Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Predicting Ventilator-Associated Pneumonia with Machine Learning

Christine Giang
Dascena, Inc.
Jacob Calvert
Dascena, Inc.
Gina Barnes
Dascena, Inc.
Anna Siefkas (anna@dascena.com)
Dascena, Inc.
Abigail Green-Saxena
Dascena, Inc.
Jana Hoffman
Dascena, Inc.
Qingqing Mao
Dascena, Inc.
Ritankar Das
Dascena, Inc.

Research Article

Keywords: Machine learning, ventilator-associated pneumonia, logistic regression, prediction

Posted Date: November 24th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-107907/v1

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

Abstract

Objective Ventilator-associated pneumonia (VAP) is the most common and fatal nosocomial infection in intensive care units (ICUs). Existing methods for identifying VAP display low accuracy, and their use may delay antimicrobial therapy. VAP diagnostics derived from machine learning methods that utilize electronic health record data have not yet been explored. The objective of this study is to compare the performance of a variety of machine learning models trained to predict whether VAP will be diagnosed during the patient stay.

Methods A retrospective study examined data from 6,129 adult ICU encounters lasting at least 48 hours following the initiation of mechanical ventilation. The gold standard was the presence of a diagnostic code for VAP. Five different machine learning models were trained to predict VAP 48 hours after initiation of mechanical ventilation. Model performance was evaluated with regard to area under the receiver operating characteristic curve (AUROC) on a 10% hold-out test set. Feature importance was measured in terms of Shapley values.

Results The highest performing model achieved an AUROC value of 0.827. The most important features for the best-performing model were the length of time on mechanical ventilation, presence of antibiotics, sputum test frequency, and most recent Glasgow Coma Scale assessment.

Discussion Supervised machine learning using patient electronic health record data is promising for VAP diagnosis and warrants further validation.

Conclusion This tool has the potential to aid the timely diagnosis of VAP.

Background

Pneumonia is one of the deadliest infections across the globe [1]. It can be acquired in community or hospital settings, as well as through the use of invasive mechanical ventilation [1, 2]. Ventilator-associated pneumonia (VAP) is notoriously difficult to diagnose due to the absence of a diagnostic gold standard, which can be attributed to the diversity of disease-causing pathogens, lung cultures that limit sampling to anatomical surfaces, clinical interpretation of pathogens extracted in lung samples, underlying health issues in individuals, and agreement of radiological scans with centers for disease control (CDC) diagnostic criteria [1-5]. VAP, which is defined as pneumonia which develops after more than 48 hours following the initiation of invasive mechanical ventilation [4], is estimated to occur in 5%-67% of intubated patients, with the highest rates among those who are hospitalized due to physical trauma [6, 7]. Mortality estimates range from 24%-76% [8]. VAP is also associated with an estimated $47,238 in additional healthcare costs per patient in the United States [6-9].

Given the high mortality rate with VAP and the vulnerability of the patient population at risk for VAP, prevention is crucial for clinical management to reduce the prevalence of VAP [6, 7]. The CDC's evidence-based recommendations to prevent VAP encompass in-hospital safety protocols that limit the

introduction of microbes into the lungs of intubated patients, such as opting for non-invasive mechanical ventilation methods and avoiding intubation when feasible, keeping levels of subglottic secretions low by frequent secretion removal, and maintaining intubated patients at an elevated chest position [10].

Diagnostic methods for VAP vary and rely on a combination of factors, including clinical assessment for infectious signs and symptoms (e.g. fever), chest radiology, lung biopsy, and/or quantitative microbiological testing of respiratory secretions [8, 11-13]. However, diagnostic accuracy for VAP continues to be poor. A recent meta-analysis by Fernando et al. found that eight common clinical criteria for VAP diagnosis lacked specificity, highlighting the need for diagnostic tools which can better inform the need for and timing of antibiotic treatment while respecting antibiotic stewardship [14].

The Predisposition, Insult, Response, Organ dysfunction (PIRO) score, originally developed for sepsis, can be utilized for VAP risk stratification in hospitalized patients [15, 16]. The Clinical Pulmonary Infection Score (CPIS) is also used as a VAP diagnostic tool for hospitalized patients [17]. However, prior research has shown these tools exhibit poor diagnostic performance. In a recent study, PIRO achieved an area under the receiver operating characteristic curve (AUROC) of 0.605 [18]. A recent meta-analysis found that CPIS demonstrated a sensitivity and specificity of 73.8% and 66.4%, respectively, for VAP detection [14]. The Confusion, Urea, Respiratory Rate, Blood Pressure (CURB-65) score, utilized for risk stratification and to predict 30 day mortality, has better accuracy than PIRO and CPIS with a c-statistic of 0.761 [19, 20]. However, the sensitivity of CURB-65 varies with the severity of pneumonia. A 2016 study showed that 36% of patients who were classified as low risk based on their CURB-65 score were ultimately hospitalized as the result of pneumonia [20].

Despite the shortcomings of existing clinical indicators and scoring tools for VAP diagnosis, there is minimal research on the application of machine learning (ML) methods using electronic health record (EHR) data to diagnose VAP. Studies of ML applications to pneumonia have focused largely on mortality prediction among known pneumonia cases. ML has been proposed as a component of other diagnostic tools, such to assist with chest radiography interpretation [21], to predict pneumonia outcomes [22], or work with electronic sensors to detect pneumonia in exhaled breath [23]. Yet ML has not been robustly investigated to develop stand-alone diagnostic tools [24-28]. The dearth of such ML research may reflect the very reason for its urgent need–the lack of a clear diagnostic gold standard. The absence of such a standard complicates the identification of pneumonia onset in retrospective EHR data, making it difficult to develop a supervised ML approach to pneumonia prediction. For VAP, the difficulty in identifying onset time in retrospective data is partially alleviated by the constraints on onset time relative to the start of mechanical ventilation, as per the definition of VAP. It is possible to determine when inpatients have reached the 48th hour from the initiation of mechanical ventilation, the minimum required time for pneumonia to be designated as “ventilator-acquired.” Using this definition, it is possible to develop and assess the ability of ML technology to accurately diagnose VAP in retrospective intensive care unit (ICU) data.

