

RESEARCH ARTICLE

Machine learning for prediction of in-hospital mortality in lung cancer patients admitted to intensive care unit

Tianzhi Huang¹, Dejin Le², Lili Yuan³, Shoujia Xu^{4*}, Xiulan Peng^{5*}

1 Department of Rehabilitation, The Second Affiliated Hospital of Jiangnan University, Wuhan, China, **2** Department of Respiratory Medicine, People's Hospital of Daye, The Second Affiliated Hospital of Hubei Polytechnic University, Daye, Hubei, China, **3** Department of Anesthesiology, The Second Affiliated Hospital of Jiangnan University, Wuhan, China, **4** Department of Orthopedics, Sinopharm Dongfeng General Hospital, Hubei University of Medicine, Shiyan, Hubei, China, **5** Department of Oncology, The Second Affiliated Hospital of Jiangnan University, Wuhan, China

☞ These authors contributed equally to this work.

* pengxlwh@163.com (XP); 513783626@qq.com (SX)



Abstract

Backgrounds

The in-hospital mortality in lung cancer patients admitted to intensive care unit (ICU) is extremely high. This study intended to adopt machine learning algorithm models to predict in-hospital mortality of critically ill lung cancer for providing relative information in clinical decision-making.

Methods

Data were extracted from the Medical Information Mart for Intensive Care-IV (MIMIC-IV) for a training cohort and data extracted from the Medical Information Mart for eICU Collaborative Research Database (eICU-CRD) database for a validation cohort. Logistic regression, random forest, decision tree, light gradient boosting machine (LightGBM), eXtreme gradient boosting (XGBoost), and an ensemble (random forest+LightGBM+XGBoost) model were used for prediction of in-hospital mortality and important feature extraction. The AUC (area under receiver operating curve), accuracy, F1 score and recall were used to evaluate the predictive performance of each model. Shapley Additive exPlanations (SHAP) values were calculated to evaluate feature importance of each feature.

Results

Overall, there were 653 (24.8%) in-hospital mortality in the training cohort, and 523 (21.7%) in-hospital mortality in the validation cohort. Among the six machine learning models, the ensemble model achieved the best performance. The top 5 most influential features were the sequential organ failure assessment (SOFA) score, albumin, the oxford acute severity of illness score (OASIS) score, anion gap and bilirubin in random forest and XGBoost model. The SHAP summary plot was used to illustrate the positive or negative effects of the top 15 features attributed to the XGBoost model.

OPEN ACCESS

Citation: Huang T, Le D, Yuan L, Xu S, Peng X (2023) Machine learning for prediction of in-hospital mortality in lung cancer patients admitted to intensive care unit. PLoS ONE 18(1): e0280606. <https://doi.org/10.1371/journal.pone.0280606>

Editor: Samuele Ceruti, Clinica Luganese Moncucco, SWITZERLAND

Received: October 7, 2022

Accepted: January 4, 2023

Published: January 26, 2023

Copyright: © 2023 Huang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting Information](#) files.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Conclusion

The ensemble model performed best and might be applied to forecast in-hospital mortality of critically ill lung cancer patients, and the SOFA score was the most important feature in all models. These results might offer valuable and significant reference for ICU clinicians' decision-making in advance.

Introduction

Lung cancer is the third most common malignancy and is reported the leading cause of cancer death in males and the second most common cancer in females, which taking up more than one-fifth of all cancer deaths worldwide [1–3]. Exceed 158,000 patients died from lung cancer in the United States in 2016, which accounted for 27% of all cancer deaths [4, 5], the prognosis remains poor although improvement has been made in the therapy of lung cancer, the 5-year survival rate for all stages combined is only 15% [6, 7]. Many lung cancer patients require admitted to intensive care unit (ICU) and respiratory failure requiring mechanical ventilation is the major reason for lung cancer patients being admitted to the ICU [8, 9]. Although progressive improvement has been made to improve the prognosis in lung cancer patients admitted to the ICUs, the mortality rate remains extremely high, the mortality rate in lung cancer patients admitted to ICU was 43% and the in-hospital mortality is 60%, and the mortality rate is higher in patients with stage IV (68%) [10]. Currently, the lack of early prediction and risk stratification for in-hospital mortality is the main challenge for ICU clinicians. The decision regarding which groups of lung cancer patients admitted to the ICU at high-risk and would have poor prognosis is based on a complex suite of considerations including therapeutic options and the wishes of patients and their family. These critically ill lung cancer patients usually have poor long-term survival and high financial cost. Hence, it's necessary to explore risk prediction models to distinguish those at high-risk of critically ill lung cancer patients admitted to ICU.

The development of artificial intelligence has led to a significant improvement in the predictive models used for estimating the risk of mortality in cancer patients. Machine learning (ML), a new type of artificial intelligence can transform measurement results into relevant predictive models, especially cancer models, based on the rapid development of large datasets and deep learning. Recently, ML have been shown to be effective in predicting lung cancer susceptibility, recurrence, and survival of malignant tumors [11–13]. However, there is still limited data relating to the in-hospital mortality risk prediction models using ML methods in patients with lung cancer in the ICU setting.

Therefore, this study aimed to develop six ML algorithm models including logistic regression, decision tree, random forest, light gradient boosting machine (GBM), extreme gradient boosting (XGBoost), and an ensemble model to predict the in-hospital mortality among lung cancer patients admitted to ICU so that individual prevention strategies for critically ill lung cancer patients could be proposed to help clinicians to make therapeutic decisions. Moreover, we also intended to compared the six ML models and determined the best model for in-hospital mortality prediction in lung patients admitted to the ICU.

Methods

Data source

This retrospective study utilized information from the eICU Collaborative Research Database (eICU-CRD) [14] and the Medical Information Mart for Intensive Care-IV (MIMIC-IV

