Original Research

DIGITAL HEALTH

Digital Health
Volume 8: 1–15
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076221120317
journals.sagepub.com/home/dhj
SAGE

Artificial intelligence–aided diagnosis model for acute respiratory distress syndrome combining clinical data and chest radiographs

Kai-Chih Pai[1], Wen-Cheng Chao[2,3,4], Yu-Len Huang[5], Ruey-Kai Sheu[5], Lun-Chi Chen[1], Min-Shian Wang[6], Shau-Hung Lin[7], Yu-Yi Yu[8,9], Chieh-Liang Wu[2,3,4] and Ming-Cheng Chan[3,8]

Abstract

Objective: The aim of this study was to develop an artificial intelligence–based model to detect the presence of acute respiratory distress syndrome (ARDS) using clinical data and chest X-ray (CXR) data.

Method: The transfer learning method was used to train a convolutional neural network (CNN) model with an external image dataset to extract the image features. Then, the last layer of the model was fine-tuned to determine the probability of ARDS. The clinical data were trained using three machine learning algorithms–eXtreme Gradient Boosting (XGB), random forest (RF), and logistic regression (LR)–to estimate the probability of ARDS. Finally, ensemble-weighted methods were proposed that combined the image model and the clinical data model to estimate the probability of ARDS. An analysis of the importance of clinical features was performed to explore the most important features in detecting ARDS. A gradient-weighted class activation mapping (Grad-CAM) model was used to explain what our CNN sees and understands when making a decision.

Results: The proposed ensemble-weighted methods improved the performances of the ARDS classifiers (XGB + CNN, area under the curve [AUC] = 0.916; RF + CNN, AUC = 0.920; LR + CNN, AUC = 0.920; XGB + RF + LR + CNN, AUC = 0.925). In addition, the ML model using clinical data to present the top 15 important features to identify the risk factors of ARDS.

Conclusion: This study developed combined machine learning models with clinical data and CXR images to detect ARDS. According to the results of the Shapley Additive exPlanations values and the Grad-CAM techniques, an explicable ARDS diagnosis model is suitable for a real-life scenario.

Keywords

Acute respiratory distress syndrome, artificial intelligence, machine learning, clinical data, chest X-ray, ensemble-weighted model

Submission date: 16 May 2022; Acceptance date: 29 July 2022

[1]College of Engineering, Tunghai University, Taichung, Taiwan
[2]Department of Critical Care Medicine, Taichung Veterans General Hospital, Taichung, Taiwan
[3]College of Medicine, National Chung Hsing University, Taichung, Taiwan
[4]Department of Automatic Control Engineering, Feng Chia University, Taichung, Taiwan
[5]Department of Computer Science, Tunghai University, Taichung, Taiwan
[6]Artificial Intelligence Studio, Taichung Veterans General Hospital, Taichung, Taiwan
[7]DDS-THU Artificial Intelligence Center, Tunghai University, Taichung, Taiwan
[8]Division of Critical Care and Respiratory Therapy, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan
[9]Institute of Emergency and Critical Care Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

Corresponding author:
Ming-Cheng Chan, 1650 Boulevard Sect. 4, Taichung, 40705, Taiwan.
Email: mingcheng.chan@gmail.com

Introduction

Acute respiratory distress syndrome (ARDS) is common among critically ill patients in the intensive care unit (ICU), but mortality remains high.[1,2] ARDS is defined by acute-onset hypoxemia with bilateral infiltrates on chest X-ray (CXR) and the exclusion of fluid overload. ARDS may result from several causes, both pulmonary and extrapulmonary. It is characterized pathologically by alveolar inflammation and flooding that may result in noncardiogenic pulmonary edema and loss of aeratio,[3–5] which may lead to the clinical presentation of hypoxemia. Thus, patients with ARDS often need mechanical ventilation. Prevention of ventilator-induced lung injury (VILI)[6] by timely recognition[7,8] of ARDS and implementation of a protective strategy[9] is key to improving the survival of ARDS patients. However, ARDS is often unrecognized and undertreated[1] in real-life practice, and there is an urgent need to assist intensive care clinicians in early recognition of ARDS and appropriate patient management.

Previous studies have developed artificial intelligence (AI) models for predicting ARDS using machine learning or deep learning algorithms. Most studies have attempted to identify ARDS based on clinical data, including vital signs, laboratory tests, ventilator-derived parameters, etc.[10–12] For example, two studies developed prediction models for ARDS severity using machine models based on the Light Gradient Boosting Machine (LightGBM), random forest (RF), and eXtreme Gradient Boosting (XGBoost).[10,12] Le et al. also developed a model for the early prediction of ARDS using XGBoost.[11] In another study, Sayed et al. developed ARDS predictive models using LightGBM, XGBoost, and RF. Overall, the models in these studies showed acceptable or good performance. However, it is difficult to help clinicians to identify ARDS accurately and explicably according to the Berlin definition of ARDs[5] because of the lack of chest radiographs.

Convolutional neural networks (CNNs) have been tested in different domains and have shown that they can be successfully trained to identify a wide range of relevant findings on chest radiographs. For example, Zaglam et al. used CXR data alone to develop a computer-aided diagnosis system for identifying ARDS. They used a semi-automatic segmentation method to extract the CXR features and then classified them using a support vector machine classifier. A total of 321 images were analyzed for modeling, and 90 images were evaluated. The results showed a sensitivity of 90.6% and a specificity of 86.5%.[13] It remains challenging to develop a classification model for ARDS based on CXRs because a large training image dataset is needed when using CNNs. A few studies have tried to diagnose ARDS using deep learning neural networks. For instance, Sjoding et al. developed a classification model for detecting ARDS findings on CXRs using a deep CNN. They used transfer learning to learn the general features of CXRs with bilateral airspace disease from large datasets, including 595,506 images, and then trained the network on 8073 radiographs annotated for ARDS. The results of the model performance were consistent with or higher than those of intensivist physicians.[14]

Recently, several studies have developed COVID-19 detection models based on CXRs and associated clinical information.[15,16] Comparing the performance of different models, the results showed that the joint models (clinical data combined with computed tomography [CT] images) outperformed the models trained on clinical data only or on CT images only and had higher areas under the curve (AUC). In view of the results of previous studies, our goal was to develop a real-time model to detect the presence of ARDS in critically ill patients by integrating clinical information from electronic health records and the features of CXR images through AI calculation.

Materials and methods

Study setting

Taichung Veterans General Hospital (TCVGH) is a teaching hospital and tertiary referral medical center in central Taiwan. TCVGH has six intensive care units with a total of 105 beds for medical, surgical, cardiovascular, and neurological critically ill patients. There are approximately 4800 ICU admissions every year, and around 70% of these patients need invasive mechanical ventilation support. Since 2018, an ARDS working group composed of pulmonary specialists and respiratory therapists, has routinely reviewed mechanically ventilated patients with or without ARDS based on the Berlin definition on working days. This research was approved by the Institutional Review Boards I & II of Taichung Veterans General Hospital (Certificate Number: SE20249B and CE20049B).

Data acquisition

The enrolled patients were those who had been reviewed as with or without ARDS by the working group from October 2018 to December 2019. They also met the following criteria: (1) age ≥ 20 years old; (2) stayed in an ICU for more than 48 h. Patients readmitted to an ICU and whose CXR images were missing were excluded. A total of 1577 patients were included in the final data analysis. The patient selection flowchart is presented in Figure 1.

