

A deep learning model to identify the fluid overload status in critically ill patients based on chest X-ray images

Xiaoyi Qin^{1*}, Wei Zhang^{2*}, Xiang Hu³, Wei Zhou⁴

1 Department of Hematology, First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China

2 Department of Thoracic Surgery, First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China

3 Department of Endocrine and Metabolic Diseases, First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China

4 Intensive Care Unit, First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China

KEY WORDS

chest X-ray, critical illness, deep learning, fluid overload

ABSTRACT

INTRODUCTION Recent studies have highlighted adverse outcomes of fluid overload in critically ill patients. Therefore, its early recognition is essential for the management of these patients.

OBJECTIVES Our aim was to propose a deep learning (DL) model using data from noninvasive chest X-ray (CXR) imaging associated with the fluid overload status.

PATIENTS AND METHODS We collected data from the Medical Information Mart for Intensive Care IV (MIMIC-IV, v. 1.0) and MIMIC Chest X-Ray (v. 2.0.0) databases for modeling, and from our hospital database for testing. The extravascular lung water index (ELWI) greater than 10 ml/kg and the global end-diastolic volume index (GEDI) greater than 700 ml/m² were used as threshold values for the fluid overload status. A DL model with a transfer learning strategy was proposed to predict the fluid overload status based on CXR images, and compared with clinical and semantic label models. Additionally, a visualization technique was adopted to determine the important areas of features in the input images.

RESULTS The DL model showed a relatively strong performance for predicting the ELWI (area under the curve [AUC] = 0.896; 95% CI, 0.819–0.972 and AUC = 0.718; 95% CI, 0.594–0.822, respectively) and the GEDI status (AUC = 0.814; 95% CI, 0.699–0.930 and AUC = 0.649; 95% CI, 0.510–0.787, respectively) in both the primary and the test cohort. The performance was better than that of the clinical and semantic label models.

CONCLUSIONS As CXR is routinely used in the intensive care unit, a simple, fast, low-cost, and noninvasive DL model based on this modality can be regarded as an effective supplementary tool for identifying fluid overload, and should be widely adopted in the clinical setting, especially when invasive hemodynamic monitoring is not available.

INTRODUCTION Fluid therapy for restoration and maintenance of tissue perfusion is a routine component of management of almost all critically ill patients.¹ Early and adequate fluid resuscitation by intravenous injection is considered crucial for the stabilization of tissue hypoperfusion, especially in patients with septic shock.² However, there is accumulating evidence that fluid overload is associated with increased mortality and can also lead to progressive organ dysfunction.^{3–5} Therefore, early recognition of the fluid overload status is essential for the management of critically

ill patients. Although determination of the actual extent of fluid overload in such patients is challenging, several methods of fluid overload quantification are worth taking into consideration, that is, clinical evaluation, cumulative fluid balance (FB), chest X-ray (CXR), ultrasound examination, bioimpedance vector analysis, and invasive hemodynamic monitoring.^{3,6}

The pulse indicator continuous cardiac output (PiCCO) system is an “all-inclusive” hemodynamic monitoring procedure for the assessment of fluid load using transpulmonary

Correspondence to:

Wei Zhou, MMed, Intensive Care Unit, First Affiliated Hospital of Wenzhou Medical University, Nan Bai Xiang Street, Ouhai District, Wenzhou, 325000 Zhejiang, China, phone: +86 0577 55579999, email: wzyyzw@yahoo.com

Received: August 11, 2022.

Revision accepted:

October 31, 2022.

Published online: January 4, 2023.

Pol Arch Intern Med. 2023;

133 (2): 16396

doi:10.20452/pamw.16396

Copyright by the Author(s). 2023

* XQ and WZha contributed equally to this work.

WHAT'S NEW?

In the present study we established a deep learning (DL) model based on chest X-ray (CXR) images to predict the fluid overload status in critically ill patients. Validated by an independent external test dataset, the DL model shows a relatively strong generalization performance. Additionally, as compared with previously reported CXR scores or our CXR label models, the DL model has certain advantages and the potential for clinical application. As CXR is routinely used in the intensive care unit, our DL model can be useful for identifying the fluid overload status, especially in the cases where invasive hemodynamic monitoring is unavailable. The present study is the first to combine data from the new MIMIC-IV database and the MIMIC-CXR database, providing a feasible ground for future research.

thermodilution technology.⁷ Extravascular lung water index (ELWI) is a measure of the volume of water accumulated in the lungs outside the pulmonary vascular system,⁷ whereas global end-diastolic volume index (GEDI), reflecting the blood volume in the 4 chambers of the heart, is a quantitative volumetric variable that directly measures cardiac preload.⁷ ELWI and GEDI have been shown to be reliable indicators of the volume status, and to have a number of advantages over traditional pressure preload parameters.⁷⁻¹⁰ However, the difficulty associated with repeated invasive procedures, complications during and after catheterization, unreliable measurements in the presence of some specific cardiopulmonary diseases, and high testing costs limit the availability of the PiCCO monitoring. Therefore, a noninvasive and readily available method to predict the volume status of critically ill patients is required.

CXR is one of the most accessible and repeatable examinations under routine conditions in the intensive care unit (ICU). Preliminary discrimination of patients with or without fluid overload during a CXR evaluation remains clinically important for ICU clinicians. Previous studies¹¹⁻¹⁸ have explored the predictive values of chest radiography for ELWI and GEDI using subjective and objective CXR scores in critically ill patients. However, the predictive performance of most CXR scoring systems was not satisfactory. In recent decades, artificial intelligence and deep learning (DL) have been widely used for research on medical imaging and have provided new prospects in the fields of medical diagnosis, treatment, and prognosis analysis.¹⁹ Using a DL strategy, CXR was employed to diagnose a wide spectrum of diseases from simply "abnormal findings" to more specific diagnoses, such as pneumonia, pneumothorax, and tuberculosis.²⁰ Nevertheless, there have been no previous studies on the applicability of DL for predicting the fluid overload status based on chest radiographs. Therefore, we proposed a DL model to explore the CXR imaging information associated with the fluid overload status based on the ELWI and GEDI values, and compared it with clinical and semantic label models.

PATIENTS AND METHODS **Data source** The data of the primary cohort were extracted from the Medical Information Mart for Intensive Care IV (MIMIC-IV, v. 1.0)^{21,22} and MIMIC Chest X-Ray (MIMIC-CXR, v. 2.0.0) databases.²²⁻²⁴ MIMIC-IV is a relational database containing real critical care data of patients admitted to the Beth Israel Deaconess Medical Center between 2008 and 2019.^{21,22} The MIMIC-CXR database is a large, publicly available dataset of chest radiographs with free-text radiology reports that contains 377 110 images corresponding to 227 835 radiographic studies.²²⁻²⁴ One of the study investigators (WZho) was allowed to download data from the databases, having completed the "Data or Specimens Only Research" course (record identity, 25222342).

Patient records and CXR images for the external test dataset were obtained from the First Affiliated Hospital of Wenzhou Medical University after approval from that institution's ethics committee (202302090921).

Informed consent of the study patients was not required because the present study did not use any protected health-related information or impacted clinical care.

Study design A flowchart of the study process is shown in [FIGURE 1](#). Patients with available PiCCO monitoring parameters assessed in the period between 24 hours before and 24 hours after the CXR examination were enrolled in the study. We excluded patients aged 16 years or younger, as well as those with repeated admissions to the ICU, the length of hospital stay shorter than 24 hours, CXR examinations performed outside the ICU, incomplete clinical data for further analysis, and comorbidities that interfered with the results of PiCCO monitoring, including pulmonary embolism, acquired or congenital absence of the lung, aortic aneurysm, congenital heart disease (eg, atrial or ventricular septal defect, tetralogy of Fallot, patent ductus arteriosus, and valvular regurgitation). The external test dataset was collected between July 1, 2017, and June 30, 2021, according to the same inclusion and exclusion criteria.

