ESC HEART FAILURE
*ESC Heart Failure* 2023; **10**: 1103–1113
Published online 29 December 2022 in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/ehf2.14282

ORIGINAL ARTICLE

# Prognostic significance of pulmonary arterial wedge pressure estimated by deep learning in acute heart failure

Yuki Saito[1]*, Yuto Omae[2], Saki Mizobuchi[1], Hidesato Fujito[1], Masatsugu Miyagawa[1], Daisuke Kitano[1], Kazuto Toyama[1], Daisuke Fukamachi[1], Jun Toyotani[2] and Yasuo Okumura[1]

[1]*Division of Cardiology, Department of Medicine, Nihon University School of Medicine, Tokyo, Japan; and* [2]*Department of Industrial Engineering and Management, College of Industrial Technology, Nihon University, Chiba, Japan*

## Abstract

**Aims** Acute decompensated heart failure (ADHF) presents with pulmonary congestion, which is caused by an increased pulmonary arterial wedge pressure (PAWP). PAWP is strongly associated with prognosis, but its quantitative evaluation is often difficult. Our prior work demonstrated that a deep learning approach based on chest radiographs can calculate estimated PAWP (ePAWP) in patients with cardiovascular disease. Therefore, the present study aimed to assess the prognostic value of ePAWP and compare it with other indices of haemodynamic congestion.

**Methods and results** We conducted a post hoc analysis of a single-centre, prospective, observational heart failure registry and analysed data from 534 patients admitted for ADHF between January 2018 and December 2019. The deep learning approach was used to calculate ePAWP from chest radiographs at admission and discharge. Patients were divided into three groups based on the ePAWP tertiles at discharge, as follows: first tertile group (ePAWP $\leq$ 11.2 mm Hg, $n$ = 178), second tertile group (11.2 $<$ ePAWP $<$ 13.5 mm Hg, $n$ = 170), and third tertile group (ePAWP $\geq$ 13.5 mm Hg, $n$ = 186). The third tertile group had a higher prevalence of atrial fibrillation and lower systolic blood pressure at admission; a lower platelet count and higher total bilirubin at both admission and discharge; and a higher left atrial diameter, peak early diastolic transmitral flow velocity, right ventricular end-diastolic diameter, and maximal inferior vena cava diameter at discharge. During the median follow-up period of 289 days, 223 (41.7%) patients reached the primary endpoint (a composite of all-cause mortality or rehospitalization for heart failure). Kaplan–Meier analysis revealed a significantly higher composite event rate in the third tertile group (log-rank test, $P$ = 0.006). Even when adjusted for clinically relevant factors, a higher ePAWP at discharge and a smaller decrease in ePAWP from admission to discharge were significantly associated with higher event rates [ePAWP at discharge: hazard ratio, 1.10; 95% confidence interval (CI), 1.02–1.19; $P$ = 0.010; and size of ePAWP decrease: hazard ratio, 0.94; 95% CI, 0.89–0.99; $P$ = 0.038].

**Conclusions** Our study suggests that ePAWP calculated by a deep learning approach may be useful for identifying and monitoring pulmonary congestion during hospitalization for ADHF.

**Keywords** Heart failure; Artificial intelligence; Pulmonary congestion

*Received: 23 August 2022; Revised: 6 November 2022; Accepted: 15 December 2022*
**Correspondence to: Yuki Saito, MD, Division of Cardiology, Department of Medicine, Nihon University School of Medicine, 30-1 Ohyaguchi-kamicho, Itabashi-ku, Tokyo 173-8610, Japan. Email: saito.yuki@nihon-u.ac.jp*

## Introduction

Acute decompensated heart failure (ADHF) has high rates of mortality and hospitalization and represents a major burden for health care systems.[1] Pulmonary congestion often develops in ADHF because of increased pulmonary arterial wedge pressure (PAWP).[2] Among patients hospitalized for heart failure (HF), a persistent high PAWP and residual pulmonary congestion at discharge are strongly associated with high mortality and readmission rates.[2] Thus, alleviation of

congestion is one of the most important therapeutic goals in patients hospitalized for ADHF,[3] and reliable detection and monitoring of pulmonary congestion before discharge are important in these patients. Nevertheless, quantitative evaluation of pulmonary congestion is usually difficult, and about half of patients admitted for ADHF are reported to be discharged with residual congestion.[4]

Chest radiography is a fast, simple, and classic method to assess elevated PAWP and pulmonary congestion in HF,[5] but the interpretation of chest radiographs is subjective and does not necessarily provide an accurate evaluation of PAWP. Recently, the further development of artificial intelligence has allowed deep learning to be applied to medical images.[6] Our prior work demonstrated that a deep learning approach based on chest radiographs can calculate estimated PAWP (ePAWP) in patients with cardiovascular diseases who have undergone right heart catheterization.[7] However, data are lacking on the prognostic relevance of this deep learning approach in ADHF. We hypothesized that our deep learning application based on chest radiographs could be used to evaluate the severity of HF and predict adverse clinical outcomes in patients with ADHF. Therefore, this study aimed to assess the prognostic value of ePAWP calculated by this deep learning approach and to compare it with other indices of haemodynamic congestion in patients with ADHF.

# Methods

## Study population

The present study was a post hoc analysis of a single-centre, prospective, observational cohort registry, the SAKURA HF REGISTRY-2 (UMIN 000043852). This registry enrolled consecutive patients with ADHF admitted to Nihon University Itabashi Hospital, Tokyo, Japan, who agreed to be followed for the collection of outcome data, as previously described.[8] The diagnosis of ADHF was based on the Framingham criteria.[9] All patients provided written informed consent. To evaluate potential prognostic factors, demographic, laboratory, and echocardiographic data were obtained at admission and discharge. We included consecutive patients enrolled in the registry between January 2018 and December 2019. A total of 684 patients were screened, but we excluded 50 patients who died during hospitalization, 93 whose prognosis could not be determined, and 7 with missing chest radiographs at admission or discharge. Thus, we analysed data from 534 patients.

The study complied with the principles of the Declaration of Helsinki. The use of patient information was approved by the Nihon University Itabashi Hospital Ethics Committee (RK-180612-2).