Model framework

Three AI models were used to generate the probability of having ARDS. The models included two major modules: judgment of ARDS based on clinical data and CXR images separately. For the clinical data, we used eXtreme

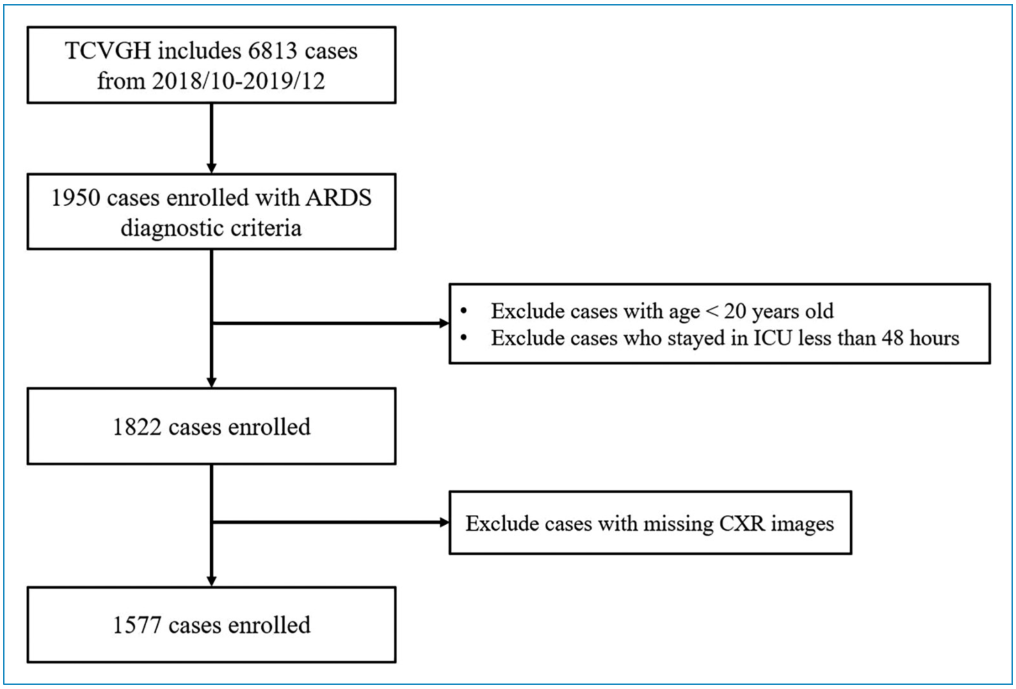

TCVGH includes 6813 cases from 2018/10-2019/12
1950 cases enrolled with ARDS diagnostic criteria
• Exclude cases with age < 20 years old
• Exclude cases who stayed in ICU less than 48 hours
1822 cases enrolled
Exclude cases with missing CXR images
1577 cases enrolled

Figure 1. Patient selection flowchart.

Gradient Boosting (XGboost), random forest (RF), and logistic regression (LR) classifiers to identify patients with ARDS. For the CXR images, we only used sample image preprocessing steps that could avoid extensive resources and improve the generalization ability of the CNN classification model. We first resized all the images to the dimension 224 × 224 pixels, then converted the pixel intensity range between 0 and 255 using histogram equalization. We then used a deep CNN to segment the lung region in the CXR images. For model development, a pretrained network model for learning the imaging characteristics of patients with bilateral airspace disease on CXRs from the ChestX-ray14 datasets was constructed using a transfer learning approach.[17,18] We then trained the network on 1577 CXRs annotated for ARDS. Finally, we created an ensemble model combining a CNN model from CXR images and three machine learning models from clinical data to identify patients with ARDS (Figure 2). More detail on the data pre-process and model training is described in the following sections.

Clinical data preprocessing

We focused on identifying ARDS during the first 48 h of admission. We extracted 48 h of data to ensure sufficient data for modeling, including ventilatory parameters, vital signs, laboratory data, and fluids (Table 1). For the ventilatory parameters, vital signs, and fluids, we categorized the features into 1–24 h (Day 1) and 25–48 h (Day 2). For the laboratory data, we extracted the mean value during 48 h.

Because all features included missing data, we needed to impute values for those missing values. The missing values for these features are imputed by this mean. For optimized algorithmic performance, we reduced the data dimensions (features) using recursive feature elimination (RFE),[19] a feature selection method that fits a model and removes the weakest feature (or features) until the specified number of features is reached. In total, 21 features were included in the final model training.

Image pre-processing

Datasets, image pre-processing, and data augmentation.

The original image dataset consists of 1577 de-identified images of chest X-rays, which were selected from CXRs taken during the first 48 h of admission. The CXR image is an effective imaging method that is used extensively in the initial diagnosis of pulmonary diseases. The original CXR size is between a resolution of 2000 × 2000 and 3000 × 3000 pixels. In our study, the images were converted from the Digital Imaging and Communications in Medicine (DICOM) format to 224 × 224 pixel, 8-bit gray-scale JPEG images.

Previous studies have developed different pre-processing methods in CXRs, such as enhancement, reduction of noise, etc. Chen et al. proposed a method for enhancing CXRs. The CXR image is divided into three subregions, and the image is enhanced using gray-level normalization.[20] Another study developed a CXR image denoising approach based on total variation regularization

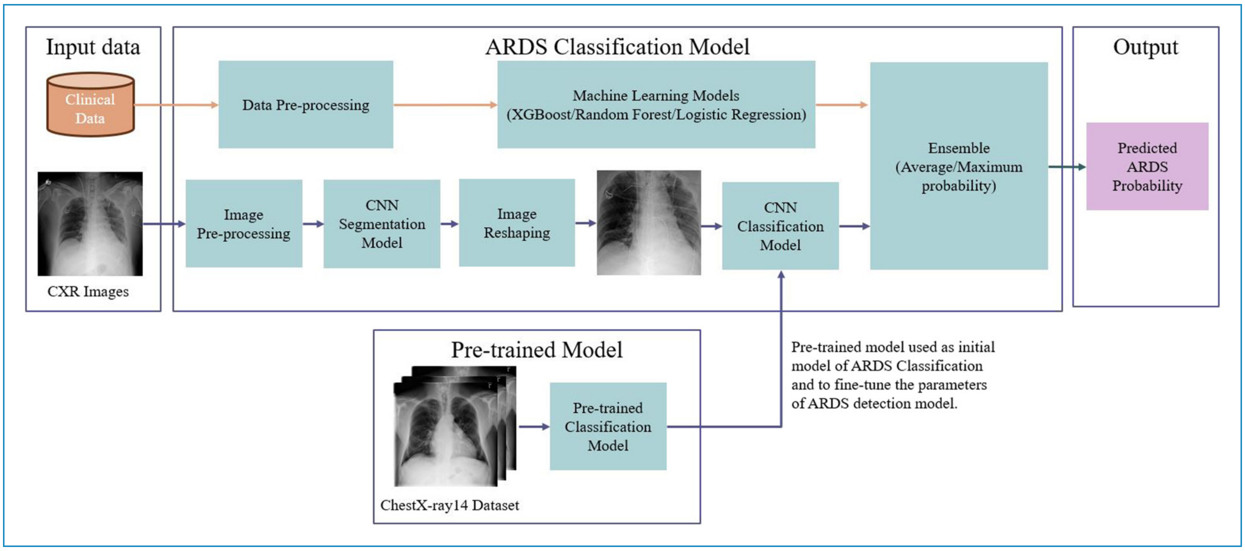

Input data
Clinical Data
CXR Images
ARDS Classification Model
Data Pre-processing
Machine Learning Models (XGBoost/Random Forest/Logistic Regression)
Image Pre-processing
CNN Segmentation Model
Image Reshaping
CNN Classification Model
Ensemble (Average/Maximum probability)
Output
Predicted ARDS Probability
Pre-trained Model
Pre-trained Classification Model
ChestX-ray14 Dataset
Pre-trained model used as initial model of ARDS Classification and to fine-tune the parameters of ARDS detection model.

Figure 2. Model framework overview.

Table 1. Clinical features.

Feature type	Unit (measurement)	Feature type	Unit (measurement)
Vital signs		Laboratory data	
Temperature	°C	PO2-A	mmHg
SBP	mmHg	Procalcitonin	ng/mL
DBP	mmHg	PCO2-A	mmHg
Pulse rate	bpm	Fluids	
Respiratory rate	breath/min	Urine output	ml
SPO2	%		
Ventilatory parameters			
FiO2	%		
Positive end-expiratory pressure	cmH_2O		
Total respiratory rate	breath/min		
Tidal volume	cc/kg		
Minute volume	L/min		
Mean airway pressure	cmH_2O		

with implementation of the Nesterov optimization method.[21] Recently, an artificial intelligence-aided diagnosis of lung disease was diagnosed by CXRs using deep learning models. The researchers used some pre-processing methods, such as thresholding, blurring, and histogram equalization.[22] In our study, the histogram equalization method was used to extend the pixel's intensity range from the original range to 0 to 255. Thus, the enhanced image has a wider range of intensity and a slightly higher contrast. Figure 3 shows examples of the original images and after image preprocessing.