Data extraction The data were extracted from MIMIC-IV and the electronic medical records system of our hospital; detailed information is presented in [TABLES 1](#) and [2](#). The Sequential Organ Failure Assessment (SOFA) score was calculated based on the predefined criteria.²⁵ FB was calculated based on the following formula: $FB = (\text{total fluid in} - \text{total fluid out}) / \text{admission weight}$.³ All CXR images in the MIMIC-CXR dataset have been assigned specific semantic labels by CheXpert with binary mapping to 0 or 1 (Uncertain-Zeros model and Uncertain-Ones model).²⁶ Of note, CheXpert is a large dataset that contains a labeler to automatically detect the presence of 14 observations in radiology reports, capturing uncertainties inherent in the interpretation of

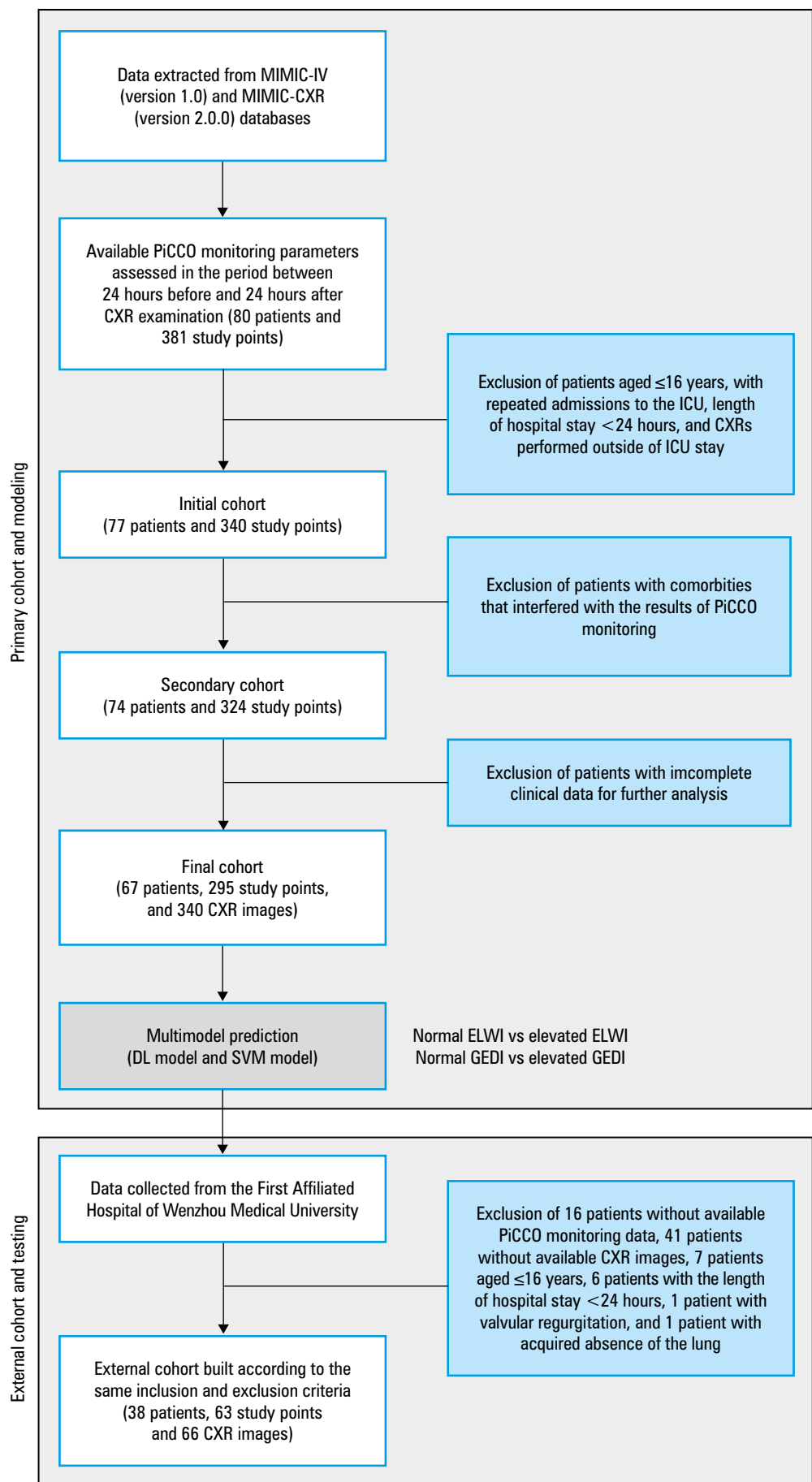


FIGURE 1 The flowchart of the study process

Abbreviations: CXR, chest X-ray; DL, deep learning; ELWI, extravascular lung water index; GEDI, global end-diastolic volume index; ICU, intensive care unit; MIMIC-CXR, Medical Information Mart for Intensive Care-Chest X-Ray; MIMIC-IV, Medical Information Mart for Intensive Care IV; PiCCO, pulse indicator continuous cardiac output; SVM, support vector machine

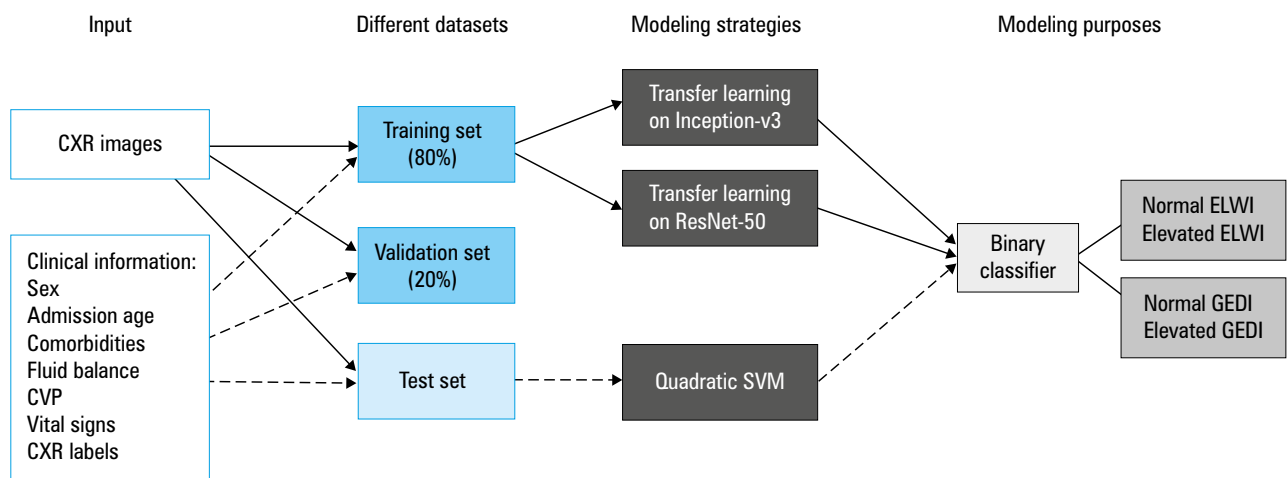


FIGURE 2 The pipelines of the deep learning (solid lines) and support vector machine (dotted lines) methods using chest X-ray images and clinical information for predicting the ELWI and GEDI statuses

Abbreviations: CVP, central venous pressure; others, see [FIGURE 1](#)

radiographs.²⁶ CXR labels of the external test dataset were obtained by consulting the radiology reports. If a label was uncertain, 2 radiologists re-read the CXR images to determine the final label by consensus.