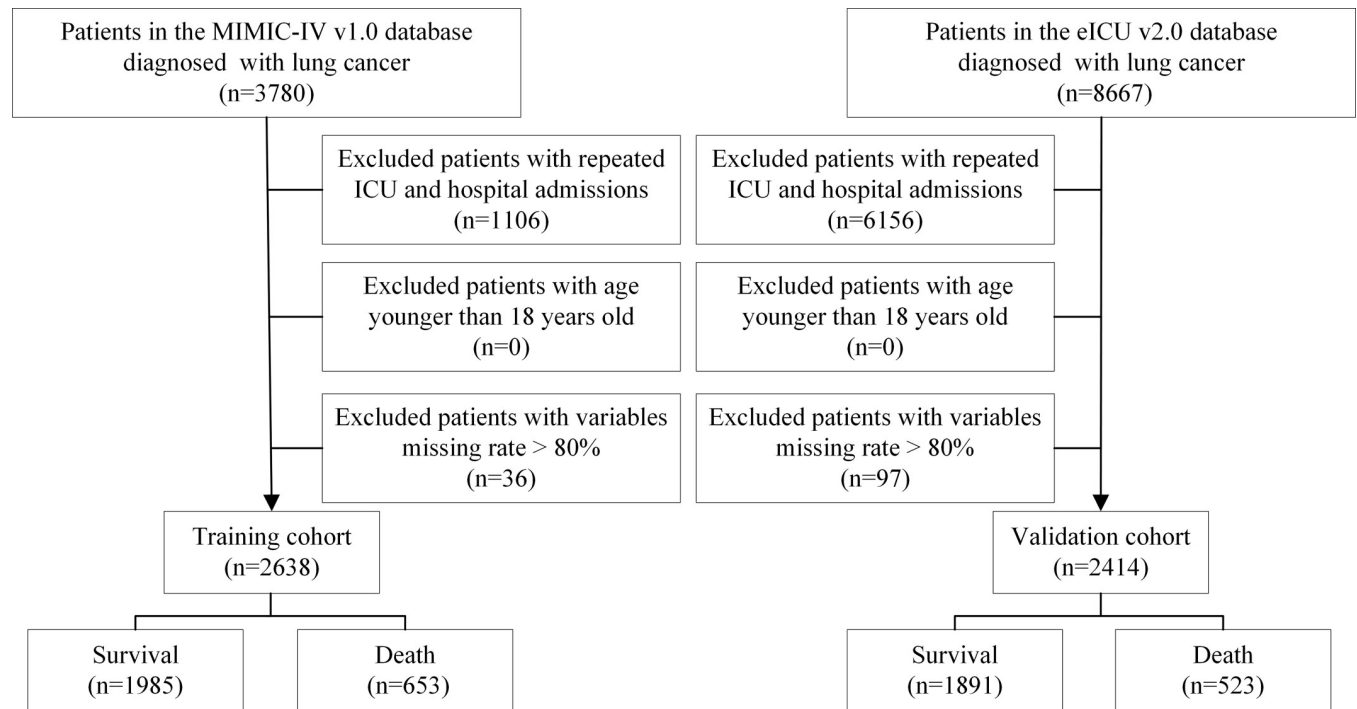


Fig 1. The flow chart of this study.

<https://doi.org/10.1371/journal.pone.0280606.g001>

version 1.0) database [15], eICU-CRD contains data of more than 200 thousand ICU admissions in 2014 and 2015 at 208 US hospitals while MIMIC-IV includes information of more than 70,000 patients admitted to the ICUs of Beth Israel Deaconess Medical Center in Boston, MA, from 2008 to 2019. Due to the data used in this study were extracted from public databases, it was exempt from the requirement for informed consent from patients and approval of the Institutional Review Board (IRB). All procedures were performed according to the ethical standards of the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. After finishing the web-based training courses (S1 Fig) and the Protecting Human Research Participants examination, we obtained permission to extract data from the eICU-CRD and MIMIC-IV database.

Cohort selection

Patients with one of the following conditions were excluded: (1) less than 18-year-old at first admission to ICU; (2) repeated ICU admissions; (3) more than 80% of personal data was missing. We randomly selected MIMIC-IV database as the training cohort and eICU-CRD database as the validation cohort. A total of 2,638 patients in the MIMIC-IV database assigned into the training cohort and 2,414 patients in the eICU-CRD database assigned into the validation cohort were finally included in this study, the detailed flowchart was shown in Fig 1.

Date collection and outcomes

Baseline characteristics and admission information: age, gender and body mass index (BMI) were calculated as described in previous studies. Comorbidities including hypertension, diabetes, chronic kidney disease, myocardial infarction, congestive heart failure, atrial fibrillation, valvular disease, chronic obstructive pulmonary disease, stroke, hyperlipidemia and liver

disease were also collected for analysis based on the recorded ICD codes in the two databases. Charlson comorbidity index (CCI) was also included. In addition, severity scores including sequential organ failure assessment (SOFA) score, the oxford acute severity of illness score (OASIS), the acute physiology score III (APSO) were collected. Acute complications during ICU including acute heart failure, acute respiratory failure, acute hepatic failure and cardiac arrest based on ICD codes, acute kidney injury based on KDIGO guideline in 48 hours [16], sepsis based on sepsis 3.0 criteria [17] were also recorded. In addition, initial vital signs and laboratory results were also measured during the first 24 hours of ICU admission.

The primary outcome was in-hospital mortality.

Statistical analysis

For all continuous covariates, the mean values and standard deviations are reported. Categorical data were expressed as frequency (percentage). The Chi-square test or Fisher's test was appropriately performed to compare the differences between groups. The baseline characteristics were reported as a training cohort and validation cohort. The comparison of baseline characteristics was performed in R software (version 4.1.0). $P < 0.05$ was considered statistically significant. Modeling work were done using Python 3.6.4.

Construction of in-hospital mortality predictive models

Logistic regression, decision tree, random forest, and two gradient boosting decision trees, including LightGBM, and XGBoost, were adopted to construct prediction models. In order to improve prediction, an ensemble model was constructed, which applied staking strategy using random forest, LightGBM and XGBoost [18]. The prediction probabilities of the three models were input into a logistic regression model to produce a final prediction. Hence, six in-hospital mortality predictive models were developed using logistic regression, decision tree, random forest, LightGBM, XGBoost and ensemble models, which each used 100 full features for each time window. Furthermore, the top 10 important features derived from random forest, lightGBM, and XGBoost model were also analysis [18].

Performance evaluation

To evaluate and compare the predictive accuracy of prediction by decision tree, random forest, LightGBM, XGBoost, ensemble model and logistic regression models. Each model was evaluated according to accuracy, recall, F1 score, and AUC (area under the receiver operating characteristic) curve [19].

SHAP analysis

To further analyze the positive or negative effect of the important features identified for in-hospital mortality prediction and investigate the relationship between, a shapely additive explanations (SHAP) analysis was performed using Python 3.7.0. The SHAP value is the assigned predicted value of each feature of the data [20].

Results

Baseline characteristics

A total of 5,052 patients were finally included in the present study, including 2,638 patients in the training cohort extracted from the MIMIC-IV database and 2,414 patients in the validation cohort extracted from the eICU-CRD database. There were 653 (24.8%) in-hospital death in

the training cohort, and 523 (21.7%) in-hospital death in the validation cohort. [Table 1](#) showed the baseline characteristics both in the training cohort and in the validation cohort.