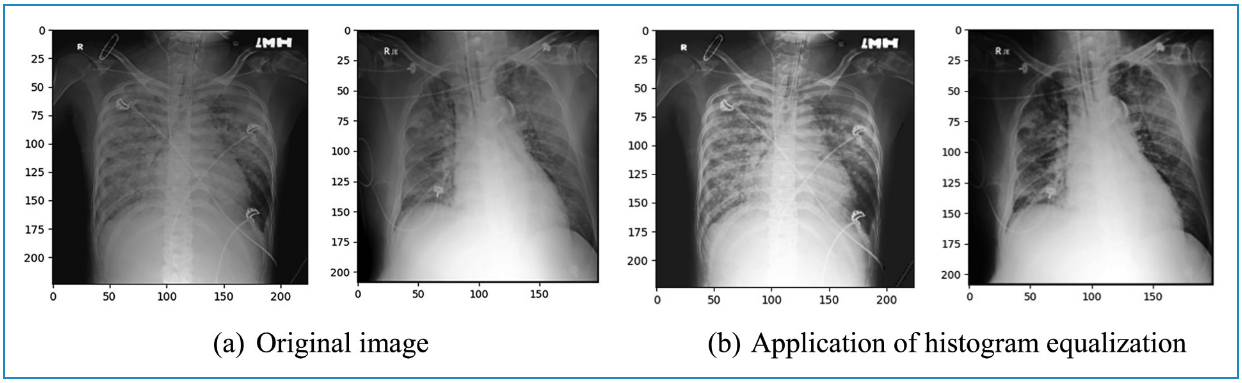
The more trainable parameters as the models' network deepens. Because we had a small number of images, this would easily lead to overfitting during model training. To solve the overfitting problem, we enhanced our training dataset using the data augmentation technique, including random rotation and scale jittering. One image in the training dataset was converted into five images. Figure 4 shows an example of an original image and five enhanced images. The augmented data were created every time a test fold changed.

Chest X-ray segmentation. We first used a segmentation model that combines the Mask R-CNN and U-Net methods to segment the CXR images,[22–24] which reduces the amount of noise on CXR images and focuses on the lung region. The segmentation model is trained by 400 CXR images labeled by a clinical expert. Afterward, we cropped the images by lung segmentation based on a bounding box. Figure 5 demonstrates the CXR input and the extracted lung region image cropped by the segmentation model. Finally, we reshaped the images to 224 × 224 pixels to fit the input shape of the CNN model training.

Deep convolutional neural network training

The deep learning model usually requires a large number of labeled images to train the model. We only had a small dataset, which is challenging to obtain better results using the CNN model. Transfer learning is a machine learning approach in which a model developed for a task is reused as the starting point for another model for a second task.[18] The benefits of transfer learning include saving resources and improving efficiency when training a new model. In our study, we trained the CNN model to identify ARDS by CXRs using transfer learning. Figure 6 presents the workflow of CNN model with transfer learning. For the proposed image classification model, we developed a CNN model to identify an ARDS event using a 121-layer dense network architecture (DenseNet121)[25] with two steps: pre-training and fine-tuning ARDS detection training. We used the weights of a DenseNet121 network trained on ImageNet as a starting point.[26] The CNN model was trained with Tensorflow and Keras in Python. For the pre-training step, the CNN model was trained to detect 14 common descriptive CXR findings (e.g. edema, infiltrate, and pleural effusion) and to extract learned features from CXRs using the ChestX-Ray14 dataset, totaling 86,524 CXRs.

In the fine-tuning step, we only retrained the parameters in the last convolutional block and subsequent layers to detect ARDS, while all the others were kept fixed after the pre-training.[14] Binary cross-entropy loss was used when adjusting model weights during training. The adaptive moment estimation (ADAM) optimizer was used to optimize the parameters of model training. The hyperparameters used in the training dataset had an initial learning



(a) Original image

(b) Application of histogram equalization

Figure 3. Examples of chest X-rays (CXRs) from the original dataset and after histogram equalization preprocessing.

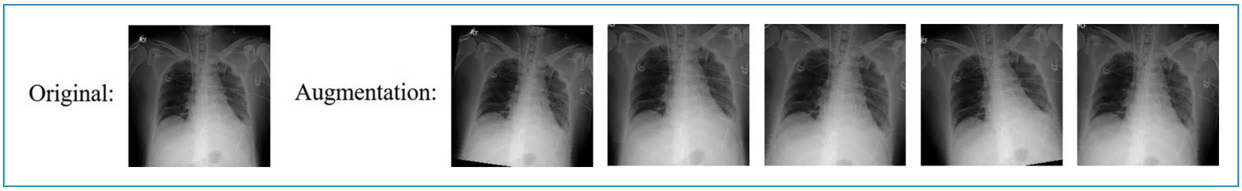


Figure 4. Original image and five enhanced images using data augmentation.

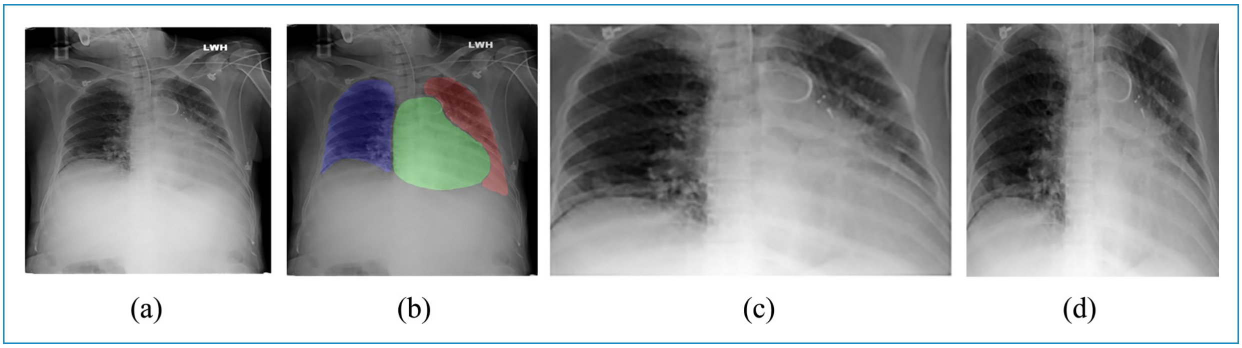


Figure 5. Example of input image (a) and extracted lung region image (b) cropped by the segmentation model (c), resulting in the reshaped image (d).

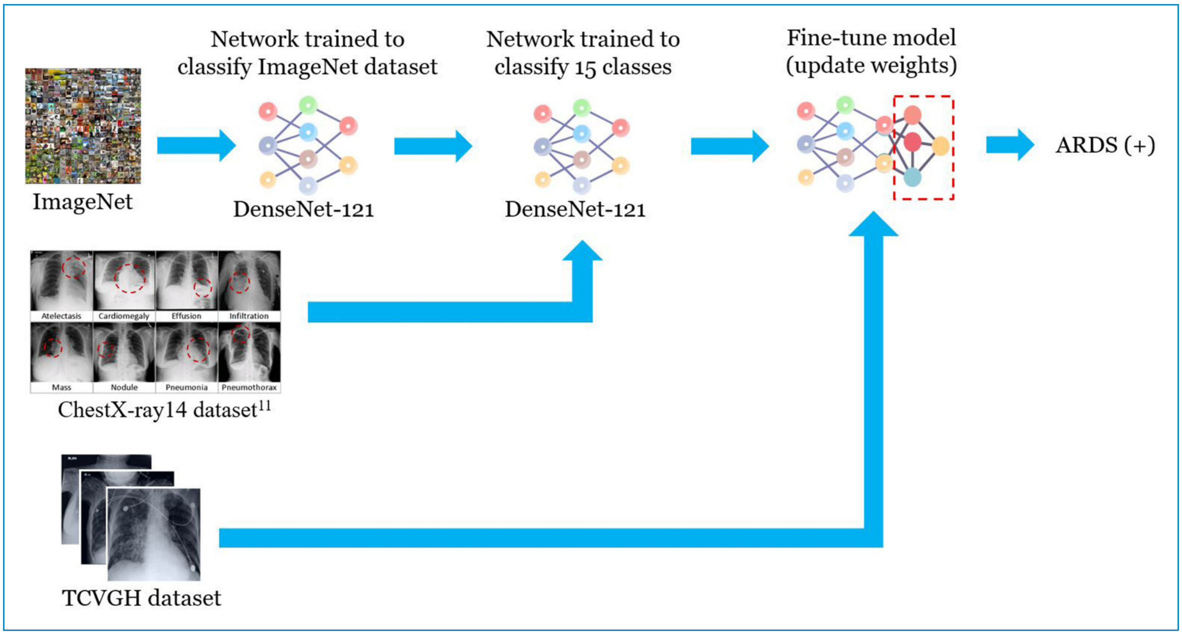


Figure 6. Workflow of the convolutional neural network (CNN) model with transfer learning.

rate of 10^{-5}, and the learning rate was reduced by a factor of 10 if the validation loss did not improve for two consecutive epochs. The model was trained with a batch size of 32 and 22 epochs to minimize computational time.