We artificially set the examination time of each CXR as a study point. The mean values of PiCCO monitoring parameters and vital statistics assessed in the period between 24 hours before and 24 hours after the corresponding study point were regarded as baseline data. We evaluated the SOFA score, central venous pressure (CVP), and physical examinations assessed or performed at the time point closest to the corresponding study point and within 24 hours before and after the corresponding study point. The same principle applied to the measurement of FB.

Pulse indicator continuous cardiac output monitoring parameters

The PiCCO monitoring parameters recorded were cardiac output, cardiac index, systemic vascular resistance index, ELWI, and GEDI. ELWI greater than 10 ml/kg and GEDI greater than 700 ml/m² were regarded as threshold values for the fluid overload status.

Chest X-ray image acquisition and preprocessing

All frontal CXR images were directly obtained from the MIMIC-CXR-JPG database (v. 2.0.0), which is wholly derived from MIMIC-CXR, providing JPG files converted from DICOM images.^{22,23,27} To help prevent the network from overfitting and memorizing the exact details of the training images, a data augmentation method with randomized operations including reflection (X or Y axis), rotation (−180 ° to +180 °), rescaling (× 0.5 to × 2), horizontal translation (−50 pixels to +50 pixels), and vertical translation (−50 pixels to +50 pixels) was used. Before entering the network for training and validation, all images were converted to an RGB

format, resized to match the size of the network, and normalized.

Development, testing, and visualization of the deep learning model

The pipeline of the DL method with CXR images for predicting the ELWI and GEDI status is shown in [FIGURE 2](#). The CXR images in the MIMIC-CXR dataset were randomly divided into training (80%) and validation (20%) sets during modeling, and those from our hospital were used as a test set. The validation set was used to monitor model performance and protect against overfitting when training. We used 2 pre-trained networks, ResNet-50 and Inception-v3, convolutional neural networks 50 and 48 layers deep, respectively, to learn the natural image features from the ImageNet dataset.^{28,29} This transfer learning strategy is suitable for use with sparse datasets as it enlarges the training data. The detailed steps of transfer learning involve loading the data and pretrained network, replacing the last 3 layers with a new, fully connected layer, softmax layer, and classification output layer, fine-tuning for our new classification problem by specifying the options of the new fully connected layer, specifying training hyperparameters, and training a new network. In addition, the optimization of stochastic gradient descent with momentum was used to train the weights with an initial learning rate of 0.0001, minibatch size of 16, drop factor of 0.1, drop period of 10, and momentum of 0.9. To avoid overfitting and improve generalization, we adopted a strategy of multiple training networks (5 models), and each network was tested on the test dataset, with performance calculated for the average output. The final prediction score (Score_{final}) was then calculated using the following formulas weighted by the combined performance of ResNet-50 and Inception-v3:

$$\text{Score}_{\text{ResNet-50 or Inception-v3}} = (\text{Score}_{\text{model-1}} + \text{Score}_{\text{model-2}} + \text{Score}_{\text{model-3}} + \text{Score}_{\text{model-4}} + \text{Score}_{\text{model-5}}) / 5;$$

TABLE 1 Baseline characteristics of the study patients

Characteristics		MIMIC-CXR dataset (n = 67)	External test dataset (n = 38)
Sex	Male	38 (56.7)	24 (63.2)
	Female	29 (43.4)	14 (36.8)
Admission age, y	≤30	5 (7.5)	0
	> 30 and ≤60	28 (41.8)	18 (47.4)
	> 60	34 (50.7)	20 (52.6)
Race or ethnicity	White	43 (64.2)	0
	Black	4 (6)	0
	Hispanic	1 (1.5)	0
	Other	19 (28.4)	38 (100)
Admission weight, kg		80 (71.1–95)	59 (53–70)
Comorbidities	Congestive heart failure	12 (17.9)	7 (18.4)
	Chronic pulmonary disease	25 (37.3)	1 (2.6)
	Chronic renal disease	3 (4.5)	4 (10.5)
Length of stay, d	Hospital	19.11 (7.27–29.77)	13.85 (6.78–26.58)
	ICU	9.75 (4.94–15.72)	7.94 (6.35–12.47)

Data are presented as medians (interquartile ranges) or numbers (percentages).

Abbreviations: see [FIGURE 1](#)

$$\text{Score}_{\text{final}} = \alpha_1 \times \text{Score}_{\text{ResNet-50}} + \alpha_2 \times \text{Score}_{\text{Inception-v3}};$$

$$\alpha_1 = \text{AUC}_{\text{ResNet-50}} / (\text{AUC}_{\text{ResNet-50}} + \text{AUC}_{\text{Inception-v3}}),$$

$$\text{and } \alpha_2 = 1.000 - \alpha_1;$$

Here, $\text{AUC}_{\text{ResNet-50}}$ and $\text{AUC}_{\text{Inception-v3}}$ represent the area under the receiver operating characteristic curve (AUC) values of ResNet-50 and Inception-v3, respectively.

The DL network is very complex and its decisions are not intuitive with regard to interpretation. To further explain the prediction process of the DL network, we adopted a visualization method, that is, the locally interpretable model-agnostic explanations (LIME) technique, to determine the important areas of features in the input images.³⁰ Red areas on the LIME map had greater importance for classification decisions, while blue areas were less important.

Statistical analysis The numerical variables were expressed as medians with interquartile ranges. Categorical variables were expressed as frequencies with percentages. The variables clinically associated with fluid overload were considered for establishing the clinical model and the semantic label model (a detailed list of variables is shown in Supplementary material, *Tables S1* and *S2*). As shown in [FIGURE 2](#), the quadratic support vector machine (SVM) method was used for further modeling and testing. Similarly, the average score calculated from 5 SVM models was regarded as the final prediction output. The performances of all models were presented with the receiver operating characteristic (ROC) curve, sensitivity, specificity, and accuracy. We set the cutoff value

of score/probability for the ROC curve to 0.5. The Hanley and McNeil test was used to evaluate the differences in the AUC values among the different models.³¹ Additionally, cumulative frequency and decision curve were used to expansively interpret the predictive performance of the DL model.

A 2-sided *P* value below 0.05 was considered significant. Statistical analyses were performed using the SPSS software, version 22.0 (SPSS, Chicago, Illinois, United States) and the MedCalc software, version 19.0.5 (MedCalc, Ostend, Belgium). The DL model was implemented using the MATLAB software, version R2020b (MathWorks, Natick, Massachusetts, United States).

RESULTS Baseline characteristics Baseline characteristics of the study patients and study points are summarized in [TABLES 1](#) and [2](#), respectively. According to the threshold values of the parameters reflecting the fluid overload status, the patients were divided into groups with normal and elevated ELWI and GEDI.

Development and testing of the deep learning model In the primary cohort, a total of 340 CXR images (272 images for training and 68 images for validation) were used to transfer learning on our binary classification problems and build the predictive DL models. Subsequently, 66 CXR images from our hospital were used to test independently the generalization performance of the DL models. As shown in [TABLE 3](#), for both the validation and independent test datasets, the DL models showed encouraging average performance for predicting the ELWI (AUC = 0.896; 95% CI, 0.819–0.972 and AUC = 0.718; 95% CI, 0.594–0.822, respectively) and the GEDI status (AUC = 0.814; 95% CI, 0.699–0.930 and AUC = 0.649; 95% CI, 0.510–0.787, respectively). The ROC curves of the DL models in the 2 datasets are plotted in [FIGURE 3](#).

In addition, [FIGURE 4A](#) and [4B](#) shows that the cumulative frequency curves of the DL scores revealed significant differences between the normal and elevated ELWI groups (*P* < 0.001 in the validation dataset and *P* = 0.005 in the test dataset) as well as between the normal and elevated GEDI groups (*P* < 0.001 in the validation dataset and *P* = 0.05 in the test dataset) in the 2 datasets.

