 doi:10.20944/preprints202206.0149.v1

Article

Early Prediction of Ventilator-Associated Pneumonia in ICU Patients Using An Interpretable Machine Learning Algorithm

Maxime Faucher BS[1], Satish Casie Chetty PhD[1*], Sepideh Shokouhi PhD[1], Gina Barnes MPH[1], Keyvan Rahmani PhD[1], Jacob Calvert PhD[1] and Qingqing Mao PhD[1]

[1] Dascena, Inc., Houston, TX, United States
* Correspondence: dchetty@dascena.com

Abstract: (1) Background: Ventilator-associated pneumonia (VAP) causes high mortality among patients with respiratory disease and imposes major burdens on healthcare infrastructure. Models that use electronic health record data to predict the onset of VAP may spur earlier treatment and improve patient outcomes. We developed and studied the performance of interpretable machine learning (ML) models that predict the onset of VAP from electronic health records (EHRs); (2) Methods: We trained Logistic Regression (LR), full feature Explainable Boosting Machine (fEBM), and eXtreme Gradient Boosting (XGBoost) ML models on data from the MIMIC- III (v1.3) database. Model performance was measured by area under the receiver operating characteristic curves (AUCs). We trained a minimal-feature EBM model (mEBM) with features derived from white blood cell (WBC) counts, duration of ventilation, and Glasgow Coma Scale (GCS). Finally, model robustness was evaluated on randomly sparsified EHR datasets; (3) Results: The fEBM model outperformed the XGBoost and LR models at 24 hours post-intubation. The mEBM model maintained an AUC of 0.893. The fEBM model performance remained robust on sparsified datasets; (4) Conclusions: Our novel interpretable ML algorithm reliably predicts the onset of VAP in intubated patients. Integration of this EBM-based model into clinical practice may enable clinicians to better anticipate and prevent VAP.

Keywords: critical care; artificial intelligence ; predictive analytics; VAP; interpretable models

1. Introduction

Ventilator-associated pneumonia (VAP), which is defined as pneumonia that develops 48 hours or longer after endotracheal intubation,[1] is the second most common healthcare-associated infection in intensive care unit (ICU) patients.[2] Although the American Thoracic Society and Infectious Disease Society of America have established guidelines for the prevention and management of VAP,[1,3] there remains a significant burden on hospital systems due to the high incidence and mortality rate of VAP.[1] Incidence rates are estimated to be between 5%-67% of mechanically ventilated patients,[4] and mortality rates range from 13% in the US[1] to 31% in Europe.[5] VAP results in prolonged hospitalization and increased healthcare costs of approximately $40,000 per patient.[6,7] Accordingly, innovative approaches to prevent VAP and to mitigate the severity of infection are a priority.

The challenges of preventing, identifying, and treating VAP relate to variability in the pathogens that cause VAP, limited non-antibiotic therapeutics for prophylaxis against VAP, and a lack of a standard diagnostic strategy for VAP.[8,9] Various machine learning (ML) methods have attempted to bridge these gaps in care for VAP detection and prediction, drawing upon a range of data types including electronic health records (EHRs), physiological measurements, and genetic profiling.[10–13] ML-based clinical tools that use individualized patient data to make personalized predictions may allow healthcare providers to stratify risk and intervene earlier to prevent VAP, thus alleviating the concomitant financial burden on the healthcare system, as well as improving patient

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0149.v1

outcomes. However, it is typically difficult for healthcare providers to decipher why a ML model makes a individual prediction.[14,15] This may lead to ethical issues in patient care by introducing bias in providers' final clinical determinations.[14] Though research on the use of ML and artificial intelligence in healthcare applications is not new, the lack of model interpretability remains a barrier to the incorporation of these tools into clinical practice.

In this study, we deepened our previous work on VAP prediction[10] by developing and evaluating three ML models for the prediction of VAP onset among intubated ICU patients. Our primary objective was to construct a ML model that was more accurate, interpretable , and robust, and provided early prediction of VAP onset. We explored the benefits and shortcomings of different balances of complexity and interpretability in our algorithm design, and evaluated the best performing model's robustness by assessing its performance when trained on sparsified EHR data. We demonstrate that a relatively simple, interpretable model requiring only 3 feature inputs can accurately predict the onset of VAP in patients 24 hours after initial intubation.

2. Materials and Methods

2.1. Dataset Processing

Data used for training and testing the models were passively extracted from the Multiparameter Intelligent Monitoring Intensive Care (MIMIC)-III version 1.3 database. MIMIC-III contains de-identified EHRs from Beth Deaconess Medical Center ICU, Boston, Massachusetts, collected between 2001 and 2012.[16] The MIMIC-III dataset is publicly available at MIMIC-III Clinical Database v1.4 (physionet.org). This study used existing, de-identified data in compliance with the Health Insurance Portability and Accountability Act and did not require institutional review board approval or informed patient consent per 45 Code of Federal Regulations 46.102.