Ensemble method

Previous studies have developed weighted and unweighted ensemble methods by combining different machine learning or deep learning models.[27,28] In our study, we experimented with different ensemble methods that create multiple models and then combined them to produce improved results. First, the ensemble average probability method was defined so that the prediction results of the classifiers based on clinical data (XGB, RF, LR) were combined with the prediction results of the image classification model, and the average of the two predictions was taken as the final prediction value. Second, the maximum probability method was defined so that the output of the maximum probability of ARDS identification was a higher probability from the image classification and clinical data classification models. For instance, if the output probability of identifying ARDS was 0.475 in the image classification model and 0.568 in the clinical data classification model, the final output probability of the ensemble method was 0.568. Third, the three classifiers using clinical data were combined using the average probability and maximum probability methods to create two different ensemble models. Finally, we combined the three machine learning models based on clinical data and the CNN model. The average probability and the maximum probability methods are also used to create two different ensemble models.

Table 2. Demographic characteristics.

	All (N = 1577)	Non-ARDS (n = 1194)	ARDS (n = 383)	*p* value
Demographic characteristics				
Age (years)	66.1 ± 15.8	66.1 ± 15.7	66.3 ± 16.3	0.823
Male, n (%)	1017 (64.5%)	762 (63.8%)	255 (66.6%)	0.357
BMI (kg/m^2)	24.3 ± 4.9	24.2 ± 4.7	24.4 ± 5.4	0.623
Admission source, n (%)				
Emergency room	1577 (100.0)	1194 (100.0)	383 (100.0)	1.000
ICU type, n (%)				
Medical	927 (58.8%)	631 (52.8%)	296 (77.3%)	<0.001
Surgical	650 (41.2%)	563 (47.2%)	87 (22.7%)	
Severity scores				
APACHE II score	25.2 ± 6.1	24.2 ± 5.7	28.3 ± 6.3	<0.001
SOFA score, Day 1	8.6 ± 3.6	7.9 ± 3.5	10.3 ± 3.4	<0.001
SOFA score, Day 3	7.4 ± 4.0	6.6 ± 3.7	9.3 ± 4.0	<0.001
SOFA score, Day 7	6.8 ± 4.0	6.3 ± 3.9	7.6 ± 4.2	<0.001
Comorbidities, n (%)				
Cardiovascular disease	444 (28.2%)	344 (28.8%)	100 (26.1%)	0.338
Cerebrovascular disease	454 (28.8%)	376 (31.5%)	78 (20.4%)	<0.001
Dementia	101 (6.4%)	77 (6.4%)	24 (6.3%)	0.994
Chronic pulmonary disease	272 (17.2%)	212 (17.8%)	60 (15.7%)	0.387
Rheumatic disease	273 (17.2%)	49 (4.1%)	35 (9.1%)	<0.001
Hepatic disease	269 (17.1%)	196 (16.4%)	73 (19.1%)	0.263
Diabetes mellitus	548 (34.7%)	395 (33.1%)	153 (39.9%)	0.017
Renal disease	484 (30.7%)	346 (29.0%)	138 (36.0%)	0.011
Malignancy	480 (30.4%)	326 (27.3%)	154 (40.2%)	<0.001
Charlson Comorbidity Index (CCI)	2.2 ± 1.4	2.1 ± 1.4	2.3 ± 1.5	0.034
Clinical outcome				
ICU length of stay (days)	14.1 ± 14.0	13.4 ± 14.2	16.3 ± 12.9	<0.001

(continued)

Table 2. Continued.

	All (N = 1577)	Non-ARDS (n = 1194)	ARDS (n = 383)	*p* value
Hospital length of stay (days)	31.6 ± 27.1	31.8 ± 28.3	31.0 ± 23.1	0.571
Ventilator days	11.5 ± 12.7	10.6 ± 12.4	14.4 ± 13.1	<0.001
Hospital mortality, n (%)	433 (27.5)	282 (23.6)	151 (39.4)	<0.001

BMI: body mass index; APACHE II: Acute Physiology and Chronic Health Evaluation score; SOFA: Sequential Organ Failure Assessment; ICU: intensive care unit.

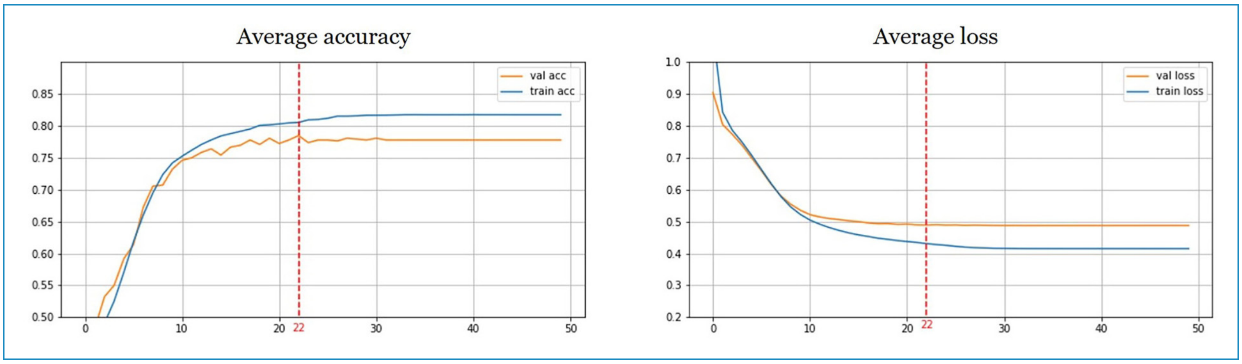


Figure 7. Convolutional neural network (CNN) model for training accuracy, validation accuracy, training loss, validation loss.

Model training and evaluation

The data were randomly split five times (5 k-fold) into training datasets (80%) and validation datasets (20%). Each datapoint was only in one of the training and validation datasets. For the model evaluation, the results of the classification performance are presented in terms of accuracy, sensitivity, specificity, and AUC. The calculations of accuracy, sensitivity, and specificity were as follows:

$$\text{Accuracy} = \frac{t_p + t_n}{t_p + t_n + f_p + f_n}$$

$$\text{Sensitivity} = \frac{t_p + t_n}{t_p + f_n}$$

$$\text{Specificity} = \frac{t_n}{t_n + f_p}$$

where true positive (t_p) is an ARDS patient classified as an ARDS patient, and false positive (f_p) is a non-ARDS patient classified as an ARDS patient. True negative (t_n) is a non-ARDS patient classified as non-ARDS, and false negative (f_n) is an ARDS patient classified as a non-ARDS patient.

Results

Cohort statistics (total, ARDS, non-ARDS)

A total of 1577 subjects requiring mechanical ventilation were enrolled. Their mean age was 66.1 ± 15.8 years and 64.5% were male. We found that 24.2% (383/1577) of the patients had ARDS. They had a higher average APACHE II score (28.3 ± 6.3 vs. 24.2 ± 5.7, $p < 0.001$) and stayed longer in the ICU (16.3 ± 12.9 vs. 0.13.4 ± 14.2, $p < 0.001$) (Table 2).

