Early Prediction of Ventilator-associated Pneumonia in Critical Care Patients: a Machine Learning Model

Yingjian Liang  
China Medical University Hospital

Chengrui Zhu  
The First Hospital of China Medical University: The First Affiliated Hospital of China Medical University

Cong Tian  
Philips Research

Qizhong Lin  
Philips Research

Zhiliang Li  
China Medical University Hospital

Zhifei Li  
China Medical University Hospital

Dongshu Ni  
China Medical University Hospital

Xiaochun Ma (cmu1hicu2002@sina.com)  
China Medical University

Research

Keywords: ventilator-associated pneumonia, MIMIC database, risk factors, predictive modeling

DOI: https://doi.org/10.21203/rs.3.rs-129953/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background: This study was performed to develop and validate machine learning models for the early detection of ventilator-associated pneumonia (VAP) in patients 24 h before the diagnosis that enables VAP patients to receive early intervention and reduces the occurrence of complications.

Patients and Methods: This study was based on the MIMIC-III dataset, which was a retrospective cohort. The random forest algorithm was applied to construct a base classifier, and the area under the receiver operating characteristic (ROC) curve (AUC), sensitivity and specificity of the prediction model were evaluated. Meanwhile, a Clinical Pulmonary Infection Score (CPIS)-based model (threshold value ≥ 3) using the same training and test data set was used as the control model.

Results: A total of 38,515 ventilation durations occurred in 61,532 ICU admissions. VAP occurred in 212 of these durations. We incorporated 42 VAP risk factors on admission and routinely measured vital characteristics and laboratory results. Five-fold cross-validation was performed to evaluate the model performance, and the model achieved an AUC of 84.4%±1.7% on validation, 74.3%±2.5% sensitivity and 70.76%±1.2% specificity 24 h before the gold standard time (at least 48 h after ventilation). Our VAP machine learning model improved the AUC of the CPIS-based model by almost 25%, and the sensitivity and specificity were also improved by almost 14% and 15%, respectively.

Conclusions: We developed and internally validated an automated model of VAP prediction in the MIMIC-III cohort. The VAP prediction model achieved high performance for AUC, sensitivity and specificity, and its performance was superior to that of the CPIS model. External validation and prospective interventional or outcome studies using this prediction model are envisioned as future work.

Background

Ventilator-associated pneumonia (VAP) is the most frequently occurring nosocomial pneumonia in critically ill patients with mechanical ventilation(1). Despite advances in the understanding of the cause of pneumonia and in the implementation of preventive measures, the rate of VAP has not declined(2). Recently, published data have shown that the incidence of VAP is 10%. The occurrence of VAP not only prolongs ventilator support but also prolongs stays in intensive care units (ICUs) and hospitals, which increases healthcare costs and results in a poorer prognosis(3–5). Therefore, in the United States and China, VAP was proposed as a quality-of-care indicator for ICUs.

Studies have shown that some risk factors are associated with VAP. Some risk factors were patient-specific factors, such as age and pre-existing disease (chronic obstructive pulmonary disease (COPD) or Glasgow coma score of 9 or less)(6–8). Others were care-related factors, such as head-of-the-bed angle, emergency intubation, aspiration, previous antibiotic treatment, and reintubation(6, 7, 9). Therefore, the early recognition of patients with a high risk for developing VAP and the subsequent prevention of its progression is of great value in critical care units.
Intensivists have been working on a VAP risk prediction model for several years. There are several available prediction models used to predict mortality in VAP patients. The models of the Acute Physiology and Chronic Health Evaluation II (APACHE II) (10); immunodeficiency, blood pressure, multilobar infiltrates on chest radiography, platelets and hospitalization 10 days before onset of VAP(11); Clinical Pulmonary Infection Score (CPIS)(12); and VAP predisposition, insult response and organ dysfunction(13) all have the ability to predict VAP mortality. However, there is no early risk prediction model for VAP.

The Medical Information Mart for Intensive Care (MIMIC)-III dataset is an open, large, single-center database that researchers around the world can use freely. It has been widely used in the development of predictive models, epidemiological studies, and educational courses. The aim of this study was to use the MIMIC database to develop and validate machine learning models for the early discrimination of patients at high risk of VAP 24 h before diagnosis and to assess its prognostic accuracy. Meanwhile, a CPIS-based model (threshold value ≥ 3) under the same training and test data set was used as a control model.

**Patients And Methods**

**Datasets**

The MIMIC-III database was used to train, validate and test models and comprises deidentified health-related data associated with 61,532 ICU stays in multiple critical care units of Beth Israel Deaconess Medical Center between 2001 and 2012(14). This database is a publicly available database constructed in compliance with the Health Insurance Portability and Accountability Act. The study protocol was approved by the ethics committee of the First Hospital of China Medical University (No. 2019-197-2).

