An Explainable AI Predictor to Improve Clinical Prognosis for Acute Respiratory Distress Syndrome

Songlu Lin[1], Meicheng Yang[2], Yuzhe Wang[1], Zhihong Wang[1*]

[1]Instrument Science and Electrical Engineering, Jilin University, Changchun, China
[2]State Key Laboratory of Digital Medical Engineering, School of Instrument Science and Engineering, Southeast University, Nanjing, China

Abstract

Acute Respiratory Distress Syndrome (ARDS) is a severe respiratory disorder characterized by the failure of the lungs and often associated with elevated death rates. Providing an accurate clinical prognosis for ARDS patients is complex due to the myriad clinical variables involved. In this study, we introduce an Explainable AI Predictor designed to improve the accuracy of prognostic predictions for ARDS. -This work outlines a solution to the challenge of short-term ARDS diagnosis, utilizing an algorithm that leverages a multi-feature fusion approach based on XGBoost learning and Optuna. The proposed algorithm enables ARDS status prediction within a critical 48-hour window, which is vitally important for life-saving interventions in clinical settings. Furthermore, -- the proposed algorithm incorporates features that enhance interpretability, assisting medical professionals in diagnosis and treatment planning. Experimental results substantiate the efficacy of our proposed method, with the algorithm achieving an overall micro-AUC of 0.832 when applied to the test set. This performance metric underscores the accuracy and predictive strength of our approach. The encouraging results emphasize the algorithm's potential to facilitate healthcare professionals in making timely and precise decisions in managing ARDS.

1. Introduction

Acute Respiratory Distress Syndrome (ARDS) is a life-threatening condition characterized by severe respiratory failure and widespread lung inflammation. It is a complex syndrome with diverse causes, including pneumonia, sepsis, and trauma, and is associated with high morbidity and mortality rates [1]. Due to its severe consequences, morbidity, mortality, and medical costs, ARDS has been a significant focus in clinical and basic research within critical care medicine. Traditional prevention and control methods typically rely on a combination of manual judgment and scoring systems, such as Apache IV. However, these methods do not adequately address the urgent need for early detection of ARDS to facilitate effective treatment.

With the increasing availability of publicly accessible electronic health records (EHRs), there are now tremendous opportunities to develop data-driven and efficient machine-learning models to diagnose disease. Therefore, in recent years, researchers have proposed various rule-based machine learning or deep learning models to achieve the goal of early prediction of ARDS using physiological data [2-4]. Zhang [5] identified binary classification for ARDS patients. The best-performing machine learning algorithm had an AUC of 0.84. Furthermore, according to Sidney Le BA [6], models predicting ARDS incidence and severity using continuous noninvasive parameters reached peak performance with AUC values of 0.79, using 9909 patients in 48 window cohorts. However, there are few studies on short-term survival prediction for confirmed ARDS patients. In addition, direct comparisons between these methods are challenging due to differences in clinical criteria, available patient variables, predictive tasks, evaluation metrics, and other factors. Model interpretation is also the key to current machine learning models.

To address this gap, we present an explainable algorithm that utilizes multi-feature fusion based on XGBoost [7] learning and the Optuna framework [8], to predict the status of ARDS patients (quick death, recovery, and long stay) within 48 hours after the clinical definition of ARDS. This method has the potential to provide timely early warning signals for patients with ARDS and can assist clinicians in making accurate diagnoses at critical moments. The proposed algorithm may herald a new step forward in the field, potentially improving the practical application of machine learning in clinical settings.

2. Methodology

2.1. Data source and cohort extraction

Data used were sourced from the Medical Information Mart for Intensive Care (MIMIC-IV) database (version 2.0) [9]. The Berlin criteria [10] were employed to identify ARDS patients with a PaO_2/FiO_2 ratio (P/F) ⩽ 300 mm Hg and PEEP ⩾ 5 cmH_2O for at least 24 hours. Due to the limited number of patients with available imaging data, our analysis did not include images. Patients with congestive heart failure were excluded based on ICD codes. All patients with ARDS required mechanical ventilation in the ICU. Both diagnosis and monitoring time were documented for each patient, with the diagnosed time being the latest of the patient's initial PEEP time and P/F ratio time in the MIMIC-IV database. Only the first ICU stay of each patient was considered. According to the study practice, monitoring was initiated 24 hours after the ARDS diagnosis. ARDS patients were classified into three categories based on their outcomes: "death" for those who passed away within 24 hours of monitoring (48 hours after ARDS was diagnosed), "recovery" for patients who fully recovered from ARDS within 24 hours post-monitoring, and (longer-term) "stay" for patients who continued to have ARDS beyond 24 hours after monitoring.

