Zappalà et al. J Anesth Analg Crit Care (2026) 6:23
https://doi.org/10.1186/s44158-025-00313-3

Journal of Anesthesia, Analgesia and Critical Care

RESEARCH

Open Access

Real-time prediction of ventilator-associated pneumonia onset in ICU: development of a dynamic machine learning model

Simone Zappalà[1], Lucrezia Rovati[2], Francesca Alfieri[1], Andrea Ancona[1], Amedeo Guzzardella[3], Giacomo Grasselli[3,4*] and Vittorio Scaravilli[3,5]

Abstract

Background Ventilator-associated pneumonia (VAP) is a preventable complication of invasive mechanical ventilation (IMV) with a significant healthcare impact. Early risk prediction is crucial, but current models lack real-time adaptability. This study develops a real-time VAP prediction model using machine learning and high-resolution EHR data from the MIMIC-III database.

Methods We analyzed 3523 ICU stays (3204 patients) from MIMIC-III, including adults who received IMV for at least 48 h. VAP was labeled based on microbiological cultures and antibiotic initiation. A real-time ensemble model of XGBoost regressors was developed to predict time to VAP onset, incorporating vital signs, ventilator data, and lab results. Two static classifiers (24 h and 48 h) were also compared.

Results VAP occurred in 595 ICU stays (16.89%), with an incidence rate of 23.77 per 1000 IMV-days. Median VAP onset was 113.5 h post-IMV. The real-time model outperformed static models with a C-index of 0.68, AUROC of 0.71, and AUPRC of 0.36. It provided a median lead time of 53 h before VAP onset, with key predictors including temperature, respiratory rate, and minute ventilation.

Conclusion We present a real-time VAP prediction model that outperforms static classifiers, providing actionable lead time for proactive microbiological surveillance. The model enables risk stratification for enhanced monitoring and, when clinically indicated, timely targeted antimicrobial therapy. Future work will focus on multicenter prospective validation and integration into ICU workflows to assess clinical utility and impact on patient outcomes.

Keywords Real-time prediction, Ventilator-associated pneumonia (VAP), Pneumonia

Background

Ventilator-associated pneumonia (VAP) is the second most common healthcare-associated infection, with an incidence rate of 1.1–11.9/1000 mechanical ventilator (MV)-days with a huge clinical, environmental, and economic burden [1]. VAP doubles the intensive care unit (ICU) length of stay and risk of death, leading to increased costs of care. The treatment of VAP is responsible for more than half of the antibiotics administered in the ICU, making VAP one of the most impactful causes of antibiotic resistance worldwide.

*Correspondence:
Giacomo Grasselli
giacomo.grasselli@unimi.it
[1] U-Care Medical srl, Turin 10129, Italy
[2] School of Medicine and Surgery, University of Milano-Bicocca, Milan 20126, Italy
[3] Department of Anesthesia and Critical Care, Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico, Milan 20019, Italy
[4] Department of Pathophysiology and Transplantation, University of Milan, Milan 20122, Italy
[5] Department of Biomedical Surgical and Dental Sciences, University of Milan, Milan 20122, Italy

Comprehensive surveillance and early-detection protocols are critical for mitigating VAP, implementing targeted therapeutic interventions, and attenuating its morbidity and healthcare system burden. Conventional diagnostic approaches demonstrate significant methodological limitations in sensitivity and specificity. In this context, machine learning (ML) may be a promising solution by analyzing real-time data from electronic health records (EHRs), including vital signs and lab results. To date, no real-time ML models for early VAP detection have been developed [2]. Existing ML models rely on clinical data and administrative codes that fail to capture the dynamic clinical condition of critically ill patients, leading to significant underreporting of VAP cases [3–5].

This study addresses these gaps by developing and validating a real-time ML model for VAP prediction in a general ICU cohort of critically ill adult patients undergoing invasive mechanical ventilation (IMV). Leveraging a comprehensive dataset from the MIMIC-III ICU database, we aimed to (1) establish a robust, clinically interpretable framework for real-time VAP risk assessment in a general ICU cohort, providing clinicians with continuous risk evaluation to support early detection, targeted surveillance, and timely intervention, and (2) demonstrate the advantages of real-time monitoring over traditional static models in improving decision-making and patient outcomes.

Methods

Data source

This study utilized the MIMIC-III dataset [6], version 1.4 [7], a publicly available resource including de-identified health-related data from 61,532 admissions of 46,476 distinct patients treated at Beth Israel Deaconess Medical Center in Boston, MA, USA, between 2001 and 2012. Of note, other open-source datasets (e.g., AmsterdamUMCdb, eICU, and HiRID) were not utilized, given the lacking robustness of microbiological data or the complete lack thereof.