**Data annotation and extraction**

A total of 38,515 ventilation sessions were identified from the MIMIC-III database and filtered according to the patient inclusion process depicted in Fig. 1. A total of 10,431 patients with mechanical ventilation longer than 24 h who were more than 18 years old were included in this study. Ventilation sessions, which were over 48 h and with pneumonia after 48 h of ventilation, were annotated as VAP according to the VAP definition(15). Others were grouped into non-VAP sessions. When VAP was diagnosed, the presence of infection at other sites was recorded.

To detect the risk of VAP early, a set of 42 variables (features) were extracted from the MIMIC-III dataset, including the worst value of the partial pressure of the arterial oxygen/fraction of inspired oxygen (PaO$_2$/FiO$_2$) ratio, white blood cell count (WBC), and body temperature in the first 24 h after ventilation, the worst value of the APACHE III and its subcomponents and the sequential organ failure assessment (SOFA) and its subcomponents in the first 24 h after admission to the ICU, age, sex, admission source (medical intensive care unit (MICU) = 1, others = 0) and type (emergency = 1, elective = 0), reintubation, aspiration, trauma/polytrauma, and pre-existing diseases (detailed information on the 42 variables is provided in Table 1).
Table 1
List of 42 variables and missing value in study cohort

| Variable                        | Overall (n = 10431) | VAP group (n = 212) | Non-VAP group (n = 10219) | p value |
|---------------------------------|---------------------|---------------------|---------------------------|---------|
| PaO$_2$/FiO$_2$                 | 876 (8.4%)          | 17 (8.02%)          | 859 (8.41%)               | 1.0     |
| WBC                             | 128 (1.23%)         | 3 (1.42%)           | 125 (1.22%)               | 0.746   |
| Body temperature                | 213 (2.04%)         | 0 (0.0%)            | 213 (2.08%)               | 0.024   |
| APACHE III                      | 0 (0.0%)            | 0 (0.0%)            | 0 (0.0%)                  | 1.0     |
| HR score                        | 47 (0.45%)          | 2 (0.94%)           | 45 (0.44%)                | 0.248   |
| MAP score                       | 49 (0.47%)          | 2 (0.94%)           | 47 (0.46%)                | 0.263   |
| Temperature score               | 257 (2.46%)         | 2 (0.94%)           | 255 (2.5%)                | 0.181   |
| RR score                        | 47 (0.45%)          | 2 (0.94%)           | 45 (0.44%)                | 0.248   |
| A-aDO$_2$/PaO$_2$ score         | 7124 (68.3%)        | 170 (80.19%)        | 6954 (68.05%)             | <0.001  |
| Hematocrit score                | 27 (0.26%)          | 0 (0.0%)            | 27 (0.26%)                | 1.0     |
| WBC score                       | 54 (0.52%)          | 0 (0.0%)            | 54 (0.53%)                | 0.629   |
| Creatinine score                | 19 (0.18%)          | 0 (0.0%)            | 19 (0.19%)                | 1.0     |
| UO score                        | 573 (5.49%)         | 20 (9.43%)          | 553 (5.41%)               | 0.021   |
| BUN score                       | 19 (0.18%)          | 0 (0.0%)            | 19 (0.19%)                | 1.0     |
| Sodium score                    | 19 (0.18%)          | 0 (0.0%)            | 19 (0.19%)                | 1.0     |
| ALB score                       | 5069 (48.6%)        | 80 (37.74%)         | 4989 (48.82%)             | 0.001   |
| Bilirubin score                 | 4153 (39.81%)       | 74 (34.91%)         | 4079 (39.93%)             | 0.156   |
| Glucose score                   | 9 (0.09%)           | 0 (0.0%)            | 9 (0.09%)                 | 1.0     |
| Acid-base score                 | 735 (7.05%)         | 6 (2.83%)           | 729 (7.13%)               | 0.014   |
| GCS score                       | 210 (2.01%)         | 5 (2.36%)           | 205 (2.01%)               | 0.62    |
| SOFA                            | 0 (0.0%)            | 0 (0.0%)            | 0 (0.0%)                  | 1.0     |
| Respiration sofa                | 1500 (14.38%)       | 15 (7.08%)          | 1485 (14.53%)             | 0.001   |
| Coagulation sofa                | 38 (0.36%)          | 0 (0.0%)            | 38 (0.37%)                | 1.0     |
| Liver sofa                      | 4153 (39.81%)       | 74 (34.91%)         | 4079 (39.92%)             | 0.156   |
| Cardiovascular sofa             | 49 (0.47%)          | 2 (0.94%)           | 47 (0.46%)                | 0.263   |
| Abbreviations | Overall (n = 10431) | VAP group (n = 212) | Non-VAP group (n = 10219) | p value |
|---------------|---------------------|--------------------|---------------------------|---------|
| CNS sofa      | 54 (0.52%)          | 3 (1.42%)          | 51 (0.5%)                 | 0.097   |
| Renal sofa    | 7 (0.07%)           | 0 (0.0%)           | 7 (0.07%)                 | 1.0     |
| Age           | 0 (0.0%)            | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Gender        | 0 (0.0%)            | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Admission source | 0 (0.0%)          | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Admission type | 0 (0.0%)           | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Reintubation  | 0 (0.0%)            | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| COPD          | 0 (0.0%)            | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Diabetes      | 0 (0.0%)            | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Pneumothorax adm | 0 (0.0%)          | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Coma adm      | 0 (0.0%)            | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Sepsis adm    | 0 (0.0%)            | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Bacteremia adm | 0 (0.0%)           | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Hypertension  | 0 (0.0%)            | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Renal failure | 0 (0.0%)            | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Liver failure | 0 (0.0%)            | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Trauma adm    | 0 (0.0%)            | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Polytrauma adm | 0 (0.0%)           | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Aspiration adm | 0 (0.0%)           | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Fracture adm  | 0 (0.0%)            | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Solid tumor   | 0 (0.0%)            | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |
| Metastatic tumor | 0 (0.0%)        | 0 (0.0%)           | 0 (0.0%)                  | 1.0     |