Model performance

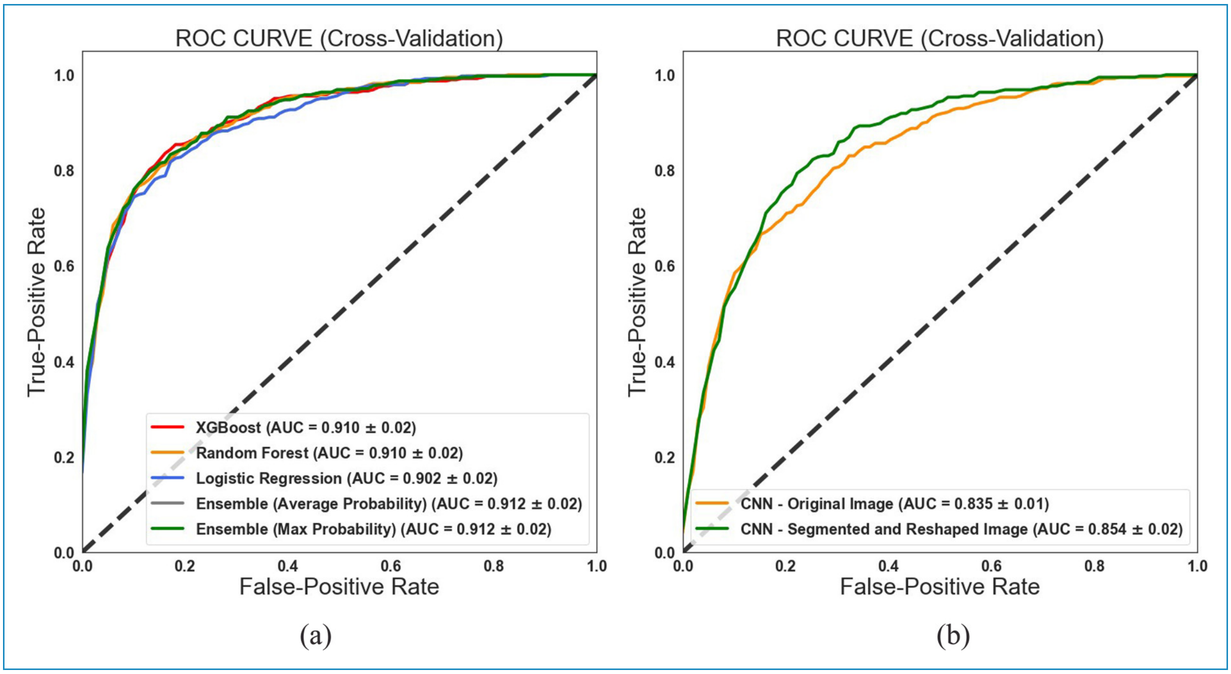
The AI models were evaluated using a five-fold cross-validation. Here, values obtained after the CNN model for training accuracy, validating accuracy, training loss, and validation loss are shown in Figure 7. This indicates the mean of training accuracy is over 80% after the 20 epochs, and the mean of training loss becomes stable after 30 epochs. The proposed model is stopped early in the 22th epoch to avoid overfitting.

The accuracy, sensitivity, specificity, and AUC are presented in Table 3. Comparing the three machine learning models constructed using clinical data, all classifiers had a good AUC value (Figure 8(a)). XGBoost had the highest model performance, with accuracy, sensitivity, specificity, and AUC of 0.848, 0.809, 0.861, and 0.910, respectively. Moreover, the ensemble of three classifiers using average probability performed the highest specificity at 0.934 but the lowest sensitivity, at 0.676. The ensemble of three classifiers using maximum probability had the highest sensitivity of 0.849 and AUC of 0.912. However, the model had the lowest accuracy: 0.821. It seems that the ensemble models by clinical data showed no significant improvement.

Table 3. ARDS classification results based on clinical data.

Data type	Classifier	Accuracy	Sensitivity	Specificity	AUC
Clinical data	XGB	0.848 ± 0.03	0.809 ± 0.03	0.861 ± 0.03	0.910 ± 0.02
	RF	0.840 ± 0.03	0.791 ± 0.03	0.855 ± 0.04	0.910 ± 0.02
	LR	0.832 ± 0.02	0.791 ± 0.05	0.845 ± 0.02	0.902 ± 0.02
	Ensemble (Average probability)	0.871 ± 0.02	0.676 ± 0.02	0.934 ± 0.02	0.912 ± 0.02
	Ensemble (Maximum probability)	0.821 ± 0.03	0.849 ± 0.04	0.812 ± 0.03	0.912 ± 0.02

ARDS: acute respiratory distress syndrome; XGB: eXtreme Gradient Boosting; RF: random forest; LR: logistic regression; AUC: area under the curve.



(a)

(b)

Figure 8. Receiver operating characteristic (ROC) curves demonstrating the performance of the machine learning models and convolutional neural network (CNN) models for ARDS classification: (a) three machine models using clinical data; (b) two CNN models. *Note.* ARDS: acute respiratory distress syndrome; AUC: area under the curve.

In the CXR image classification, the model that used image segmentation and the reshape pre-processing method performed better than the model that used the original image (Table 4): the accuracy, sensitivity, specificity, and AUC were 0.791, 0.760, 0.802, and 0.854, respectively. These two types of machine learning models (based on clinical data and CXR) achieved good classification performance in identifying ARDS.

We further combined the two classification models' output probabilities of identifying patients with and without ARDS. We experimented with combining ensemble methods (average probability and maximum probability). Both methods performed better than the models that only used clinical data or images. The ensemble average probability method presented excellent results for sensitivity (0.846 ± 0.02), specificity (0.863 ± 0.02), accuracy (0.859 ± 0.02), and AUC (0.920 ± 0.02) (Table 5). The mean accuracy, sensitivity, and specificity of the combined models were above 0.8. The mean AUC values of the three combined models (XGB, RF, and LR) were above 0.920 (Figure 9(a)). The maximum probability method resulted in sensitivity values of 0.922, 0.914, and 0.927, respectively, and specificity values of 0.731, 0.715, and 0.718, respectively (Table 5). The mean AUC values were 0.907, 0.901, and 0.911, respectively (Figure 9(b)). Moreover, we combined three machine learning models based on clinical data and a CNN model based on CXRs and experimented with combining ensemble methods (see Table 5). The average

Table 4. ARDS classification results based on CXRs.

Data type	Classifier	Accuracy	Sensitivity	Specificity	AUC
Original image	CNN	0.743 ± 0.02	0.783 ± 0.05	0.729 ± 0.02	0.835 ± 0.01
Segmented and reshaped image		0.791 ± 0.03	0.760 ± 0.04	0.802 ± 0.04	0.854 ± 0.02

ARDS: acute respiratory distress syndrome; CXRs: chest X-ray; CNN: convolutional neural network; AUC: area under the curve.

Table 5. ARDS classification results from two AI models combining clinical data and CXRs.

Classifier	Ensemble method	Accuracy	Sensitivity	Specificity	AUC
XGB + CNN	Ensemble (Average probability)	0.859 ± 0.02	0.846 ± 0.02	0.863 ± 0.02	0.920 ± 0.02
	Ensemble (Maximum probability)	0.777 ± 0.03	0.922 ± 0.02	0.731 ± 0.03	0.909 ± 0.02
RF + CNN	Ensemble (Average probability)	0.852 ± 0.02	0.836 ± 0.01	0.857 ± 0.03	0.919 ± 0.02
	Ensemble (Maximum probability)	0.763 ± 0.03	0.914 ± 0.02	0.715 ± 0.04	0.906 ± 0.02
LR + CNN	Ensemble Average probability)	0.854 ± 0.02	0.849 ± 0.01	0.856 ± 0.03	0.920 ± 0.02
	Ensemble (Maximum probability)	0.769 ± 0.03	0.927 ± 0.04	0.718 ± 0.04	0.911 ± 0.02
XGB + RF + LR + CNN	Ensemble (Average probability)	0.855 ± 0.03	0.830 ± 0.02	0.863 ± 0.03	0.925 ± 0.02
	Ensemble (Maximum probability)	0.749 ± 0.04	0.935 ± 0.04	0.689 ± 0.04	0.915 ± 0.02

ARDS: acute respiratory distress syndrome; AI: artificial intelligence; CXRs: chest X-rays; XGB: eXtreme Gradient Boosting; RF: random forest; LR: logistic regression; AUC: area under the curve.

probability method results performed more improvement with AUC (0.925 ± 0.02).

Model interpretation

We further explored why the model made the correct classifications and the most important features in making the classification. Figure 7 presents the global feature importance computed by SHAP values[25] after training the ARDS classifiers using XGBoost from the clinical dataset. A total of 15 important features were demonstrated, including 8 ventilator features, 5 vital sign features, and 2 laboratory features. The mean positive end-expiratory pressure (PEEP) value and SPO_2 on the first day are associated with a high risk of ARDS.