The MIMIC-IV database recorded 50,920 patients' information, after defining 4337 ICU patients who have ARDS, as shown in Figure 1.

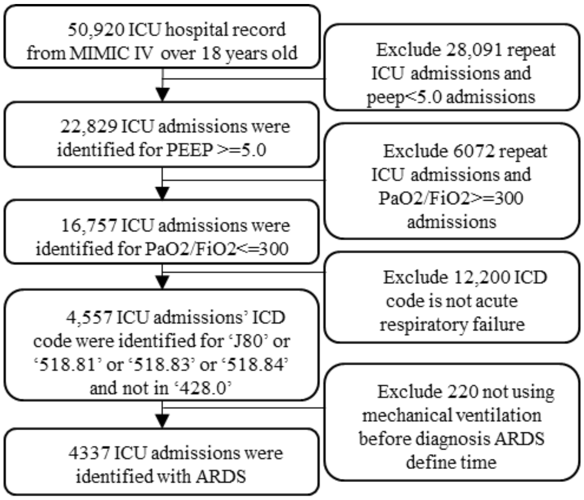
50,920 ICU hospital record from MIMIC IV over 18 years old
Exclude 28,091 repeat ICU admissions and peep<5.0 admissions
22,829 ICU admissions were identified for PEEP >=5.0
Exclude 6072 repeat ICU admissions and PaO2/FiO2>=300 admissions
16,757 ICU admissions were identified for PaO2/FiO2<=300
Exclude 12,200 ICD code is not acute respiratory failure
4,557 ICU admissions' ICD code were identified for 'J80' or '518.81' or '518.83' or '518.84' and not in '428.0'
Exclude 220 not using mechanical ventilation before diagnosis ARDS define time
4337 ICU admissions were identified with ARDS

Figure 1. MIMIC- IV dataset screening process

This research processes the unrefined data through the subsequent two stages.

(1) Remove the continuous variates whose missing values account for more than 80%.

(2) Impute missing data with the cosine similarity and median strategy. We use the following formula (1) to calculate the cosine similarity:

$$similarity = \frac{A \cdot B}{||A||\ ||B||} = \frac{\sum_{i=1}^{n} A_i * B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2} * \sqrt{\sum_{i=1}^{n}(B_i)^2}} \qquad (1)$$

A and B represent different times of the same variables of the same patients. A_i and B_i represent the components of vectors A and B respectively.

Variables for which cosine similarity cannot be calculated are filled with the median.

2.2. Feature extraction

According to clinically relevant studies [2, 11, 12], this study performed data preprocessing and feature selection. After Manny-Whitney U test, we incorporated 21 features into the final model, including static variables and continuous variables, including Acute Physiology Score III (APS III score), pH, PaO_2 (mm Hg), $PaCO_2$ (mm Hg), Blood Urea Nitrogen (mg/dL), Non-Invasive SBP (mm Hg), Non-Invasive MBP (mm Hg), Non-Invasive DBP (mm Hg), Base excess (mmol/L), SBP (mm Hg), MBP (mm Hg), DBP (mm Hg), Respiratory rate (breaths/min), PEEP, PaO_2 (%), FiO_2 (%), Non-calcium (mg/dL), Bicarbonate (mmol/L), GCS verbal, GCS motor, GCS eyes. The continuous variables were employed in three summarization methods:

(1) Mean is represented by the average value.

(2) Variance is computed when multiple data sets were available for the same patient at the same stage, to capture variability.

(3) Rate of change is calculated when there were multiple data sets, indicating the speed or magnitude of change over time.

2.3. Classification

2.3.1 Model training

Due to the unbalanced sample size across categories, The MIMIC-IV database was divided into a training set (80%) and an internal test set (20%) using stratified sampling, in a random manner.

The framework of the proposed algorithm for the prognosis prediction of ARDS is shown in Figure 2.

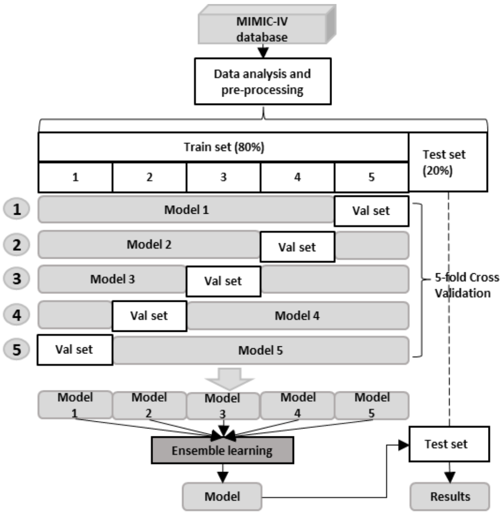

Figure 2. The framework of the proposed algorithm

Raw patient data is analyzed first for helping us to get more effective information in feature extraction. Then the processed data is divided into a train set and a test set. After that, the XGBoost classifier receives training data after feature extraction as inputs and tunes the hyperparameter automatically using an Optuna optimizer. Meanwhile, the 5-fold cross-validation method is used to verify the stability of this work and finally uses soft voting [13] strategy exports an ensemble model for comparison.