The inclusion/exclusion criteria consisted of encounters of adult (age ≥ 18 years) patients who were mechanically ventilated during their hospital stay and who had *some* clinical variable (any of the vitals and labs or Boolean indicators in **Table S1**) measured after the initiation of mechanical ventilation (MV) (**Figure 1**). Patients with a diagnosis of pneumonia at the time of admission were identified as community-acquired pneumonia (CAP) patients and were excluded from the study. Four corresponding sets of encounters were created in preparation of the four related prediction tasks, i.e., predicting a diagnosis of VAP on the basis of the first k hours of data following the initiation of MV, where k takes the values 12, 24, 36, and 48. The set of encounters for the task of predicting VAP onset after k hours consisted of all encounters meeting the preceding criteria, which lasted at least k hours following the initiation of MV.

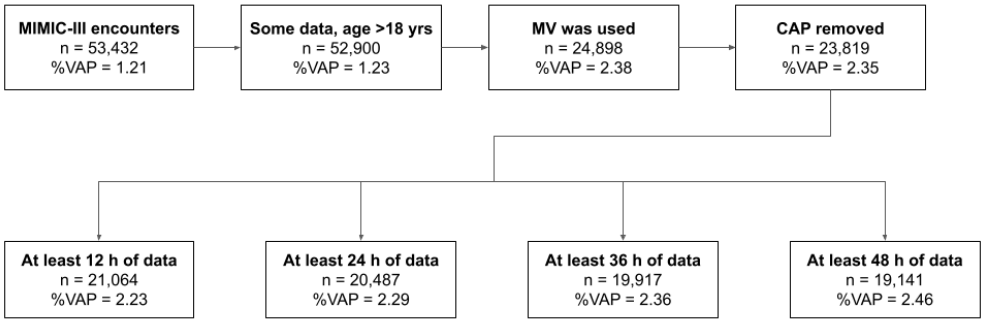

Figure 1. Flow chart of inclusion criteria, patient numbers, and prevalence of VAP. Patients were required to have at least one measurement of vital signs and lab tests. Abbreviations: N = number of patients; %VAP = percent of patients with VAP; MV=mechanical ventilation; CAP=community-acquired pneumonia; ICU=intensive care unit.

 doi:10.20944/preprints202206.0149.v1

2.2. Gold Standard

The presence of the International Classification of Disease revision nine (ICD-9) code 997.31 was used to identify encounters with VAP onset.[17] Due to the difficulties with retrospectively labeling the time of VAP onset, we were unable to assign a time of VAP onset to encounters with a diagnosis of VAP.

2.3. ML Tasks and Methods

The performance of three different ML models were developed and compared; an eXtreme Gradient Boosting (XGBoost) model,[18] a full feature Explainable Boosting Machine (fEBM) model,[19] and an Elastic Net Logistic Regression (LR) model. XGBoost is a popular and versatile decision tree–based ensemble learning method that has been applied to a variety of clinical prediction tasks.[20–24] EBM is designed to perform as accurately as state-of-the-art opaque models such as XGBoost while remaining highly interpretable.

An XGBoost, fEBM, and LR model were trained for each of the four prediction tasks mentioned above: Given encounter data from the first k hours following intubation, predict—at the k^{th} hour—whether or not VAP will be diagnosed at any later time during the encounter (k=12, 24, 36, 48 hours). We emphasize that, because VAP is defined as pneumonia which develops at least 48 hours after intubation, predictions are nominally made *at least* 48 - k hours in advance of onset.

Encounter data were processed in the following way. Raw measurements of clinical variables were binned into 1-hour intervals and measurements of the same clinical variable were averaged within bins to produce a representative value for each hour. Six summary statistics—minimum, maximum, median, first, last, and average—were calculated for each representative vital sign and lab test over the time window of k hours following intubation. The data were partitioned via random stratified split into a set for training and hyperparameter tuning (80% of encounters) and a hold-out validation set (20% of encounters). Missing measurements were imputed using median imputation, applied separately to the training and testing encounters to ensure there was no data leakage. All features were normalized to take values between zero and one.

All models were implemented in Python. Model hyperparameters were chosen via grid search cross-validation with four folds. For LR, the type of penalty (e.g., L1, L2, Elastic Net), the Elastic Net mixing parameter, alpha (controls the strength of regularization), learning rate, and eta0 (initial learning rate) hyperparameters were tuned. For XGBoost, the maximum tree depth, lambda (controls the strength of L2 regularization), gamma (controls the reduction in loss needed to split a leaf), and column subsampling parameters were tuned. For fEBM, the parameters concerning learning rate, minimum samples per leaf, and maximum number of leaves were tuned. The source code for data processing and model training is provided in the Supplementary Materials (Code S1).

Across all time windows, a total of 12 models were developed and their performance was assessed using the area under the receiver operating characteristic (AUC) curve. The XGBoost, fEBM, and LR model which performed best, in an average sense, across values of k, was chosen to create a *minimal input* model.