The data were selected for their real-world, diverse ICU population, suitable for developing generalizable models despite being from a single center. A stratified random split was used for development and evaluation, ensuring balance for VAP onset. The split was performed at the ICU stay level to avoid data leakage. Given the MIMIC-III dataset's inclusion of over 60,000 ICU admissions, using the entire dataset maximized statistical power and ensured broad ICU population representation.

As this study was based on de-identified, publicly available MIMIC-III data, ethical approval was not required, no study protocol was developed, and the study was not registered.

The full code cannot be released due to proprietary constraints and a pending patent (European no. 24202572.4). Interested researchers may contact the corresponding author to request a demo for external validation.

Improved labeling approach

Ventilator data transmitted to the EHR allowed each hour of an ICU stay to be classified as "ventilated" or "not ventilated." We further distinguished between invasive and noninvasive ventilation, as previously documented [8]. Continuous hours of IMV without breaks ≥ 48 h were defined as unique IMV events [9].

Microbiologically confirmed cultures from the lower respiratory tract (i.e., bronchoaspirate or bronchoalveolar lavage) leading to antimicrobial regimen introduction or modification were considered for VAP labeling. The first concurrent collection of positive cultures and antimicrobial administration marked VAP onset, following SEPSIS-3 [10]. Both pathogens and antimicrobials were selected from a predefined list (see supplementary materials). Infectious episodes were thus labeled as VAP if their onset occurred after > 48 h of IMV and < 72 h following the IMV event termination. Only the first VAP episode was considered for the analyses, and patients were categorized into VAP-positive and VAP-negative cohorts.

Patient population and inclusion criteria

We selected all adults (age ≥ 18) who underwent IMV for > 48 h. Patients treated with extracorporeal membrane oxygenation were excluded.

Statistical analysis

To validate the accuracy of our VAP labeling, we compared patients' cohorts, considering demographics, comorbidities, reasons for ICU admission, length of hospital stay, inhospital mortality, and IMV duration. Data are shown as median [interquartile range]. We used the Mann–Whitney *U*-test for continuous variables and chi-square or Fisher's exact tests for categorical variables. We applied Weibull regression to compare 90-day mortality. To identify VAP risk factors, we used Fine-Gray competing-risks regression with death as the competing event and demographics, comorbidities, and admission reasons as covariates. Sub-hazard ratios (sHR) and 95% confidence intervals (CI) (sHR, [CI95%]) were computed for each covariate. Additionally, cumulative incidence functions (CIFs) were plotted. Significance was set at $p < 0.001$.

Medical variable selection and feature engineering

We selected medical variables based on their clinical relevance in critically ill ventilated patients. Values outside predefined intervals were discarded (Table S1).

To capture intra-hour variability in high-frequency vital signs, we calculated within-hour mean, minimum, maximum, and standard deviation values, producing an hourly time series for each original vital sign and its derived metrics. We then summarized these hourly measurements over multiple time windows (e.g., the past 6, 12, and 24 h) to capture both acute changes and longer-term trends in the patient's condition. This approach, similar to how clinicians consider both recent vital sign changes and overall trajectory when assessing patient stability, has been used in prior healthcare prediction models [3]. As the patient's clinical course evolves, we recalculate these features every hour, so each prediction reflects the most current assessment of their condition.

Models development and training

We developed three distinct models for VAP prediction:

- A 24-h single-shot model, predicting VAP using data collected after 24 h of ventilation
- A 48-h single-shot model, predicting VAP using data collected after 48 h of ventilation
- A real-time regression model, providing continuous hourly time-to-VAP predictions and thus estimating the time remaining until VAP onset.

The single-shot models served as baselines. They were fed by the same feature engineering pipeline as the real-time model, leveraging all the medical variables commonly used in risk assessment scores (e.g., SOFA, APACHE) while excluding the Glasgow Coma Scale (GCS), comorbidities, and vasopressor therapy due to their reliance on clinical notes.

The real-time regression model treats each hour beyond the 48-h from intubation as a prediction time point. At every hour, the model evaluates the patient's current clinical state, with the target time (i.e., hours) until the VAP event.

Our model prevents data leakage by defining VAP strictly through objective criteria—positive cultures and antibiotic initiation—ensuring no predictors serve as proxies for the outcome. Additionally, outcome assessment remains blinded as no clinical notes or predictor variables influence the labeling process.

To handle the complexity of VAP prediction and address the class imbalance, we constructed an ensemble of boosted trees. Each model was trained on a balanced subset of ICU stays to improve accuracy and minimize overfitting. The models were trained on separate datasets with distinct training and validation splits, ensuring diverse learning configurations within the ensemble.