Model performance

Six models, logistic regression, decision tree, random forest, LightGBM, XGBoost, and ensemble models were used to predict in-hospital mortality using all the features. As can be seen in [Table 2](#), the traditional model logistic regression exhibited the worst predictive ability, followed by decision tree, random forest, XGBoost, LightGBM. And the ensemble model showed the best predictive ability with the highest accuracy (0.89), recall (0.80), F1 score (0.82) and AUC (0.92) in training cohort. And the results in the validation cohort similar to the results in the training cohort ([Table 2](#)). In addition, we also performed ROC analysis to further confirm the in-hospital mortality predictive ability of these six models, as shown in [Fig 2A and 2B](#), the logistic regression model depicted the worst predictive ability, followed by decision tree, random forest, XGBoost, LightGBM. And the ensemble model showed the best predictive performance both in the training cohort and in the validation cohort.

Feature importance analysis

To clarify the important features that impacts on model output, the feature importance analysis was conducted. The top 15 features derived from random forest, lightGBM, and XGBoost model were shown in [Fig 3](#). In random forest model, SOFA score was the most influential feature, followed by albumin, OASIS score, anion gap, bilirubin, mechanical ventilation, acute respiratory failure, APACHE II score, length of hospital, BUN, WBC, respiratory rate, vasopressors usage and RDW, and these features also had important on random forest model ([Fig 3A](#)). For lightGBM model, anion gap played the most important role in prediction in-hospital mortality, moreover, SOFA score, OASIS score, albumin, length of hospital, bilirubin, WBC, platelet, BUN, heart rate, MCH, APACHE II score, creatinine and MCV also plays important role in prediction ([Fig 3B](#)). Furthermore, in terms of XGBoost model, SOFA score had the most influence on in-hospital mortality prediction, followed by anion gap, bilirubin, OASIS score, albumin, white blood cell, bicarbonate, length of hospital, acute respiratory failure, RDW, temperature, creatinine, platelet, MCHC and BMI ([Fig 3C](#)). Moreover, the feature importance analysis derived from random forest, lightGBM, and XGBoost model were also conducted in validation cohort in [S2–S4 Figs](#). And the results were coincided with the result of the training cohort.

SHAP analysis

In order to manifest an overall positive or negative impact on model output, and to analyze the similarities and differences of important characteristics of critically ill lung cancer with different severities, the SHAP summary chart was used. As shown in [Fig 4](#), SOFA score ranked the first in importance among the top 20 features of the XGBoost model, and the higher the SOFA score, the higher probability of in-hospital mortality development, indicating that SOFA score should be observed first in in-hospital mortality prediction.

Taking the XGBoost model with excellent performance for predicting dead/survival using all features as an example, combined with the SHAP analysis method, a representative dead patient and a survival patient were selected to illustrate the effect of features on the prediction ability. As shown in [Fig 5](#), for predicting dead patients, SOFA score plays a major positive role in the prediction results, the SHAP value of final model predicted for this patient is 0.96, which is beyond than 0, thus successfully predicting the patient as an in-hospital died patient. For predicting survival patients, anion gap plays a major positive role in the prediction results, SOFA score played a major negative role in predicting outcomes, the SHAP value of final

Table 1. Comparisons of baseline characteristics in all cohorts.

| Characteristics | Training cohort | Validation cohort | P value |
|--|-----------------|-------------------|---------|
| N | 2638 | 2414 | |
| Age, years old | 68.9 ± 12.1 | 66.5 ± 11.3 | <0.001 |
| Gender, male, n (%) | 1361 (51.6) | 1272 (52.7) | 0.451 |
| BMI, kg/m ² | 26.1 ± 7.6 | 26.0 ± 7.3 | 0.138 |
| Ethnicity, n (%) | | | <0.001 |
| White | 1924 (72.9) | 1950 (80.8) | |
| Black | 302 (11.4) | 247 (10.2) | |
| Others | 412 (14.7) | 217 (9.0) | |
| Tumor type, n (%) | | | <0.001 |
| Primary | 1804 (68.4) | 1889 (78.3) | |
| Metastatic | 834 (31.6) | 525 (21.7) | |
| Critical care procedure, n (%) | | | |
| Mechanical ventilation | 901 (34.2) | 738 (30.6) | 0.007 |
| Continuous renal replacement therapy | 35 (1.3) | 20 (0.8) | 0.117 |
| Vasopressors | 718 (27.2) | 359 (14.9) | <0.001 |
| Comorbidities, n (%) | | | |
| Myocardial infarction | 339 (12.9) | 186 (7.7) | <0.001 |
| Congestive heart failure | 561 (21.3) | 250 (10.4) | <0.001 |
| Hypertension | 1090 (41.3) | 1223 (50.7) | <0.001 |
| Diabetes | 608 (23.0) | 493 (20.4) | 0.026 |
| Chronic kidney disease | 462 (17.5) | 171 (7.1) | <0.001 |
| Liver disease | 198 (7.5) | 21 (0.9) | <0.001 |
| Chronic obstructive pulmonary disease | 1034 (39.2) | 796 (33.0) | <0.001 |
| Stroke | 246 (9.3) | 127 (5.3) | <0.001 |
| Atrial fibrillation | 720 (27.3) | 248 (10.3) | <0.001 |
| Hyperlipidemia | 888 (33.7) | 102 (4.2) | <0.001 |
| Charlson comorbidity index, points | 9.4 (2.7) | 6.9 (1.7) | <0.001 |
| Acute complications during ICU, n (%) | | | |
| Acute heart failure | 247 (9.4) | 100 (4.1) | <0.001 |
| Acute respiratory failure | 765 (29.0) | 904 (37.4) | <0.001 |
| Acute hepatic failure | 31 (1.2) | 6 (0.2) | <0.001 |
| Acute kidney injury | 582 (22.1) | 363 (15.0) | <0.001 |
| Sepsis | 1244 (47.2) | 434 (18.0) | <0.001 |
| Cardiac arrest | 84 (3.2) | 58 (2.4) | 0.111 |
| Score system, points | | | |
| Sequential organ failure assessment | 4.4 ± 1.4 | 3.8 ± 1.3 | <0.001 |
| Oxford acute severity of illness score | 32.3 ± 9.0 | 24.8 ± 10.0 | <0.001 |
| Acute physiology score III | 48.3 ± 12.8 | 47.9 ± 21.5 | 0.594 |
| Vital signs | | | |
| Systolic blood pressure, mmHg | 123.7 ± 24.1 | 123.3 ± 24.5 | 0.507 |
| Diastolic blood pressure, mmHg | 69.4 ± 17.3 | 69.6 ± 16.4 | 0.661 |
| Mean arterial pressure, mmHg | 83.9 ± 17.9 | 83.7 ± 18.3 | 0.657 |
| Heart rate, bpm | 94.8 ± 21.6 | 97.3 ± 21.8 | <0.001 |
| Respiratory rate, bpm | 20.9 ± 6.6 | 21.3 ± 6.3 | 0.057 |
| Temperature, °C | 36.7 ± 0.7 | 36.8 ± 0.7 | 0.001 |
| SpO ₂ , % | 96.1 ± 4.3 | 95.3 ± 6.0 | <0.001 |
| Laboratory values | | | |