2.4. Minimal Input Model

Following the identification of fEBM as the best-performing model, its most important features were identified by ranking them in terms of the percentage of absolute total score that they contributed to test set predictions. By inspection, the overwhelming majority of important features were derived from just three clinical variables—white blood cell count (WBC), MV hours, and Glasgow Coma Scale (GCS). A new mEBM model was then trained with features derived from just these variables. The performance of the minimal input model was assessed at the previously used time windows of predefined length.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0149.v1

2.5. Robustness of the model

The robustness of the fEBM model to missing measurements or less frequent measurement of clinical variables was evaluated by comparing the performance of the fEBM models trained on non-sparsified data to their performance when trained on the sparsified data, when applied to sparsified datasets. All clinical variables were sparsified, except for demographic information, MV hours, and comorbidity data. To sparsify the data, we returned to the data processing step and, before binning measurements and computing summary statistics for the clinical variables, each measurement with probability r (the rate of sparsifying, which took values r = 0.25, 0.50, and 0.75) was deleted, independently of all other measurements. This resulted in three new sets of encounter data, which were then processed as before.

3. Results

3.1. Patient Characteristics

Out of 53,432 patient encounters in the dataset, there were 23,819 ICU encounters in which the patient was mechanically ventilated and had a VAP diagnosis. Of these, 21,064 had 12 hours of patient data, 20,487 had 24 hours of data, 19,917 had 36 hours of data, and 19,141 had 48 hours of data. Among the inclusion criteria, the requirement of MV excluded the greatest number of encounters (from 52,900 to 24,898) and was associated with an increase in VAP prevalence of roughly 1%, compared to the VAP prevalence among MIMIC-III encounters. **Figure 1** illustrates the flow chart of inclusion criteria with patient counts and VAP prevalence for each time window of patient data**.** Patient demographics in the dataset with k = 48 h of data are shown in **Table 1** (demographics of patients with k = 12, 24 & 36 h of data are presented in **Table S2**).

Table 1. Patient demographics for k = 48 h. Abbreviations: VAP = ventilator-associated pneumonia.

	Characteristic	**VAP Positive** n = 470	**VAP Negative** n = 18671
Age	<30	26 (5.5%)	706 (3.8%)
	30-49	78 (16.6%)	2647 (14.2%)
	50-59	98 (20.9%)	3403 (18.2%)
	60-69	99 (21.1%)	4402 (23.6%)
	70-79	95 (20.2%)	4303 (23%)
	80+	74 (15.7%)	3210 (17.2%)
Gender	Male	282 (60%)	11183 (59.9%)
	Female	188 (40%)	7488 (40.1%)
Ethnicity	White	314 (66.8%)	13352 (71.5%)
	Black/African-American	43 (9.1%)	1345 (7.2%)
	Asian	18 (3.8%)	395 (2.1%)
	Hispanic/Latino	12 (2.6%)	590 (3.2%)
	Unknown/Other	83 (17.7%)	2989 (16%)

3.2. Comparison of model performance at different prediction windows

Performance of the LR, fEBM, and XGBoost models for the 12, 24, 36, and 48 hour windows (12 models in total) was measured by AUC (**Figure 2**). Overall, the performance of all models increased with increasing time windows. In terms of model comparison by AUC, the fEBM model outperformed the LR models at all timepoints and had slightly better performance than the XGBoost models at 12, 24 and 36 h. The exception was the 48-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0149.v1

hour window, where XGBoost had marginally better performance than fEBM. The performance of the models was also evaluated by the area under the precision recall curves (AUPRCs); the EBM model had slightly better performance than the XGBoost model at the 24 h timepoint (Supplementary Figure 1).