*Abbreviations:* PaO$_2$/FiO$_2$ the partial pressure of arterial oxygen/ fraction of inspired oxygen, WBC white blood cell count, APACHE III Acute Physiology and Chronic Health Evaluation III, HR heart rate, MAP mean arterial pressure, RR respiratory rate, A-aDO$_2$/PaO$_2$ pulmonary alveolus-arterial difference of oxygen pressure/ partial pressure of oxygen, UO urine output, BUN blood urea nitrogen, ALB albumin, GCS Glasgow Coma Scale, SOFA sequential organ failure assessment, CNS central nervous system, COPD chronic obstructive pulmonary disease, adm admission
Data splitting and sampling

Figure 2 describes the pipeline of model training, validation and testing. The included dataset was split into a training dataset and test dataset for five-fold cross-validation, in which four folds were used as the training dataset and the remaining fold was used as the test dataset, with the folds being mutually exclusive. To identify the optimal hyperparameter of the model, two-fold cross-validation was performed on the training dataset, and then the model was retrained using the optimal hyperparameter on the entire training dataset to learn the model parameters. Due to an extreme imbalance between the number of non-VAP and VAP patients, the negative dataset was divided into 100 subgroups for resampling. Stratified sampling was used to ensure an even class distribution.

Data preprocessing

Figure S1 shows the data preprocessing steps. For numeric variables, if a patient did not have a measurement, the missing value was filled by using the median interpolation of the whole cohort (Table 1 shows the count and percentage of missing data in the VAP group and non-VAP group; Fisher’s exact test was used to test the significance). For categorical variables with $d$ categories, raw data were mapped to a $d$-dimensional vector, where each dimension corresponded to a different category, with the exception that categorical variables with two categories (e.g., sex=\{F, M\}) were sufficiently mapped to \{0, 1\}. Next, both numeric and categorical data were normalized for the training dataset, which required min-max feature scaling to adjust variable values measured on different scales.

Model development and performance measurement

Since non-VAP instances were much higher than VAP instances, we split the non-VAP instances into 100 subgroups with mutual exclusivity. One subgroup of non-VAP instances was combined with the VAP dataset to train one model; then, 100 models were combined based on the performance average or major voting as the final model. The ensemble method was applied on 100 subgroups of non-VAP instances in combination with VAP instances as shown in Fig. 3.

The random forest algorithm was applied to construct the classifier, which was widely used as a classifier for clinical datasets. The area under the receiver operating characteristic (ROC) curve (AUC), accuracy, sensitivity and specificity of the prediction model were evaluated. Meanwhile, we used a CPIS-based model (threshold $\geq 3$) for the early detection of VAP as the benchmark model to compare with our machine learning model, and the performance of the classification model was evaluated using the same training and test datasets. The performance is described as the mean $\pm$ SD to indicate the performance distribution of subgroups, and the SD was used to determine whether any overfitting of the model occurred in certain datasets.

The Bayes search method was applied to fine-tune the hyperparameters of the base classifier on the validation set. In the random forest classifier, the optimal number of estimators for the hyperparameter was adjusted to 104, which was randomly obtained via Bayes search in the range from 1 to 300.
Statistical analysis

For analysis of the clinical characteristics of both the VAP and non-VAP groups, numeric variables are described as medians with interquartile ranges (IQRs; represented by the 25th and 75th percentile values), and categorical variables are described as counts with percentages. To compare the two groups, we used Fisher's exact test for categorical variables and the Mann-Whitney U-test for numeric variables. A $p$-value less than 0.05 was considered statistically significant.

Results

According to the screening criteria in Fig. 1, 38,515 ventilation sessions were included with 212 VAP sessions between 2001 and 2012 in the MIMIC-III cohort, and the incidence density was 2 per 1,000 ventilator-days. The median time on mechanical ventilation from endotracheal intubation to the first VAP episode was 130 h (IQR, 76.3 to 204.1 h). None of these VAP patients had other site infections. The missing counts and percentages of the 42 variables in the overall, VAP, and non-VAP groups are shown in Table 1. Compared with the overall study cohort, the non-VAP group had significantly higher missing albumin (ALB) and acid-base scores for the APACHE III, and respiration scores for SOFA. However, the VAP group had a higher missing percentage of pulmonary alveolus-arterial difference of oxygen pressure/partial pressure of oxygen ($A-a\text{DO}_{2}/\text{PaO}_{2}$) and urine output.