## Deep learning approach for estimated pulmonary arterial wedge pressure

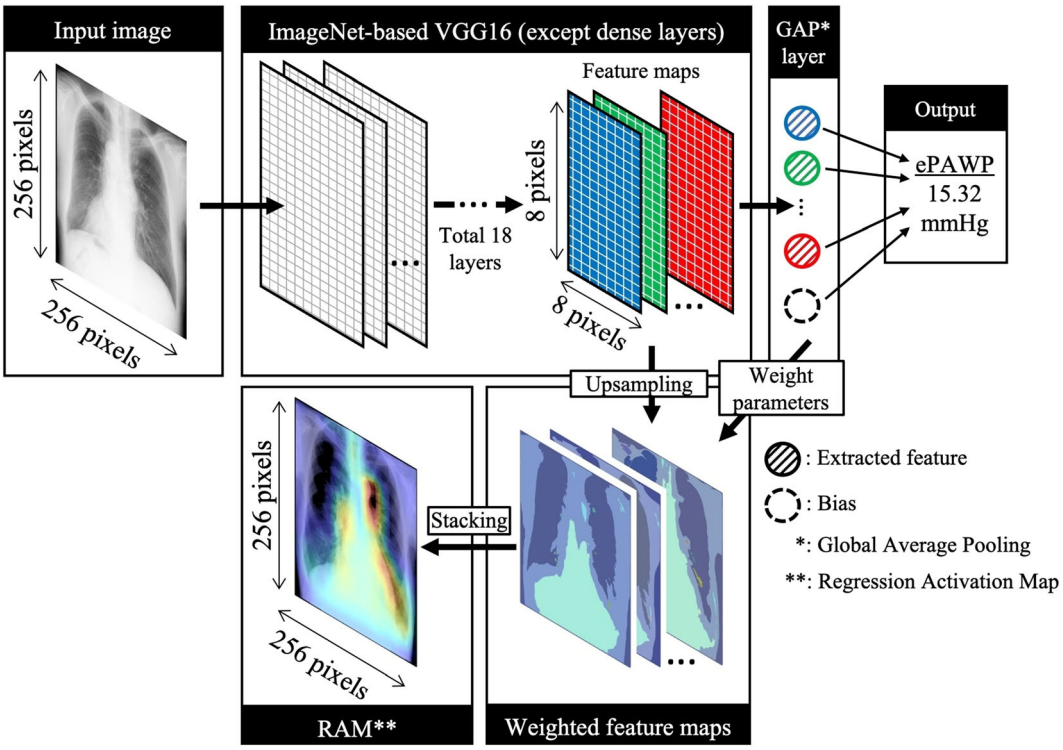
For this study, we used the previously developed ePAWP deep learning model based on chest radiographs.[7] Briefly, we constructed a regression convolutional neural network (CNN) based on the VGG16 model, one of the popular models for transfer learning.[10] Digital Imaging and Communications in Medicine images of the chest radiographs were transformed to images with a size of 256 × 256 pixels. After analysis by the CNN based on VGG16, the convoluted images were transitioned to a global average pooling (GAP) layer. We set the identity function as the activation function of the output layer to estimate PAWP as a numeric value. Furthermore, we used the GAP layer to create a regression activation map to visualize how our deep learning model estimated PAWP. The model was trained by cross-validating 748 data and was tested in 188 data. In the test data, the ePAWP generated by the deep learning model was significantly correlated with PAWP obtained by right heart catheterization ($r$ = 0.62, $P < 0.001$). The details of the model structures are shown in *Figure 1*.

We calculated ePAWP with the deep learning model by using the chest radiographs performed at admission and discharge and divided patients into three groups based on the ePAWP tertiles at discharge. In addition, we calculated the decrease in ePAWP by subtracting the ePAWP at discharge from that at admission. We also calculated the cardiothoracic ratio (CTR) for the assessment of cardiac chamber size from the chest radiographs performed at discharge. At admission, 296 (55.4%) chest radiographs were obtained in anteroposterior view, and 238 (44.6%) were in posteroanterior view. At discharge, 153 (28.7%) chest radiographs were obtained in anteroposterior view, and 381 (71.3%) were in posteroanterior view.

## Laboratory tests and echocardiography

Routine laboratory data were collected at admission and discharge. Echocardiography was performed before discharge by experienced sonographers according to the American Society of Echocardiography guidelines.[11] Left ventricular (LV) diastolic diameter and left atrial diameter were measured in the parasternal long-axis view, and LV ejection fraction (LVEF) was measured with the modified Simpson method or the Teichholz method. The right ventricular end-diastolic diameter (RVDd) was measured at the basal ventricular level of the right ventricle in end diastole. LV diastolic function was measured as the ratio of early transmitral flow velocity to mitral annular velocity (E/e'), which was calculated by transmitral Doppler flow and tissue Doppler imaging. The tricuspid regurgitation pressure gradient (TRPG) was measured by continuous-wave Doppler imaging. The severity of valvular

**Figure 1** Structure of the regression convolutional neural network. The model is based on the VGG16 model. ePAWP, estimated pulmonary arterial wedge pressure; GAP, global average pooling; RAM, regression activation map.




regurgitation was graded by using colour flow Doppler images. The diameter of the inferior vena cava (IVC) was measured from the subcostal view.

## Follow-up and clinical outcomes

Patients were prospectively followed up until death or December 2020. The primary endpoint was the composite of all-cause mortality or rehospitalization for HF.

## Statistical analysis

Continuous variables are presented as medians [interquartile range (IQR)], and categorical variables as numbers (percentages). Statistical differences between continuous variables were compared by a one-way analysis of variance followed by the post hoc Tukey–Kramer test or the Kruskal–Wallis test followed by the Steel–Dwass test. Comparisons of the ePAWP values at admission and discharge were performed with the paired *t*-test. Categorical variables were compared with the $\chi^2$ test with Bonferroni correction. The Kaplan–Meier method was used to analyse patient survival, and the log-rank test was used to compare group differences. Correlations between variables were evaluated based on Pearson's correlation coefficients. Univariate Cox proportional hazards analyses were performed to evaluate the association between ePAWP at discharge, ePAWP decrease, and event incidences. Hazard ratios with 95% confidence intervals (CIs) were calculated.

We compared the prognostic value of ePAWP with the Meta-Analysis Global Group in Chronic Heart Failure (MAGGIC) risk score, one of the established risk stratification scores in patients with ADHF. The MAGGIC risk score was calculated on the basis of 13 clinically relevant factors that can affect risk [age, sex, body mass index (BMI), systolic blood pressure, LVEF, creatinine level, current smoking, type 2 diabetes, chronic obstructive pulmonary disease, New York Heart Association class, duration of HF $>$ 18 months, beta-blocker use, and angiotensin-converting enzyme inhibitor use].[12]

We performed a multivariable analysis by constructing four multivariate models, as follows: In Model 1, we adjusted for the effects of clinically relevant factors that can affect risk, including age, sex, log N-terminal pro-brain natriuretic peptide at discharge, hypertension, type 2 diabetes, haemoglobin at discharge, and estimated glomerular filtration rate at discharge; in Model 2, for LV function variables (LVEF, LV end-diastolic diameter, left atrial diameter, and E/e'); in Model 3, for right ventricular function variables (RVDd, TRPG, and maximal IVC diameter); in Model 4, for the MAGGIC risk score; and in Model 5, for CTR.