We used recursive feature elimination (RFE) with k-fold cross validation, where folds were created at the ICU stay level to prevent data leakage, ensuring that each ICU stay appeared in only onefold. Early stopping was applied if the validation performance plateaued. Each ensemble member produced an independent prediction. We averaged these predictions to form the final score. Training used balanced sampling, label perturbation, and feature pruning to improve stability. The full list of derived features used in the model can be found in Table S2 (see Online Supplement). None of these features requires subjective interpretation, as all are based on objective, quantitatively measured clinical data.

Evaluation metrics

To assess regression performance for close-to-event cases, we calculated the mean absolute error (MAE) and root-mean-squared error (RMSE) for VAP occurring within 72 h. Additionally, we computed concordance indices (C-index) to measure the concordance between predicted times and observed events, accounting for censoring in cases where the event was not observed within the threshold.

Since the output ranges from 0 to 72 h until the next event (capped at 72), we transformed it using the following calculation: 1−prediction/72 to create a probability score for VAP risk, allowing direct comparison with baseline single-shot classifiers. We calculated its area under the receiver operating characteristic (AUROC) and area under the precision-recall curve (AUPRC) and 95% CI by bootstrapping.

We report performance at a fixed 70% sensitivity to illustrate a clinically meaningful trade-off: capturing most true VAP cases while containing alarm burden. This mirrors reporting practices in ICU early-warning literature and provides a clear benchmark for comparison. For implementation, the operating point should be site-specific and may be selected using decision-curve analysis to reflect local resource constraints and acceptable false-alarm rates (10.1126/scitranslmed.aab3719, 10.1016/j.compbiomed.2016.05.003,10.1177/0272989X06295361). We evaluated specificity, precision, F1 score, and time-to-event predictions at this threshold. Any alarms occurring after VAP onset were considered false negatives.

To address model fairness, we computed the average odds difference across demographic groups to ensure that the model's performance was consistent and unbiased. No recalibration on the validation cohort was employed.

Software and libraries

The analysis was conducted using Python (version 3.10.9) and R (version 4.4.2). For data manipulation, we used the *pandas* library in Python to process and prepare the dataset. Demographic analysis and statistical tests, including Fisher's exact test and Mann–Whitney *U* tests, were performed using the *SciPy* library in Python. Weibull regression analysis, Fine and Gray's competing risks model, and cumulative incidence function (CIF) curves were computed using the *cmprsk* package in R. Machine learning models were developed using *XGBoost*, a powerful gradient boosting framework recognized for its high performance, scalability, and built-in capability to manage missing values effectively. As tree-based models do not require rescaling or standardization of predictors, no transformations were applied to the input variables.

Results

Cohort characteristics

We analyzed 3523 ICU stays from 3204 patients in MIMIC-III. Inclusion criteria were adults (≥18 years), invasive mechanical ventilation (ECMO excluded), ≥48 h of ventilation, and available microbiology and antibiotic timestamps for VAP adjudication (see Fig. 1 for the full inclusion flowchart). Five hundred ninety-five (16.89%) VAP-positive cases were detected, with an incidence rate of 23.77 [21.80, 25.58] VAP/1000 IMV-days. The total number of prediction points was 146,600 for the training cohort and 38,290 for the validation cohort.

VAP occurred after a median time of 113.5 [72–197] IMV hours. The microbial agents responsible for the VAP episodes are shown in Table S3 (see Additional results). Gram-negative pathogens were more prevalent (72%), with Enterobacterales species being the most represented. Among gram-positive germs, *Staphylococcus aureus* was the most prevalent agent (93%).

Patients with VAP had a significantly longer LOS (18.68 vs. 10.17 days, $p<0.001$) and IMV duration (373.53 vs. 168.64 h, $p<0.001$). The Weibull regression model revealed that VAP was significantly associated with 90-day mortality, with a HR of 1.32 (1.18, 1.48), $p<0.001$. Among all covariates, cancer (*HR* 1.62 [1.45, 1.83], $p<0.001$) and a neurological reason for admission (*HR* 1.46 [1.27, 1.68], $p<0.001$) showed the greatest impact on 90-day mortality (Table S4).

Patients' cohort characteristics and Fine-Gray competing risks regression are summarized in Table 1. Older age, male gender, and Black or Afro-American patients had higher sHR for VAP. In contrast, cancer and congestive heart failure were associated with reduced sHR of VAP. Among reasons for admission, neurological and trauma were associated with significantly increased sHR of VAP (see Figs. S1, S2, S2, S3, S4, and S5 of the supplementary materials for CIF).

For the list of medical variables, along with their median and interquartile range (IQR) distribution, as well as the median and IQR gap between consecutive measurements, please refer to Table S5.

The dataset was split into a training cohort of 2818 ICU stays (471 positive VAP, 16.71%) and a validation cohort of 705 ICU stays (124 positive VAP, 17.59%). Event distribution was statistically comparable between the two subsets (Fisher's exact test $p=0.57$). The comparison of demographics between the development and evaluation cohorts is provided in Table S6.