Univariate analysis indicated that compared to that of the control group, the VAP group of the study cohort had a significantly different admission source and type ($p < 0.001$); specifically, the VAP group had a significantly higher ratio of patients from the MICU, and only one VAP patient was not transferred from emergency. Compared to the control group, the VAP group also had significantly higher total APACHE III ($p = 0.031$) and SOFA ($p = 0.034$) scores on admission. Most of the sub items also showed significant differences between the VAP and control groups, except for 1 sub item of the SOFA, the renal score ($p = 0.471$) (see details in Table 2). The worst value of the PaO$_2$/FiO$_2$ ratio in the first 24 h after ventilation was significantly deteriorated ($p < 0.001$) in the VAP group compared with that in the control group. The reintubation ratio was not significantly different ($p = 0.823$) between the VAP group and the non-VAP group, whereas the VAP group demonstrated a significantly higher ratio in aspiration ($p = 0.004$). For the pre-existing diseases, there was no difference between the VAP group and the non-VAP group, except for hypertension.
| Demographic and clinical characteristics of study cohort in MIMIC III | Overall (n = 10431) | VAP group (n = 212) | Non-VAP group (n = 10219) | p value |
|---|---|---|---|---|
| Age | 66.3 (53.1–76.0) | 66.3 (52.1–77.8) | 66.3 (53.1–76.0) | 0.387 |
| Gender (M = 1) | 5937 (56.9%) | 109 (51.42%) | 5828 (57.0%) | 0.107 |
| Admission source (MICU = 1) | 4089.0 (39.2%) | 154 (72.64%) | 3935.0 (38.51%) | <0.001 |
| Admission type (Emergency = 1) | 9504.0 (91.11%) | 211 (99.53%) | 9293.0 (90.94%) | <0.001 |
| APACHE III | 50.0 (37.0–66.0) | 53.0 (40.75–64.0) | 50.0 (37.0–66.0) | 0.031 |
| SOFA | 6.0 (4.0–9.0) | 6.0 (4.0–8.0) | 6.0 (4.0–9.0) | 0.034 |
| Reintubation | 3294.0 (31.58%) | 65.0 (30.66%) | 3229.0 (30.6%) | 0.823 |
| Aspiration adm | 32.0 (0.31%) | 4.0 (1.89%) | 28.0 (0.27%) | 0.004 |
| PaO$_2$/FiO$_2$ | 240.0 (178.5–315.37) | 194.17 (149.98–256.5) | 241.0 (179.37–316.67) | <0.001 |
| WBC | 12.8 (9.2–17.7) | 13.2 (9.8–18.0) | 12.8 (9.2–17.7) | 0.121 |
| Body temperature | 37.83 (37.33–38.44) | 37.92 (37.28–38.57) | 37.83 (37.33–38.44) | 0.255 |
| APACHE III HR score | 5.0 (0.0–7.0) | 5.0 (0.0–7.0) | 5.0 (0.0–7.0) | 0.028 |
| MAP score | 15.0 (7.0–15.0) | 15.0 (7.0–15.0) | 15.0 (7.0–15.0) | 0.197 |
| Temperature score | 0.0 (0.0–2.0) | 0.0 (0.0–2.0) | 0.0 (0.0–2.0) | <0.001 |
| RR score | 6.0 (0.0–6.0) | 6.0 (0.0–9.0) | 6.0 (0.0–6.0) | 0.001 |
| A-aDO$_2$/PaO$_2$ score | 0.0 (0.0–0.0) | 0.0 (0.0–0.0) | 0.0 (0.0–0.0) | <0.001 |
| Hematocrit score | 3.0 (3.0–3.0) | 3.0 (3.0–3.0) | 3.0 (3.0–3.0) | 0.269 |
| WBC score | 0.0 (0.0–1.0) | 0.0 (0.0–1.0) | 0.0 (0.0–1.0) | 0.083 |
| Creatinine score | 0.0 (0.0–7.0) | 0.0 (0.0–7.0) | 0.0 (0.0–7.0) | 0.484 |
| UO score | 5.0 (0.0–7.0) | 5.0 (0.0–7.0) | 5.0 (0.0–7.0) | 0.013 |
| BUN score | 7.0 (2.0–11.0) | 7.0 (2.0–11.0) | 7.0 (2.0–11.0) | 0.094 |
|                               | Overall (n = 10431) | VAP group (n = 212) | Non-VAP group (n = 10219) | p value |
|-------------------------------|----------------------|---------------------|---------------------------|---------|
| Sodium score                 | 0.0 (0.0–2.0)        | 0.0 (0.0–2.0)       | 0.0 (0.0–2.0)             | 0.016   |
| ALB score                    | 0.0 (0.0–6.0)        | 0.0 (0.0–6.0)       | 0.0 (0.0–6.0)             | 0.087   |
| Bilirubin score              | 0.0 (0.0–0.0)        | 0.0 (0.0–0.0)       | 0.0 (0.0–0.0)             | 0.008   |
| Glucose score                | 0.0 (0.0–3.0)        | 3.0 (0.0–3.0)       | 0.0 (0.0–3.0)             | 0.017   |
| Acid-base score              | 3.0 (1.0–6.0)        | 3.0 (1.0–5.0)       | 3.0 (1.0–6.0)             | 0.077   |
| GCS score                    | 0.0 (0.0–0.0)        | 0.0 (0.0–0.0)       | 0.0 (0.0–0.0)             | 0.017   |
| **SOFA**                     |                      |                     |                           |         |
| Respiration sofa             | 3.0 (0.0–3.0)        | 3.0 (2.0–3.0)       | 3.0 (0.0–3.0)             | 0.014   |
| Coagulation sofa             | 0.0 (0.0–1.0)        | 0.0 (0.0–1.0)       | 0.0 (0.0–1.0)             | < 0.001 |
| Liver sofa                   | 0.0 (0.0–1.0)        | 0.0 (0.0–0.0)       | 0.0 (0.0–1.0)             | 0.004   |
| Cardiovascular sofa          | 1.0 (1.0–3.0)        | 1.0 (1.0–3.0)       | 1.0 (1.0–3.0)             | 0.022   |
| CNS sofa                     | 0.0 (0.0–1.0)        | 0.0 (0.0–0.0)       | 0.0 (0.0–1.0)             | 0.008   |
| Renal sofa                   | 1.0 (0.0–2.0)        | 1.0 (0.0–2.0)       | 1.0 (0.0–2.0)             | 0.471   |
| **Pre-existing Diseases**    |                      |                     |                           |         |
| COPD                         | 101.0 (0.97%)        | 2.0 (0.94%)         | 99.0 (0.97%)              | 1.0     |
| Diabetes                     | 2698.0 (25.87%)      | 60.0 (28.3%)        | 2638.0 (25.81%)           | 0.428   |
| Hypertension                 | 4943.0 (47.39%)      | 80.0 (37.74%)       | 4863.0 (47.59%)           | 0.004   |
| Solid tumor                  | 313.0 (3.0%)         | 8.0 (3.77%)         | 305.0 (2.98%)             | 0.537   |
| Metastatic tumor             | 418.0 (4.01%)        | 5.0 (2.36%)         | 413.0 (4.04%)             | 0.286   |
| Renal failure                | 1721.0 (16.5%)       | 26.0 (12.26%)       | 1695.0 (16.59%)           | 0.111   |
| Liver failure                | 1653.0 (15.85%)      | 22.0 (10.38%)       | 1631.0 (15.96%)           | 0.028   |
| Pneumothorax adm             | 19.0 (0.18%)         | 0.0 (0.0%)          | 19.0 (0.19%)              | 1.0     |
| Coma adm                     | 6.0 (0.06%)          | 0.0 (0.0%)          | 6.0 (0.06%)               | 1.0     |
| Sepsis adm                   | 548.0 (5.25%)        | 10.0 (4.72%)        | 538.0 (5.26%)             | 0.876   |
| Bacteremia adm               | 11.0 (0.11%)         | 0.0 (0.0%)          | 11.0 (0.11%)              | 1.0     |
| Trauma adm                   | 202.0 (1.94%)        | 0.0 (0.0%)          | 202.0 (1.98%)             | 0.037   |
|                  | Overall (n = 10431) | VAP group (n = 212) | Non-VAP group (n = 10219) | p value |
|------------------|---------------------|---------------------|---------------------------|---------|
| Polytrauma adm   | 45.0 (0.43%)        | 0.0 (0.0%)          | 45.0 (0.44%)              | 1.0     |
| Fracture adm     | 27.0 (0.26%)        | 0.0 (0.0%)          | 27.0 (0.26%)              | 1.0     |