**Table 1** Clinical characteristics of patients stratified according to the tertiles of estimated pulmonary arterial wedge pressure at discharge

| Item | All patients ($N$ = 534) | First tertile group (ePAWP ≤ 11.2 mm Hg) ($n$ = 178) | Second tertile group (11.2 < ePAWP < 13.5 mm Hg) ($n$ = 170) | Third tertile group (ePAWP ≥ 13.5 mm Hg) ($n$ = 186) | $P$ value |
|---|---|---|---|---|---|
| *Demographic variables* | | | | | |
| Age, years | 77 (68–84) | 78 (68–85) | 77 (68–82) | 76 (67–83) | 0.37 |
| Male, *n* (%) | 332 (62.1) | 105 (58.9) | 111 (65.3) | 116 (62.3) | 0.47 |
| Body mass index at admission | 24.0 (21.2–26.9) | 23.0 (21.0–25.4) | 24.6 (21.7–27.2)* | 24.6 (20.7–27.8)* | 0.002 |
| Systolic blood pressure at admission, mm Hg | 139 (120–163) | 148 (125–168) | 144 (121–167) | 132 (113–156)*† | <0.001 |
| Type 2 diabetes, *n* (%) | 218 (40.8) | 72 (40.4) | 71 (41.7) | 75 (40.3) | 0.95 |
| Hypertension, *n* (%) | 388 (72.6) | 125 (70.2) | 129 (75.8) | 134 (72.0) | 0.47 |
| COPD, *n* (%) | 44 (8.2) | 15 (8.4) | 13 (7.6) | 16 (8.6) | 0.94 |
| Atrial fibrillation, *n* (%) | 249 (46.6) | 64 (35.9) | 91 (53.5)* | 94 (50.5)* | 0.001 |
| Ischaemic heart disease, *n* (%) | 146 (27.3) | 55 (30.9) | 49 (28.8) | 42 (22.5) | 0.17 |
| HFpEF, *n* (%) | 187 (35.0) | 32 (17.9) | 34 (20.0) | 27 (14.5) | 0.67 |
| NYHA class at admission | | | | | |
| II, *n* (%) | 69 (12.9) | 24 (13.4) | 25 (14.7) | 20 (10.7) | 0.51 |
| III, *n* (%) | 291 (54.5) | 93 (52.2) | 88 (51.7) | 110 (59.1) | 0.28 |
| IV, *n* (%) | 174 (32.6) | 61 (34.2) | 57 (33.5) | 56 (30.1) | 0.66 |
| ePAWP at admission, mm Hg | 17.6 (14.8–21.3) | 15.5 (13.4–21.0) | 17.5 (15.0–21.3)* | 18.7 (16.6–21.6)*† | <0.001 |
| ePAWP decrease, mm Hg | 4.4 (2.1–8.6) | 6.3 (3.6–11.0) | 5.0 (2.8–9.0)* | 2.6 (0.5–4.9)*† | <0.001 |
| CTR at discharge, % | 54 (50–60) | 51 (18–55) | 54 (51–59)* | 59 (54–63)*† | <0.001 |
| MAGGIC risk score | 27 (22–30) | 27 (22–30) | 26 (22–29) | 27 (23–30) | 0.45 |
| *Medications at discharge* | | | | | |
| ACE-I or ARB, *n* (%) | 385 (72.1) | 137 (76.9) | 124 (72.9) | 124 (66.6) | 0.08 |
| Beta-blocker, *n* (%) | 442 (82.7) | 135 (75.8) | 153 (90.0)* | 154 (82.8) | 0.001 |
| Diuretics, *n* (%) | 434 (81.2) | 140 (78.6) | 140 (82.3) | 154 (82.8) | 0.54 |
| *Laboratory variables at admission* | | | | | |
| Haemoglobin, g/dL | 11.6 (9.7–13.3) | 11.6 (9.8–13.3) | 11.6 (9.8–13.6) | 11.7 (9.6–13.4) | 0.87 |
| Platelet count, $\times 10^3/\mu L$ | 190 (150–246) | 204 (159–267) | 189 (153–253) | 183 (145–224)* | 0.02 |
| Total bilirubin, mg/dL | 0.71 (0.44–1.07) | 0.59 (0.39–0.96) | 0.70 (0.45–0.99) | 0.84 (0.52–1.30)*† | <0.001 |
| AST, U/L | 30 (22–46) | 29 (22–48) | 29 (22–44) | 30 (22–46) | 0.89 |
| ALT, U/L | 22 (13–38) | 23 (13–39) | 20 (13–38) | 21 (13–37) | 0.90 |
| Albumin, g/dL | 3.6 (3.2–3.9) | 3.5 (3.1–3.8) | 3.7 (3.3–4.0) | 3.6 (3.2–3.9) | 0.05 |
| Sodium, mEq/L | 141 (139–143) | 141 (139–143) | 141 (139–143) | 141 (138–143) | 0.85 |
| BUN, mg/dL | 23.2 (17.4–35.5) | 22.4 (16.8–34.5) | 23.1 (17.0–31.4) | 24.6 (18.3–36.8) | 0.17 |
| Cr, mg/dL | 1.16 (0.87–1.71) | 1.14 (0.83–1.82) | 1.14 (0.91–1.49) | 1.21 (0.91–1.71) | 0.62 |
| eGFR, mL/min/1.73 $m^2$ | 43.3 (28.1–59.9) | 43.0 (26.9–61.5) | 46.5 (30.3–58.2) | 39.9 (27.3–60.2) | 0.55 |
| NT-proBNP, pg/mL | 5172 (2543–11 057) | 5289 (2836–12 134) | 4380 (2113–9578) | 5604 (2493–11 010) | 0.25 |
| *Laboratory variables at discharge* | | | | | |
| Haemoglobin, g/dL | 11.4 (10.1–12.9) | 11.2 (9.8–12.8) | 11.5 (10.2–12.7) | 11.4 (10.1–13.1) | 0.45 |
| Platelet count, $\times 10^3/\mu L$ | 189 (149–250) | 202 (156–255) | 193 (150–254) | 181 (136–229)* | 0.012 |
| Total bilirubin, mg/dL | 0.53 (0.37–0.74) | 0.45 (0.34–0.60) | 0.51 (0.36–0.73) | 0.63 (0.45–0.81)*† | <0.001 |
| AST, U/L | 22 (17–30) | 22 (16–30) | 22 (16–31) | 22 (18–28) | 0.90 |