This exploratory analysis examined the suitability of ML methods for the prediction of VAP. To meet this goal, we developed and assessed a variety of ML approaches for the following two prediction tasks:

Intubation task. Among ICU encounters lasting at least 48 hours following the initiation of mechanical ventilation, predict whether or not the given encounter will be associated with a diagnosis of VAP at any later time during the patient's stay, with predictions generated at the 48th hour following intubation.

Admission task. Among all ICU encounters lasting at least *k* hours, classify whether or not the given encounter will be associated with a diagnosis of VAP at any time, with classifications made at the kth hour following admission. Note that, for this task, classifications could be made prior to initiation of mechanical ventilation.

For the intubation task, 20 possible models were explored: five different model types using statistics of certain required measurements from the most recent 6, 12, 24, or 48 hours relative to the prediction time. The same five model types were explored for the admission task, with *k* taking values 6, 12, 24, or 48 hours after ICU admission time; this also resulted in 20 models. A variety of model types were chosen in order to identify the most appropriate ML methods to address these novel tasks, including both simple linear models and complex, non-linear models, to evaluate the potential effectiveness of different machine learning models for VAP prediction.

Methods

Data processing

For all models, data were extracted from the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC)-III version 1.3 dataset collected from the ICU at Beth Israel Deaconess Medical Center in Boston, Massachusetts [29]. MIMIC-III contains EHR data (including lab results) and clinical notes on over 40,000 individual patient encounters. All MIMIC-III data were passively extracted from the patient EHR, were de-identified, and collected in compliance with the Health Insurance Portability and Accountability Act.

Intubation task

Data were included from encounters of patients aged 18 years or older, with a minimum of one observation of each of the following vital signs and lab tests: diastolic blood pressure (DiasBP), creatinine, Glasgow Coma Scale (GCS), heart rate (HR), oxygen saturation (SpO_2), platelet count, respiratory rate (RR), systolic blood pressure (SysBP), temperature, hematocrit, and white blood cell count (WBC). Hematocrit has been shown to improve other pneumonia-related predictions [30]. Community-acquired pneumonia patients were identified by the presence of a pneumonia diagnosis at admission and were excluded. All encounters for this task were required to involve at least one period of invasive mechanical ventilation. As VAP is defined as pneumonia developing after 48 hours following intubation, encounters were required to last at least 48 hours after intubation. All mechanically ventilated patients in this dataset met this requirement. ML models were compared against the CURB-65 [19, 20], VAP PIRO

[15], and CPIS scoring systems [17]. To facilitate the comparison with CURB-65, we required encounters to include at least one measurement of blood urea nitrogen (BUN). These exclusion steps are summarized in **Figure 1.** For this task, the windows of data used to generate predictions were calculated backwards from the 48th hour following the initiation of ventilation. That is, a 12-hour intubation task model used the 12 hours of data up to and including 48 hours after initiation of mechanical ventilation, or hours 37 to 48 after the initiation of mechanical ventilation. All windows for this intubation task included data from an identical number of patients.

Admission task

Identical exclusion criteria were applied for the admission task as were applied to the intubation task, with the exception of the initiation of mechanical ventilation requirement, which was not applied. For this task, the windows of data used to generate predictions were calculated forward from the time of ICU admission. For example, a 12-hour admission task model used the first 12 hours of data after a patient was admitted to the ICU, after which point a VAP risk prediction was generated. Patients were required to have a length of stay as long as or longer than the prediction window being examined; the number of patients included in the experiments therefore varied by prediction window (**Figure 1**).

For both tasks, we extracted patient baseline and time-varying clinical measurements for each encounter. Baseline data included age, and a boolean value for the presence of any relevant comorbidities or symptoms at time of admission (bacteremia, cirrhosis, congestive heart failure, fever, intracranial hemorrhage, renal failure, respiratory distress, respiratory failure, sepsis, subarachnoid hemorrhage, and shortness of breath). We additionally included an indicator for acute respiratory distress syndrome (ARDS), as pneumonia is associated with ARDS [31]. Time-varying clinical measurements included the required vital signs and laboratory tests, as well as urine output (evaluated as number of urine measurements over duration of stay) and blood culture information (evaluated as the order of any tests during the relevant window, and as the total test count during the window). We further included the hour of the initiation of mechanical ventilation and the number of accumulated mechanical ventilation hours at the time of prediction.

Raw measurements were binned into one-hour intervals and averaged within bins to produce a single, representative value for each hour. Missing values were imputed based on median values, which were determined using only data in a training set. This process did not allow information from the hold-out test set to influence the imputation. We calculated six summary statistics (minimum, maximum, median, first, last, and average) of each vital sign and laboratory tests over a variable-length window (**Table 1**). Specifically, for each window length , we calculated the statistics for the hours preceding and including the 48th hour after the initiation of mechanical ventilation (in the case of the intubation task) or the 48th hour after admission. Age, the number of total urine output events, and the number of blood culture tests were kept in their raw form. Boolean indicators were added for the presence of antibiotics, sputum labs, blood culture labs, comorbidities and symptoms listed above, and ARDS **(Table 1)**. All variables were then concatenated into one vector for each encounter.

Table 1: Data included as input to the algorithm

Required vitals and labs	Boolean Indicators	Optional Measures
- Systolic BP - Dias BP - HR - Respiratory Rate - Temperature - Hematocrit - SpO2 - BUN - GCS - Platelet Count - WBC - Creatinine	- Antibiotics - Sputum labs - Blood culture labs - Any of cirrhosis, congestive heart failure, fever, bacteremia, intracranial hemorrhage, renal failure, respiratory distress, respiratory failure, sepsis, subarachnoid hemorrhage, shortness of breath - Acute respiratory distress syndrome (ARDS)	- Age - Total urine output events - Number of blood culture tests - Number of sputum tests - Number of MV hours

Gold Standard

The International Classification of Diseases (ICD) Revision 9 code 997.31 was the gold standard definition of VAP. Literature assessing the accuracy of ICD codes for VAP identification remains limited [32]. However, studies have suggested that, while sensitivity of administrative coding may be only moderate for VAP identification, specificity and NPV are quite high [33, 34].