**Abbreviations:** MICU medical intensive care unit, APACHE III Acute Physiology and Chronic Health Evaluation III, PaO$_2$/FiO$_2$ the partial pressure of arterial oxygen/fraction of inspired oxygen, WBC white blood cell count, HR heart rate, MAP mean arterial pressure, RR respiratory rate, A-aDO$_2$/PaO$_2$ pulmonary alveolus-arterial difference of oxygen pressure/partial pressure of oxygen, UO urine output, BUN blood urea nitrogen, ALB albumin, GCS Glasgow Coma Scale, SOFA sequential organ failure assessment, CNS central nervous system, COPD chronic obstructive pulmonary disease, adm admission

Figure 4 shows that the AUC of the optimal performance corresponding to the random forest model was 84.4%±1.7% on validation in the pure testing datasets, and the sensitivity and specificity approached 74.3%±2.5% and 70.7.7%±1.2%, respectively. When using the same test datasets, the best performance of the CPIS-based model was AUC = 0.59, sensitivity = 0.6, and specificity = 0.55. Figure 5 shows the feature importance of the optimal random forest model, which indicates the contribution ranking of the features to the prediction value in the model. Admission source, APACHE III and SOFA scores together with their sub items, age, and the worst value of body temperature, PaO$_2$/FiO$_2$ ratio, and WBC in the initial 24 h after ventilation were the top 10 important features, which contributed over 46% in total to the prediction value. Respiration items of the SOFA were the highest contributor to the total SOFA score (4%) which indicated the significance of respiration for organ failure.