(Continued)

Table 1. (Continued)

| | | | |
|--|------------------|------------------|--------|
| White blood cell, $\times 10^9/L$ | 11.5 \pm 8.7 | 12.2 \pm 9.7 | 0.011 |
| Hemoglobin, g/dL | 10.7 \pm 2.2 | 11.0 \pm 2.3 | <0.001 |
| Platelet, $\times 10^9/L$ | 251.2 \pm 81.6 | 235.6 \pm 71.4 | <0.001 |
| Hematocrit, % | 32.7 \pm 6.2 | 33.6 \pm 6.8 | <0.001 |
| Mean corpuscular volume, fL | 90.6 \pm 7.7 | 90.3 \pm 7.1 | 0.184 |
| Mean corpuscular hemoglobin, pg | 29.5 \pm 2.9 | 29.5 \pm 2.6 | 0.418 |
| Mean corpuscular hemoglobin concentration, g/L | 32.6 \pm 1.7 | 32.8 \pm 1.5 | <0.001 |
| Red blood cell, $\times 10^{12}/L$ | 3.6 \pm 0.7 | 3.7 \pm 0.8 | <0.001 |
| Red cell distribution width, % | 15.7 \pm 2.5 | 16.1 \pm 2.6 | <0.001 |
| Albumin, g/dL | 3.1 \pm 0.7 | 3.0 \pm 0.7 | <0.001 |
| Bilirubin, mmol/L | 1.2 \pm 0.4 | 0.9 \pm 0.3 | <0.001 |
| Anion gap, mEq/L | 14.8 \pm 3.6 | 10.8 \pm 4.0 | <0.001 |
| Bicarbonate, mEq/L | 24.1 \pm 4.7 | 25.8 \pm 5.2 | <0.001 |
| Glucose, mg/dL | 136.5 \pm 54.3 | 141.4 \pm 51.7 | 0.019 |
| Blood urea nitrogen, mg/dL | 24.3 \pm 8.9 | 23.9 \pm 8.6 | 0.378 |
| Creatinine, mg/dL | 1.2 \pm 0.5 | 1.1 \pm 0.4 | 0.457 |
| Calcium, mg/dL | 8.6 \pm 1.0 | 8.7 \pm 1.0 | <0.001 |
| Chloride, mmol/L | 101.7 \pm 6.3 | 100.6 \pm 6.5 | <0.001 |
| Potassium, mmol/L | 4.3 \pm 0.7 | 4.2 \pm 0.7 | <0.001 |
| Sodium, mmol/L | 137.4 \pm 5.1 | 136.2 \pm 5.5 | <0.001 |
| Prothrombin time, s | 15.4 \pm 5.2 | 15.6 \pm 5.7 | 0.329 |
| Activated partial thromboplastin time, s | 35.2 \pm 10.3 | 36.3 \pm 11.1 | 0.062 |
| International normalized ratio | 1.4 \pm 0.5 | 1.4 \pm 0.4 | 0.500 |
| Length of hospital, days | 7.4 (4.3, 12.1) | 6.9 (4.1, 11.7) | 0.012 |
| Length of ICU, days | 2.0 (1.1, 4.0) | 2.1 (1.1, 4.0) | 0.490 |
| Death, n (%) | 653 (24.8) | 523 (21.7) | 0.010 |

<https://doi.org/10.1371/journal.pone.0280606.t001>

Table 2. Performance of the prediction models using all features.

| Model | Accuracy | Recall | F1 score | AUC |
|---------------------|----------|--------|----------|------|
| Training cohort | | | | |
| Logistic regression | 0.83 | 0.68 | 0.71 | 0.83 |
| Decision tree | 0.85 | 0.73 | 0.75 | 0.86 |
| Random forest | 0.84 | 0.70 | 0.78 | 0.89 |
| XGBoost | 0.86 | 0.77 | 0.79 | 0.90 |
| LightGBM | 0.88 | 0.80 | 0.81 | 0.91 |
| Ensemble model | 0.89 | 0.80 | 0.82 | 0.92 |
| Validation cohort | | | | |
| Logistic regression | 0.79 | 0.79 | 0.70 | 0.60 |
| Decision tree | 0.86 | 0.74 | 0.75 | 0.85 |
| Random forest | 0.84 | 0.84 | 0.80 | 0.92 |
| XGBoost | 0.87 | 0.87 | 0.77 | 0.91 |
| LightGBM | 0.88 | 0.80 | 0.82 | 0.92 |
| Ensemble model | 0.89 | 0.89 | 0.88 | 0.93 |

<https://doi.org/10.1371/journal.pone.0280606.t002>

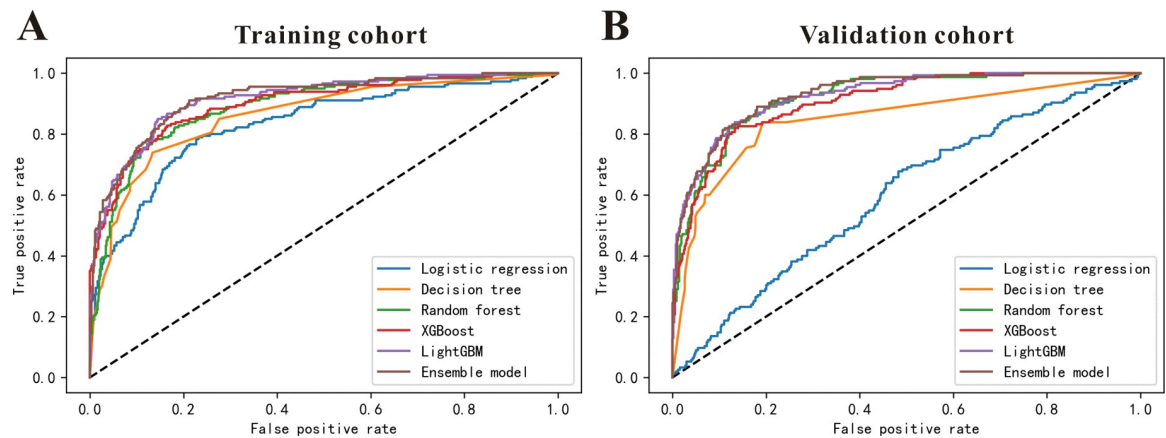


Fig 2. The performance of the six in-hospital mortality predictive models. ROC curves of the six prediction models using all features for predicting in-hospital mortality (A) in training cohort and (B) in the validation cohort.