(*Continues*)

**Table 1** (continued)

| Item | All patients ($N$ = 534) | First tertile group (ePAWP ≤ 11.2 mm Hg) ($n$ = 178) | Second tertile group (11.2 < ePAWP < 13.5 mm Hg) ($n$ = 170) | Third tertile group (ePAWP ≥ 13.5 mm Hg) ($n$ = 186) | $P$ value |
|---|---|---|---|---|---|
| ALT, U/L | 16 (11–27) | 16 (10–28) | 16 (10–28) | 16 (11–25) | 0.88 |
| Albumin, g/dL | 3.4 (3.0–3.7) | 3.3 (2.9–3.7) | 3.4 (3.1–3.8) | 3.3 (3.0–3.7) | 0.12 |
| Sodium, mEq/L | 140 (138–142) | 141 (138–142) | 141 (139–142) | 140 (137–142) | 0.14 |
| BUN, mg/dL | 23.7 (17.5–33.5) | 24.7 (17.0–35.2) | 23.7 (17.7–32.5) | 23.2 (17.5–34.0) | 0.91 |
| Cr, mg/dL | 1.08 (0.85–1.57) | 1.07 (0.81–1.72) | 1.10 (0.87–1.53) | 1.06 (0.85–1.52) | 0.86 |
| NT-proBNP, pg/mL | 1769 (902–3561) | 1549 (763–2198) | 1675 (814–4031) | 2148 (1168–3642) | 0.21 |

ACE-I, angiotensin-converting enzyme inhibitor; ALT, alanine aminotransferase; ARB, angiotensin II receptor blocker; AST, aspartate aminotransferase; BUN, blood urea nitrogen; COPD, chronic obstructive pulmonary disease; Cr, creatinine; CTR, cardiothoracic ratio; eGFR, estimated glomerular filtration rate; ePAWP, estimated pulmonary arterial wedge pressure; HFpEF, heart failure with preserved ejection fraction; MAGGIC, Meta-Analysis Global Group in Chronic Heart Failure; NT-proBNP, N-terminal pro-brain natriuretic peptide; NYHA, New York Heart Association.

Values are shown as median (interquartile range) unless otherwise indicated. For multiple comparisons, analysis of variance was used for symmetrical continuous variables, the Kruskal–Wallis test was used for asymmetric continuous variables, and the $\chi^2$ test was used for categorical variables. All pairwise comparisons were performed with the Tukey–Kramer test for symmetrical continuous variables, the Steel–Dwass test for asymmetric continuous variables, and the $\chi^2$ test with Bonferroni correction for categorical variables.

*$P < 0.05$ vs. first tertile group.

†$P < 0.05$ vs. second tertile group.

All statistical analyses were performed with JMP 16.2 (SAS Institute, Cary, NC, USA). For all analyses, a $P$ value of $<0.05$ was considered statistically significant.

# Results

## Patient characteristics

The clinical characteristics of the study patients are presented in *Tables 1* and *2*. In our cohort, the median age is 77 (IQR, 68–84), 62.1% of patients are male, 146 (27.3%) patients had ischaemic heart disease, and 249 (46.6%) patients had atrial fibrillation. The median value of LVEF is 48 (IQR, 34–61) %, and 208 (39.3%) patients had moderate or severe mitral regurgitation (MR). The median value of ePAWP at discharge was 12.3 (IQR, 10.6–14.5) mm Hg. The ePAWP tertiles at discharge were 11.2 and 13.5 mm Hg, so we used these values to divide patients into three groups, as follows: first tertile group (ePAWP ≤ 11.2 mm Hg, $n$ = 178), second tertile group (11.2 < ePAWP < 13.5 mm Hg, $n$ = 170), and third tertile group (ePAWP ≥ 13.5 mm Hg, $n$ = 186). Patients in the third tertile group, that is, the one with the highest ePAWP, had a higher prevalence of atrial fibrillation and lower systolic blood pressure at admission; lower platelet count and higher total bilirubin at both admission and discharge; higher CTR at discharge; and higher left atrial diameter, peak early diastolic E, RVDd, and maximal IVC diameter at discharge (all $P < 0.05$). This group also had higher ePAWP at admission and a smaller decrease in ePAWP from admission to discharge (all $P < 0.05$) (*Table 1*). Furthermore, the prevalence of moderate or severe MR and tricuspid regurgitation (TR) was significantly higher in the third tertile group (*Table 2*). We also investigated the correlation between ePAWP and echocardiographic parameters (E/A, E/e', and maximal IVC diameter). ePAWP was significantly and positively correlated with E/A ($r$ = 0.19, $P < 0.001$), E/e' ($r$ = 0.12, $P$ = 0.005), and maximal IVC diameter ($r$ = 026, $P < 0.001$).

## Changes in estimated pulmonary arterial wedge pressure during hospitalization

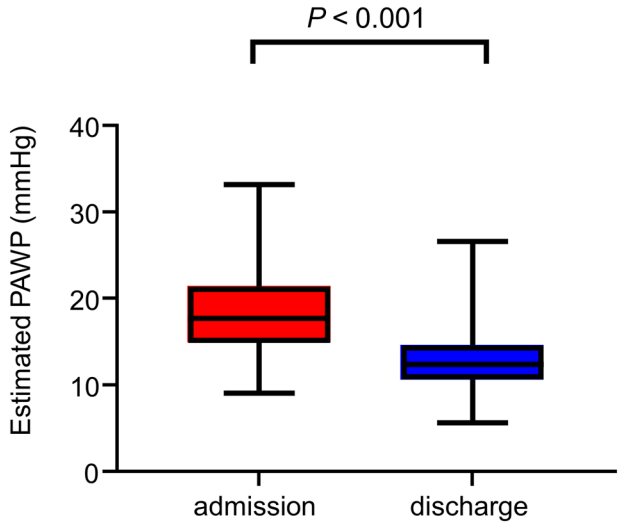
Among the total 534 patients, the median values of ePAWP at admission and discharge were 17.6 (IQR, 14.8–12.3) mm Hg and 12.3 (IQR, 10.6–14.5) mm Hg, respectively. The ePAWP significantly decreased during hospitalization ($P < 0.001$; *Figure 2*). *Figure 3* shows the images from a representative case at admission and discharge.