**Discussion**

In this retrospective cohort study, we developed and validated a machine learning model for the early detection of VAP in patients 24 h before the diagnosis. The final predictive AUC showed a good performance (AUC: 84.4%, sensitivity: 74.3%, specificity: 70.7%), as an AUC value between 0.75 and 0.92 indicates good diagnostic capability(16). Additionally, our VAP machine learning model improved the AUC of the CPIS-based model by almost 25%, and the sensitivity and specificity were improved by almost 14% and 15%, respectively. Our predictive model can provide risk stratification for VAP patients within independently-defined patient groups. Prevention guidelines have been developed to allow higher-risk patients to benefit from more aggressive strategies or adjuvant therapy (semirecumbent position, oral hygiene). Additionally, a longer prediction lead time will increase the likelihood that a patient can benefit from early intervention.

The CPIS is a method used to diagnose VAP and in timely manner. There are several clinical indicators in the CPIS that describe VAP; therefore, this score could be used as a reference to help physicians provide better and faster treatments for patients. According to the proposed model based on the CPIS for the early detection of VAP, when the threshold was equal to 6, a clear difference could be observed between
the existence and non-existence of pulmonary infection (17). However, in our MIMIC III cohort data, a CPIS score of 6 did not show a good performance similar to that mentioned in the reference article. In contrast, when the CPIS was 6, the CPIS-based model exhibited the worst performance. Thus, different score thresholds were tested to determine the best performance; Figure S2 shows that when score was equal or greater than 3, the CPIS-based model had the best performance.

The typical acute respiratory distress syndrome (ARDS) manifestations include increased pulmonary vascular permeability, pulmonary edema and alveolar trapping, which lead to refractory hypoxia and decreased pulmonary compliance (18). The optimal mechanical ventilation strategy for these patients is to decrease tidal volume and increase positive end-expiratory pressure, which is associated with the highest PaO$_2$/FiO$_2$ ratio (19–21). The relationship between ARDS and subsequent development of VAP is complex. In mechanically-ventilated patients, cyclic stretch of lung cells induces acidification of the milieu, which promotes bacterial growth (22). Injurious mechanical ventilation may promote the lungs to release cytokines (23, 24). In addition, alveolar macrophages and neutrophils exhibit reduced bacterial phagocytosis and killing, thereby affecting lung and systemic antibacterial defenses (23, 25).

We found that APACHE III and SOFA scores greatly contributed to the final predictive model. The APACHE scoring system is used to describe the severity of illness and predict the outcome of critically ill patients. The APACHE II and III are widely employed scores in the ICU (26, 27), and the overall goodness-of-fit of the two predictive models was similar. APACHE III expanded the acute physiology score (APS) project compared to APACHE II, and based on APACHE II, APS added six parameters: blood urea nitrogen (BUN), total bilirubin, blood glucose, ALB, artery CO$_2$ partial pressure (PaCO$_2$) and urine output. These six parameters are more responsive to clinical practice (28, 29). The APACHE II was better at predicting risk for surgical patients and patients with gastrointestinal disease (28), while the APACHE III score was a good predictor of internal medical conditions and nosocomial pneumonia (29, 30).

In our study, the control group included patients with mechanical ventilation for 24 h rather than patients with 48 h of mechanical ventilation. The reasons are as follows: we selected the worst values of the body temperature, PaO$_2$/FiO$_2$ ratio, and WBC during the initial 24 h after ventilation and the worst values of the APACHE III and SOFA scores in the first 24 h after admission to the ICU as VAP predictors. If we had included patients with 48 h of mechanical ventilation in the control group, some non-VAP patients would be missed. The purpose of the model is to predict whether VAP can occur in patients with mechanical ventilation, which is more consistent with our original intention and clinical reality. Additionally, some references support this grouping scheme (31).

The major limitation of this study was the annotation of VAP sessions. We annotated the VAP session by the VAP definition, i.e., ventilation sessions that were over 48 h and with pneumonia after 48 h of ventilation. With this strategy, we could not only identify VAP sessions but also query the recorded time in the chart event table from the MIMIC-III database. The limitation of this annotation procedure was the high false negative rate due to potentially less VAP diagnoses recorded in the chart event table by the nurse. Another protocol of VAP annotation is to use the ICD-9 code in MIMIC-III. However, in MIMIC-III
cohort, since the diagnostic information did not link to the exact diagnosis time, we were unable to query the precise PaO$_2$/FiO$_2$ ratio, WBC, and body temperature variables.