The results of the summary plot of SHAP values combining feature importance with feature effects (Figure 10) explain the relationship between the features and the risk of ARDS. The position on the y-axis is defined by the feature and by the SHAP value on the x-axis. The color represents the feature value from low (blue) to high (red). We found that lower PEEP values reduce the risk of ARDS, while a larger PEEP value increases the risk. A higher value in the total respiratory rate (TOTRR) measured on day 1 and day 2 also denotes a higher risk of ARDS. Moreover, the mean airway pressure (MAPS), FIO2 also present a positive effect on ARDS identification.

Contrary to the results on PEEP, TOTRR, MAPS, and FIO2, lower values of SPO_2, PaO_2 (PO2_A), and tidal volume (VTEXH) represent a high risk of ARDS. In terms of the measurement of vital signs, the results reveal that pulse rate and respiratory rate are important features, which is consistent with previous studies. Rapid breathing (tachypnea) and a rapid heartbeat (tachycardia) are signs and symptoms of ARDS.

We further explored the model interpretability using the gradient-weighted class activation mapping (Grad-CAM) technique to generate class activation maps (CAMs). We report two cases of false-negative CXRs (Figure 11) that demonstrate the color visualization approach to CXRs using the Grad-CAM technique. The results show that the original images resulted in a false negative, and we observed that the CNN model was not focused on the lungs. In contrast with the CNN model using the original images, the segmented image CNN model found true positives in the same two cases where the CNN model focused on the right lungs. The results indicate that the proposed CNN model with segmented images is associated more with clinical experts' diagnosis.

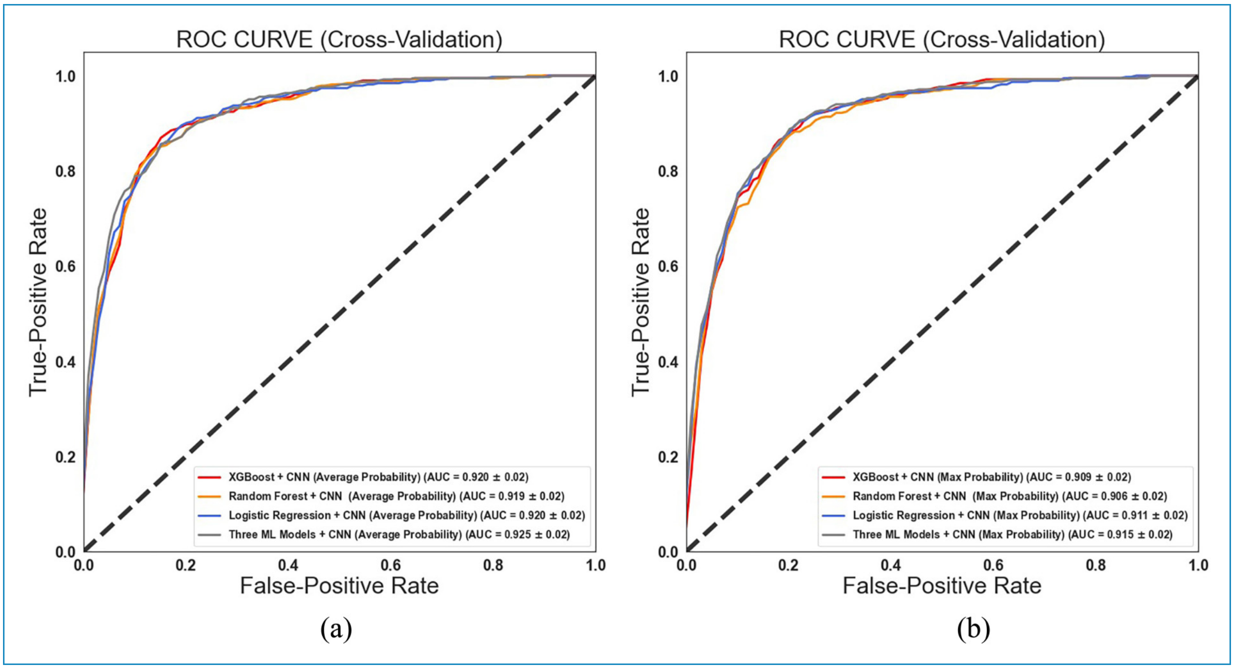


Figure 9. Receiver operating characteristic (ROC) curves demonstrating the performance of two ensemble-weighted models: (a) average probability; (b) maximum probability. *Note.* XGB: eXtreme Gradient Boosting; CNN: convolutional neural network; AUC: area under the curve.

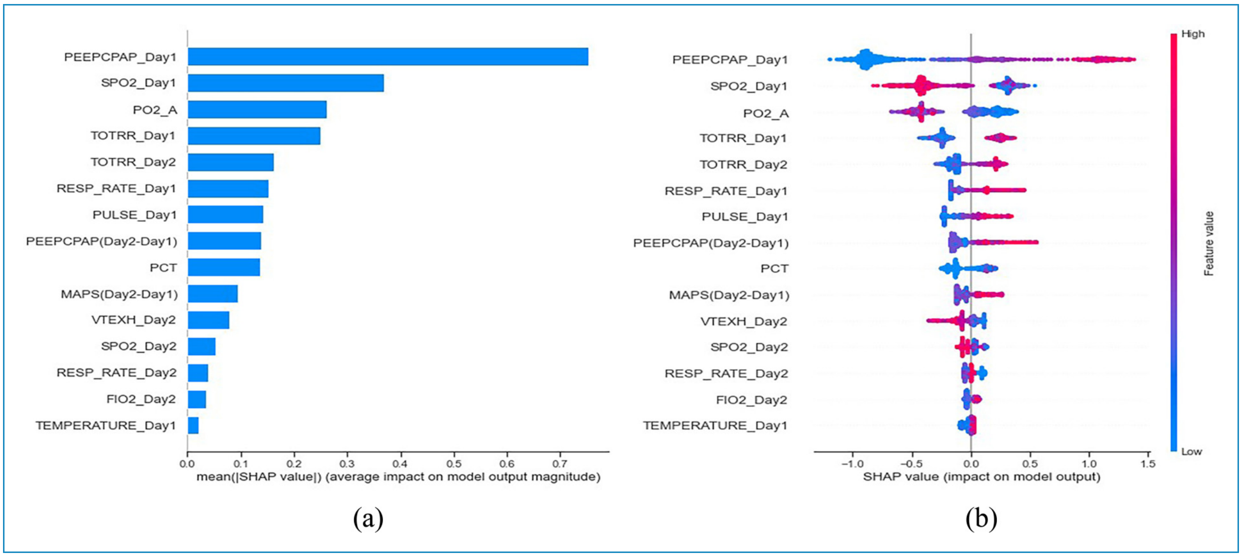


Figure 10. Feature importance (a) and summary plot (b) of SHAP values.

CXR presentation is one of the essential criteria for defining ARDS.[5] To explore the effectiveness of the proposed ensemble model, a Venn diagram was used to illustrate the classification efficiencies of these two classifiers. Here, the XGBoost classifier and the CNN classifier are represented by red and green, respectively. Figure 12(a) represents the number of objects with ARDS "correctly" classified by XGBoost and CNN (true positive). We found that an additional 54 objects were correctly classified by CNN. On the other hand, Figure 12(b) represents the number of objects with non-ARDS "correctly" classified by XGBoost and CNN (true negative). We found that an additional 72 objects were correctly classified by CNN. Approximately 8% (126/1577) of objects were misclassified by clinical data classifiers but correctly classified by CNN. It seems that the proposed ensemble model