**Table 2** Echocardiographic data of patients stratified according to the tertiles of estimated pulmonary arterial wedge pressure at discharge

| Item | All patients ($N$ = 534) | First tertile group (ePAWP ≤ 11.2 mm Hg) ($n$ = 178) | Second tertile group (11.2 < ePAWP < 13.5 mm Hg) ($n$ = 170) | Third tertile group (ePAWP ≥ 13.5 mm Hg) ($n$ = 186) | $P$ value |
|---|---|---|---|---|---|
| *Echocardiographic variables at discharge* | | | | | |
| LVDd, mm | 52 (46–59) | 50 (44–56) | 53 (47–59)* | 53 (46–62)* | 0.005 |
| LVEF, % | 48 (34–61) | 49 (39–61) | 47 (33–61) | 47 (32–61) | 0.29 |
| LAD, mm | 45 (40–50) | 42 (36–46) | 45 (41–50)* | 48 (42–53)*† | <0.001 |
| E, cm/s | 86 (66–105) | 78 (57–95) | 88 (68–106)* | 89 (70–09)* | <0.001 |
| E/e' ratio | 13.9 (10.6–17.9) | 13.4 (9.9–17.4) | 14.4 (11.7–17.8) | 14.0 (10.6–18.7) | 0.24 |
| RVDd, mm | 32 (27–37) | 30 (26–35) | 32 (27–37) | 34 (28–40)*† | <0.001 |
| Moderate or severe MR, *n* (%) | 208 (39.3) | 53 (29.9) | 76 (44.7)* | 79 (43.6)* | 0.006 |
| Moderate or severe TR, *n* (%) | 157 (29.7) | 37 (20.9) | 41 (24.1) | 79 (43.6)*† | <0.001 |
| TRPG, mm Hg | 27 (20–34) | 26 (19–32) | 25 (19–33) | 28 (21–37) | 0.07 |
| Maximal IVC diameter, mm | 15 (12–19) | 14 (12–18) | 15 (12–19) | 17 (14–22)*† | <0.001 |

E, peak early diastolic transmitral flow velocity; e', peak early diastolic mean mitral annular velocity; ePAWP, estimated pulmonary arterial wedge pressure; IVC, inferior vena cava; LAD, left atrial diameter; LVDd, left ventricular end-diastolic diameter; LVEF, left ventricular ejection fraction; MR, mitral regurgitation; RVDd, right ventricular end-diastolic diameter; TR, tricuspid regurgitation; TRPG, tricuspid regurgitation pressure gradient.
Values are shown as median (interquartile range) unless otherwise indicated. For multiple comparisons, analysis of variance was used for symmetrical continuous variables, the Kruskal–Wallis test was used for asymmetric continuous variables, and the $\chi^2$ test was used for categorical variables. All pairwise comparisons were performed with the Tukey–Kramer test for symmetrical continuous variables, the Steel–Dwass test for asymmetric continuous variables, and the $\chi^2$ test with Bonferroni correction for categorical variables.
*$P < 0.05$ vs. first tertile group.
†$P < 0.05$ vs. second tertile group.

**Figure 2** In-hospital change in estimated PAWP in 534 patients with acute decompensated heart failure. Estimated PAWP was significantly lower at discharge than at admission. PAWP, pulmonary arterial wedge pressure.




## Clinical outcomes

During the median follow-up period of 289 days (IQR, 90–390 days), 223 (41.7%) patients reached the primary endpoint (all-cause mortality, 31 patients; rehospitalization for HF, 192 patients). The Kaplan–Meier analysis revealed a significantly higher composite event rate in the third tertile group (log-rank test, $P$ = 0.006; *Figure 4*).

## Ability of estimated pulmonary arterial wedge pressure to predict adverse events

The univariate Cox proportional hazards regression analysis showed that both higher ePAWP at discharge and a smaller decrease in ePAWP were significantly associated with the risk of composite events (Supporting Information, *Table S1*). In addition, the following factors were also significantly associated with the risk of composite events: higher age, MAGGIC risk score, CTR, blood urea nitrogen, log (N-terminal pro-brain natriuretic peptide), left atrial diameter, E/e', and TRPG; lower BMI, haemoglobin, platelet count, aspartate aminotransferase, and sodium; and the presence of hypertension, atrial fibrillation, or ischaemic aetiology. Even after adjusting for clinically relevant factors, both higher ePAWP at discharge and smaller decrease in ePAWP were significantly associated with higher event rates (ePAWP at discharge: hazard ratio, 1.10; 95% CI, 1.02–1.19; $P$ = 0.010; and change in ePAWP: hazard ratio, 0.94; 95% CI, 0.89–0.99; $P$ = 0.038; *Table 3*). The association between higher ePAWP and higher risk of composite events also remained significant in multivariate analysis, even after adjusting for factors related to left and right ventricular function and the conventional HF risk score, that is, the MAGGIC risk score (*Table 3*).

**Figure 3** Images from a 44-year-old man with acute decompensated heart failure. Estimation of PAWP from chest radiograph images at admission (A) and discharge (B). The regression activation map shows that, at admission, the deep learning model focused on pulmonary congestion; the estimated PAWP at admission was 19.1 mm Hg (A). At discharge, deep learning focused on the left side of the cardiac area, and the estimated PAWP was 10.7 mm Hg (B). NT-proBNP, N-terminal pro-brain natriuretic peptide; PAWP, pulmonary arterial wedge pressure.

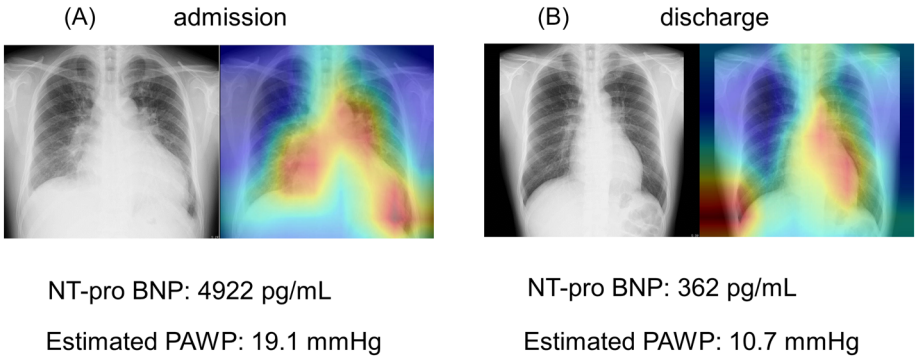



**Figure 4** Kaplan–Meier curves of event-free survival in 534 patients stratified according to the tertile of ePAWP at discharge. CI, confidence interval; ePAWP, estimated pulmonary arterial wedge pressure; HR, hazard ratio.