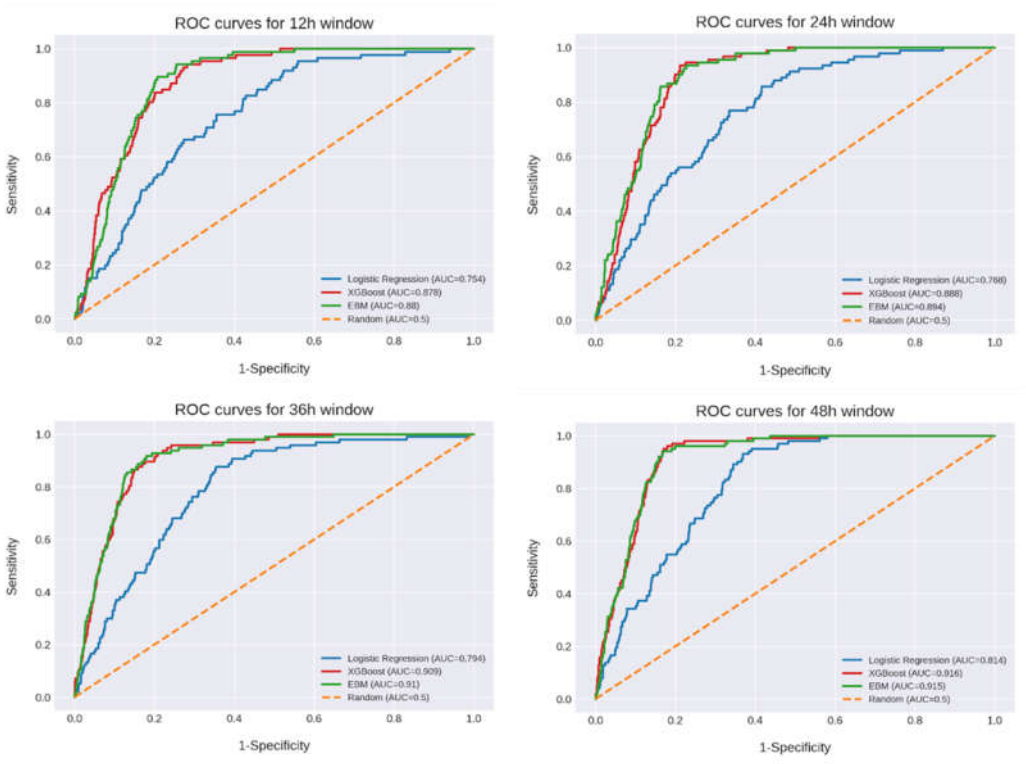

Figure 2. ROC curves for LR, fEBM and XGB models at k = 12, 24, 36, 48 hours. Abbreviations: ROC - Receiver Operating Characteristic; fEBM - full feature Explainable Boosting Machine; XGB - XGBoost; AUC - Area Under the Curve.

3.3. Feature Importance and Minimal Feature Model

The feature importance plot for the fEBM model at k = 24 h (**Figure 3**) shows the summary statistics calculated from WBC count, MV hours and GCS to be the most important features contributing to the prediction of VAP. Pairwise interactions were observed between MV hours, WBC, blood count and weight at k = 24 h. Similar features and pairwise interactions were seen at the other time windows analyzed (**Figure S2**).

Using only features derived from WBC, MV and GCS, an mEBM model was developed that had performance comparable to the fEBM model at all time windows **(Figure 4**). This is referred to as the minimal or few feature EBM model (mEBM), although it is perhaps more accurate to say that it uses a relatively minimal set of clinical variables- WBC first, MV hours (value) and GCS last. Plots of predicted VAP risk versus the value of three representative features respectively derived from WBC, MV hours, and GCS are provided in Figure S3. These plots highlight the complicated, nonlinear relationships mEBM learns between the feature values and risk of VAP onset.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0149.v1

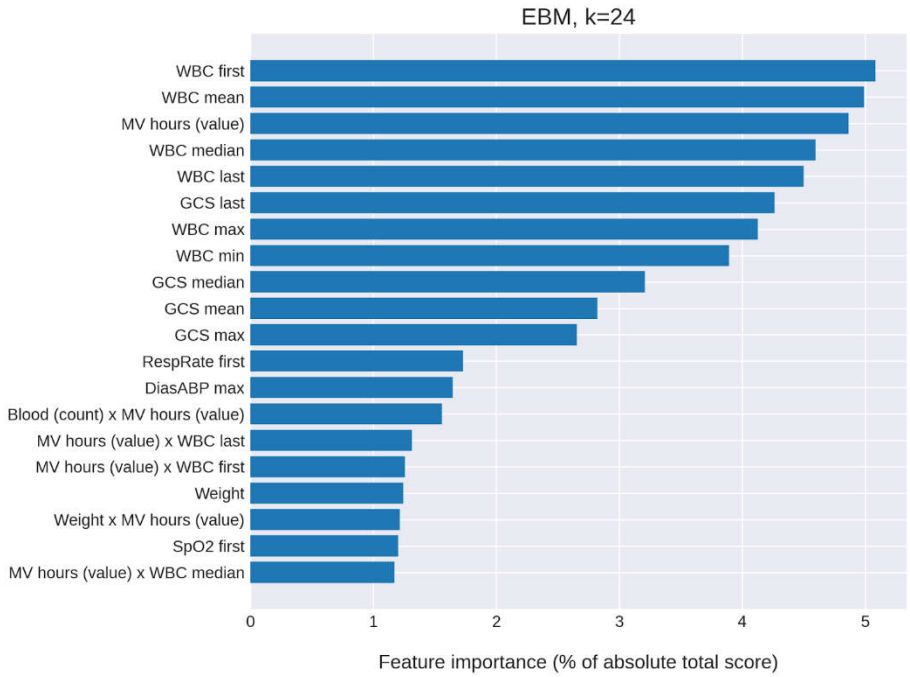

Figure 3. Feature importance plot for fEBM at *k*=24 h. Abbreviations: MV: mechanical ventilation; WBC: White Blood Cell ; GCS: Glasgow Coma Scale ; ResRate: Respiratory Rate; DiasABP: Diastolic Blood Pressure; SpO_2:Peripheral Capillary Oxygen Saturation.