2.3.2 Model evaluation

The overall discriminative ability of the models is compared by using the area under the curve (Micro-AUC [14]), and the optimal model is selected based on the highest micro-AUC value, which was calculated as follow formula (2). The classification and prediction performances of each model were compared using ACC to determine their effectiveness in identifying positive instances, which was calculated as follow formula (3). In the formula (2) and (3), n is the number of categories, in this study was divided into 3 categories: recovery, death, and stay, where TP, TN, FP, and FN represent the numbers of true positives, true negatives, false positives, and false negatives, respectively.

$$Micro - AUC = \frac{1}{n}\sum_{i=1}^{n} AUC_i \tag{2}$$

$$Accuracy\ (ACC) = \frac{1}{n} * \frac{TP+TN}{TP+FP+TN+FN} \tag{3}$$

2.3.3 Model explanation

SHAP [15], a widely recognized technique in the field of Explainable Artificial Intelligence (XAI), holds a prominent position for elucidating model behaviour. This method is particularly invaluable in medical contexts where algorithmic expertise may be lacking, enabling medical practitioners to seamlessly grasp the intricacies of model operations. Using the SHAP interpreter, we conducted a careful factor analysis of the model, further substantiating our investigative approach.

3. Results and Discussions

3.1. Data analysis and pre-processing

Statistical analysis was performed based on the ARDS patients in the MIMIC-IV database. Table 1 shows the number of ICU and hospital survivors in each category. The hospital mortality rate for ICU patients diagnosed with severe ARDS was 38.78%, with an overall ICU mortality rate of 26.52% and a hospital mortality rate of 32.60%. The median length of stay in the ICU was 9.66 days, and the median duration of ARDS in the cohort was 26.00 hours. The data show clear trends: approximately 5% of all cases resulted in death, while 48.99% were associated with hospitalization and 45.74% indicated recovery. Those who recovered or required hospitalization had less severe symptoms. Conversely, death was correlated with greater disease severity. In addition, the population requiring hospitalization tended to have severe ARDS, while those who recovered had milder symptoms.

Table 1. Statistical analysis of the MIMIC-IV database

ARDS patients	Recovery	Death	Stay
ICU non-survival	289	234	629
ICU survival	1686	3	1496
Hospital non-survival	435	237	751
Hospital survival	1540	0	1374

3.2. Model results

A side-by-side comparison of the five individual XGBoost models trained with 5-fold cross-validation and the ensemble model was performed to assess their performance. The results are presented in Table 2, with the ACC and micro-AUC. It was found that there is no significant difference between the results obtained by the individual models, indicating the stability of our algorithm. Moreover, it can be observed that the ensemble model achieved the highest ACC and micro-AUC simultaneously. This outcome confirms that employing an ensemble approach leads to improved predictions compared to using a single model.

Table 2. Performance of different models on the test set.

Model		Micro-AUC	ACC
Individual XGBoost Model	1	0.807	0.635
	2	0.807	0.612
	3	0.815	0.628
	4	0.824	0.654
	5	0.827	0.657
Ensemble model		0.832	0.692

The SHAP interpretation analysis was conducted on the optimal model, revealing that the overall impact of the model on the short-term prognosis of the three states was summarized. As shown in Figure 3, notably, the APS III score was found to have the greatest influence among the factors examined in three statuses, and GCS eyes, SBP, $SpO_{2,}$ and the rate change of temperature also have a significant difference in the model that allows the model to make classification predictions. The inherent explanatory nature of the model serves to elucidate, in juxtaposition, the unique attributes inherent to ARDS pathology, while simultaneously explaining the discernible substantial disparities in transient physiological fluctuations observed among individuals afflicted with ARDS.