<https://doi.org/10.1371/journal.pone.0280606.g002>

model predicted for this patient is -1.23, which is less than 0, thus successfully predicting the survival patient.

Discussion

In this retrospective study, we developed and validated machine learning algorithms based on clinical features based on largely public database MIMIC-IV and eICU-CRD, to predict in-hospital mortality of critically ill lung cancer patients. The lightGBM model exhibited the best performance for single model prediction, whereas the RF + ensemble model an ensemble model was constructed, which applied staking strategy using random forest, LightGBM and XGBoost exhibited the greatest AUC among the models we tested. Using advanced machine learning techniques, we could identify some important clinical features associated with in-hospital mortality such as SOFA score, anion gap, albumin, OASIS score and acute respiratory failure. These results have some implications and require further consideration.

ICU-related in-hospital mortality for lung cancer is ranked highest among the solid tumors and the in-hospital mortality in lung cancer patients admitted to ICU is discrepancy according to the lung cancer stage. Previous studies reported that the ICU mortality of extensive or advanced lung cancer patients over 50%. Park et al. investigated patients in Korea who had been newly diagnosed with lung cancer between 2008 and 2010 and indicated that the in-hospital mortality was 58.3% in those advanced critically ill lung cancer patients [21]. In addition, Song et al. analyzed the advanced lung cancer patients, including stage IIIB or IV non-small cell lung cancer and extensive-stage small cell lung cancer, admitted to the ICU and found before and after 2011, the in-hospital mortality was 82.4% and 65.9% [22]. In this study, our result manifested a similar result to Adam et al. [23] report a 20% in-hospital mortality rate in stage I non-small cell lung cancer. This maybe due to the vast majority of the type of the lung cancer were primary but not metastatic, so the in-hospital mortality in the present study is lower than those with advanced critically ill lung cancer patients. Unfortunately, it is difficult for clinicians to identify patients at high risk of in-hospital death in the ICU. Therefore, developing and promoting reliable prediction models is particularly urgent for identifying these patients and providing them with timely and effective interventions to improve their prognosis.

Currently, given the increasing applicability and effectiveness of supervised machine learning algorithms in predictive disease modeling, the breadth of research seems to progress

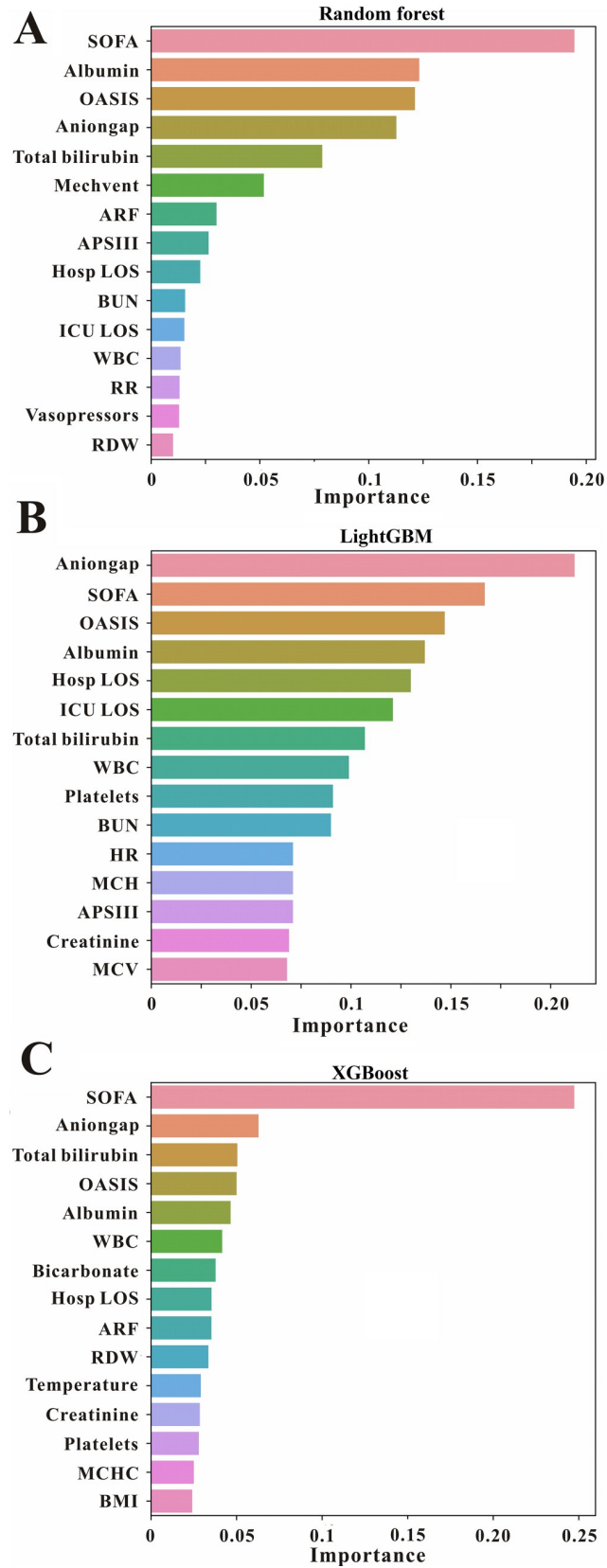


Fig 3. The important features of different models. The top 15 features derived from (A) random forest, (B) lightGBM, and (C) XGBoost model.