Models performance

Real-time regressor performance

In terms of time-to-event metrics, the regressor achieved a concordance index (C-index) of 0.75 [0.75, 0.76] in training and 0.68 [0.68, 0.69] in testing, indicating a solid level of predictive concordance. The mean

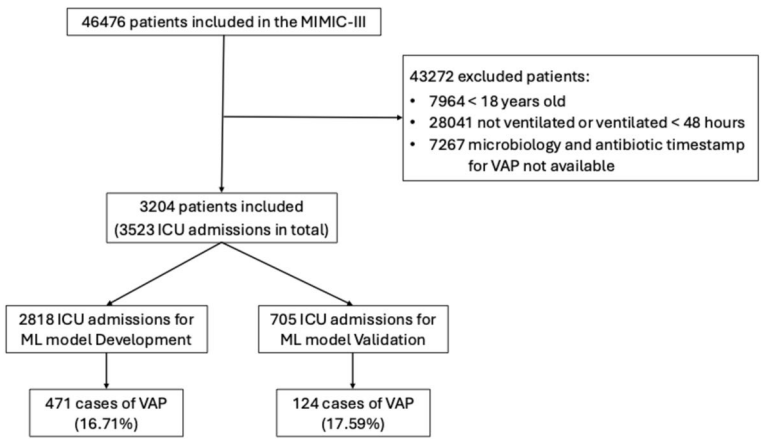

Fig. 1 Cohort derivation and dataset split

Table 1 Comparison of demographic, clinical characteristics, and results of Fine and Gray's competing risks regression analysis for VAP

		Population's statistics			Fine-Gray model			
Demographic		**No VAP**	**VAP**	***p*-value**	**sHR (95% *CI*)**	**Coefficient (β^2)**	**SE (β^2)**	***p*-value**
Patients (*n*)		2928	595					
Sex (male) (*n*, %)		1619 (55%)	377 (63%)	< 0.001	1.32 (1.11, 1.56)	0.28	0.087	0.001
Age (years)		63.9 (16.4)	61.2 (17.3)	< 0.001	0.99 (0.99, 1.0)	−0.01	0.003	0.002
Ethnicity (*n*, %)	White or Caucasian	1896 (65%)	364 (61%)	0.101	0.94 (0.78, 1.14)	−0.06	0.098	0.550
	Black or African American	223 (8%)	63 (11%)	0.021	1.42 (1.05, 1.93)	0.35	0.155	0.023
	Other/unknown	809 (28%)	168 (28%)	0.763				
Comorbidity (*n*, %)	Chronic kidney disease	574 (20%)	122 (21%)	0.612	1.14 (0.92, 1.41)	0.13	0.108	0.220
	Diabetes mellitus	897 (31%)	182 (31%)	1	1.04 (0.86, 1.25)	0.04	0.096	0.710
	Cancer	338 (12%)	36 (6%)	< 0.001	0.58 (0.41, 0.81)	−0.55	0.174	0.002
	Congestive heart failure	910 (31%)	141 (24%)	< 0.001	0.81 (0.65, 0.99)	−0.22	0.106	0.042
	Coronary artery disease	695 (24%)	119 (20%)	0.049	0.95 (0.76, 1.19)	−0.05	0.113	0.670
	Hypertension	1723 (59%)	344 (58%)	0.648	1.09 (0.9, 1.31)	0.08	0.095	0.390
	COPD	801 (27%)	139 (23%)	0.047	0.93 (0.76, 1.12)	−0.08	0.098	0.430
Reason for admission (*n*, %)	Cardiovascular	355 (16%)	73 (15%)	0.631	1.3 (0.98, 1.73)	0.26	0.146	0.073
	Neurological	311 (14%)	86 (18%)	0.047	1.57 (1.21, 2.04)	0.45	0.134	< 0.001
	Respiratory	444 (20%)	57 (12%)	< 0.001	0.85 (0.64, 1.14)	−0.16	0.148	0.280
	Sepsis	131 (6%)	18 (4%)	0.061	0.84 (0.51, 1.37)	−0.17	0.25	0.490
	Surgery	114 (5%)	26 (5%)	0.822	1.31 (0.86, 2.01)	0.27	0.216	0.200
	Trauma	146 (7%)	68 (14%)	< 0.001	2.24 (1.71, 2.95)	0.81	0.139	< 0.001
	Other	791 (36%)	173 (36%)	0.958	1.15 (0.93, 1.42)	0.14	0.108	0.190

VAP ventilator-associated pneumonia, *COPD* chronic obstructive pulmonary disease, *sHR* sub-hazard ratio, *CI* confidence intervals, *SE* standard error