The decision curves of the DL models showed that when threshold probability was greater than 15.2% for predicting the ELWI status or between 36.8% and 83.2% for predicting the GEDI status, the DL models showed greater benefits than either treat-all or treat-none strategies ([FIGURE 4C](#) and [4D](#)). More importantly, the decision curves revealed an advantage of a maximum of 0.6 to 0.7 net benefit for the DL model with a broad range of threshold probability.

Comparison of the deep learning model with clinical and semantic label models The clinical model and the semantic label model were built for comparison with the DL model. Sex, age on admission,

TABLE 2 Baseline characteristics of the study points

Characteristics		MIMIC-CXR dataset (n = 295)	External test dataset (n = 63)
Fluid balance, ml/kg		69.27 (20.97–137.03)	45.53 (5.76–78.40)
SOFA score		7 (5–10)	10 (7–12)
CVP, mm Hg		13 (9–16)	12 (8–15)
Physical examination results	Bilateral lung crackles/wheezing	106 (35.9)	42 (66.7)
	General/lower extremity edema	199 (67.5)	17 (27)
Vital signs	Heart rate, bpm	95.51 (82.06–106.07)	113 (97–136)
	Mean blood pressure, mm Hg	79.36 (72.19–89.13)	88 (80–100)
	Respiratory rate, breaths/min	22.16 (19.02–25.38)	25 (20–30)
	SpO ₂ , %	97.58 (96.05–98.84)	98 (96–100)
	Glucose, mg/dl	130.07 (115.91–150.09)	167.4 (136.8–234.0)
PiCCO monitoring parameters	CO, l/min	6.53 (5.38–8.20)	5.54 (4.12–7.3)
	CI, l/min/m ²	3.48 (2.94–4.20)	3.13 (2.46–4.18)
	SVRI, dynes s cm ⁻⁵ /m ²	1637.5 (1240.38–2001)	2021 (1531–2654)
	ELWI, ml/kg	10.33 (8.4–15.67)	11.9 (9.9–15.5)
	Normal ELWI (≤10 ml/kg)	139 (47.1)	17 (27)
	Elevated ELWI (>10 ml/kg)	156 (52.9)	46 (73)
	GEDI, ml/m ²	762.33 (607.67–981.78)	775 (653–903)
	Normal GEDI (≤700 ml/m ²)	116 (39.3)	20 (31.7)
	Elevated GEDI (>700 ml/m ²)	179 (60.7)	43 (68.3)
CXR labels (Uncertain-Zeros) ^a	Cardiomegaly	59 (20)	19 (30.2)
	Lung edema	70 (23.7)	26 (41.3)
	Pleural effusion	135 (45.8)	5 (7.9)
CXR labels (Uncertain-Ones) ^a	Cardiomegaly	74 (25.1)	–
	Lung edema	91 (30.8)	–
	Pleural effusion	142 (48.1)	–

Data are presented as medians (interquartile ranges) or numbers (percentages).

a CXR labels in the MIMIC-CXR dataset were assigned by CheXpert with binary mapping to 0 or 1 (Uncertain-Zeros model and Uncertain-Ones model).

SI conversion factors: to convert glucose to mmol/l, multiply by 0.0555.

Abbreviations: CI, cardiac index; CO, cardiac output; SOFA, Sequential Organ Failure Assessment; SVRI, systemic vascular resistance index; others, see **FIGURES 1** and **2**

comorbidities (congestive heart failure, chronic pulmonary disease, and chronic renal disease), FB, CVP, and physical examination results (bilateral lung crackles / wheeze and general or lower extremity edema) were included as covariates in the clinical model. Three radiographic features (cardiomegaly, lung edema, and pleural effusion) related to fluid overload were incorporated into the semantic model. Moreover, the combined model trained by both clinical and semantic factors involved variables statistically significant in the univariable analysis when modeling, which are indicated in bold in Supplementary material, *Tables S1* and *S2*.

As shown in **TABLE 3** and **FIGURE 3**, the absolute values of AUC in the DL models were significantly higher than those in the clinical models for predicting the ELWI and the GEDI status. A similar improvement over the semantic model and the combined model was also observed in the 2 datasets. A detailed comparison

of AUCs is presented in Supplementary material, *Table S3*.

Visualization of the deep learning model To better understand the predictive principle of the DL network, we adopted the image interpretability technique to produce a smooth heatmap of the important areas by calculating the importance of rectangular features and upsampling the resulting map. **FIGURE 5** shows the simplified heat map of the DL model, as well as the corresponding prediction score and semantic labels. There may be several unreliable important areas on the heat map that interfered with the final DL score. However, the heat map still mainly focused on the lung and heart fields. Similarly, when clinicians read CXR images to evaluate the fluid overload status, they also focus on radiographic features of the lung and heart fields. As shown in **FIGURE 5**, the patients with lung edema or pleural effusion are often more likely to

TABLE 3 Predictive performance of various models for identifying the fluid overload status in the primary and test cohorts

Models	Cohorts	ELWI				GEDI			
		Sensitivity	Specificity	Accuracy	AUC	Sensitivity	Specificity	Accuracy	AUC
Clinical model	Primary	0.844 (0.665–0.941)	0.519 (0.324–0.708)	0.695 (0.560–0.805)	0.716 (0.579–0.854)	0.861 (0.697–0.948)	0.652 (0.428–0.828)	0.780 (0.650–0.873)	0.871 (0.783–0.959)
	Test	0.587 (0.433–0.727)	0.294 (0.114–0.560)	0.508 (0.380–0.635)	0.583 (0.407–0.759)	0.651 (0.490–0.786)	0.250 (0.096–0.494)	0.524 (0.395–0.650)	0.515 (0.368–0.663)
Semantic label model (Uncertain-Zeros) ^a	Primary	0.656 (0.468–0.808)	0.370 (0.201–0.575)	0.525 (0.392–0.655)	0.532 (0.398–0.664)	0.889 (0.730–0.964)	0.217 (0.083–0.442)	0.627 (0.491–0.747)	0.713 (0.581–0.823)
	Test	0.804 (0.656–0.901)	0 (0–0.229)	0.587 (0.456–0.708)	0.512 (0.383–0.640)	0.977 (0.862–0.999)	0.050 (0.003–0.269)	0.683 (0.552–0.791)	0.552 (0.421–0.677)
Semantic label model (Uncertain-Ones) ^a	Primary	0.563 (0.379–0.732)	0.630 (0.425–0.799)	0.593 (0.458–0.717)	0.556 (0.420–0.685)	0.889 (0.730–0.964)	0.217 (0.083–0.442)	0.627 (0.491–0.747)	0.658 (0.523–0.777)
	Test	0.674 (0.519–0.800)	0.588 (0.335–0.806)	0.651 (0.520–0.764)	0.672 (0.542–0.785)	0.977 (0.862–0.999)	0.050 (0.003–0.269)	0.683 (0.552–0.791)	0.515 (0.385–0.642)
Clinical + semantic label model (Uncertain-Zeros) ^a	Primary	0.625 (0.438–0.783)	0.778 (0.573–0.906)	0.695 (0.560–0.805)	0.709 (0.577–0.820)	0.750 (0.575–0.873)	0.565 (0.349–0.761)	0.678 (0.542–0.790)	0.810 (0.687–0.900)
	Test	0.435 (0.293–0.588)	0.353 (0.153–0.614)	0.413 (0.292–0.544)	0.682 (0.552–0.793)	0.721 (0.561–0.842)	0.300 (0.128–0.543)	0.587 (0.456–0.708)	0.554 (0.423–0.679)
Clinical + semantic label model (Uncertain-Ones) ^a	Primary	0.719 (0.530–0.856)	0.667 (0.460–0.828)	0.695 (0.560–0.805)	0.675 (0.541–0.792)	0.750 (0.575–0.873)	0.565 (0.349–0.761)	0.678 (0.542–0.790)	0.792 (0.667–0.887)
	Test	0.500 (0.351–0.649)	0.529 (0.285–0.761)	0.508 (0.380–0.635)	0.505 (0.376–0.634)	0.721 (0.561–0.842)	0.300 (0.128–0.543)	0.587 (0.456–0.708)	0.552 (0.421–0.677)
DL model	Primary	0.842 (0.681–0.934)	0.833 (0.646–0.937)	0.838 (0.725–0.913)	0.896 (0.819–0.972)	0.833 (0.680–0.925)	0.692 (0.481–0.849)	0.779 (0.659–0.867)	0.814 (0.699–0.930)
	Test	0.796 (0.652–0.893)	0.647 (0.386–0.847)	0.758 (0.634–0.851)	0.718 (0.594–0.822)	0.644 (0.487–0.777)	0.619 (0.387–0.810)	0.636 (0.508–0.749)	0.649 (0.510–0.787)