In future work, to overcome the limitation of annotation, we need to define a protocol that will collect information on not only VAP diagnosis but also the charting time of VAP diagnosis. Further, external validation and prospective interventional or outcome studies using this prediction model are envisioned as future work.

**Conclusions**

We developed and internally validated an automated model of VAP prediction in the MIMIC-III cohort. The VAP prediction model achieved high performance for the AUC, sensitivity and specificity, and its performance was superior to that of the CPIS-based model. External validation and prospective interventional or outcome studies using this prediction model are envisioned as future work.

**Abbreviations**

VAP: ventilator-associated pneumonia; ICU: intensive care unit; APACHE III: Acute Physiology and Chronic Health Evaluation III; MIMIC: Medical Information Mart for Intensive Care; SOFA: sequential organ failure assessment; PaO$_2$/FiO$_2$: the partial pressure of arterial oxygen/ fraction of inspired oxygen; WBC: white blood cell count; COPD: chronic obstructive pulmonary disease; ARDS: acute respiratory distress syndrome; CPIS: Clinical Pulmonary Infection Score

**Declarations**

**Ethics approval and consent to participate:**

The study protocol was approved by the ethics committee of the first hospital of China Medical University (no. 2019-197-2).

**Consent for publication:**

Not applicable.

**Availability of data and materials:**

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

**Competing interests:**

Cong Tian and Qizhong Lin are currently employed by Philips Research China. The remaining authors declare that they have no competing interests.
Funding:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' contributions:

Y.L., X.M., and C.Z. participated in the conception of the idea, conduct of the study, and preparation of the manuscript. C.T. and Q.L. provided engineering expertise. ZL.L., ZF.L. and D.N. participated in additional analysis and critical revision. All authors reviewed and approved the submitted form of this manuscript.

Acknowledgements:

Not applicable.

References

1. Urli T., Perone G., Acquarolo A., Zappa S., Antonini B., Ciani A. Surveillance of infections acquired in intensive care: usefulness in clinical practice. J Hosp Infect. 2002;52(2):130-5.

2. Wang Y., Eldridge N., Metesky M. L., Verzier N. R., Meehan T. P., Pandolfi M. M., et al. National trends in patient safety for four common conditions, 2005-2011. N Engl J Med. 2014;370(4):341-51.

3. Kollef M. H., Hamilton C. W., Ernst F. R. Economic impact of ventilator-associated pneumonia in a large matched cohort. Infect Control Hosp Epidemiol. 2012;33(3):250-6.

4. Craven D. E., Lei Y., Ruthazer R., Sarwar A., Hudcova J. Incidence and outcomes of ventilator-associated tracheobronchitis and pneumonia. Am J Med. 2013;126(6):542-9.

5. Mathai A. S., Phillips A., Kaur P., Isaac R. Incidence and attributable costs of ventilator-associated pneumonia (VAP) in a tertiary-level intensive care unit (ICU) in northern India. J Infect Public Health. 2015;8(2):127-35.

6. Akca O., Koltka K., Uzel S., Cakar N., Pembeci K., Sayan M. A., et al. Risk factors for early-onset, ventilator-associated pneumonia in critical care patients: selected multiresistant versus nonresistant bacteria. Anesthesiology. 2000;93(3):638-45.

7. Galal Y. S., Youssef M. R., Ibrahiem S. K. Ventilator-Associated Pneumonia: Incidence, Risk Factors and Outcome in Paediatric Intensive Care Units at Cairo University Hospital. J Clin Diagn Res. 2016;10(6):SC06-11.

8. Tejerina E., Frutos-Vivar F., Restrepo M. I., Anzueto A., Abroug F., Palizas F., et al. Incidence, risk factors, and outcome of ventilator-associated pneumonia. Journal of critical care. 2006;21(1):56-65.

9. Rello J., Allegri C., Rodriguez A., Vidalur L., Sirgo G., Gomez F., et al. Risk factors for ventilator-associated pneumonia by Pseudomonas aeruginosa in presence of recent antibiotic exposure. Anesthesiology. 2006;105(4):709-14.
10. Zhou X. Y., Ben S. Q., Chen H. L., Ni S. S. A comparison of APACHE II and CPIS scores for the prediction of 30-day mortality in patients with ventilator-associated pneumonia. Int J Infect Dis. 2015;30:144-7.

11. Mirsaeidi M., Peyrani P., Ramirez J. A., Improving Medicine through Pathway Assessment of Critical Therapy of Hospital-Acquired Pneumonia Investigators. Predicting mortality in patients with ventilator-associated pneumonia: The APACHE II score versus the new IBMP-10 score. Clin Infect Dis. 2009;49(1):72-7.

12. Huang K. T., Tseng C. C., Fang W. F., Lin M. C. An early predictor of the outcome of patients with ventilator-associated pneumonia. Chang Gung Med J. 2010;33(3):274-82.