Machine Learning Methods and Comparators

For each prediction task and each window length, we trained and tested five ML models: logistic regression, multilayer perceptron, random forest, support vector machines, and gradient boosted trees. We favored variety in our choice of ML methods due to the absence of VAP prediction literature. The logistic regression and support vector machines models were chosen as representative linear models and the random forest and gradient boosted trees models were chosen as representative ensemble learning and tree-based methods. The multilayer perceptron model was included in lieu of neural network models with

more layers, as there were too few training examples to effectively train such models. Except for the gradient boosted trees model, which was created using the XGBoost Python package, the models were implemented using the scikit-learn Python package.

We compared performance of the machine learning algorithm to the CURB-65, VAP PIRO, and CPIS scores for evaluating pneumonia severity. CPIS performance was estimated from the literature [14] as it could not be calculated in our dataset. We implemented CURB-65 and VAP PIRO in our dataset [18, 19]. CURB-65 values were calculated for each hour according to the number of the following which were true: BUN > 19 ml/dL, Respiratory Rate ≥ 30, Systolic BP < 90 mmHg/Diastolic BP ≤ 60 mmHg, and age ≥ 65. We tried several variations of assigning a CURB-65 score to a temporal window, including using its maximum, average, and last values over the window. As the results were similar in each case, we reported its average over the window. PIRO is a four-variable score based on predisposition, insult, response, and organ dysfunction. The score is measured by assigning one point in four areas: detection of a comorbidity (Chronic obstructive pulmonary disease, immunocompromised, heart failure, cirrhosis, or chronic renal failure), bacteremia, a systolic BP < 90 mmHg, and ARDS.

The data were partitioned uniformly-at-random into a set for training and hyperparameter tuning (90%) and a 10% hold-out test set, against which all trained models were evaluated for final performance metrics in the last step. For each task and for each window length k, each model was trained using four-fold grid search cross-validation on the 90% training set. After searching the space of model hyperparameter values, the hyperparameters that produced the best cross-validation performance in terms of AUROC were chosen. Each model was then tested on the 10% hold-out test set. Feature importance was measured through Shapley additive explanation (SHAP) values to assess similarities or differences in the features used to generate predictions across model types.

Results

In total, 6,126 patients were included in the intubation task experiments. Of those, 524 received a diagnosis of VAP during their stay, resulting in a VAP prevalence in the intubation task patient population of 8.55% and of 4.97% in the admission task patient population. Those who were diagnosed with VAP had a higher prevalence of ARDS, a greater number of sputum labs performed, and were on average older when compared to those without VAP (**Table 2**). Additional comorbidity information for the intubation task population is presented in **Supplementary Table 1**.

Table 2. Demographic and comorbidity information for the experimental population.

	Characteristic	VAP Positive n = 524	VAP Negative n = 5602
Age	<30	27 (5.15%)	215 (3.84%)
	30-49	81 (15.46%)	735 (13.12%)
	50-59	105 (20.04%)	972 (17.35%)
	60-69	120 (22.90%)	1383 (24.69%)
	70-79	107 (20.42%)	1185 (21.15%)
	80+	71 (13.55%)	984 (17.57%)
ARDS	Yes	35 (6.68%)	284 (5.07%)
	No	489 (93.32%)	5318 (94.93%)
Sputum Test Performed	Yes	497 (94.85%)	2644 (47.20%)
	No	27 (5.15%)	2958 (52.80%)
Gender	Male	313 (59.73%)	3172 (56.62%)
	Female	211 (40.27%)	2430 (43.38%)
Ethnicity	White	354 (5.15%)	4072 (72.69%)
	Black/African-American	44 (8.40%)	484 (8.64%)
	Asian	16 (3.05%)	134 (2.39%)
	Hispanic/Latino	11 (2.10%)	209 (3.73%)
	Unknown/Other	99 (18.89%)	703 (12.55%)

Prediction of VAP 48 hours after intubation

The gradient boosted trees models generally demonstrated better performance than other model types for all prediction windows, with the exception of the models trained to use summary statistics from a 48-hour window (**Table 3**), where logistic regression demonstrated the highest performance. The best AUROC was recorded by XGBoost using summary statistics from a six-hour window. Multilayer perceptron models demonstrated lower performance than all other models, particularly when using 48 hours of data. All models outperformed the CURB-65 and PIRO scores at all prediction times.

Table 3. AUROC results on the hold-out test set of models trained to predict VAP 48 hours after intubation, using summary statistics from the previous *k* hours of patient data.

k (hours)	6	12	24	48
Logistic regression	0.779	0.691	0.721	0.808
Multilayer perceptron	0.733	0.722	0.711	0.765
Random forest	0.804	0.781	0.805	0.821
Support vector machines	0.802	0.741	0.755	0.808
XGBoost	0.827	0.787	0.823	0.799
CURB-65	0.503	0.498	0.498	0.498
PIRO	0.565	0.555	0.566	0.557

Receiver operating characteristic (ROC) curves for all intubation task models at the time windows and are presented in **Figure 2;** these models are annotated with the performance of several common clinical criteria in the diagnosis of VAP, as determined by meta-analysis [14]. All models meet or exceed the reported performance of all of the clinical criteria. ROC curves for the remaining time points are presented in **Supplementary Figure 1**.

An assessment of SHAP plots for the intubation task models showed a high degree of overlap in the features identified as most important for generating predictions (**Supplemental Figures 2-6**), with key recurrent features including length of time for mechanical ventilation, sputum culture measures, and clinical measures related to SpO_2 and GCS.

Prediction of VAP after admission

Of the models trained for the admission task, each performed best when using summary statistics from a 48-hour period between admission and the time of prediction (**Table 4**). Overall best performance was obtained by the logistic regression using 48 hours of data. **Figure 3** provides comparisons of ROC curves for these windows, with annotated comparison to common clinical criteria. Most models meet or exceed the reported performance of existing clinical criteria at 6 hours, and all did at 48 hours. The $k = 12$ and $k = 24$ cases are shown in **Supplementary Figure 7**.

