et al. Appropriateness criteria for cardiovascular imaging use in heart failure: report of literature review. Eur Heart J Cardiovasc Imaging 2015;16:147-53.

4. Nagueh SF, Bhatt R, Vivo RP, et al. Echocardiographic evaluation of hemodynamics in patients with decompensated systolic heart failure. Circ Cardiovasc Imaging 2011;4:220-7.

5. Maggioni AP, Dahlström U, Filippatos G, et al. EURObservational Research Programme: regional differences and 1-year follow-up results of the Heart Failure Pilot Survey (ESC-HF Pilot). Eur J Heart Fail 2013;15:808-17.

6. Pocock SJ, Ariti CA, McMurray JJ, et al. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. Eur Heart J 2013;34:1404-13.

7. Yancy CW, Januzzi JL, Allen LA, et al. 2017 ACC expert consensus decision pathway for optimization of heart failure treatment: answers to 10 pivotal issues about heart failure with reduced ejection fraction: a report of the American College of Cardiology Task Force on Expert Consensus Decision Pathways. J Am Coll Cardiol 2018;71:201-30.

8. Ponikowski P, Voors AA, Anker SD, et al; ESC Scientific Document Group. 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure: the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure of the European Society of Cardiology (ESC) developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J 2016;37:2129-200.

9. Thomas JT, Kelly RF, Thomas SJ, et al. Utility of history, physical examination, electrocardiogram, and chest radiograph for differentiating normal from decreased systolic function in patients with heart failure. Am J Med 2002;112:437-45.

10. Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med 2018;15:e1002686.

11. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. Annu Rev Biomed Eng 2017;19:221-48.

12. Seah JC, Tang JS, Kitchen A, Gaillard F, Dixon AF. Chest radiographs in congestive heart failure: visualizing neural network learning. Radiology 2019;290:514-22.

13. Matsumoto T, Kodera S, Shinohara H, et al. Diagnosing heart failure from chest X-ray images using deep learning. Int Heart J 2020;61:781-6.

14. Raghunath S, Cerna AEU, Jing L, et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. Nat Med 2020;26:886-91.

15. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23, 2018; Salt Lake City, UT:7132-7141, https://doi.org/10.1109/CVPR.2018.00745.

16. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 2012;25:1097-105.

17. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence—enabled electrocardiogram. Nat Med 2019;25:70-4.

18. Dignam JJ, Zhang Q, Kocherginsky. The use and interpretation of competing risks regression models. Clin Cancer Res 2012;18:2301-8.

19. Mekov E, Miravitlles M, Petkov R. Artificial intelligence and machine learning in respiratory medicine. Exp Res Respir Med 2020;14:559-64.

20. Nam JG, Kim M, Park J, et al. Development and validation of a deep learning algorithm detecting 10 common abnormalities on chest radiographs. Eur Respir J 2021;57.

21. Kusunose K, Hirata Y, Tsuji T, Kotoku Ji, Sata M. Deep learning to predict elevated pulmonary artery pressure in patients with suspected pulmonary hypertension using standard chest X ray. Sci Rep 2020;10:1-8.

22. Wang TJ, Levy D, Benjamin EJ, Vasan RS. The epidemiology of asymptomatic left ventricular systolic dysfunction: implications for screening. Ann Intern Med 2003;138:907-16.

23. Yeboah J, Rodriguez CJ, Stacey B, et al. Prognosis of individuals with asymptomatic left ventricular systolic dysfunction in the Multi-Ethnic Study of Atherosclerosis (MESA). Circulation 2012;126:2713-9.

24. Echouffo-Tcheugui JB, Erqou S, et al. Assessing the risk of progression from asymptomatic left ventricular dysfunction to overt heart failure: a systematic overview and meta-analysis. JACC Heart Fail 2016;4:237-48.

25. de Lemos JA, McGuire DK, Khera A, et al. Screening the population for left ventricular hypertrophy and left ventricular systolic dysfunction using natriuretic peptides: results from the Dallas Heart Study. Am Heart J 2009;157:746-53. e742.

26. Ng L, Loke IW, Davies JE, et al. Community screening for left ventricular systolic dysfunction using plasma and urinary natriuretic peptides. J Am Coll Cardiol 2005;45:1043-50.

27. Costello-Boerrigter LC, Boerrigter G, Redfield MM, et al. Amino-terminal pro—B-type natriuretic peptide and B-type natriuretic peptide in the general community: determinants and detection of left ventricular dysfunction. J Am Coll Cardiol 2006;47:345-53.

Supplementary Material

To access the supplementary material accompanying this article, visit the online version of the *Canadian Journal of Cardiology* at www.onlinecjc.ca and at https://doi.org/10.1016/j.cjca.2021.12.019.