<https://doi.org/10.1371/journal.pone.0280606.g003>

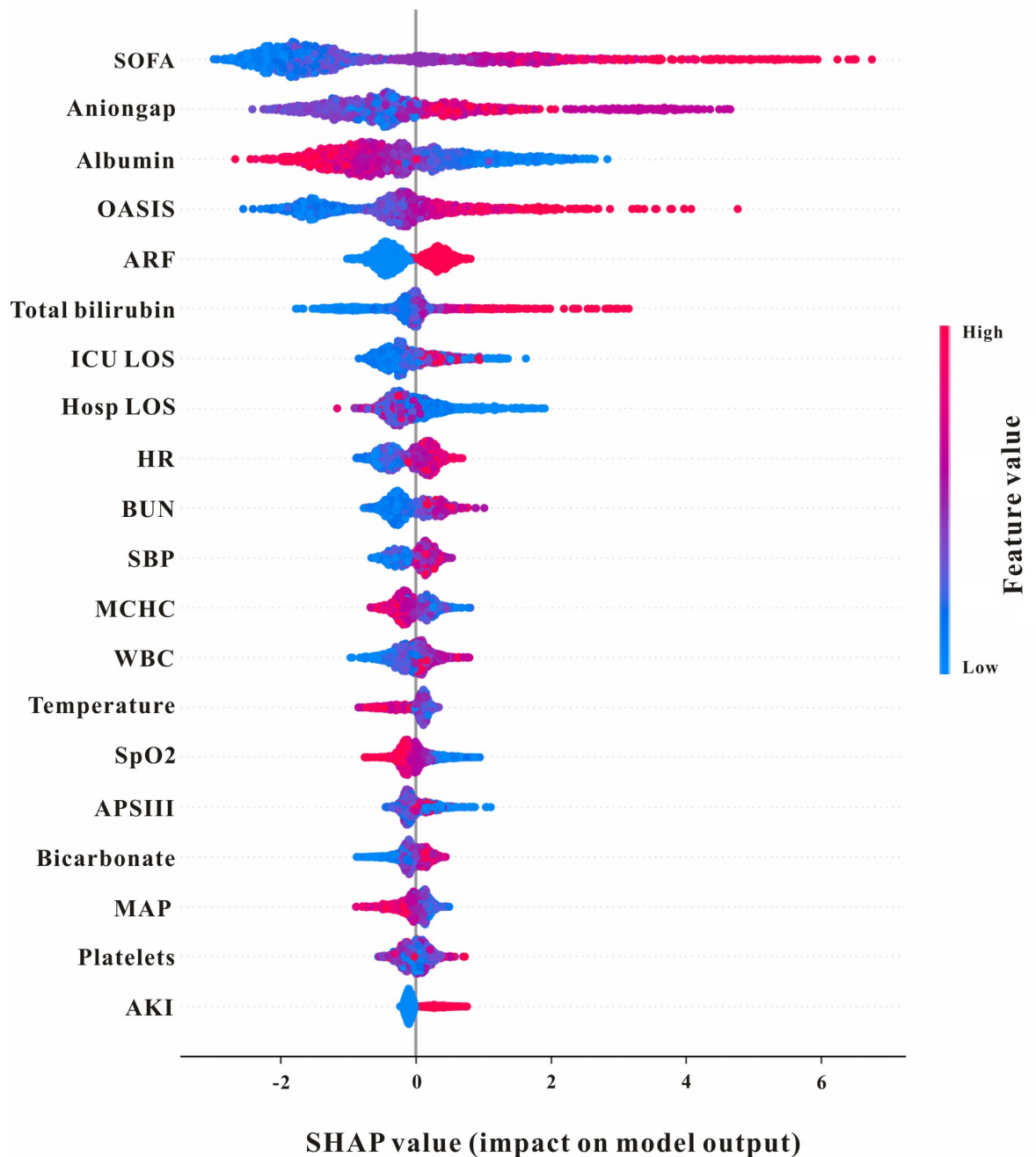


Fig 4. SHAP summary plot of the features of the XGBoost model. The higher the SHAP value of a feature, the higher the probability of in-hospital mortality development. A dot is created for each feature attribution value for the model of each patient, and thus one patient is allocated one dot on the line for each feature. Dots are colored according to the values of features for the respective patient and accumulate vertically to depict density. Red represents higher feature values, and blue represents lower feature values.

<https://doi.org/10.1371/journal.pone.0280606.g004>

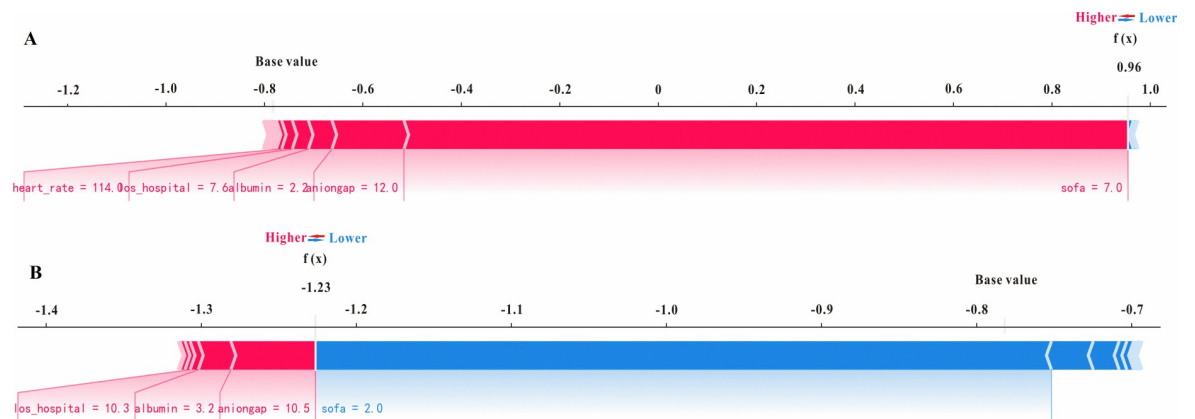


Fig 5. The SHAP force plots. The two representative SHAP force plots of a (A) dead and (B) survival patient. SHAP force plots are effective in interpreting the prediction value of the model in critical instances. The contribution of each feature to the output predicted value is shown with arrows with their force associated with the shapley values. Red arrows indicate features increasing the prediction results (i.e., yield values) to reach the predicted value (output value). Blue arrows show features decreasing the prediction values to reach the same output value. The arrows with positive and negative effects on yield values compensate on a point which is the prediction (output) value.

<https://doi.org/10.1371/journal.pone.0280606.g005>

[24, 25]. The well-known supervised learning classifiers, including support vector machine, random forest, convolutional neural network, and decision tree, have been gradually applied to clinical practice [26, 27]. With the help of machine learning classification, it showed that the machine learning-assisted decision-support model has more advantages than the traditional linear regression model. In this study, we used six different machine learning methods (logistic regression, decision tree, random forest, LightGBM, XGBoost, and ensemble models) to build predictive models. Four popular metrics (ROC, F1 score, accuracy and recall) were used to evaluate the performance of these algorithms. There is no doubt that the results showed that the ensemble model (which combined random forest, LightGBM and XGBoost) achieved the best performance and predictive stability, which was consistent with previous reported [18]. Apart from this, lightGBM model achieved the best predictive performance. The lightGBM modeling is a novel technique that has been widely adopted in tumors survival prediction but not been widely adopted in critical care research [28, 29]. Otaguro et al. evaluated data from patients who underwent intubation for respiratory failure and received mechanical ventilation in ICU and use three learning algorithms (Random Forest, XGBoost, and LightGBM) to predict successful extubation, the result demonstrated that lightGBM exhibited the best overall performance [30]. Moreover, Yang et al. adopted nine machine learning models to predict in-hospital mortality in critically ill patients with hypertension and found that among nine machine learning models, the lightGBM model had the best predictive ability [31].