The results are expressed as ratios with 95% CIs.

a CXR labels in the MIMIC-CXR dataset were assigned by CheXpert with binary mapping to 0 or 1 (Uncertain-Zeros model and Uncertain-Ones model).

Abbreviations: AUC, area under the receiver operating characteristic curve; others, see [FIGURE 1](#)

have fluid overload. However, the results of our study showed that the performance of the DL model was superior to that of the cardiopulmonary semantic features identified on CXR images for predicting the fluid overload status, which was attributed to better discrimination of image details by the DL network.

DISCUSSION Fluid overload is an almost universal finding in critically ill patients due to the emphasis on the importance of early fluid resuscitation and the difficulty for most clinicians in accurately controlling the fluid intake to meet only the actual demand.^{3,6,32} Moreover, recent studies have highlighted the adverse outcomes of fluid overload, while a restrictive fluid strategy has been shown to significantly reduce the duration of ventilation and short-term mortality in critically ill patients.^{3-6,33} Therefore, early recognition of

fluid overload has become a primary component of the management of critically ill patients. As a substitute for pulmonary artery catheterization, with accurate assessment of the constant and dynamic hemodynamic statuses, PiCCO monitoring has been used extensively in the management of critically ill patients.³⁴ Furthermore, the parameters obtained from the PiCCO system, especially ELWI and GEDI, are considered effective and precise for quantitative evaluation of the fluid status of critically ill patients.^{7,33} ELWI is a marker reflecting the volume of water accumulated in the lungs outside the pulmonary vascular system, corresponding to the sum of interstitial, intracellular, alveolar, and lymphatic fluid, not including pleural effusion.^{7,33} ELWI, assessed using transpulmonary thermodilution technology, was shown to be closely associated with the gold standard gravimetric measurement in experimental animal studies.³⁵⁻³⁷

FIGURE 3

Comparison of ROC curves between different models for predicting the ELWI and GEDI statuses; **A** – ELWI in the primary cohort; **B** – ELWI in the test cohort; **C** – GEDI in the primary cohort; **D** – GEDI in the test cohort

Abbreviations: ROC, receiver operating characteristic; U, uncertain; others, see [FIGURE 1](#)