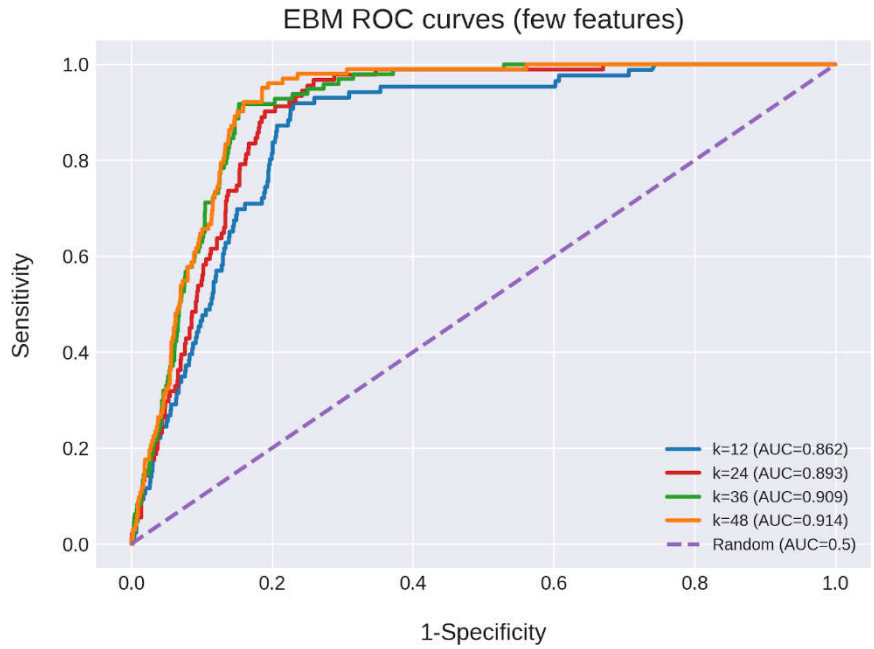

Figure 4. Minimal-feature EBM (mEBM) model at all time prediction windows. Abbreviations: ROC: Receiver Operating Characteristic; EBM: Explainable Boosting Machine; AUC: Area Under the Curve.

3.4. Robustness of the model

When the fEBM model was trained separately on non-sparsified data and on data sparsified at rate *r* (*r* = 0.25, 0.50, or 0.75), and tested on data sparsified at rate *r*, the differences in AUCs were small. Results for the full feature model are shown in **Table 2**; results for the mEBM model showed a very similar pattern . The table shows that, despite deleting up to 75% of all measurements of the clinical variables (except for MV hours and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0149.v1

demographic information) in the training set, both EBM models performed essentially as well as they did when applied to the original data.

Table 2. Comparison of AUCs for fEBM models trained with k = 24 h on non-sparsified and sparsified data, and tested on sparsified data (rates of r = 0.25, 0.50, and 0.75). Abbreviation: AUC - Area Under the Receiver Operating Characteristic. .

		Rate of data sparsifying, r			
		0	0.25	0.50	0.75
AUC	Model trained on non-sparsified data	0.893	0.893	0.889	0.891
	Model trained on sparsified data	0.893	0.892	0.890	0.889

4. Discussion

The objective of this study was to build upon our previous work[10] by developing a more interpretable and clinically relevant ML model to accurately predict the onset of VAP among intubated ICU patients. Previously, we have developed models that predicted VAP using data collected within 48 hours post-intubation.[10] Here we demonstrated that a relatively simple and more interpretable model can outperform more complex and less interpretable models, using shorter time windows of prediction data. Our results also demonstrate that the model is resilient to perturbations in the test data and maintains robust performance when validated on sparse datasets, indicating that although it was developed on a dataset of ICU encounters, the model may generalize to less intensive care settings.

Prediction of VAP may provide adequate warning to clinicians before its onset and using individualized data may broaden clinicians' options for providing preventive measures and for personalizing such strategies based on a patient's risk profile. For example, intubated patients receiving treatment for cardiac arrest may have a reduced risk of developing VAP if they are given a two-day course of antibiotics prior to VAP development[5] and other patient populations may get preventative benefits from the administration of probiotics,[6] nebulized antibiotics,[7] or other therapeutics, such as N-acetyl-cysteine.[8] Prevention of VAP prior to its occurrence may also encourage antibiotic stewardship, which is imperative to reduce mortality resulting from VAP caused by multidrug resistant bacteria and inappropriate administration of antibiotics.[25–27]

Our study advances the field in several key ways. In our prior research, we developed an ML model that utilizes readily available EHR data to make predictions of VAP at the 48th hour of intubation, the earliest time at which VAP can be diagnosed.[10] The previous study used windows of vital sign and lab data (6, 12, 18 hours etc.) calculated backward from the 48th hour and made a prediction of VAP at the 48th hour. While that model may be useful in classifying patients at risk of developing VAP beyond the 48th hour, its utility may be improved by providing an earlier prediction to allow time for preventative measures to be taken. To address this, we developed our models to make predictions of VAP after different time windows of post-intubation data collection, from 12 to 48 hours. We observed only a marginal increase in performance with increasing time windows of data collection. The benefit afforded to patients by an early prediction may outweigh the slight improvement in predictive performance of the model at a later time point. We demonstrate that a prediction made at 24 h post-intubation strikes a good balance between the advantage of an early prediction and model performance. An early, accurate prediction at 24 h would allow clinicians to more closely monitor and implement prophylactic measures for patients predicted to develop VAP.