We employed visualization function in SHAP to find the effect of the specific value of each variable on model output. There are some factors contributing most including SOFA score, anion gap, albumin and so on. SOFA score is an useful tool to quantify the degree of organ dysfunction or failure present on ICU admission which has been widely used for in-hospital mortality prediction in the ICU settings [32–35]. And SOFA score was reported to exhibit better performance than other score systems in predicting infection-related in-hospital mortality in ICU patients, the higher the SOFA score, the higher the risk of in-hospital mortality [36]. Anion gap (AG) is commonly used to classify acid-base disorders and to diagnose various conditions. Recently, AG has been reported to associated with in-hospital mortality in ICU patients. Hu et al. indicated that AG was related to in-hospital mortality in intensive care patients with sepsis [37]. Moreover, Chen et al. demonstrated that AG could significantly

predict ICU mortality for aortic aneurysm patients [38]. Hypoalbuminemia is almost associated with worse prognosis. And low albumin level was usually related to higher risk of in-hospital mortality in ICU settings [39]. Moreover, SHAP force plots of a dead and a survival patient (Fig 5) were selected to further verify the effect of features on the prediction ability and the results further confirmed the SOFA score, anion gap, albumin, etc. features have positive or negative effect on the output of these predictive models.

We should acknowledge some limitations of this research. First, the retrospective and observational nature of our study may lead to inevitable selection bias. Second, the data used in this study were based on public databases MIMIC-IV and eICU-CRD, an external validation is required to prevent overfitting. Third, the data did not include any information on the pathologic and radiologic finding of lung cancer. We could not differentiate between small cell carcinoma and non-small cell carcinoma, the algorithm model is skewed because important medical information about molecular diagnosis.

Conclusions

In the present study, we applied six machine learning methods to predict in-hospital mortality in critically ill lung cancer patients. We demonstrated that the ensemble model achieved the best predictive performance and the lightGBM model exhibited the best performance for single model prediction. And the SOFA score, anion gap and albumin are the most important factors which impacted on the output of the machine learning models in predicting in-hospital mortality of critically ill patients with lung cancer. Our study obtained clinical feature interpretations to provide clinicians in ICU with some information for reference in clinical prognosis prediction.

Declarations

Ethics approval and consent to participate

The study was ethically approved by an affiliated of the Massachusetts Institute of Technology (No.27653720). All patients-related information in the database is anonymous, so there is no need to obtain the informed consent of the patients. This study is described in conformity to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement, and was managed to conform to the tenets of the Declarations of Helsinki.

Supporting information

S1 Fig. The web-based training courses.

(TIF)

S2 Fig. The importance of all features derived from random forest in the validation set.

(TIF)

S3 Fig. The importance of all features derived from lightGBM in the validation set.

(TIF)

S4 Fig. The importance of all features derived from XGBoost in the validation set.

(TIF)

Author Contributions

Conceptualization: Lili Yuan.

Data curation: Tianzhi Huang, Dejin Le.

Formal analysis: Dejin Le, Lili Yuan.

Methodology: Dejin Le.

Project administration: Shoujia Xu, Xiulan Peng.

Software: Lili Yuan.

Supervision: Xiulan Peng.

Validation: Shoujia Xu, Xiulan Peng.

Writing – original draft: Tianzhi Huang.

Writing – review & editing: Shoujia Xu.

References

1. Ferlay J, Colombet M, Soerjomataram I, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer*. 2019; 144:1941–53. <https://doi.org/10.1002/ijc.31937> PMID: 30350310
2. Bade BC, Dela Cruz CS. Lung Cancer 2020: Epidemiology, Etiology, and Prevention. *Clin Chest Med*. 2020; 41(1):1–24. <https://doi.org/10.1016/j.ccm.2019.10.001> PMID: 32008623
3. Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. *Lancet*. 2021; 398(10299):535–554. [https://doi.org/10.1016/S0140-6736\(21\)00312-3](https://doi.org/10.1016/S0140-6736(21)00312-3) PMID: 34273294
4. Barta JA, Powell CA, Wisnivesky JP. Global Epidemiology of Lung Cancer. *Ann Glob Health*. 2019; 85(1):8. <https://doi.org/10.5334/aogh.2419> PMID: 30741509
5. Mattiuzzi C, Lippi G. Current Cancer Epidemiology. *J Epidemiol Glob Health*. 2019; 9(4):217–222. <https://doi.org/10.2991/jegh.k.191008.001> PMID: 31854162
6. Jemal A, Chu KC, Tarone RE. Recent trends in lung cancer mortality in the United States. *J Natl Cancer Inst* 2001; 93:277–283. <https://doi.org/10.1093/jnci/93.4.277> PMID: 11181774
7. Jemal A, Siegel R, Xu J, et al. Cancer statistics, 2010. *CA Cancer J Clin* 2010; 60:277–300. <https://doi.org/10.3322/caac.20073> PMID: 20610543
8. Soubani AO, Ruckdeschel JC. The outcome of medical intensive care for lung cancer patients: the case for optimism. *J Thorac Oncol*. 2011; 6(3):633–8. <https://doi.org/10.1097/JTO.0b013e318200f9eb> PMID: 21266923
9. Lai CC, Ho CH, Chen CM, Chiang SR, Chao CM, Liu WL, et al. Risk factors and mortality of adults with lung cancer admitted to the intensive care unit. *J Thorac Dis*. 2018; 10(7):4118–4126. <https://doi.org/10.21037/jtd.2018.06.165> PMID: 30174856
10. Reichner CA, Thompson JA, O'Brien S, Kuru T, Anderson ED. Outcome and code status of lung cancer patients admitted to the medical ICU. *Chest*. 2006; 130(3):719–23. <https://doi.org/10.1378/chest.130.3.719> PMID: 16963668
11. Gould MK, Huang BZ, Tammemagi MC, Kinar Y, Shiff R. Machine Learning for Early Lung Cancer Identification Using Routine Clinical and Laboratory Data. *Am J Respir Crit Care Med*. 2021; 204(4):445–453. <https://doi.org/10.1164/rccm.202007-2791OC> PMID: 33823116
12. Duan S, Cao H, Liu H, Miao L, Wang J, Zhou X, et al. Development of a machine learning-based multi-mode diagnosis system for lung cancer. *Aging (Albany NY)*. 2020; 12(10):9840–9854. <https://doi.org/10.18632/aging.103249> PMID: 32445550
13. Li D, Li Z, Ding M, Ni R, Wang J, Qu L, et al. Comparative analysis of three data mining techniques in diagnosis of lung cancer. *Eur J Cancer Prev*. 2021; 30(1):15–20. <https://doi.org/10.1097/CEJ.0000000000000598> PMID: 32868638
14. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci Data*. 2018; 5:180178. <https://doi.org/10.1038/sdata.2018.178> PMID: 30204154
15. Johnson A, Bulgarelli L, Pollard T, Horng S, Celi L A, Mark R. MIMIC-IV (version 1.0). *PhysioNet*. 2021. Available from: <https://doi.org/10.13026/s6n6-xd98>.
16. Kellum JA, Lameire N. Diagnosis, evaluation, and management of acute kidney injury: a KDIGO summary (Part 1). *Crit Care*. 2013; 17:204. <https://doi.org/10.1186/cc11454> PMID: 23394211