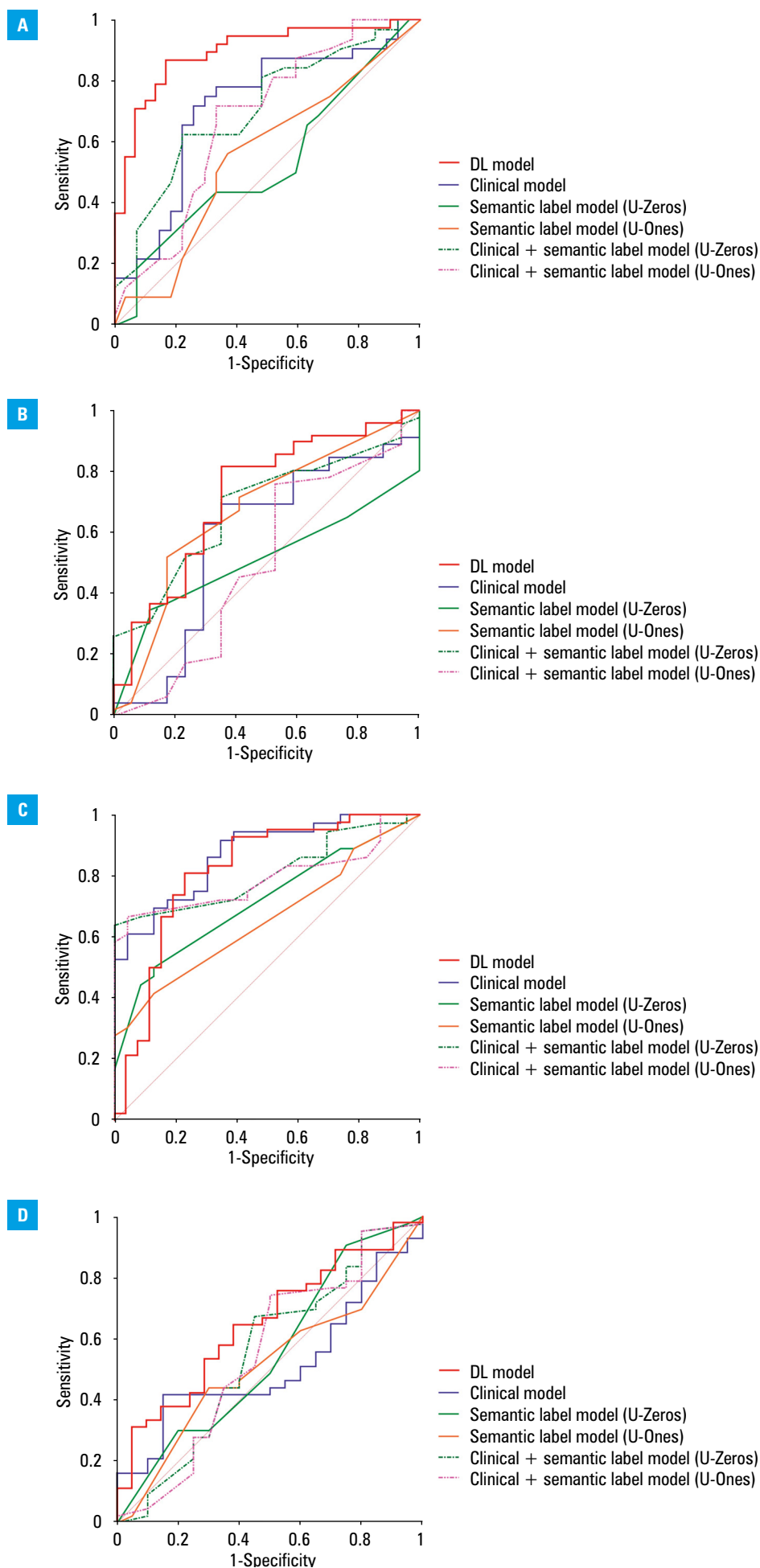


FIGURE 4

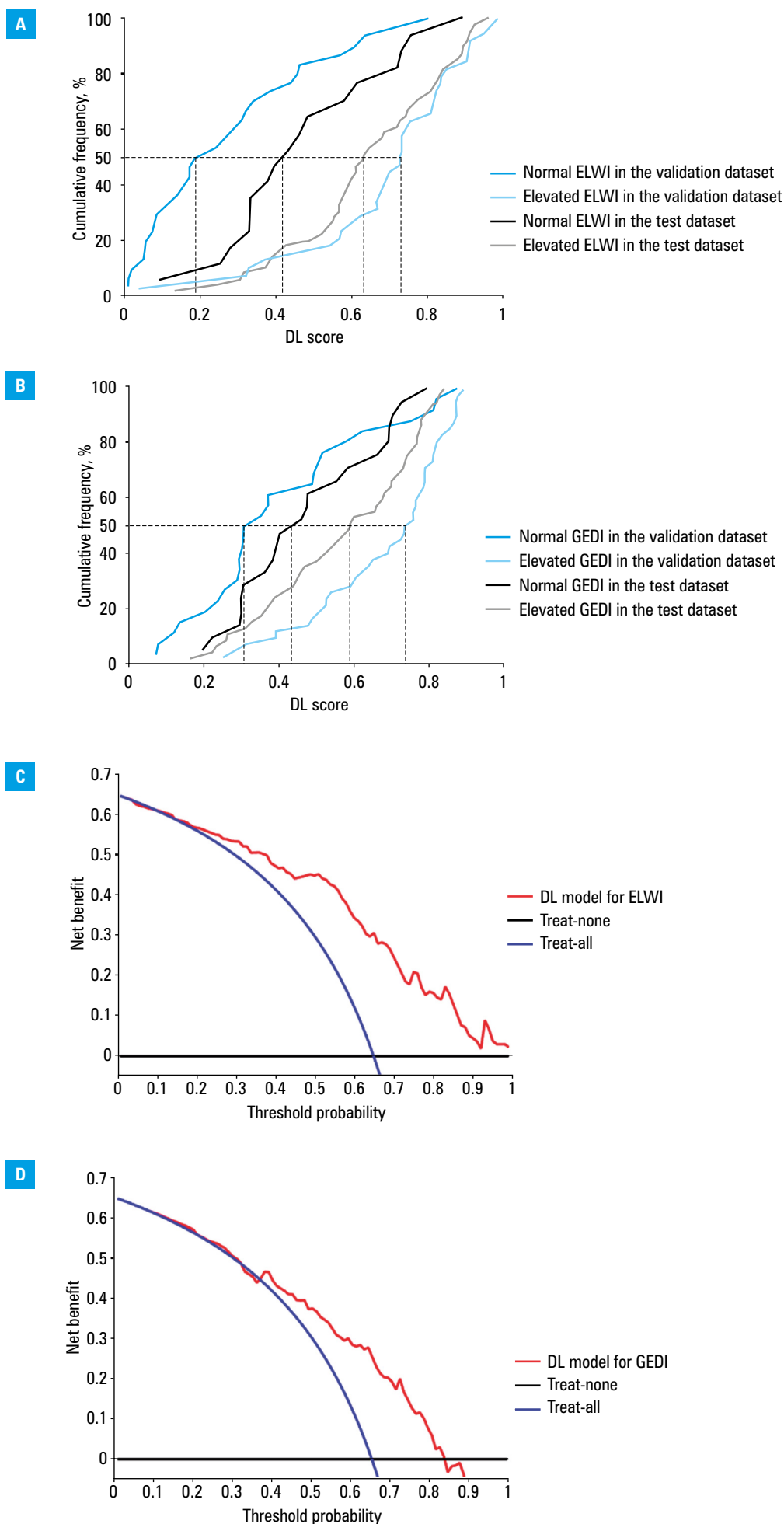
Predictive performance of the deep learning (DL) models;

A, B – cumulative frequency curves of the DL scores for predicting the **(A)** ELWI and **(B)** GEDI status in the 2 datasets. Dotted lines represent the DL scores corresponding to the median of the cumulative frequency.

C, D – decision curves of the DL models for predicting the **(C)** ELWI and **(D)** GEDI status

Abbreviations: see

FIGURE 1



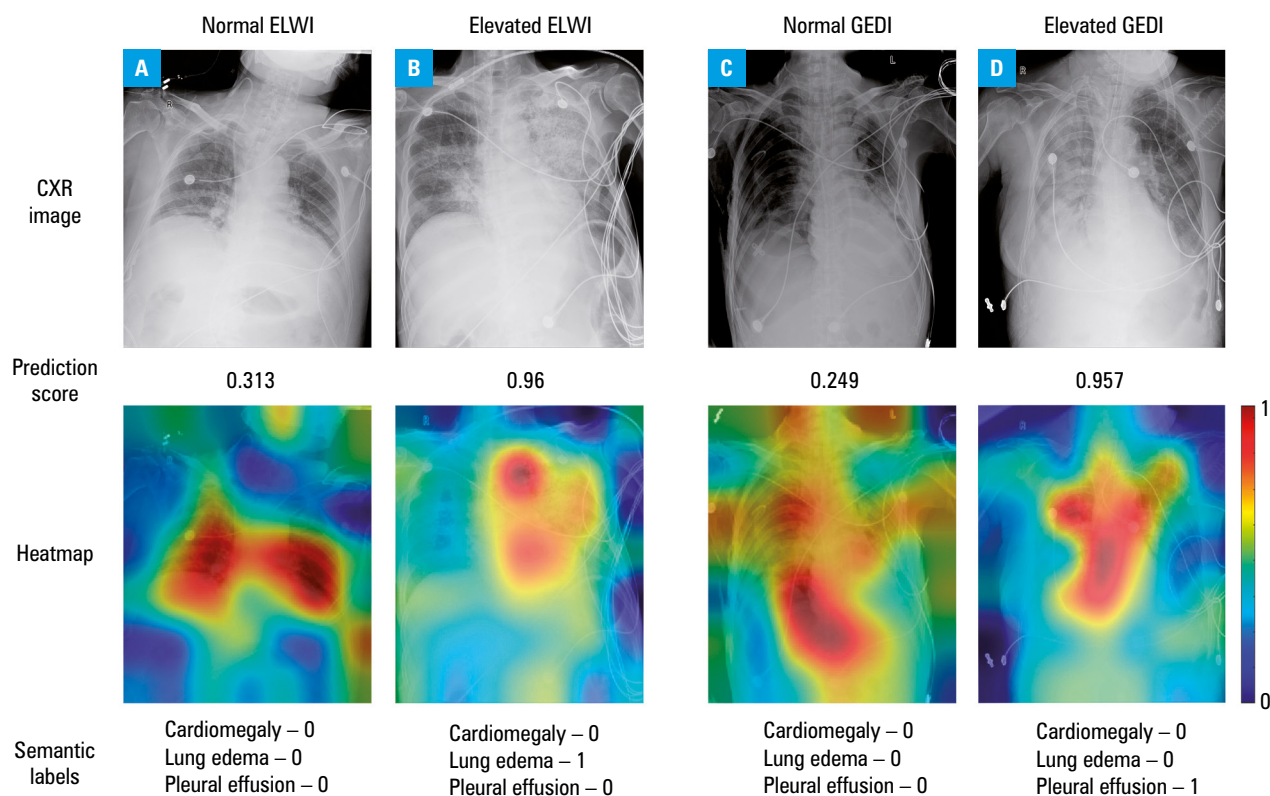


FIGURE 5 Heatmap of the deep learning model along with the corresponding prediction score and the semantic labels. Chest X-ray images of (A) normal ELWI, (B) elevated ELWI, (C) normal GEDI, and (D) elevated GEDI come from the external test dataset.