Table 4: AUROC results on the hold-out test set of models trained to predict VAP *k* hours after ICU admission.

k (hours)	6	12	24	48
Logistic regression	0.777	0.731	0.716	0.877
Multilayer perceptron	0.599	0.772	0.800	0.833
Random forest	0.729	0.680	0.731	0.829
Support vector machines	0.793	0.751	0.777	0.858
XGBoost	0.748	0.780	0.769	0.850
CURB-65	0.481	0.496	0.506	0.517
PIRO	0.584	0.595	0.599	0.622

As with the intubation task, SHAP plots demonstrated that features most important for generating VAP predictions were similar across model types and prediction windows (**Supplementary Figures 8-12**), with mechanical ventilation hours, GCS, and sputum again being important features across many models. This finding further supports that a wide range of models may be suitable for this prediction task.

Discussion

Our retrospective results demonstrate the success of ML models trained to predict VAP in two use cases (**Tables 3** and **4**). Due to the novelty of the task, a variety of models and prediction windows were explored, all of which demonstrated strong predictive performance. In both use cases, the test set performance of ML models significantly exceeds the reported performance of classic clinical indicators of VAP [14] and does so with the potential of advance warning (**Figures 1** and **3**, **Supplementary Figures 1** and **4**). Additionally, simple, interpretable models such as logistic regression demonstrate strong performance for both tasks. Given the morbidity and mortality associated with VAP [35] and with VAP treatment delays [36], these models may have important implications for improving patient care and outcomes, subject to future external and prospective validation.

Despite the urgent need for better VAP diagnostics and the popularity of ML applications in healthcare, relatively little effort has been devoted to application of ML to EHR data for the purpose of predicting VAP. Several methods have been developed for identifying community acquired pneumonia using neural networks and genetic algorithms [37-39] and one study predicted hospital acquired pneumonia in patients with schizophrenia [40]. Concerning VAP, studies have examined the accuracy of electronic nose (e-nose) sniffers for screening of potential VAP cases. These devices use ML methods to analyze exhaled breath for metabolites that may be suggestive of VAP, and some have demonstrated strong discrimination for identifying the presence of VAP [41, 42]. However, prospective validation of an e-nose tool found that sensitivity and specificity were insufficient for general clinical use [43]. In this context, our study provides a valuable characterization of a variety of ML methods applied to two VAP prediction use cases.

The first use case corresponds to the intubation task, which predicted VAP at the first time VAP can be diagnosed, the 48th hour following the initiation of mechanical ventilation. While the highest overall AUROC was demonstrated by XGBoost using data from the six-hour window leading up to and including the prediction time, all models met or exceeded the performance of existing VAP identification methods. An alert at or before VAP onset by any of these methods is likely to improve identification of VAP, potentially overcoming limitations in diagnostic criteria that may lead to both under- and over-treatment with antibiotics [14, 44, 45]. It is worth noting that XGBoost demonstrated decreasing performance over longer data collection windows for the intubation task. This may be due to the fact that, with increasing time from the point at which predictions are made, many of the model inputs lose clinical relevance to the current patient state, decreasing the overall relevance of the inputs. However, other methods, such as random forest, demonstrated increased performance with longer data windows. The relatively low performance of the multilayer perceptron model may be because a single layer perceptron can only classify linearly separable sets of vectors. Since the data used here have at least 82 dimensions, it's likely that the classes are not linearly separable. Although we tried a multilayered neural network, we had insufficient training data.

The second use case corresponds to the admission task, which aims to predict VAP a fixed number of hours (6, 12, 24, or 48) following admission to the ICU. In the first three of these cases, alerts generated by the models give advance warning of VAP, as patients cannot have been ventilated for the 48 hours required for a VAP diagnosis. Further, these models have no requirement that patients be ventilated for algorithm alerts to be generated, although they incorporate information about current ventilation status. Therefore, these models may be able to identify patients at high risk of developing VAP should they become ventilated before ventilation occurs. For this task, the best-performing model was a logistic regression model using summary statistics generated from the entire 48-hour window after admission (**Table 4**, **Figure 3**). All models demonstrated performance meeting or exceeding methods current methods for VAP identification. For this task, models generally demonstrated increasing performance when using larger windows of data and when generating predictions further into the patient stay.

Models addressing these two use cases exhibit different strengths, with potential implications for future clinical use. The intubation task is, by definition, applied to a high-risk population of patients. In many clinical settings, the positive predictive value of an intubation task alert is therefore likely to be high and an intubation-based ML system likely to identify the vast majority of VAP cases. Such a system could therefore meaningfully improve the timeliness of antimicrobial administration, improving patient outcomes. In contrast, the admission task is designed for application to all ICU patients, regardless of ventilation status. Therefore, in addition to providing early identification of patients for whom additional monitoring is warranted, the admission task alert may provide clinicians with the opportunity to consider non-invasive methods of ventilation for high-risk patients [46, 47], preventing VAP entirely. The success of both approaches in this retrospective study supports the potential for ML methods to meet a wide range of clinical needs relating to VAP treatment, identification, and prevention.

For both use cases, the strong performance of simple linear models (e.g. logistic regression) has important implications. Logistic regression models are readily interpretable, with the relative importance of each input feature measurable as the relative magnitude of its coefficient. While many machine learning methods are viewed as “black boxes,” linear models are in contrast far more transparent, with clear similarities to tabular scores already commonly used in clinical practice. This feature of simple models may increase trust in the model, increasing its utility [48].