Table 2 Performance metrics of the ensemble model for time-to-event prediction

Dataset	**C-index [95% CI]**	**MAE [95% CI]**	**MSE [95% CI]**	**RMSE [95% CI]**
Development	0.754 [0.75, 0.76]	32.980 [32.81, 33.16]	1293.227 [1283.04, 1303.02]	35.961 [32.80, 33.15]
Validation	0.682 [0.68, 0.69]	36.176 [35.82, 36.52]	1533.136 [1511.45, 1555.67]	39.155 [35.79, 36.53]

C-index concordance index, *CI* confidence interval, *MAE* mean absolute error, *MSE* mean-squared error, *RMSE* root-mean-squared error

absolute error (MAE) was 32.98 [32.81, 33.16] h on the training set and increased to 36.18 [35.82, 36.52] h on the test set, while the root-mean-square error (RMSE) values were 35.96 [32.80, 33.15] and 39.16 [35.79, 36.53] for training and testing, respectively (Table 2).

Feature importance analysis is shown in Fig. S6. The most influential features were ventilation duration, temperature dynamics, renal function, and white blood cell count-related features (48 h maximum of hourly differences, 48 h minimum, 48 h maximum).

Average odds differences are provided in Table S7 and Fig. S7, while the performance metrics for individual demographic groups are detailed in Table S8.

Comparison with baseline classifiers

When transformed into a binary classifier, the regressor achieved a test AUROC of 0.71 [0.66, 0.76] and an AUPR of 0.36 [0.27, 0.44], outperforming the 24-h and 48-h single-shot models, which both had an AUROC of 0.62 [0.56, 0.68]. When thresholded to maintain a sensitivity of at least 0.70, the regressor achieved a specificity of 0.60, precision of 0.28, and an F1 score of 0.40, while the single-shot models had lower specificity (0.41 and 0.47), precision (0.20 and 0.21), and F1 scores (0.31 and 0.33). Additionally, the regressor provided an estimated lead time (i.e., time from prediction to VAP event) of 53 (7–120) h. A detailed comparison of performance metrics is presented in Table 3 and Fig. 2, while the corresponding curves are provided in the supplementary materials (Figs. S8 and S9).

Discussion

We described an innovative ML framework to develop a real-time VAP prediction model, including continuous vital signs, ventilator data, blood gas analyses, and lab exams to capture dynamic temporal pathophysiological

Table 3 Performance metrics of predicting models

Model	Dataset	AUROC	AUPR	Fixed working point			
				Sensitivity	Specificity	Precision	F1 score
24	Development	0.87 [0.85, 0.88]	0.65 [0.61,0.69]	0.94	0.47	0.27	0.42
	Validation	0.62 [0.56, 0.68]	0.28 [0.19, 0.35]	0.71	0.41	0.20	0.31
48	Development	0.85 [0.83, 0.87]	0.67 [0.63, 0.71]	0.90	0.55	0.29	0.43
	Validation	0.62 [0.56, 0.68]	0.27 [0.18, 0.33]	0.71	0.47	0.21	0.33
Real time[a]	Development	0.83 [0.81, 0.85]	0.51 [0.46, 0.56]	0.85	0.62	0.31	0.45
	Validation	0.71 [0.66, 0.76]	0.36 [0.27, 0.44]	0.71	0.60	0.28	0.40

[a] The fixed cutoff is computed to achieve 0.7 sensitivity in the validation cohort

patterns. The model demonstrated robust predictive capabilities, superior performance metrics compared to static prediction models, and clinically meaningful prediction time windows.

VAP represents a frequent but potentially preventable complication of IMV [1, 11] affecting approximately 25% of patients requiring intubation [12], prolonging IMV and ICU length of stay, and increasing mortality [13]. VAP is a primary driver of antimicrobial resistance development through sustained pathogen selective pressure, extending the VAP impacts beyond single patient care to the overall healthcare system [14]. These challenges have prompted extensive research, initially focusing on scoring systems (PIRO, CPIS) [15, 16] and subsequently evolving into more sophisticated mathematical modeling approaches [17, 18]. A recent meta-analysis collates this evidence [2], demonstrating the superiority of ML algorithms compared to traditional scoring tools. However, significant methodological constraints limit the clinical implementation of these ML models, which rely on ICD codes and thus fail to incorporate timing information, offering no temporal context for predicting VAP and neglecting the dynamic nature of the patient's condition over time.

The present work tries to address these limitations. The model provides a continuous dynamic evaluation of VAP risk throughout the entire IMV period, adapting to changing patient conditions [19, 20] by incorporating historical and real-time data from monitors, mechanical ventilators, blood gas analyses, and lab exams. Two static predictive models utilizing the same dataset and model development framework provided almost nil VAP prediction capabilities, highlighting the added value of integrating continuous, real-time data.