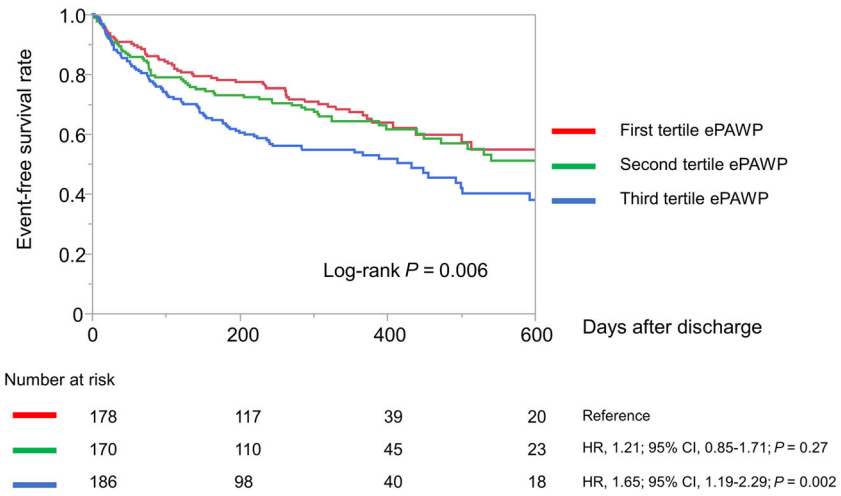



**Table 3** Cox proportional hazards regression analysis to assess the effect of both estimated pulmonary arterial wedge pressure at discharge and the decrease in estimated pulmonary arterial wedge pressure on composite risk

| | ePAWP at discharge (per 1 mm Hg increase) | | | ePAWP decrease (per 1 mm Hg change) | | |
|---|---|---|---|---|---|---|
| Variables | Hazard ratio | 95% CI | *P* | Hazard ratio | 95% CI | *P* |
| Model 1 | 1.10 | 1.02–1.19 | 0.010 | 0.94 | 0.89–0.99 | 0.038 |
| Model 2 | 1.06 | 1.01–1.11 | 0.011 | 0.98 | 0.95–1.01 | 0.30 |
| Model 3 | 1.07 | 1.02–1.12 | 0.002 | 0.99 | 0.96–10.2 | 0.68 |
| Model 4 | 1.07 | 1.03–1.12 | <0.001 | 0.97 | 0.95–1.00 | 0.10 |
| Model 5 | 1.07 | 1.02–1.12 | 0.003 | 0.98 | 0.95–1.00 | 0.21 |

CI, confidence interval; ePAWP, estimated pulmonary arterial wedge pressure.
The primary endpoint was defined as all-cause mortality and rehospitalization for heart failure. Model 1: adjusted for age, sex, and clinically relevant factors (log N-terminal pro-brain natriuretic peptide, haemoglobin, and estimated glomerular filtration rate at discharge; hypertension; and type 2 diabetes). Model 2: adjusted for age, sex, and left ventricular function variables (left ventricular ejection fraction, left ventricular end-diastolic diameter, left atrial diameter, and E/e'). Model 3: adjusted for age, sex, and right ventricular function variables (right ventricular end-diastolic diameter, tricuspid regurgitation pressure gradient, and maximal inferior vena cava diameter). Model 4: adjusted for Meta-Analysis Global Group in Chronic Heart Failure risk score. Model 5: adjusted for age, sex, and cardiothoracic ratio.

### Combination of estimated pulmonary arterial wedge pressure and conventional heart failure risk score for risk stratification

Although MAGGIC risk score showed a significant association with composite events (hazard ratio, 1.08; 95% CI, 1.05–1.10; $P < 0.001$), higher ePAWP remained associated with a higher risk of composite events even after adjustment for MAGGIC risk score (hazard ratio, 1.07; 95% CI, 1.03–1.12; $P < 0.001$). To evaluate the incremental prognostic value of ePAWP and MAGGIC risk score, we stratified patients into four groups by using the higher ePAWP tertile at discharge (i.e. 13.5 mm Hg) as the cutoff for high and low ePAWP and the median MAGGIC risk score (i.e. 26) for high and low MAGGIC risk score. The combined ePAWP and MAGGIC risk score subgroups had significantly different probabilities of composite events (log-rank test, $P < 0.001$; *Figure 5*). In particular, in the high MAGGIC risk score subgroups, the subgroup with high ePAWP had a significantly higher rate of composite events than the subgroup with low ePAWP (log-rank test, $P = 0.028$).

### Comparison of prognostic value between estimated pulmonary arterial wedge pressure and cardiothoracic ratio

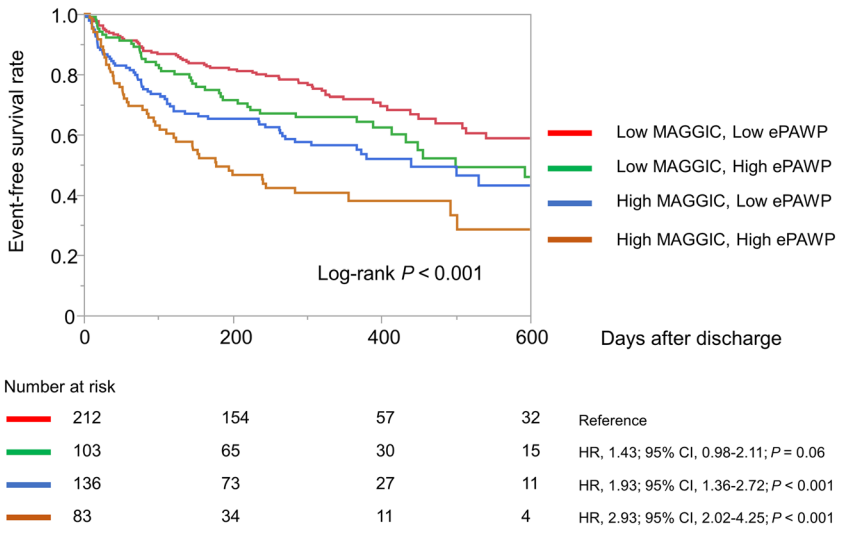
CTR was significantly correlated with ePAWP ($r = 0.47$, $P < 0.001$; Supporting Information, *Figure S1*), and univariate analysis revealed that CTR was significantly associated with the risk of composite events (hazard ratio, 1.03; 95% CI, 1.01–1.05; $P < 0.001$; Supporting Information, *Table S1*). Even after adjustment for age and sex, CTR remained associated with the risk of composite events (hazard ratio, 1.02; 95% CI, 1.00–1.04; $P = 0.01$). However, when ePAWP was included in the adjustment, CTR did not remain associated with the clinical outcomes (hazard ratio, 1.00; 95% CI, 0.98–1.03; $P = 0.50$); in contrast, ePAWP remained associated with the risk of composite events (hazard ratio, 1.07; 95% CI, 1.02–1.12; $P = 0.003$; *Table 3*).