There are several limitations to this study. Because the exact onset time of VAP could not be determined retrospectively from this dataset, it was not possible to determine the degree of advance warning provided by the models. For the best-performing models, the cumulative duration of mechanical ventilation, presence of antibiotics administration, and the ordering of a sputum test were the most important features, along with statistics of GCS and creatinine (**Supplementary Figures 2** and **4**). It may be that, while by definition VAP may not be diagnosed until 48 hours after the initiation of mechanical ventilation, clinicians may suspect VAP is developing and order sputum tests or administer antibiotics within this time window. In these cases, while the classifications made by the models may be technically considered predictions, the alerts they would provide would not necessarily lead to a significant change in care. These results may also be limited by the use of ICD codes for the VAP gold standard, which may fail to accurately capture all patients who experienced VAP during their hospitalization. However, the generally high specificity of codes for hospital-acquired infections supports that our positive class consisted of true positive VAP cases. Another limitation is that these models were trained and tested on data from a single institution, which may limit generalizability. Model performance on novel patient populations or specific subpopulations cannot be inferred. Finally, due to the retrospective nature of this study, the impact that these algorithms may have on patient care and outcomes in a live clinical setting cannot be determined. These limitations underscore the need for future additional and prospective validation.

Conclusions

The development of accurate and timely diagnostic tools for ventilator associated pneumonia has been limited, despite the prevalence, mortality, and costliness associated with VAP. Machine learning may be a key contributor for future management of VAP risk associated with mechanical intubation, with a variety of machine learning methods demonstrating suitability for this prediction task. Future work is necessary for further validation of machine learning algorithms for VAP prediction.

List Of Abbreviations

ARDS: acute respiratory distress syndrome

AUROC: area under the receiver operating characteristic

BUN: blood urea nitrogen

CAP: community acquired pneumonia

CDC: Centers for Disease Control

CPIS: Clinical Pulmonary Infection Score

CURB-65: Confusion, Urea, Respiratory Rate, Blood Pressure

DiasBP: diastolic blood pressure

EHR: electronic health record

GCS: Glasgow Coma Scale

HR: heart rate

ICD: International Classification of Diseases

ICU: intensive care unit

MIMIC-III: Multiparameter Intelligent Monitoring in Intensive Care-III

ML: machine learning

MV: mechanical ventilation

PIRO: Predisposition, Insult, Response, Organ dysfunction

ROC: receiver operating characteristic

RR: respiratory rate

SpO_2: oxygen saturation

SysBP: systolic blood pressure

VAP: ventilator-associated pneumonia

WBC: white blood cell

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Data was obtained from Multiparameter Intelligent Monitoring in Intensive Care (MIMIC)-III version 1.3 dataset collected from the ICU at Beth Israel Deaconess Medical Center in Boston, Massachusetts [29]. MIMIC-III is a publicly available database, comprised of a dataset that was de-identified, and collected in compliance with the Health Insurance Portability and Accountability Act.

Competing interests

All authors who have affiliations listed with Dascena (Houston, Texas, U.S.A) are employees or contractors of Dascena.

Funding

No funding to report.

Authors' contributions

CG and JC conceived and designed the study. CG performed the data analysis and prepared the figures. CG, GB, and AS drafted the manuscript. All authors revised the manuscript for intellectual content.

Acknowledgements

Not applicable.

Authors' information (optional)

Not provided

References

1. Wagh H, Acharya D. Ventilator Associated Pneumonia – an Overview. 2009;2(2):4.
2. Klompas M. Interobserver variability in ventilator-associated pneumonia surveillance. *Am J Infect Control*. 2010;38(3):237-239. doi:10.1016/j.ajic.2009.10.003
3. HealthManagement.org. Radiology Management, ICU Management, Healthcare IT, Cardiology Management, Executive Management. HealthManagement. Accessed October 2, 2020. https://healthmanagement.org/c/icu/issuearticle/controversies-in-ventilator-associated-pneumonia-diagnosis
4. Bergin SP, Coles A, Calvert SB, et al. PROPHETIC: Prospective Identification of Pneumonia in Hospitalized Patients in the ICU. *Chest*. Published online June 29, 2020. doi:10.1016/j.chest.2020.06.034

5. Mortality and the Diagnosis of Ventilator-associated Pneumonia: A New Direction. *Am J Respir Crit Care Med*. 1998;157(2):349-350. doi:10.1164/ajrccm.157.2.ed16-97
6. 6. Torres A, Niederman MS, Chastre J, et al. International ERS/ESICM/ESCMID/ALAT guidelines for the management of hospital-acquired pneumonia and ventilator-associated pneumonia: Guidelines for the management of hospital-acquired pneumonia (HAP)/ventilator-associated pneumonia (VAP) of the European Respiratory Society (ERS), European Society of Intensive Care Medicine (ESICM), European Society of Clinical Microbiology and Infectious Diseases (ESCMID) and Asociación Latinoamericana del Tórax (ALAT). *Eur Respir J*. 2017;50(3). doi:10.1183/13993003.00582-2017
7. 7. Timsit J-F, Esaied W, Neuville M, Bouadma L, Mourvillier B. Update on ventilator-associated pneumonia. *F1000Research*. 2017;6. doi:10.12688/f1000research.12222.1
8. 8. Miller DF. Ventilator-Associated Pneumonia. Published online 2018:6.
9. 9. Estimating the Additional Hospital Inpatient Cost and Mortality Associated With Selected Hospital-Acquired Conditions | Agency for Health Research and Quality. Accessed October 6, 2020. https://www.ahrq.gov/hai/pfp/haccost2017-results.html
10. Guidelines for the Management of Adults with Hospital-acquired, Ventilator-associated, and Healthcare-associated Pneumonia. *Am J Respir Crit Care Med*. 2005;171(4):388-416.
11. 11. Wang G, Ji X, Xu Y, Xiang X. Lung ultrasound: a promising tool to monitor ventilator-associated pneumonia in critically ill patients. *Crit Care*. 2016;20. doi:10.1186/s13054-016-1487-y
12. 12. Mayhall CG. Ventilator-Associated Pneumonia or Not? Contemporary Diagnosis - Volume 7, Number 2—April 2001 - Emerging Infectious Diseases journal - CDC. doi:10.3201/eid0702.700200
13. 13. Kalanuria AA, Zai W, Mirski M. Ventilator-associated pneumonia in the ICU. *Crit Care*. 2014;18(2):208. doi:10.1186/cc13775
14. Fernando SM, Tran A, Cheng W, et al. Diagnosis of ventilator-associated pneumonia in critically ill adult patients—a systematic review and meta-analysis. *Intensive Care Med*. Published online April 18, 2020:1-10. doi:10.1007/s00134-020-06036-z
15. 15. Marshall JC. The PIRO (predisposition, insult, response, organ dysfunction) model. *Virulence*. 2014;5(1):27-35. doi:10.4161/viru.26908
16. 16. Lisboa T, Diaz E, Sa-Borges M, Socias A, Sole-Violan J, Rodríguez A, Rello J. The ventilator-associated pneumonia PIRO score: a tool for predicting ICU mortality and health-care resources use in ventilator-associated pneumonia. Chest. 2008 Dec;134(6):1208-1216. doi: 10.1378/chest.08-1106. Epub 2008 Sep 8. PMID: 18779186.
17. 17. Zilberberg MD, Shorr AF. Ventilator-Associated Pneumonia: The Clinical Pulmonary Infection Score as a Surrogate for Diagnostics and Outcome. *Clin Infect Dis*. 2010;51(Supplement_1):S131-S135. doi:10.1086/653062
18. 18. Furtado GH, Wiskirchen DE, Kuti JL, Nicolau DP. Performance of the PIRO Score for Predicting Mortality in Patients with Ventilator-Associated Pneumonia. *Anaesth Intensive Care*. 2012;40(2):285-291. doi:10.1177/0310057X1204000211