One critical component of the utility of our model in clinical settings is its ability to reliably outperform previously developed methods to predict imminent onset of VAP in intubated patients. These methods include risk scores used to guide these predictions—e.g., the Clinical Pulmonary Infection Score (CPIS), the VAP predisposition, insult, response, and organ dysfunction (PIRO) score, and the VAP Acute Physiology and

 doi:10.20944/preprints202206.0149.v1

Chronic Health Evaluation II (APACHE-II) score—as well as previously published ML-based methods.[2,12,28–31] This outperformance is particularly notable because of our models' relative simplicity and availability of input data compared to those used in several of these previous studies. This improves the potential of our model to be adopted and routinely used by healthcare facilities. When compared to our own previously published model for predicting VAP, this updated approach consistently predicted diagnosis more accurately despite increased constraints on time and quantity of input data.[10]

A remarkable characteristic of the mEBM model is that it uses features that can be derived from the observations of just three clinical variables: GCS (last), MV hours, and WBC (first) (**Figure 3**, **Suppl. Figure 3**). These clinical variables are known to be associated with increased risk of VAP.[30–32] For example, WBC indicates a strong immune response to severe infection and is a common sign of VAP[30–32] and GCS is a known proxy for aspiration risk and subsequent acquisition of pneumonia.[31] In this sense, the mEBM model is both simple and clinically plausible, and it is natural to wonder if this model can be reduced to existing wisdom or a score that can be calculated by hand. In fact, the individual relationships between these features and VAP risk can be extremely complex (**Figure S3**). Because of this result and the relatively poor performance of the simpler LR models (**Figure 2**), it seems unlikely that the mEBM model could be appreciably simplified without a substantial loss of performance.

Moreover, it seems that there is no benefit from using a more complicated model than either a mEBM or fEBM model. The evidence for this claim is provided partly by the inability of the XGBoost models to consistently outperform the EBM models (**Figures 2** and **4**) and partly by the relatively small contributions from pairwise feature interactions to fEBM (**Figure 3**). In principle, due to the hyperparameters used in our experiments, XGBoost explored a class of candidate models which contained those explored by both EBM models. In a sense, XGBoost could incorporate interactions between up to five features, while EBM models can only consider pairwise interactions between features. Nevertheless, XGBoost and mEBM and fEMB performed roughly equally well.

One of the major advantages of our model is its high interpretability. Although complex algorithms like XGBoost, Random Forests and Artificial Neural Networks have gained increasing popularity and widespread use in predictive modeling, they suffer from a lack of intelligibility. Unlike these opaque models, EBM (considered to be a *glassbox* model) is both "globally" interpretable, in the sense that the relationships between features and predicted risk are explicit (e.g., **Figure S3**), and "locally" interpretable, in the sense that it can produce feature importances for individual predictions.[19,32] This has significant implications in a healthcare setting, since it would give clinicians a more comprehensive understanding of what features contribute most to a prediction for a given patient. Importantly, the increased interpretability of the model does not come at the cost of accuracy; our fEBM model has comparable or better performance than the more complex XGBoost model at all data collection windows (**Figure 2**). It is also worth noting that the difference in performance of the fEBM model between the 24 h (AUC = 0.894) and 48 h (AUC = 0.915) time windows is small.

With the widespread use of ML in real-world applications, there has been an increasing effort to develop tools and methods to ensure the resilience of classifiers in the face of perturbations to test data.[33] Documentation of data relevant to healthcare-associated infections in hospital systems are often inconsistent,[34] posing a major challenge when developing and implementing ML models with inputs reliant on EHR data.[15] To account for potential missingness of data in real world settings, we performed random data sparsification at varying degrees on all clinical variables except MV hours and demographic information (since they are routinely collected at all institutions), and assessed the performance of our model. Our results demonstrate high performance of the model upon random data sparsification (**Table 2**), providing evidence that our model would have robust performance in a variety of hospital ICU settings and potentially settings with less intensive care.

 doi:10.20944/preprints202206.0149.v1

There are several limitations to this study. First, our algorithms were designed only to predict intubated patients' acquisition of VAP and not the severity of the disease or risk of subsequent mortality; this prediction also did not stratify patients by degree of risk (e.g. high versus low, odds, or percentiles). Second, the ICD-9 code for VAP (997.31) was first introduced in 2008, at the beginning of the time span during which MIMIC-III data were collected. Since the code had not been widely recognized and implemented for a long period of time prior to data collection, some diagnoses of VAP from earlier years in the dataset may not have been encoded. There are also limitations to the accuracy of studying healthcare-associated infections based on administrative coding data, with studies contemporary to MIMIC-III data collection reporting only moderate positive predictive values for correct diagnosis coding.[35,36]

Future developments on this work should include developing and validating algorithms on datasets where correct coding or diagnosis of infection has been confirmed via physician adjudication. They may also benefit from including more refined data on patients' infections, e.g. the pathogenic agent causing the infection, the location of the infection within the patient's respiratory tract, or the phenotypic antibiotic resistance profile of the pathogen itself. This may enhance the value of the algorithm as a clinical decision support tool, as it may guide treatment strategies. Following retrospective research, the clinical utility of this tool would need to be evaluated in a real-world setting by conducting a prospective clinical study in collaboration with physicians at multiple medical institutions.