## Discussion

### Major findings

The present study investigated the clinical and prognostic value of ePAWP in patients with ADHF by using a deep learning approach based on chest radiographs. The study has three major findings. First, ePAWP at discharge is associated with the severity of HF assessed by echocardiography. The third tertile group had a larger left atrium, higher E wave, more severe MR, larger IVC diameter, and more severe TR, suggesting that patients with higher ePAWP have higher left atrial and right atrial pressure. Furthermore, ePAWP was significantly and positively correlated with E/A, E/e', and maximal IVC diameter. Second, ePAWP at discharge is strongly associated with adverse clinical outcomes, including all-cause mortality and rehospitalization for HF. In addition, ePAWP at discharge shows an incremental prognostic value when combined with

**Figure 5** Probabilities of composite events in 534 patients with acute decompensated heart failure. Patients were divided into four groups by using the higher ePAWP tertile at discharge (13.5 mm Hg) as the cutoff for high and low ePAWP and the median MAGGIC risk score (26) as the cutoff for high and low MAGGIC score. CI, confidence interval; ePAWP, estimated pulmonary arterial wedge pressure; HR, hazard ratio; MAGGIC, Meta-Analysis Global Group in Chronic Heart Failure.



a conventional HF risk score. Last, ePAWP is significantly lower at discharge than at admission, and a smaller decrease is associated with adverse clinical outcomes. Our findings suggest that ePAWP obtained by a deep learning approach can be used as a reliable indicator of residual pulmonary congestion, which reflects the severity of HF, and can markedly improve risk stratification at discharge in patients hospitalized for ADHF.

## Pulmonary congestion and acute decompensated heart failure

In patients with developing ADHF, elevated PAWP leads to pulmonary congestion. Although HF treatment may alleviate clinical congestion, pulmonary congestion may persist and not be adequately resolved during hospitalization. About half of patients admitted for ADHF are reported to be discharged with residual congestion,[4] and residual pulmonary congestion at discharge is a significant prognostic marker in patients hospitalized with HF.[13] Although assessment of pulmonary congestion is crucial for the management of HF, quantitative assessment is often difficult.

## Artificial intelligence and chest radiographs

Chest radiography is a simple and useful tool for screening pulmonary congestion and elevated PAWP. However, it has some limitations, such as a large interobserver variability. With the development of artificial intelligence and deep learning in recent years, interest has grown in using these technologies to evaluate chest radiographs in patients with HF. Previous clinical studies revealed that by analysing chest radiograph images of patients with HF, a CNN model can identify elevated pulmonary pressure in patients with pulmonary hypertension and elevated PAWP.[14,15] These results demonstrate that deep learning models can be trained to estimate elevated left-sided heart pressure from chest radiograph images. In our previous study, we applied regression CNN to quantitatively estimate PAWP by deep learning from chest radiographs in patients with cardiovascular disease.[7] Our findings showed the potential usefulness of this approach because the PAWP estimated by the model from chest radiograph image data correlated with the actual PAWP measured by right heart catheterization.[7] However, we did not investigate the clinical significance of this method.

## Prognostic value of estimated pulmonary arterial wedge pressure calculated by deep learning

In the present study, we tested the prognostic ability of this deep learning model to predict adverse clinical outcomes in ADHF. The relationship between ePAWP and echocardiographic parameters showed that higher ePAWP is associated with parameters of not only left atrial but also right atrial pressure. This finding is reasonable because right atrial pressure can be affected by post-pulmonary hypertension caused by elevated left atrial pressure and volume status. These data suggest that ePAWP at discharge reflects the comprehensive severity of HF. In the present study, ePAWP at discharge was strongly associated with adverse clinical outcomes, including all-cause mortality and rehospitalization for HF, and even after adjusting for echocardiographic parameters, it was significantly associated with clinical outcomes. Our study also revealed that ePAWP at discharge provides additional prognostic value when combined with the MAGGIC risk score, a well-established comprehensive HF risk score. Patients with a high MAGGIC risk score and a high ePAWP had the highest risk of poor outcomes, whereas those with a low MAGGIC risk score and a low ePAWP had significantly better clinical outcomes. In particular, ePAWP could predict the differential prognosis in patients with a high MAGGIC risk score. Thus, ePAWP may be useful for further risk stratification in patients at high risk according to conventional HF risk scores and may be able to better identify patients at high risk of rehospitalization for HF.

## Comparison of prognostic relevance between estimated pulmonary arterial wedge pressure and cardiothoracic ratio

CTR is a widely used method to evaluate cardiac chamber size, and concordant with PAWP. In the present study, CTR was associated with adverse events in univariate analysis but did not remain a significant predictor for worse clinical outcomes in a multivariate model including ePAWP. Furthermore, ePAWP showed predictive value beyond CTR. The potential reason for these results is that the deep learning model assesses not only cardiac size but also pulmonary congestion to estimate PAWP. Dash *et al*. reported that CTR, alveolar oedema, interstitial oedema, and left atrial size each contributed independently to the prediction of PAWP in chest radiographs.[5] Therefore, CTR may not be sufficient to assess the level of PAWP and the severity of HF. In the present study, the regression activation map showed that the deep learning approach assesses both pulmonary congestion and cardiac size, and this may be the possible explanation for our results.

## Clinical implications

Chest radiography can be used as a simple, standardized, and inexpensive diagnostic tool. The evaluation of PAWP by a deep learning approach based on chest radiographs is objective and

repeatable. Our results show the usefulness of this method as a prognostic marker in patients hospitalized for ADHF. Furthermore, we consider that ePAWP can also be used to monitor the severity of congestion and the therapeutic effect of HF treatment during hospitalization. Our results show that ePAWP is significantly lower at discharge than at admission and that a smaller decrease is associated with adverse clinical outcomes. These data suggest that ePAWP can detect improvement of pulmonary congestion with diuretic therapy and that less improvement of ePAWP may be a marker for poor clinical prognosis after discharge. We believe that this study represents an important step towards applying a deep learning approach in the clinical assessment of acute and chronic pulmonary congestion in HF.