19. Lim WS, van der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax*. 2003;58(5):377-382. doi:10.1136/thorax.58.5.377
20. Sharp AL, Jones JP, Wu I, et al. CURB-65 Performance among Admitted and Discharged Emergency Department Patients with Community-acquired Pneumonia. Acad Emerg Med. 2016;23(4):400-405. doi:10.1111/acem.12929
21. Jaiswal AK, Tiwari P, Kumar S, Gupta D, Khanna A, Rodrigues JJPC. Identifying pneumonia in chest X-rays: A deep learning approach. Measurement. 2019 Oct 1;145:511–8.
22. Kang SY, Cha WC, Yoo J, et al. Predicting 30-day mortality of patients with pneumonia in an emergency department setting using machine-learning models. Clin Exp Emerg Med. 2020;7(3):197-205. doi:10.15441/ceem.19.052
23. Wilson AD. Advances in electronic-nose technologies for the detection of volatile biomarker metabolites in the human breath. Metabolites. 2015;5(1):140-163. Published 2015 Mar 2. doi:10.3390/metabo5010140
24. 24. Cooper GF, Aliferis CF, Ambrosino R, et al. An evaluation of machine-learning methods for predicting pneumonia mortality. *Artif Intell Med*. 1997;9(2):107-138. doi:10.1016/S0933-3657(96)00367-3
25. 25. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15*. ACM Press; 2015:1721-1730. doi:10.1145/2783258.2788613
26. 26. Rajpurkar P, Irvin J, Zhu K, et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *ArXiv171105225 Cs Stat*. Published online December 25, 2017. Accessed October 6, 2020. http://arxiv.org/abs/1711.05225
27. 27. Toğaçar M, Ergen B, Cömert Z. A Deep Feature Learning Model for Pneumonia Detection Applying a Combination of mRMR Feature Selection and Machine Learning Models. *IRBM*. Published online November 1, 2019. doi:10.1016/j.irbm.2019.10.006
28. 28. Chandra TB, Verma K. Pneumonia Detection on Chest X-Ray Using Machine Learning Paradigm. In: Chaudhuri BB, Nakagawa M, Khanna P, Kumar S, eds. *Proceedings of 3rd International Conference on Computer Vision and Image Processing*. Advances in Intelligent Systems and Computing. Springer; 2020:21-33. doi:10.1007/978-981-32-9088-4_3
29. 29. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3(1):160035. doi:10.1038/sdata.2016.35
30. Ahn JH, Choi EY. Expanded A-DROP score: a new scoring system for the prediction of mortality in hospitalized patients with community-acquired pneumonia. Scientific reports. 2018 Oct 1;8(1):1-9.
31. ARDS - Symptoms and causes [Internet]. Mayo Clinic. [cited 2020 Oct 26]. Available from: https://www.mayoclinic.org/diseases-conditions/ards/symptoms-causes/syc-20355576