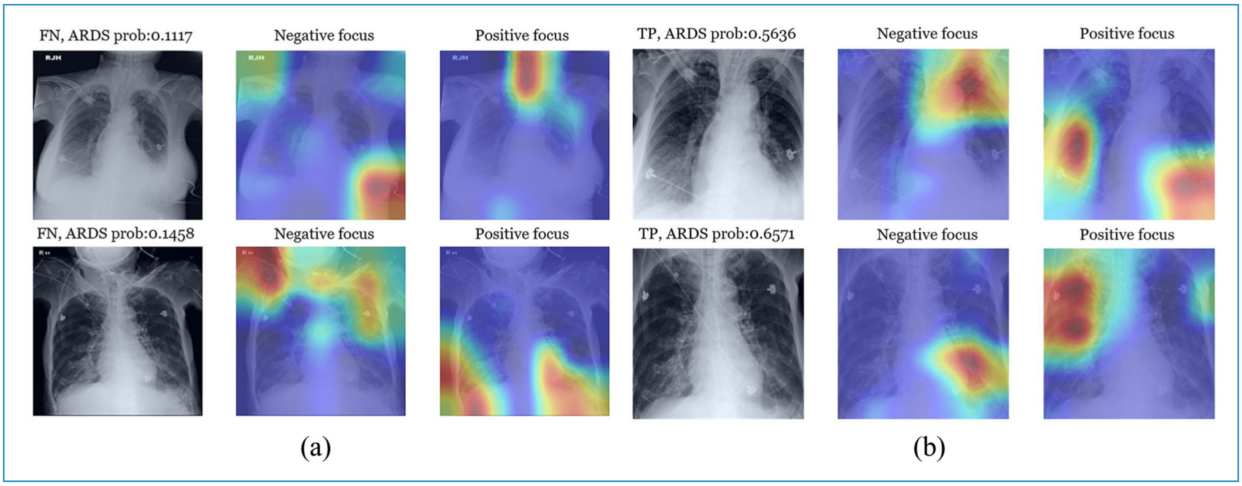


Figure 11. Comparison of acute respiratory distress syndrome (ARDS) classification models based on original data and segmented image from two cases: (a) Color visualization of a false negative on original images. (b) Color visualization of true positive on segmented images.

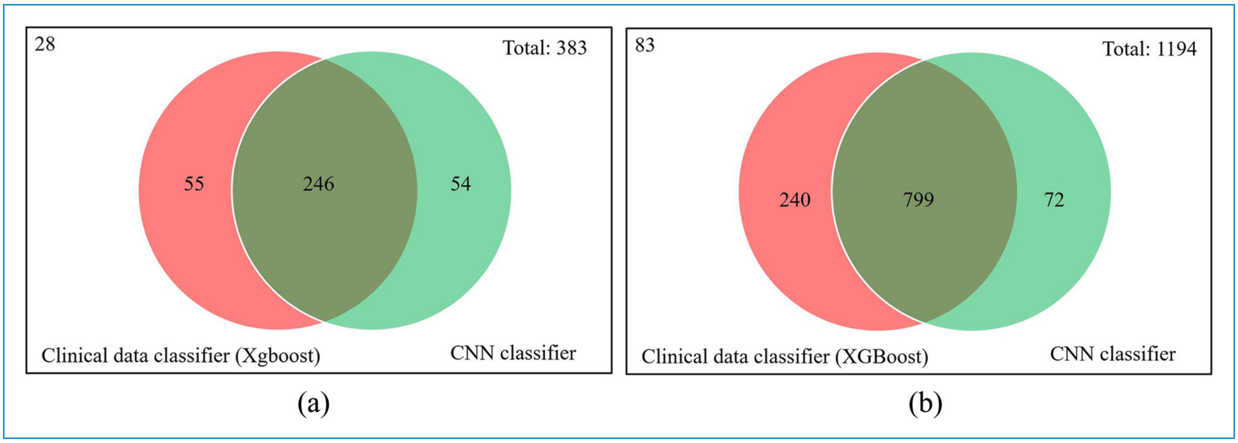


Figure 12. Venn diagrams represent the effectiveness of classification by XGBoost and convolutional neural network (CNN) classifier. (a) True Positive, (b) True Negative.

performed a marginally improvement compared to the clinical data and image models in diagnosing ARDS.

Discussion

ARDS is a common disorder among critically ill patients requiring mechanical ventilation support. Sometimes, it is not recognized immediately and treated appropriately. We have successfully developed a machine learning model to classify patients with or without ARDS. We used clinical data and CXR features with ensemble models to enhance sensitivity and specificity. The black box of the models and the important features relevant to clinical viewpoints were also explored. Our novel model design can be integrated into clinical practice in the future and remind clinicians to implement the optimal bundle of actions for ARDS at the right time.

Burnout is common among ICU healthcare professionals, and workload is one of the major contributing factors.[29] Burnout has been associated with self-reported medical errors.[30] It is an especially important issue during the COVID-19 pandemic, when ICUs face an influx of patients and exposure to the coronavirus.[31] When faced with huge amounts of data, identifying useful information and making correct and timely decisions continue to be a challenge to critical care physicians. Because of human cognitive limitations, even the most knowledgeable and experienced clinicians have difficulty dealing with variables on a continuous basis. The judicious application of AI technology can help clinicians deal with information overload.[32] In addition, the development and application of AI to facilitate decision support in clinical practice may not only alleviate burnout syndrome among ICU healthcare professionals but also improve treatment outcomes by making the right decisions and avoiding unnecessary medical errors.[33] Although the introduction of machine learning to the ICU is still in its infancy, a growing body of applications for outcome prediction, patient monitoring, and decision support have been developed in the past few years, and these will reshape the practice of critical care medicine and ICU management in the coming future.

ARDS is a highly prevalent clinical disorder with high morbidity and mortality rates in critically ill patients. The application of machine learning algorithms for the diagnosis and management of ARDS has emerged in previous years. ARDS is clinically and biologically heterogeneous, and phenotyping by experimental biomarkers potentially offers insights for prognosis and treatment.[34] The application of machine learning models by using readily available clinical data can help classify ARDS phenotypes with high accuracy.[35] This may enable rapid phenotype identification at the bedside for prediction of prognosis and possibly decisions about intervention. Another important aspect in real-life practice is that ARDS is often unrecognized and undertreated.[1] By using routinely collected clinical variables and numerical representations of radiological reports from the Medical Information Mart for Intensive Care III (MIMIC-III) database, Sidney et al. demonstrated that supervised machine learning predictions may help predict patients with ARDS up to 48 h prior to onset.[11] Singhal et al. also showed the ability of an interpretable machine learning algorithm for the early prediction of ARDS in COVID-19 patients.[36] As CXR presentation is one of the criteria for ARDS definition,[5] we found that combining clinical variables and images into machine learning algorithms has the highest ability to predict ARDS.

Our results are in line with the results that show the important clinical features associated with ARDS.[37,38] Tidal volume is an important characteristic of ARDS, and we found that tidal volumes on day 2 is an important feature, which is consistent with previous studies.[7,8] Previous studies also indicated that tidal volume is associated with the risk of ICU mortality.[8] We further compared the CNN model performance with the image segmentation model. The results showed an improvement in the accuracy of identifying ARDS using the image segmentation model. The proposed CNN model performance seems lower than previous studies because of the small size of our dataset. We only collected 1577 CXRs as our dataset. Compared with previous studies using a similar model framework, they developed the ARDS classification model using 8073 CXRs as a training dataset and performed an excellent AUC (0.92, 95% CI 0.89–0.94).[14] Moreover, the results of Venn diagram indicated that there are some differences between the pattern of the clinical data classifier and the CNN model. Importantly, the combined machine learning model aims to provide related clinical and CXR information to identify ARDS for clinicians, which could result in shorter time to diagnosis and treatment decision-making.

However, this study has some limitations. First, there is little similar research to compare with our results. Our results are in line with a similar study by Jabbour et al., who developed machine learning models combining CXRs and clinical data to identify acute respiratory failure.[39] The results presented in our combined model with clinical data and images had better sensitivity, specificity, and AUC than the separate clinical data model and image models. We found the evidence that seems to suggest the benefit of combined machine learning models and could be comparable with a clinician's judgment. On the other hand, we only used the simple CXR image enhancement method in our image pre-processing for considering the efficiency of model clinical implementation. Moreover, we also used limited clinical features that avoid model complexity and overfitting and would be unlikely to leak labels. We found better results even with a limited set of data. Second, we collected only a small sample size and conducted our study at a single center. External validation of multiple centers should be conducted in the future. Based on the privacy protection of medical data, the federated learning approach can be considered and implemented.[40]

Conclusion

We have successfully developed a novel machine learning model to classify patients with or without ARDS based on the combined features of clinical data and CXRs. The model has been designed for a real-life scenario. We have started to integrate the model into our clinical practice and are conducting a study to investigate its impact on the outcomes of ARDS patients.