13. Lisboa T., Diaz E., Sa-Borges M., Socias A., Sole-Violan J., Rodriguez A., et al. The ventilator-associated pneumonia PIRO score: a tool for predicting ICU mortality and health-care resources use in ventilator-associated pneumonia. Chest. 2008;134(6):1208-16.

14. Johnson A. E., Pollard T. J., Shen L., Lehman L. W., Feng M., Ghassemi M., et al. MIMIC-III, a freely accessible critical care database. Sci Data. 2016;3:160035.

15. Kalil A. C., Metersky M. L., Klompas M., Muscedere J., Sweeney D. A., Palmer L. B., et al. Executive Summary: Management of Adults With Hospital-acquired and Ventilator-associated Pneumonia: 2016 Clinical Practice Guidelines by the Infectious Diseases Society of America and the American Thoracic Society. Clin Infect Dis. 2016;63(5):575-82.

16. Jones C. M., Athanasiou T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. Ann Thorac Surg. 2005;79(1):16-20.

17. Bickenbach J., Marx G. Diagnosis of pneumonia in mechanically ventilated patients: what is the meaning of the CPIS? Minerva Anestesiologica. 2013;79(12):1406-14.

18. Force Ards Definition Task, Ranieri V. M., Rubenfeld G. D., Thompson B. T., Ferguson N. D., Caldwell E., et al. Acute respiratory distress syndrome: the Berlin Definition. Jama. 2012;307(23):2526-33.

19. Forel J. M., Voillet F., Pulina D., Gacouin A., Perrin G., Barrau K., et al. Ventilator-associated pneumonia and ICU mortality in severe ARDS patients ventilated according to a lung-protective strategy. Crit Care. 2012;16(2):R65.

20. Guo L., Wang W., Zhao N., Guo L., Chi C., Hou W., et al. Mechanical ventilation strategies for intensive care unit patients without acute lung injury or acute respiratory distress syndrome: a systematic review and network meta-analysis. Crit Care. 2016;20(1):226.

21. Bellani G., Laffey J. G., Pham T., Fan E., Brochard L., Esteban A., 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(8):788-800.

22. Pugin J., Dunn-Siegrist I., Dufour J., Tissieres P., Charles P. E., Comte R. Cyclic stretch of human lung cells induces an acidification and promotes bacterial growth. Am J Respir Cell Mol Biol. 2008;38(3):362-70.

23. Chollet-Martin S., Jourdain B., Gibert C., Elbim C., Chastre J., Gougerot-Pocidalo M. A. Interactions between neutrophils and cytokines in blood and alveolar spaces during ARDS. American journal of...
24. Dreyfuss D., Ricard J. D. Acute lung injury and bacterial infection. Clin Chest Med. 2005;26(1):105-12.
25. Munford R. S., Pugin J. Normal responses to injury prevent systemic inflammation and can be immunosuppressive. American journal of respiratory and critical care medicine. 2001;163(2):316-21.
26. Knaus W. A., Wagner D. P., Draper E. A., Zimmerman J. E., Bergner M., Bastos P. G., et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. Chest. 1991;100(6):1619-36.
27. Kaya E., Dervisoglu A., Polat C. Evaluation of diagnostic findings and scoring systems in outcome prediction in acute pancreatitis. World J Gastroenterol. 2007;13(22):3090-4.
28. Beck D. H., Taylor B. L., Millar B., Smith G. B. Prediction of outcome from intensive care: a prospective cohort study comparing Acute Physiology and Chronic Health Evaluation II and III prognostic systems in a United Kingdom intensive care unit. Critical care medicine. 1997;25(1):9-15.
29. Markgraf R., Deutschinoff G., Pientka L., Scholten T. Comparison of acute physiology and chronic health evaluations II and III and simplified acute physiology score II: a prospective cohort study evaluating these methods to predict outcome in a German interdisciplinary intensive care unit. Critical care medicine. 2000;28(1):26-33.
30. Cunnion K. M., Weber D. J., Broadhead W. E., Hanson L. C., Pieper C. F., Rutala W. A. Risk factors for nosocomial pneumonia: comparing adult critical-care populations. American journal of respiratory and critical care medicine. 1996;153(1):158-62.
31. Potes C., Conroy B., Xu-Wilson M., Newth C., Inwald D., Frassica J. A clinical prediction model to identify patients at high risk of hemodynamic instability in the pediatric intensive care unit. Crit Care. 2017;21(1):282.

Figures
Figure 1

Our predictive model Feature importance of VAP. The feature importance of the optimal random forest model indicates the feature's contribution to VAP prediction.
Figure 1

Study profile. MIMIC, Medical Information Mart for Intensive Care; VAP, ventilator-associated pneumonia; ICU, intensive care unit
Figure 1

Imbalanced dataset model. The non VAP dataset was divided into 100 subgroups, one of which was combined with the VAP dataset to train the model, and then 100 models were combined into the final model.