ICD-based approaches for identifying VAP suffer from confirmation bias, lack temporal precision, and fail to capture VAP onset in real time, limiting their clinical applicability [3–5]. Contrarily, by integrating microbiological data (i.e., positive cultures) and therapeutic regimens (i.e., initiation of antibacterial therapy), our model established a precise temporal localization of VAP events, aligning with previous reports [21, 22] and guidelines [10]. Our results align with prior epidemiology [21, 23, 24]: VAP incidence 15% and 23.77 [21.80–25.58] per 1000 IMV-days. Similarly, the pathogens recognized as responsible for VAP adhere to recent reports. Fine–Gray multivariate analysis supported our VAP identification methodology while accounting for immortal-time bias. It also showed that VAP-positive patients were more often older, male, Black/African American, and admitted for neurological or traumatic disease. Finally, applying a Weibull regression analysis, we observed that VAP-positive patients had longer IMV and increased mortality.

A previous model, the VAP risk index [25], offers a similar ICD-free approach but does not provide continuous risk prediction.

A key finding is the model's lead time—the interval between alarm and VAP. As a binary classifier, it achieved > 0.75 sensitivity and specificity with a median anticipation exceeding 2 days. Such a wide, actionable predictive window may facilitate preventive measures—e.g., increased microbiological surveillance (see Fig. 3) and timelier diagnostic assessment or treatment consideration, potentially leading to better management, shorter ICU stays, and reduced healthcare costs [26]. Finally, the analysis of average odds difference (AOD) revealed no significant bias in terms of ethnicity, sex, and age group, indicating that the model's performance was equitable across these demographic variables.

The model has several limitations. First, while the model demonstrates statistically significant improvement over static baselines (*AUROC* 0.71 vs 0.62), the overall predictive performance (C-index 0.68, *PPV* 0.28) remains insufficient to guide therapeutic decisions such as antibiotic initiation. The model's clinical utility is therefore restricted to risk stratification for enhanced microbiological surveillance rather than direct treatment decisions. Prospective validation studies are essential to determine whether this surveillance-based