Acknowledgments: The authors also would like to thank the DDS-THU AI center for support of this project.

Conflict of interest: The authors have no conflicts of interest to declare.

Contributorship: Study concept and design: KCP, WCC, YLH, RKS, LCC, CLW, and MCC; Acquisition of data: KCP, MSW, YYY; Analysis and interpretation of data: KCP, WCC, YLH, LCC, MSW, and SHL; Writing—original draft preparation, KCP and MCC; Supervision: WCC, RKS; Project administration, CLW. All authors agree and are responsible for the content of the manuscript. All authors read and approved the final manuscript.

Ethical approval: Not applicable.

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Ministry of Science and Technology, Taiwan (grant number 109-2321-B-075A-001, 110-2321-B-075A-001).

Guarantor: MCC.

Institutional review board statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by Institutional Review Board of Taichung Veterans General Hospital under case number SE20249B and CE20049B.

ORCID iD: Kai-Chih Pai https://orcid.org/0000-0002-4379-1186

References

1. Bellani G, Laffey JG, Pham T, et al. Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries. *JAMA* 2016; 315: 788–800.
2. Rubenfeld GD, Caldwell E, Peabody E, et al. Incidence and outcomes of acute lung injury. *N Engl J Med* 2005; 353: 1685–1693.
3. Ware LB and Matthay MA. The acute respiratory distress syndrome. *N Engl J Med* 2000; 342: 1334–1349.
4. Thompson BT, Chambers RC and Liu KD. Acute respiratory distress syndrome. *N Engl J Med* 2017; 377: 562–572.
5. ARDS Definition Task Force, Ranieri VM, Rubenfeld GD, et al. Acute respiratory distress syndrome: the Berlin definition. *JAMA* 2012; 307: 2526–2533.
6. Beitler JR, Malhotra A and Thompson BT. Ventilator-induced lung injury. *Clin Chest Med* 2016; 37: 633–646.
7. Chan MC, Chao WC, Liang SJ, et al. First tidal volume greater than 8 mL/kg is associated with increased mortality in complicated influenza infection with acute respiratory distress syndrome. *J Formos Med Assoc* 2019;118:378–385.
8. Needham DM, Yang T, Dinglas VD, et al. Timing of low tidal volume ventilation and intensive care unit mortality in acute respiratory distress syndrome. A prospective cohort study. *Am J Respir Crit Care Med* 2015; 191: 177–185.
9. Acute Respiratory Distress Syndrome Network, Brower RG, Matthay MA, et al. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *N Engl J Med* 2000; 342: 1301–1308.
10. Sayed M, Riaño D and Villar J. Novel criteria to classify ARDS severity using a machine learning approach. *Crit Care* 2021; 25: 150.
11. Le S, Pellegrini E, Green-Saxena A, et al. Supervised machine learning for the early prediction of acute respiratory distress syndrome (ARDS). *J Crit Care* 2020; 60: 96–102.
12. Sayed M, Riaño D and Villar J. Predicting duration of mechanical ventilation in acute respiratory distress syndrome using supervised machine learning. *J Clin Med* 2021; 10: 3824. Published 26 August 2021.
13. Zaglam N, Jouvet P, Flechelles O, et al. Computer-aided diagnosis system for the acute respiratory distress syndrome from chest radiographs. *Comput Biol Med* 2014; 52: 41–48.
14. Sjoding MW, Taylor D, Motyka J, et al. Deep learning to detect acute respiratory distress syndrome on chest radiographs: a retrospective study with external validation. *Lancet Digit Health* 2021; 3: e340–e348.
15. Ahsan MME, Alam T, Trafalis T, et al. Deep MLP-CNN model using mixed-data to distinguish between COVID-19 and non-COVID-19 patients. *Symmetry (Basel)* 2020; 12: 1526.
16. Mei X, Lee HC, Diao Ky, et al. Artificial intelligence–enabled rapid diagnosis of patients with COVID-19. *Nat Med* 2020; 26: 1224–1228.
17. Wang X, Peng Y, Lu L, et al. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017; 3462–3471.
18. Pan SJ and Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2010; 22: 1345–1359.
19. Guyon I, Weston J, Barnhill S, et al. Gene selection for cancer classification using support vector machines. *Mach Learn* 2002; 46: 389–422.
20. Shuyue C, Honghua H, Yanjun Z, et al. Study of automatic enhancement for chest radiograph. *J Digit Imaging* 2006; 19: 371–375
21. Thanh DNH, Kalavathi P, Thanh LT, et al. Chest X-ray image denoising using Nesterov optimization method with total variation regularization. *Procedia Comput. Sci*. 2020; 171: 1961–1969.
22. Wang D, Mo J, Zhou G, et al. An efficient mixture of deep and machine learning models for COVID-19 diagnosis in chest X-ray images. *PLoS One* 2020; 15: e0242535. PMID: 33201919; PMCID: PMC7671547
23. He K, Gkioxari G, Dollar P, et al. Mask R-CNN. *IEEE Trans Pattern Anal Mach Intell* 2020; 42: 386–397. Epub June 5, 2018. PMID: 29994331.
24. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells W, et al. (eds) *Medical image computing and computer-assisted intervention - MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science*. New York, NY: Springer, Cham, 2015, vol. 9351, pp. 234–241.
25. Huang G, Liu Z and Weinberger KQ. Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016: 2261–2269.
26. Russakovsky O., et al. Imagenet large scale visual recognition challenge. *Int J Comput Vis* 2015; 115: 211–252.
27. Mahendran N, Vincent DR, Srinivasan K, et al. Sensor-assisted weighted average ensemble model for detecting major depressive disorder. *Sensors (Basel)*. 2019; 19: 4822. Published November 6, 2019.
28. Ju C, Bibaut A and van der Laan M. The relative performance of ensemble methods with deep convolutional neural networks for image classification. *J Appl Stat* 2018; 45: 2800–2818.
29. Moss M, Good VS, Gozal D, et al. A critical care societies collaborative statement: burnout syndrome in critical care

health-care professionals. A call for action. *Am J Respir Crit Care Med* 2016; 194: 106–113.

30. Menon NK, Shanafelt TD, Sinsky CA, et al. Association of physician burnout with suicidal ideation and medical errors. *JAMA Netw Open* 2020; 3: e2028780.
31. Caillet A, Coste C, Sanchez R, et al. Psychological impact of COVID-19 on ICU caregivers. *Anaesth Crit Care Pain Med* 2020; 39: 717–722.
32. Gutierrez G. Artificial intelligence in the intensive care unit. *Crit Care* 2020; 24: 101. Published March 24, 2020.
33. Jacobs SS, Lindell KO, Collins EG, et al. Patient perceptions of the adequacy of supplemental oxygen therapy. Results of the American thoracic society nursing assembly oxygen working group survey. *Ann Am Thorac Soc* 2018; 15: 24–32.
34. Calfee CS, Delucchi K, Parsons PE, et al. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med* 2014; 2: 611–620. .
35. Sinha P, Churpek MM and Calfee CS. Machine learning classifier models can identify acute respiratory distress syndrome phenotypes using readily available clinical data. *Am J Respir Crit Care Med* 2020; 202: 996–1004. .
36. Singhal L, Garg Y, Yang P, et al. eARDS: a multi-center validation of an interpretable machine learning algorithm of early onset Acute Respiratory Distress Syndrome (ARDS) among critically ill adults with COVID-19. *PLoS One* 2021;16: e0257056. Published September 24, 2021.
37. Štrumbelj E and Kononenko I. Explaining prediction models and individual predictions with feature contributions. *Knowl Inf Syst* 2014; 41: 647–665.
38. Yang P, Wu T, Yu M, et al. A new method for identifying the acute respiratory distress syndrome disease based on noninvasive physiological parameters. *PLoS One* 2020; 15: e0226962. Published February 5, 2020.
39. Jabbour S, Fouhey D, Kazerooni E, Wiens J, Sjoding MW. Combining chest X-rays and electronic health record (EHR) data using machine learning to diagnose acute respiratory failure. *J Am Med Inform Assoc* 2022; 29(6): 1060–1068.
40. Dayan I, Roth HR, Zhong A, et al. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nat Med*. 2021; 27: 1735–1743.