### Limitations

The present study has several limitations. First, it was a single-centre study with only a moderate sample size. Second, the study population was limited to patients with ADHF, so the study did not evaluate whether this method can be applied to patients with chronic HF being treated in outpatient clinics. Third, we focused mainly on the evaluation of pulmonary congestion at discharge and excluded patients who died during hospitalization, so the association between ePAWP at admission and in-hospital mortality is unclear. Last, different approaches in performing chest radiography (anteroposterior view and posteroanterior view) may have affected the results. In our population, 55% of chest radiographs were obtained in anteroposterior view at admission, whereas 71% of chest radiographs were obtained in posteroanterior view at discharge. Our deep learning model can calculate ePAWP from both views, but we cannot prove that the deep learning model calculates precisely equal values. A multi-centre study with a larger sample size is needed to validate our findings. Nevertheless, we believe that this study provides novel insights into the application of artificial intelligence and deep learning in HF management.

## Conclusions

Our study suggests that ePAWP calculated by a deep learning approach may be a useful tool to identify high-risk patients with HF and to monitor pulmonary congestion in patients hospitalized for ADHF.

## Acknowledgements

None.

## Conflict of interest

None declared.

## Funding

This research received no external funding.

## Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Figure S1.** Correlation between estimated pulmonary arterial wedge pressure and cardiothoracic ratio. CTR, cardiothoracic ratio; PAWP, pulmonary arterial wedge pressure.
**Table S1.** Univariate Cox proportional hazard analysis for risk of composite events in 534 patients hospitalized for acute decompensated heart failure.

## References

1. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, Falk V, González-Juanatey JR, Harjola VP, Jankowska EA, Jessup M, Linde C, Nihoyannopoulos P, Parissis JT, Pieske B, Riley JP, Rosano GMC, Ruilope LM, Ruschitzka F, Rutten FH, van der Meer P. 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**: 2129–2200.
2. Gheorghiade M, Follath F, Ponikowski P, Barsuk JH, Blair JE, Cleland JG, Dickstein K, Drazner MH, Fonarow GC, Jaarsma T, Jondeau G, Sendon JL, Mebazaa A, Metra M, Nieminen M, Pang PS, Seferovic P, Stevenson LW, van Veldhuisen DJ, Zannad F, Anker SD, Rhodes A, McMurray JJ, Filippatos G. Assessing and grading congestion in acute heart failure: a scientific statement from the Acute Heart Failure Committee of the Heart Failure Association of the European Society of Cardiology and endorsed by the European Society of Intensive Care Medicine. *Eur J Heart Fail*. 2010; **12**: 423–433.
3. McMurray JJ, Adamopoulos S, Anker SD, Auricchio A, Böhm M, Dickstein K, Falk V, Filippatos G, Fonseca C, Gomez-Sanchez MA, Jaarsma T, Køber L, Lip GY, Maggioni AP, Parkhomenko A, Pieske BM, Popescu BA, Rønnevik PK, Rutten FH, Schwitter J, Seferovic P, Stepinska J, Trindade PT, Voors AA, Zannad F, Zeiher A, Task Force for the Diagnosis and Treatment of Acute and Chronic

Heart Failure 2012 of the European Society of Cardiology, Bax JJ, Baumgartner H, Ceconi C, Dean V, Deaton C, Fagard R, Funck-Brentano C, Hasdai D, Hoes A, Kirchhof P, Knuuti J, Kolh P, McDonagh T, Moulin C, Popescu BA, Reiner Z, Sechtem U, Sirnes PA, Tendera M, Torbicki A, Vahanian A, Windecker S, McDonagh T, Sechtem U, Bonet LA, Avraamides P, Ben Lamin HA, Brignole M, Coca A, Cowburn P, Dargie H, Elliott P, Flachskampf FA, Guida GF, Hardman S, Iung B, Merkely B, Mueller C, Nanas JN, Nielsen OW, Orn S, Parissis JT, Ponikowski P. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *Eur J Heart Fail*. 2012; **14**: 803–869.

4. Ambrosy AP, Pang PS, Khan S, Konstam MA, Fonarow GC, Traver B, Maggioni AP, Cook T, Swedberg K, Burnett JC Jr, Grinfeld L, Udelson JE, Zannad F, Gheorghiade M, EVEREST Trial Investigators. Clinical course and predictive value of congestion during hospitalization in patients admitted for worsening signs and symptoms of heart failure with reduced ejection fraction: findings from the EVEREST trial. *Eur Heart J*. 2013; **34**: 835–843.
5. Dash H, Lipton MJ, Chatterjee K, Parmley WW. Estimation of pulmonary artery wedge pressure from chest radiograph in patients with chronic congestive cardiomyopathy and ischaemic cardiomyopathy. *Br Heart J*. 1980; **44**: 322–329.
6. Kagiyama N, Shrestha S, Farjo PD, Sengupta PP. Artificial intelligence: practical primer for clinical research in cardiovascular disease. *J Am Heart Assoc*. 2019; **8**: e012788.
7. Saito Y, Omae Y, Fukamachi D, Nagashima K, Mizobuchi S, Kakimoto Y, Toyotani J, Okumura Y. Quantitative estimation of pulmonary artery wedge pressure from chest radiographs by a regression convolutional neural network. *Heart Vessels*. 2022; **37**: 1387–1394.
8. Mizobuchi S, Saito Y, Fujito H, Miyagawa M, Kitano D, Toyama K, Fukamachi D, Okumura Y. Prognostic importance of improving hepatorenal function during hospitalization in acute decompensated heart failure. *ESC Hear Fail*. 2022; **9**: 3113–3123.
9. McKee PA, Castelli WP, McNamara PM, Kannel WB. The natural history of congestive heart failure: the Framingham study. *N Engl J Med*. 1971; **285**: 1441–1446.
10. Theckedath D, Sedamkar RR. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. *SN Comput Sci*. 2020; **1**: 79.
11. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, Flachskampf FA, Foster E, Goldstein SA, Kuznetsova T, Lancellotti P, Muraru D, Picard MH, Rietzschel ER, Rudski L, Spencer KT, Tsang W, Voigt JU. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J Am Soc Echocardiogr*. 2015; **28**: 1–39.
12. Pocock SJ, Ariti CA, McMurray JJ, Maggioni A, Køber L, Squire IB, Swedberg K, Dobson J, Poppe KK, Whalley GA, Doughty RN. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. *Eur Heart J*. 2013; **34**: 1404–1413.
13. Coiro S, Rossignol P, Ambrosio G, Carluccio E, Alunni G, Murrone A, Tritto I, Zannad F, Girerd N. Prognostic value of residual pulmonary congestion at discharge assessed by lung ultrasound imaging in heart failure. *Eur J Heart Fail*. 2015; **17**: 1172–1781.
14. 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*. 2020; **10**: 19311.
15. Hirata Y, Kusunose K, Tsuji T, Fujimori K, Kotoku J, Sata M. Deep learning for detection of elevated pulmonary artery wedge pressure using standard chest X-ray. *Can J Cardiol*. 2021; **37**: 1198–1206.