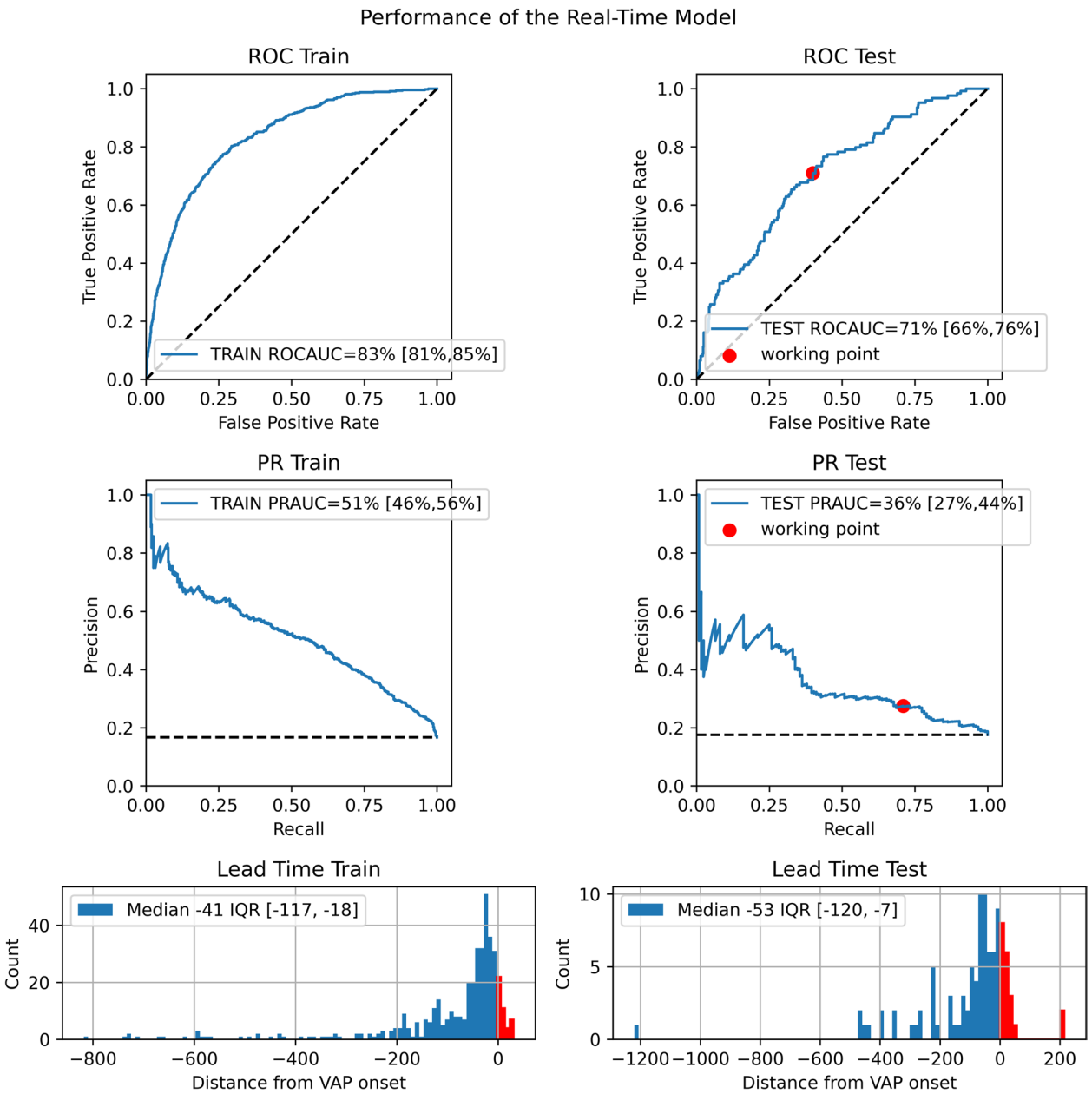

Fig. 2 Performance of the real-time time-to-event predictor as a classification model. *ROC*, receiving operator characteristics; *PR*, precision recall; *IQR*, interquartile range. Red bars in histograms indicate late alarms (false negatives)

approach translates into improved patient outcomes, reduced inappropriate antibiotic use, and cost-effectiveness in real-world ICU settings. Second, while it showed robust performance on the MIMIC-III dataset, this constitutes only an internal validation using a development and validation split. Its generalizability must be confirmed through external validation in independent cohorts. Third, the model was trained on MIMIC-III, which includes patients admitted between 2001 and 2012, a period when modern multiplex polymerase chain reaction techniques for rapid pathogen identification were unavailable. This historical limitation could affect the model's generalizability to contemporary ICU populations, where faster and more sensitive diagnostic modalities are now routinely used. Future validation in more recent cohorts with advanced diagnostic methods is needed to confirm the model's applicability in current practice.