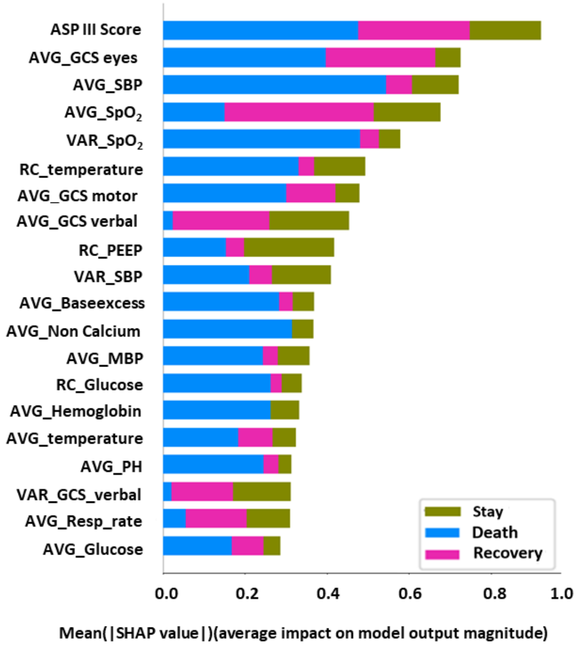

Figure 3. Model explanation of three diagnosis status

In Figure 3, ‘AVG’ represents the mean value of the variables, ‘VAR’ represents the variance of the variables, and ‘RC’ represents the rate of change of the variables.

4. Conclusion

This paper presents a solution to the challenge of predicting a short-term diagnosis of ARDS, which proposed an algorithm that leverages multi-feature fusion based XGBoost learning and Optuna. The algorithm enables the status prediction of ARDS within a crucial 48-hour window, which is of paramount importance in clinical settings for life-saving interventions. Moreover, the model incorporates interpretability features to assist in clinical diagnosis and treatment. The results of our experiments demonstrate the efficacy of the proposed method. When applied to the test set, the algorithm achieves an overall micro-AUC of 0.832. This performance metric indicates the accuracy and predictive power of our approach in identifying. The promising results highlight the potential of our algorithm to aid healthcare professionals in timely and accurate decision-making for ARDS management.

References

[1] M. R. Suchyta *et al.*, "Increased mortality of older patients with acute respiratory distress syndrome," *Chest,* vol. 111, no. 5, pp. 1334-1339, 1997.

[2] E. Schwager *et al.*, "Utilizing machine learning to improve clinical trial design for acute respiratory distress syndrome," *NPJ Digital Medicine,* vol. 4, no. 1, p. 133, 2021.

[3] J. Máca *et al.*, "Past and present ARDS mortality rates: a systematic review," *Respiratory Care,* vol. 62, no. 1, pp. 113-122, 2017.

[4] E. Papoutsi *et al.*,"Association between ventilatory ratio and mortality persists in patients with ARDS requiring prolonged mechanical ventilation," *Intensive Care Medicine,* pp. 1-2, 2023.

[5] W. Zhang *et al.*, "To Establish an Early Prediction Model for Acute Respiratory Distress Syndrome in Severe Acute Pancreatitis Using Machine Learning Algorithm," *Journal of Clinical Medicine,* vol. 12, no. 5, p. 1718, 2023.

[6] S. Le *et al.*, "Supervised machine learning for the early prediction of acute respiratory distress syndrome (ARDS)," *Journal of Critical Care,* vol. 60, pp. 96-102, 2020.

[7] T. Chen *et al.*, "Xgboost: extreme gradient boosting," *R Package Version 0.4-2,* vol. 1, no. 4, pp. 1-4, 2015.

[8] T. Akiba *et al.*, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2623-2631.

[9] A. E. Johnson *et al.*, "MIMIC-IV, a freely accessible electronic health record dataset," *Scientific Data,* vol. 10, no. 1, p. 1, 2023.

[10] N. D. Ferguson *et al.*, "The Berlin definition of ARDS: an expanded rationale, justification, and supplementary material," *Intensive Care Medicine,* vol. 38, pp. 1573-1582, 2012.

[11] W. Wu *et al.*,"Developing and evaluating a machine-learning-based algorithm to predict the incidence and severity of ARDS with continuous non-invasive parameters from ordinary monitors and ventilators," *Computer Methods and Programs in Biomedicine,* vol. 230, p. 107328, 2023.

[12] M. A. Matthay *et al.*, "Acute respiratory distress syndrome," *Nature Reviews Disease Primers,* vol. 5, no. 1, p. 18, 2019.

[13] S. Kumari, D. Kumar, and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier," *International Journal of Cognitive Computing in Engineering,* vol. 2, pp. 40-46, 2021.

[14] X.-Z. Wu and Z.-H. Zhou, "A unified view of multi-label performance measures," in *International Conference on Machine Learning,* 2017: PMLR, pp. 3780-3788.

[15] D. Slack *et al.*,"Fooling lime and shap: Adversarial attacks on post hoc explanation methods," in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 2020, pp. 180-186.

Address for correspondence:

Zhihong Wang
#938 West Minzhu Street, Jilin University, Changchun, China.
Email: zhwang@jlu.edu.cn