Abbreviations: see [FIGURE 1](#)

GEDI is used to assess the sum of intracardiac blood volume using transpulmonary thermodilution technology.⁷ Similarly to ELWI, GEDI has been shown to be a reliable indicator of cardiac preload and fluid responsiveness in animal models.^{38,39} Recent studies have confirmed that ELWI- and GEDI-guided fluid management can improve clinical outcomes in critically ill patients, including lower cumulative FB, improved short-term mortality, and shorter duration of mechanical ventilation and ICU stay.^{7,33,40,41} On the basis of the clinical decision tree, elevated ELWI (>10 ml/kg) and GEDI (>700 ml/m²) values indicate an increase in the risk of further fluid overload, and fluid removal should be initiated at the post-shock phase. However, limitations of the transpulmonary thermodilution technology may lead to unreliable measurements of ELWI and GEDI in the cases of pulmonary vascular occlusion, lung resection, heterogeneous lung injury, and application of positive end-expiratory pressure.^{7,33} We eliminated these interferences as much as possible in this study by excluding patients with related comorbidities. In addition, the difficulty of repeated invasive procedures, complications during and after catheterization, and high testing costs may limit the feasibility of PiCCO monitoring in the clinical setting, especially in developing countries/territories. Therefore, we hypothesized that a low-cost DL method combined with repeatable and non-invasive CXR instead of routine PiCCO monitoring may be useful for predicting the fluid overload status in critically ill patients.

With the rapid development of ultrasound and computed tomography technologies, the clinical role of CXR has been gradually diminishing. However, as most ICU patients are at high risk of adverse events during transportation, the preliminary information that can be provided by portable CXR is still important to ICU clinicians. Most studies on the correlation between ELWI and CXR were performed in the 20th century. Early reports confirmed a positive correlation between the CXR score and the ELWI value using the double indicator dilution technique in critically ill patients.^{11-13,42} Similarly, recent studies have shown that a standardized CXR score can improve diagnostic accuracy for predicting the severity of pulmonary edema represented by ELWI using the new transpulmonary thermodilution method.^{17,18} Nevertheless, several studies have shown the opposite, namely, that CXR does not correlate with ELWI or GEDI assessments of lung water and volume status.^{15,16} The correlations between CXR findings and ELWI values drawn from prior studies were inconsistent, with correlation coefficients ranging from 0.35 to 0.83.^{11-13,18} Similarly, the predictive performance of the semantic label model in our study was also not satisfactory. These discrepancies may be due to the fact that the CXR score and CXR label methods focus mainly on the low-order visual features of CXR images via subjective judgments of radiologists, which can lead to over- or underestimation of the ELWI and GEDI values.

Several previous studies using the MIMIC-CXR database for modeling reported a DL model of CXR developed for accurately and automatically detecting pulmonary edema, atelectasis, consolidation, cardiomegaly, pneumothorax, and pleural effusion.⁴³⁻⁴⁵ As compared with previously reported CXR scores or our CXR label models, the DL model of CXR has certain advantages and the potential for clinical application. First, feature extraction by traditional machine learning mainly relies on manual extraction, which is only effective for simple tasks, whereas the DL algorithm can automatically extract abstract features. Moreover, fine-tuning a network with transfer learning is usually much easier and faster than training a network with randomly initialized weights from scratch, which is useful even with scarce training data. Second, the DL algorithm has significant superiority in terms of recognition capabilities of high-order visual features for image details, and the identified important areas are visually displayed on a smooth heatmap. Third, a simple, fast, low-cost, and noninvasive DL strategy can be used as an effective supplementary tool for CXR reports. Therefore, it should be developed for clinical application, especially in the settings where invasive hemodynamic monitoring is unavailable.

This study has several strengths. Through combining data from the new MIMIC-IV database and the MIMIC-CXR database, it provided a feasible idea for future research. A multimodel comparative analysis was performed by including patient records, clinical examinations, CXR labels, and CXR images. We used 2 completely independent datasets from different ethnic backgrounds for modeling and testing. Due to the differences in the distribution of the 2 datasets, that is, “data mismatch,” it was difficult for the training model to fit the test dataset well. However, from another perspective, this strategy also further verified the true strength and generalization performance of the DL model. The calculation strategy for averaging 2 mature neural networks and multiple training values increased the credibility of the results.

Nevertheless, this study also has limitations. First, to reduce the size of image storage and facilitate the smooth progress of our research, the DICOM format had to be replaced by the JPG format, which inevitably caused a loss of image information. During the conversion in the MIMIC-CXR-JPG database, a set of standardization processes has been implemented to preserve the image quality as much as possible. Second, due to the insufficient application of PiCCO monitoring in the MIMIC-IV database and our hospital, the small sample size may have led to overfitting of the training model. Data augmentation can enlarge the training samples to some extent, but further collection of more raw data is still preferable. Third, as the Note module is temporarily unavailable in MIMIC-IV (v. 1.0), data from lung ultrasound and echocardiography, which have been

shown to be effective methods of fluid overload quantification, were not included in the comparative analysis. Therefore, further studies are warranted to supplement the results of our study.

Conclusions In this study, a DL model based on CXR images was established to predict the fluid overload status in critically ill patients. Validated by an independent external test dataset, the DL model showed a relatively strong generalization performance, which was better than that of the clinical and semantic label models. As CXR is routinely used in the ICU, a simple, fast, low-cost, and noninvasive DL model could constitute an effective supplementary tool for identifying the fluid overload status. Such a model would be particularly beneficial in the settings where invasive hemodynamic monitoring is unavailable.

SUPPLEMENTARY MATERIAL

Supplementary material is available at www.mp.pl/paim.

ARTICLE INFORMATION

ACKNOWLEDGMENTS None.

FUNDING This work was supported by a grant from the Research Incubation Project of the First Affiliated Hospital of Wenzhou Medical University (FHY2019088; to WZho) and the Science and Technology Program of Wenzhou (Y2020097; to WZho).

CONTRIBUTION STATEMENT XYQ and WZho conceived and designed this study; WZho designed data collection processes and forms; XYQ, WZha, and XH collected and assembled the data; XYQ and WZho participated in image processing and data analysis; XYQ and WZha wrote the first draft of the paper; all authors critically revised the paper and approved the final manuscript.

CONFLICT OF INTEREST None declared.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), allowing third parties to copy and redistribute the material in any medium or format and to remix, transform, and build upon the material, provided the original work is properly cited, distributed under the same license, and used for noncommercial purposes only. For commercial use, please contact the journal office at pamw@mp.pl.

HOW TO CITE Qin X, Zhang W, Hu X, Zhou W. A deep learning model to identify the fluid overload status in critically ill patients based on chest X-ray images. *Pol Arch Intern Med.* 2023; 133: 16396. doi:10.20452/pamw.16396

REFERENCES

- 1 Vincent JL. Fluid management in the critically ill. *Kidney Int.* 2019; 96: 52-57. [↗](#)
- 2 Levy MM, Evans LE, Rhodes A. The surviving sepsis campaign bundle: 2018 update. *Crit Care Med.* 2018; 46: 997-1000. [↗](#)
- 3 Claire-Del Granado R, Mehta RL. Fluid overload in the ICU: evaluation and management. *BMC Nephrol.* 2016; 17: 109. [↗](#)
- 4 Messmer AS, Zingg C, Müller M, et al. Fluid overload and mortality in adult critical care patients—a systematic review and meta-analysis of observational studies. *Crit Care Med.* 2020; 48: 1862-1870. [↗](#)
- 5 Vaara ST, Korhonen AM, Kaukonen KM, et al. Fluid overload is associated with an increased risk for 90-day mortality in critically ill patients with renal replacement therapy: data from the prospective FINNAKI study. *Crit Care.* 2012; 16: R197. [↗](#)
- 6 O'Connor ME, Prowle JR. Fluid overload. *Crit Care Clin.* 2015; 31: 803-821. [↗](#)
- 7 Oren-Grinberg A. The PiCCO Monitor. *Int Anesthesiol Clin.* 2010; 48: 57-85. [↗](#)
- 8 Hu W, Lin CW, Liu BW, et al. Extravascular lung water and pulmonary arterial wedge pressure for fluid management in patients with acute respiratory distress syndrome. *Multidiscip Respir Med.* 2014; 9: 3. [↗](#)
- 9 Redondo FJ, Padilla D, Villarejo P, et al. The global end-diastolic volume (GEDV) could be more appropriate to fluid management than central venous pressure (CVP) during closed hyperthermic intrabdominal chemotherapy with CO₂ circulation. *J Invest Surg.* 2018; 31: 321-327. [↗](#)
- 10 Kapoor PM, Bhardwaj V, Sharma A, Kiran U. Global end-diastolic volume an emerging preload marker vis-a-vis other markers – have we reached our goal? *Ann Card Anaesth.* 2016; 19: 699-704. [↗](#)