5. Conclusions

We present a machine learning–based approach to reliably predict future diagnosis of VAP among intubated patients, using data collected early in the course of intubation. This approach improves upon our previous work in the field, outperforming our own and others' previously published models. If prospectively validated, this model could provide clinicians with an interpretable and robust tool to assist in the prevention of VAP and the substantial financial burdens it incurs. Future directions in this field include refining both the criteria for inclusion or exclusion of input data and the balance of simplicity versus interpretability, to provide clinicians a tool that is as accurate and integratable as possible.

Supplementary Materials: Table S1: Feature inputs for models; Table S2: Demographic information of patients included in time windows of k = 12, 24, 36 hours; Figure S1: Precision recall (PR) curves for LR, fEBM, and XGB models at k = 12, 24, 36, 48 hours; Figure S2: Feature importance plots for the fEBM model at k = 12, 36, 48 hours; Figure S3: Contribution of the top features to fEBM scores; Code S1: Source code for data processing and model training.

Author Contributions: Conceptualization, Satish Casie Chetty PhD, Maxime Faucher and Jacob Calvert; Data curation, Maxime Faucher, Keyvan Rahmani and Jacob Calvert; Formal analysis, Maxime Faucher, Keyvan Rahmani and Jacob Calvert; Investigation, Satish Casie Chetty PhD, Maxime Faucher, Sepideh Shokouhi, Keyvan Rahmani, Jacob Calvert and Qing Qing Mao; Methodology, Maxime Faucher and Qing Qing Mao; Supervision, Qing Qing Mao; Visualization, Maxime Faucher and Jacob Calvert; Writing – original draft, Satish Casie Chetty PhD, Sepideh Shokouhi, Gina Barnes and Jacob Calvert; Writing – review & editing, Satish Casie Chetty PhD, Sepideh Shokouhi, Gina Barnes, Keyvan Rahmani and Jacob Calvert. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Institutional Review Board Statement: This study used existing, de-identified data in compliance with the Health Insurance Portability and Accountability Act and did not require institutional review board approval or informed patient consent per 45 Code of Federal Regulations 46.102.

Data Availability Statement: Data used in this study was obtained from the publicly available MIMIC-III dataset. The source code for data processing and model training is available in Supplementary Code S1.

Acknowledgments: We gratefully acknowledge Zohora Iqbal PhD and Misty Attwood PhD for their support in preparing this manuscript.

Conflicts of Interest: All authors who have affiliations listed with Dascena (Houston, Texas, U.S.A) are employees or contractors of Dascena. J. Calvert and Q. Mao have stock options at Dascena.