32. 32. Goto M, Ohl ME, Schweizer ML, Perencevich EN. Accuracy of administrative code data for the surveillance of healthcare-associated infections: A systematic review and meta-analysis. Clin Infect Dis. 2014;58(5):688-696. doi:10.1093/cid/cit737
33. Stevenson KB, Khan Y, Dickman J, et al. Administrative coding data, compared with CDC/NHSN criteria, are poor indicators of health care-associated infections. Am J Infect Control. 2008;36(3):155-164. doi:10.1016/j.ajic.2008.01.004
34. Verelst S, Jacques J, Van Den Heede K, et al. Validation of Hospital Administrative Dataset for adverse event Screening. Qual Saf Heal Care. 2010;19(5):e25-e25. doi:10.1136/qshc.2009.034306
35. Monegro AF, Muppidi V, Regunath H. Hospital Acquired Infections. In: *StatPearls*. StatPearls Publishing; 2020. Accessed October 5, 2020. http://www.ncbi.nlm.nih.gov/books/NBK441857/
36. Grief SN, Loza JK. Guidelines for the Evaluation and Treatment of Pneumonia. Prim Care. 2018;45(3):485-503. doi:10.1016/j.pop.2018.04.001
37. Heckerling PS, Gerber BS, Tape TG, Wigton RS. Selection of predictor variables for pneumonia using neural networks and genetic algorithms. *Methods Inf Med*. 2005;44(1):89-97.
38. Heckerling PS, Gerber BS, Tape TG, Wigton RS. Use of genetic algorithms for neural networks to predict community-acquired pneumonia. *Artif Intell Med*. 2004;30(1):71-84. doi:10.1016/S0933-3657(03)00065-4
39. Er O, Sertkaya C, Temurtas F, Tanrikulu AC. A comparative study on chronic obstructive pulmonary and pneumonia diseases diagnosis using neural networks and artificial immune system. *J Med Syst*. 2009;33(6):485-492. doi:10.1007/s10916-008-9209-x
40. Kuo KM, Talley PC, Huang CH, Cheng LC. Predicting hospital-acquired pneumonia among schizophrenic patients: a machine learning approach. *BMC Med Inform Decis Mak*. 2019;19(1):42. doi:10.1186/s12911-019-0792-1
41. Liao Y-H, Shih C-H, Abbod MF, Shieh J-S, Hsiao Y-J. Development of an E-nose system using machine learning methods to predict ventilator-associated pneumonia. *Microsyst Technol*. Published online March 16, 2020. doi:10.1007/s00542-020-04782-0
42. Liao Y-H, Wang Z-C, Zhang F-G, Abbod MF, Shih C-H, Shieh J-S. Machine Learning Methods Applied to Predict Ventilator-Associated Pneumonia with Pseudomonas aeruginosa Infection via Sensor Array of Electronic Nose in Intensive Care Unit. *Sensors*. 2019;19(8):1866. doi:10.3390/s19081866
43. Schnabel RM, Boumans MLL, Smolinska A, et al. Electronic nose analysis of exhaled breath to diagnose ventilator-associated pneumonia. *Respir Med*. 2015;109(11):1454-1459. doi:10.1016/j.rmed.2015.09.014
44. Camargo LFA, De Marco FV, Barbas CSV, et al. Ventilator associated pneumonia: comparison between quantitative and qualitative cultures of tracheal aspirates. *Crit Care*. 2004;8(6):R422-R430. doi:10.1186/cc2965
45. Nussenblatt V, Avdic E, Berenholtz S, et al. Ventilator-associated pneumonia: overdiagnosis and treatment are common in medical and surgical intensive care units. *Infect Control Hosp Epidemiol*. 2014;35(3):278-284. doi:10.1086/675279

46. Brochard L. Mechanical ventilation: invasive versus noninvasive. *Eur Respir J.* 2003;22(47 suppl):31s-37s. doi:10.1183/09031936.03.00050403
47. Makhabah DN, Martino F, Ambrosino N. Noninvasive Mechanical Ventilation in Patients with High-Risk Infections in Intermediate Respiratory Care Units and on the Pneumology Ward. *Noninvasive Vent High-Risk Infect Mass Casualty Events*. Published online May 29, 2013:329-332. doi:10.1007/978-3-7091-1496-4_37
48. Shortliffe EH, Sepulveda MJ. Clinical decision support in the era of artificial intelligence. JAMA. 2018 Dec 4;320(21):2199-2200. PMID: 30398550

Figures

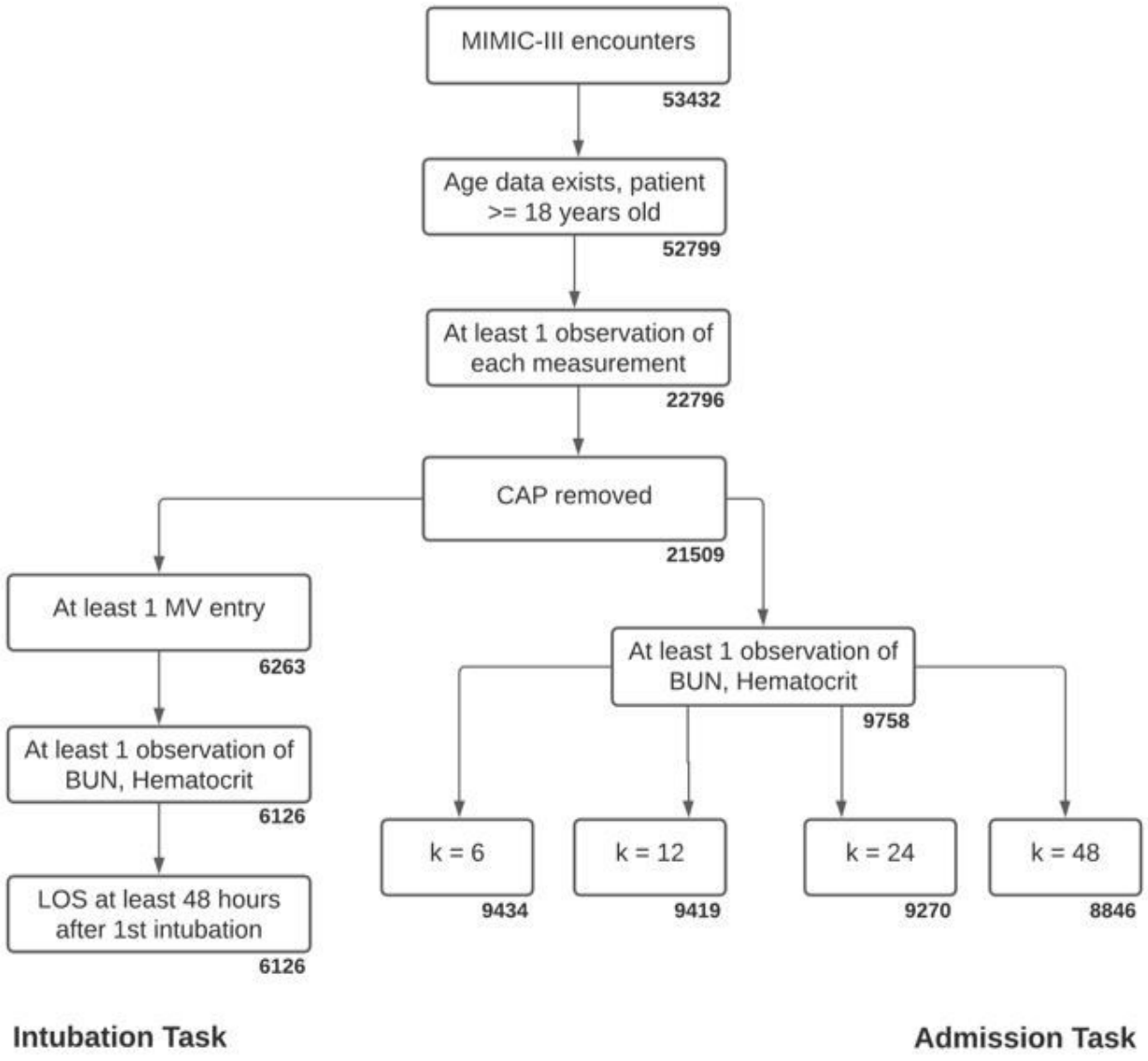

Figure 1

Patient exclusion numbers for the intubation task. Abbreviations: blood urea nitrogen (BUN); community acquired pneumonia (CAP); mechanical ventilation (MV).