17. Bonomi MR, Smith CB, Mhango G, Wisnivesky JP. Outcomes of elderly patients with stage IIIB-IV non-small cell lung cancer admitted to the intensive care unit. *Lung Cancer*. 2012; 77:600–604. <https://doi.org/10.1016/j.lungcan.2012.05.103> PMID: 22709929
18. Gao W, Wang J, Zhou L, Luo Q, Lao Y, Lyu H, et al. Prediction of acute kidney injury in ICU with gradient boosting decision tree algorithms. *Comput Biol Med*. 2021; 140:105097. <https://doi.org/10.1016/j.combiomed.2021.105097> PMID: 34864304
19. Linden A. Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis. *J Eval Clin Pract*. 2006; 12(2):132–9. <https://doi.org/10.1111/j.1365-2753.2005.00598.x> PMID: 16579821
20. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neur In*. 2017; 30:4755–4774.
21. Park J, Kim WJ, Hong JY, Hong Y. Clinical outcomes in patients with lung cancer admitted to intensive care units. *Ann Transl Med*. 2021; 9(10):836. <https://doi.org/10.21037/atm-21-298> PMID: 34164470
22. Song JH, Kim S, Lee HW, Lee YJ, Kim MJ, Park JS, et al. Effect of intensivist involvement on clinical outcomes in patients with advanced lung cancer admitted to the intensive care unit. *PLoS One*. 2019; 14(2):e0210951. <https://doi.org/10.1371/journal.pone.0210951> PMID: 30759088
23. Adam AK, Soubani AO. Outcome and prognostic factors of lung cancer patients admitted to the medical intensive care unit. *Eur Respir J*. 2008; 31(1):47–53. <https://doi.org/10.1183/09031936.00031607> PMID: 17715168
24. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol*. 2019; 19(1):64. <https://doi.org/10.1186/s12874-019-0681-4> PMID: 30890124
25. Hueman M, Wang H, Liu Z, Henson D, Nguyen C, Park D, et al. Expanding TNM for lung cancer through machine learning. *Thorac Cancer*. 2021; 12(9):1423–1430. <https://doi.org/10.1111/1759-7714.13926> PMID: 33713568
26. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak*. 2019; 19(1):281. <https://doi.org/10.1186/s12911-019-1004-8> PMID: 31864346
27. Dinh A, Miertschin S, Young A, Mohanty SD. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. *BMC Med Inform Decis Mak*. 2019; 19(1):211. <https://doi.org/10.1186/s12911-019-0918-5> PMID: 31694707
28. Osman MH, Mohamed RH, Sarhan HM, Park EJ, Baik SH, Lee KY, et al. Machine Learning Model for Predicting Postoperative Survival of Patients with Colorectal Cancer. *Cancer Res Treat*. 2022; 54(2):517–524. <https://doi.org/10.4143/crt.2021.206> PMID: 34126702
29. Gong X, Zheng B, Xu G, Chen H, Chen C. Application of machine learning approaches to predict the 5-year survival status of patients with esophageal cancer. *J Thorac Dis*. 2021; 13(11):6240–6251. <https://doi.org/10.21037/jtd-21-1107> PMID: 34992804
30. Otaguro T, Tanaka H, Igarashi Y, Tagami T, Masuno T, Yokobori S, et al. Machine Learning for Prediction of Successful Extubation of Mechanical Ventilated Patients in an Intensive Care Unit: A Retrospective Observational Study. *J Nippon Med Sch*. 2021; 88(5):408–417. https://doi.org/10.1272/jnms.JNMS.2021_88-508 PMID: 33692291
31. Yang B, Xu S, Wang D, Chen Y, Zhou Z, Shen C. ACEI/ARB Medication During ICU Stay Decrease All-Cause In-hospital Mortality in Critically Ill Patients With Hypertension: A Retrospective Cohort Study Based on Machine Learning. *Front Cardiovasc Med*. 2022; 8:787740. <https://doi.org/10.3389/fcvm.2021.787740> PMID: 35097006
32. Ferreira FL, Bota DP, Bross A, Mélot C, Vincent JL. Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA*. 2001; 286(14):1754–8. <https://doi.org/10.1001/jama.286.14.1754> PMID: 11594901
33. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med*. 1996; 22(7):707–10. <https://doi.org/10.1007/BF01709751> PMID: 8844239
34. Vincent JL, de Mendonça A, Cantraine F, Moreno R, Takala J, Suter PM, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine. *Crit Care Med*. 1998; 26(11):1793–800. <https://doi.org/10.1097/00003246-199811000-00016> PMID: 9824069
35. Cárdenas-Turanzas M, Ensor J, Wakefield C, Zhang K, Wallace SK, Price KJ, et al. Cross-validation of a Sequential Organ Failure Assessment score-based model to predict mortality in patients with cancer admitted to the intensive care unit. *J Crit Care*. 2012; 27(6):673–80. <https://doi.org/10.1016/j.jcrc.2012.04.018> PMID: 22762932

36. Raith EP, Udy AA, Bailey M, McGloughlin S, MacIsaac C, Bellomo R, et al. Australian and New Zealand Intensive Care Society (ANZICS) Centre for Outcomes and Resource Evaluation (CORE). Prognostic Accuracy of the SOFA Score, SIRS Criteria, and qSOFA Score for In-Hospital Mortality Among Adults With Suspected Infection Admitted to the Intensive Care Unit. *JAMA*. 2017; 317(3):290–300. <https://doi.org/10.1001/jama.2016.20328> PMID: 28114553
37. Hu T, Zhang Z, Jiang Y. Albumin corrected anion gap for predicting in-hospital mortality among intensive care patients with sepsis: A retrospective propensity score matching analysis. *Clin Chim Acta*. 2021; 521:272–277. <https://doi.org/10.1016/j.cca.2021.07.021> PMID: 34303712
38. Chen Q, Chen Q, Li L, Lin X, Chang SI, Li Y, et al. Serum anion gap on admission predicts intensive care unit mortality in patients with aortic aneurysm. *Exp Ther Med*. 2018; 16(3):1766–1777. <https://doi.org/10.3892/etm.2018.6391> PMID: 30186400
39. Padkins M, Breen T, Anavekar N, Barsness G, Kashani K, Jentzer JC. Association Between Albumin Level and Mortality Among Cardiac Intensive Care Unit Patients. *J Intensive Care Med*. 2021; 36(12):1475–1482. <https://doi.org/10.1177/0885066620963875> PMID: 33016174