- 11 Laggner A, Kleinberger G, Haller J, et al. Bedside estimation of extravascular lung water in critically ill patients: comparison of the chest radiograph and the thermal dye technique. *Intensive Care Med.* 1984; 10: 309-313. [↗](#)
- 12 Halperin BD, Feeley TW, Mihm FG, et al. Evaluation of the portable chest roentgenogram for quantitating extravascular lung water in critically ill adults. *Chest.* 1985; 88: 649-652. [↗](#)
- 13 Sibbald WJ, Warshawski FJ, Short AK, et al. Clinical studies of measuring extravascular lung water by the thermal dye technique in critically ill patients. *Chest.* 1983; 83: 725-731. [↗](#)
- 14 Sivak ED, Richmond BJ, O'Donovan PB, Borkowski GP. Value of extravascular lung water measurement vs portable chest x-ray in the management of pulmonary edema. *Crit Care Med.* 1983; 11: 498-501. [↗](#)
- 15 Saugel B, Ringmaier S, Holzapfel K, et al. Physical examination, central venous pressure, and chest radiography for the prediction of transpulmonary thermodilution-derived hemodynamic parameters in critically ill patients: a prospective trial. *J Crit Care.* 2011; 26: 402-410. [↗](#)
- 16 Martin GS, Eaton S, Mealer M, Moss M. Extravascular lung water in patients with severe sepsis: a prospective cohort study. *Crit Care.* 2005; 9: R74. [↗](#)
- 17 Hammon M, Dankerl P, Voit-Höhne HL, et al. Improving diagnostic accuracy in assessing pulmonary edema on bedside chest radiographs using a standardized scoring approach. *BMC Anesthesiol.* 2014; 14: 94. [↗](#)
- 18 Brown LM, Calfee CS, Howard JP, et al. Comparison of thermodilution measured extravascular lung water with chest radiographic assessment of pulmonary oedema in patients with acute lung injury. *Ann Intensive Care.* 2013; 3: 25. [↗](#)
- 19 Chrzan R, Wojciechowska W, Terlecki M, et al. The role of artificial intelligence technology analysis of high-resolution computed tomography images in predicting the severity of COVID-19 pneumonia. *Pol Arch Intern Med.* 2022; 132: 16332. [↗](#)
- 20 Aggarwal R, Sounderajah V, Martin G, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ Digit Med.* 2021; 4: 65. [↗](#)
- 21 Johnson A, Bulgarelli L, Pollard T, et al. MIMIC-IV (version 1.0). *PhysioNet.* 2021.
- 22 Goldberger AL, Amaral LA, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation.* 2000; 101: E215-E220. [↗](#)
- 23 Johnson AEW, Pollard TJ, Berkowitz SJ, et al. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Sci Data.* 2019; 6: 317. [↗](#)
- 24 Johnson A, Pollard T, Mark R, et al. MIMIC-CXR Database (version 2.0.0). *PhysioNet.* 2019.
- 25 Ferreira FL, Bota DP, Bross A, et al. Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA.* 2001; 286: 1754-1758. [↗](#)
- 26 Irvin J, Rajpurkar P, Ko M, et al. CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison. *Proceedings of the AAAI Conference on Artificial Intelligence.* 33: 590-597. [↗](#)
- 27 Johnson A, Lungren M, Peng Y, et al. MIMIC-CXR-JPG - chest radiographs with structured labels (version 2.0.0). *PhysioNet.* 2019.
- 28 He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016; 770-778. [↗](#)
- 29 Christian S, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2016; 2818-2826.
- 30 Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 2016; 1135-1144. [↗](#)
- 31 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982; 143: 29-36. [↗](#)
- 32 Oczkowski S, Alshamsi F, Belley-Cote E, et al. Surviving Sepsis Campaign Guidelines 2021: highlights for the practicing clinician. *Pol Arch Intern Med.* 2022; 132: 16290. [↗](#)
- 33 Jozwiak M, Teboul JL, Monnet X. Extravascular lung water in critical care: recent advances and clinical applications. *Ann Intensive Care.* 2015; 5: 38. [↗](#)
- 34 Duan J, Cong LH, Wang H, et al. Clinical evaluation compared to the pulse indicator continuous cardiac output system in the hemodynamic assessment of critically ill patients. *Am J Emerg Med.* 2014; 32: 629-633. [↗](#)
- 35 Katzenelson R, Perel A, Berkenstadt H, et al. Accuracy of transpulmonary thermodilution versus gravimetric measurement of extravascular lung water. *Crit Care Med.* 2004; 32: 1550-1554. [↗](#)
- 36 Rossi P, Wanecsek M, Rudehill A, et al. Comparison of a single indicator and gravimetric technique for estimation of extravascular lung water in endotoxemic pigs. *Crit Care Med.* 2006; 34: 1437-1443. [↗](#)
- 37 Kirov MY, Kuzkov VV, Kuklin VN, et al. Extravascular lung water assessed by transpulmonary single thermodilution and postmortem gravimetry in sheep. *Crit Care.* 2004; 8: R451-R458. [↗](#)
- 38 Renner J, Gruenewald M, Quaden R, et al. Influence of increased intra-abdominal pressure on fluid responsiveness predicted by pulse pressure variation and stroke volume variation in a porcine model. *Crit Care Med.* 2009; 37: 650-658. [↗](#)
- 39 Renner J, Meybohm P, Gruenewald M, et al. Global end-diastolic volume during different loading conditions in a pediatric animal model. *Anesth Analg.* 2007; 105: 1243-1249. [↗](#)
- 40 Zhong YB, Wang J, Shi F, et al. ICU management based on PiCCO parameters reduces duration of mechanical ventilation and ICU length of stay in patients with severe thoracic trauma and acute respiratory distress syndrome. *Ann Intensive Care.* 2016; 6: 113. [↗](#)
- 41 Morisawa K, Fujitani S, Homma Y, et al. Can the global end-diastolic volume index guide fluid management in septic patients? A multicenter randomized controlled trial. *Acute Med Surg.* 2019; 7: e468. [↗](#)
- 42 Baudendistel L, Shields JB, Kaminski DL. Comparison of double indicator thermodilution measurements of extravascular lung water (EVLW) with radiographic estimation of lung water in trauma patients. *J Trauma.* 1982; 22: 983-988. [↗](#)
- 43 Kuo PC, Tsai CC, López DM, et al. Recalibration of deep learning models for abnormality detection in smartphone-captured chest radiograph. *NPJ Digit Med.* 2021; 4: 25. [↗](#)
- 44 Horng S, Liao R, Wang X, et al. Deep learning to quantify pulmonary edema in chest radiographs. *Radiol Artif Intell.* 2021; 3: e190228. [↗](#)
- 45 Huang T, Yang R, Shen L, et al. Deep transfer learning to quantify pleural effusion severity in chest X-rays. *BMC Med Imaging.* 2022; 22: 100. [↗](#)