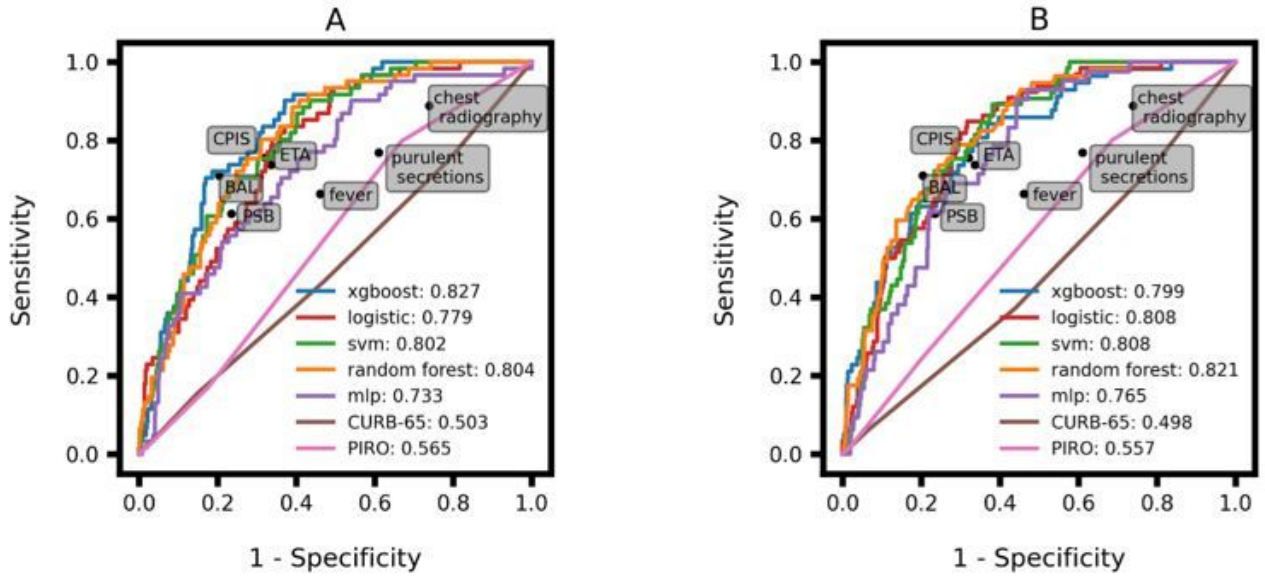

Figure 2

ROC curve comparison for intubation task models, with summary statistics calculated from the 6 hours (A) and 48 hours (B) of data preceding the time of prediction.

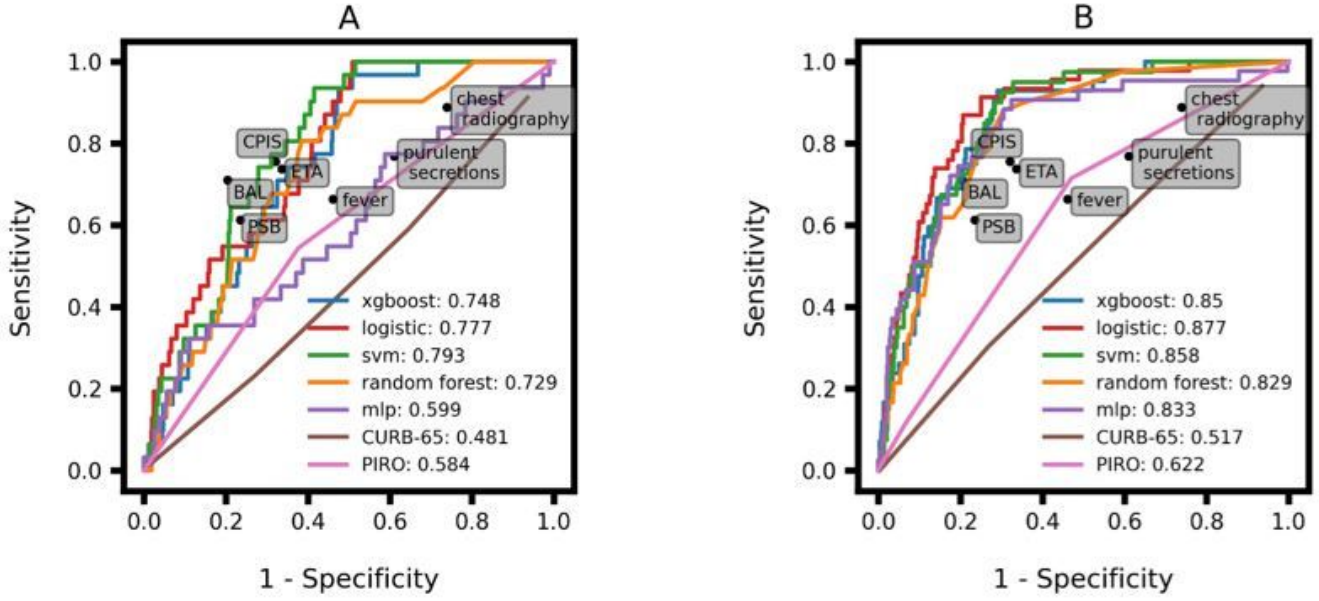

Figure 3

ROC curve comparison for admission task models with summary statistics calculated from the 6 (A) and 48 (B) hours of data preceding the time of prediction.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials.docx