References

1. Kalil AC, Metersky ML, Klompas M, et al. 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: e61–e111.
2. Goodwin TR, Demner-Fushman D. Deep Learning from Incomplete Data: Detecting Imminent Risk of Hospital-acquired Pneumonia in ICU Patients. *AMIA Annu Symp Proc* 2020; 2019: 467–476.
3. Tablan OC, Anderson LJ, Besser RE, et al. Guidelines for preventing health-care-associated pneumonia, 2003; recommendations of CDC and the Healthcare Infection Control Practices Advisory Committee;[pt. II-III].
4. Timsit J-F, Esaied W, Neuville M, et al. Update on ventilator-associated pneumonia. *F1000Research*; 6. Epub ahead of print 2017. DOI: 10.12688/f1000research.12222.1.
5. Martin-Loeches I, Rodriguez AH, Torres A. New guidelines for hospital-acquired pneumonia/ventilator-associated pneumonia: USA vs. Europe. *Curr Opin Crit Care* 2018; 24: 347–352.
6. Muscedere JG, Day A, Heyland DK. Mortality, attributable mortality, and clinical events as end points for clinical trials of ventilator-associated pneumonia and hospital-acquired pneumonia. *Clin Infect Dis* 2010; 51: S120–S125.
7. Kollef MH, Hamilton CW, Ernst FR. Economic impact of ventilator-associated pneumonia in a large matched cohort. *Infect Control Hosp Epidemiol* 2012; 33: 250–256.
8. Depuydt P, De Bus L. Controversies in Ventilator-Associated Pneumonia Diagnosis. *ICU Manag Pract*; 16, https://healthmanagement.org/c/icu/issuearticle/controversies-in-ventilator-associated-pneumonia-diagnosis (2016).
9. Guidelines for the Management of Adults with Hospital-acquired, Ventilator-associated, and Healthcare-associated Pneumonia. *Am J Respir Crit Care Med* 2005; 171: 388–416.
10. Giang C, Calvert J, Rahmani K, et al. Predicting ventilator-associated pneumonia with machine learning. *Medicine (Baltimore)* 2021; 100: e26246.
11. Liao Y-H, Wang Z-C, Zhang F-G, et al. Machine Learning Methods Applied to Predict Ventilator-Associated Pneumonia with Pseudomonas aeruginosa Infection via Sensor Array of Electronic Nose in Intensive Care Unit. *Sensors*; 19. Epub ahead of print 2019. DOI: 10.3390/s19081866.
12. Chen C-Y, Lin W-C, Yang H-Y. Diagnosis of ventilator-associated pneumonia using electronic nose sensor array signals: solutions to improve the application of machine learning in respiratory research. *Respir Res* 2020; 21: 45.
13. Cai Y, Zhang W, Zhang R, et al. Combined Use of Three Machine Learning Modeling Methods to Develop a Ten-Gene Signature for the Diagnosis of Ventilator-Associated Pneumonia. *Med Sci Monit Int Med J Exp Clin Res* 2020; 26: e919035-1-e919035-13.
14. Grote T, Berens P. On the ethics of algorithmic decision-making in healthcare. *J Med Ethics* 2020; 46: 205–211.
15. Rose S. Machine Learning for Prediction in Electronic Health Data. *JAMA Netw Open* 2018; 1: e181404–e181404.
16. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016; 3: 160035.
17. Martini RP, Yanez ND, Treggiari MM, et al. Implementation of the TaperGuard™ endotracheal tube in an unselected surgical population to reduce postoperative pneumonia. *BMC Anesthesiol* 2020; 20: 211.
18. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery, pp. 785–794.
19. Nori H, Jenkins S, Koch P, et al. InterpretML: A Unified Framework for Machine Learning Interpretability. *ArXiv190909223 Cs Stat*, http://arxiv.org/abs/1909.09223 (2019, accessed 27 August 2021).
20. Le S, Pellegrini E, Green-Saxena A, et al. Supervised machine learning for the early prediction of acute respiratory distress syndrome (ARDS). *J Crit Care* 2020; 60: 96–102.
21. Allen A, Mataraso S, Siefkas A, et al. A Racially Unbiased, Machine Learning Approach to Prediction of Mortality: Algorithm Development Study. *JMIR Public Health Surveill* 2020; 6: e22400.
22. Lam C, Siefkas A, Zelin NS, et al. Machine Learning as a Precision-Medicine Approach to Prescribing COVID-19 Pharmacotherapy with Remdesivir or Corticosteroids. *Clin Ther*. Epub ahead of print 29 March 2021. DOI: 10.1016/j.clinthera.2021.03.016.
23. Radhachandran A, Garikipati A, Zelin NS, et al. Prediction of short-term mortality in acute heart failure patients using minimal electronic health record data. *BioData Min* 2021; 14: 23.
24. Rahmani K, Garikipati A, Barnes G, et al. Early prediction of central line associated bloodstream infection using machine learning. *Am J Infect Control*; 0. Epub ahead of print 20 August 2021. DOI: 10.1016/j.ajic.2021.08.017.
25. Cotoia A, Spadaro S, Gambetti G, et al. Pathogenesis-Targeted Preventive Strategies for Multidrug Resistant Ventilator-Associated Pneumonia: A Narrative Review. *Microorganisms* 2020; 8: 821.
26. Arayasukawat P, So-ngern A, Reechaipichitkul W, et al. Microorganisms and clinical outcomes of early- and late-onset ventilator-associated pneumonia at Srinagarind Hospital, a tertiary center in Northeastern Thailand. *BMC Pulm Med* 2021; 21: 47.
27. Hosamirudsari H, Forghani S, Akbarpour S. Multi-drug resistant ventilator associated pneumonia: risk factors and outcomes. *Can J Infect Control* 2018; 33: 5.

28. Schurink CAM, Nieuwenhoven CAV, Jacobs JA, et al. Clinical pulmonary infection score for ventilator-associated pneumonia: accuracy and inter-observer variability. *Intensive Care Med* 2004; 30: 217–224.
29. Erratum. *Clin Infect Dis* 2010; 51: 1114–1114.
30. Furtado GH, Wiskirchen DE, Kuti JL, et al. Performance of the PIRO Score for Predicting Mortality in Patients with Ventilator-Associated Pneumonia. *Anaesth Intensive Care* 2012; 40: 285–291.
31. Schnabel RM, Boumans MLL, Smolinska A. Electronic nose analysis of exhaled breath to diagnose ventilator-associated pneumonia. *Respir Med* 2015; 109: 1454–1459.
32. Caruana R, Lou Y, Gehrke J, et al. 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. Epub ahead of print 2015. DOI: 10.1145/2783258.2788613.
33. Sehwag V, Bhagoji AN, Song L, et al. Analyzing the Robustness of Open-World Machine Learning. In: *Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security - AISec'19*. London, United Kingdom: ACM Press, pp. 105–116.
34. Stamm AM, Bettacchi CJ. A comparison of 3 metrics to identify health care-associated infections. *Am J Infect Control* 2012; 40: 688–691.
35. Stevenson KB, Khan Y, Dickman J. Administrative coding data, compared with CDC/NHSN criteria, are poor indicators of health care-associated infections. *Am J Infect Control* 2008; 36: 155–164.
36. Verelst S, Jacques J, Van Den Heede K. Validation of Hospital Administrative Dataset for adverse event Screening. *Qual Saf Heal Care*; 19. Epub ahead of print 2010. DOI: 10.1136/qshc.2009.034306.