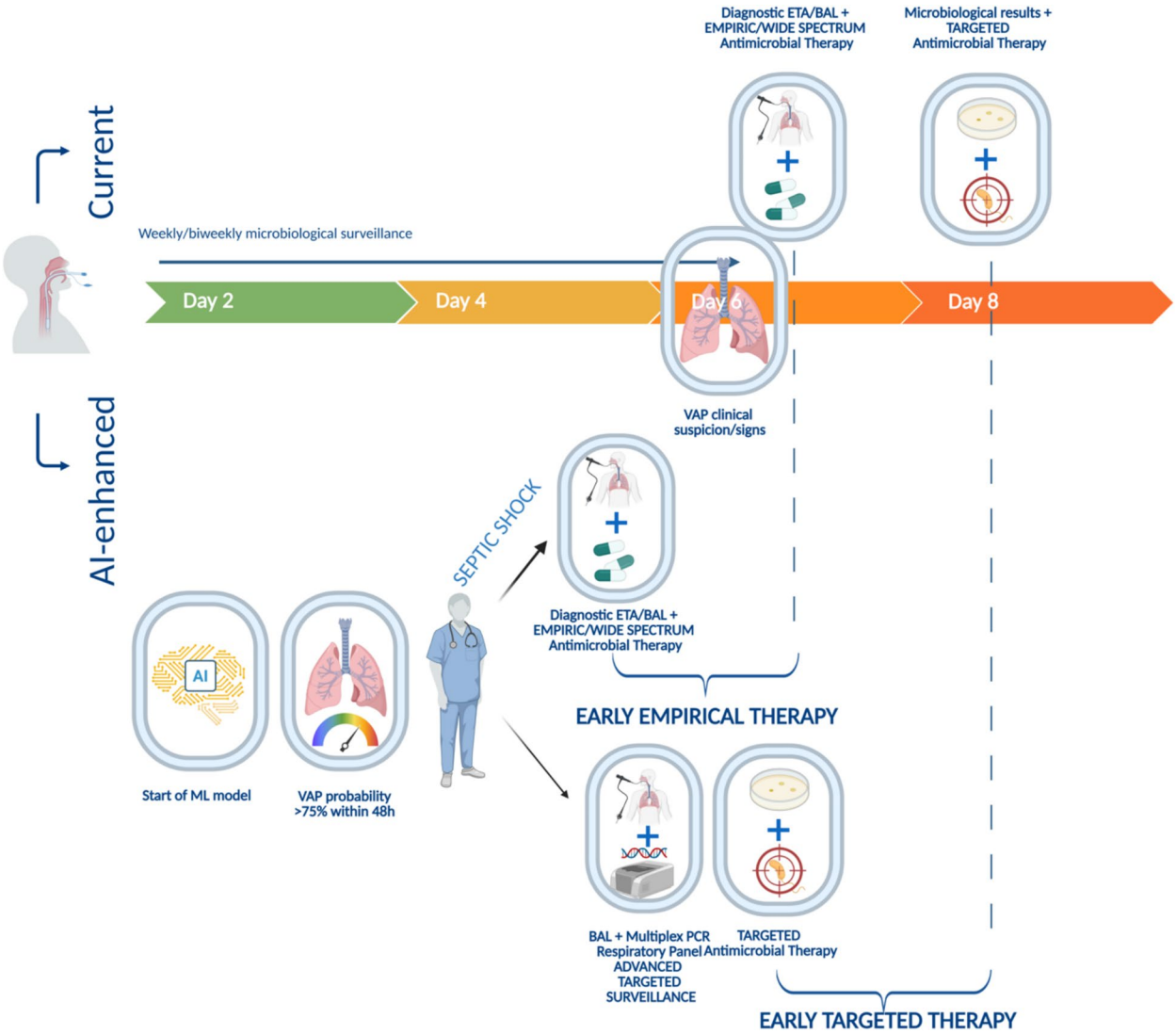

Fig. 3 Possible clinical implementation of a real-time predicting machine learning model. We envision the application of our model within the following framework. After 48 h from intubation, the ML model starts providing the attending physician with a probability of developing VAP within 48 h. Whether the chance of VAP would top a predefined threshold (e.g., > 75), two possible options are offered to the clinician. First, microbiological surveillance by collecting bronchoalveolar lavage, and a fast-turnaround multiplex PCR respiratory panel will provide immediate (i.e., 6 h) information on respiratory pathobionts' most common resistance mechanism. With this information, whether clinical signs of VAP develop, the clinician can initiate targeted, appropriate antimicrobial therapy, avoid broad-spectrum empiric therapy, and limit antibiotic selective pressure and selection of multidrug resistance agents. Otherwise, in the most severe cases, or whether clinical signs of VAP might be more complicated to document (e.g., immunocompromised, transplant recipient, ARDS patient), targeted antibiotic therapy can instead be initiated with clinically meaningful anticipation. Created in BioRender, Scaravilli, V. (2025), https://BioRender.com/gguxkjc

Conclusions

In conclusion, we describe an innovative ensemble model of XGBoost regressors for real-time VAP risk stratification, leveraging historical and real-time features to ensure transparency for healthcare providers. The model can be directly connected via an API to the electronic health record (EHR) system, enabling automatic risk time-to-event prediction without external support. No expertise or manual input is required from users, as the system autonomously processes data from the EHR, ensuring potential integration within clinical workflows. Prospective validation studies are essential to determine whether this surveillance-based approach translates into improved patient outcomes, reduced inappropriate antibiotic use, and cost-effectiveness. Future work should focus on prospective validation in multicenter ICU settings, incorporating real-time deployment and clinician feedback. Further refinement of the model can be achieved by integrating novel biomarkers or imaging data to enhance predictive accuracy.

Abbreviations

AOD	Average odds difference
AUPRC	Area under the precision-recall curve
AUROC	Area under the receiver operating characteristic
CI	Confidence interval
CIF	Cumulative incidence functions
EHR	Electronic health records
GCS	Glasgow Coma Scale
ICU	Intensive care unit
IQR	Interquartile range
LOS	Length of stay
IMV	Invasive mechanical ventilation
MAE	Mean absolute error
ML	Machine learning
RFE	Recursive feature elimination
RMSE	Root-mean-squared error
sHR	Sub-hazard ratios
VAP	Ventilator-associated pneumonia

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s44158-025-00313-3.

Supplementary Material 1: Supplementary materials: Supplementary Tables. Table S 1 Definition of Medical Variables, Units, and Data Entry Boundaries. Availability measures the percentages of ICU stays where the medical variable is present. TV=Tidal Volume, MV=Minute Volume, PEEP= Positive end-expiratory pressure. Table S2 List of Medical Variables and Associated Derived Features. Suffix _md stand for distance from previous measurement, _missing indicator of missing hour _XXh_statistic stand for the specific statistic in the previous XX hours. TV=Tidal Volume, MV=Minute Volume, PEEP= Positive end-expiratory pressure. Table S3 Organisms and their counts in VAP detection. Table S4 Coefficients, standard errors, z-values, and p-values from the Weibull regression model assessing 90-day mortality, including VAP and other clinical variables. Table S5 Summary of medical variables measured in ICU patients, including availability (number of ICU stays with available data), units of measurement (UOM), median values, interquartile ranges (IQR), and the gap between consecutive measurements. TV=Tidal Volume, MV=Minute Volume, PEEP= Positive end-expiratory pressure. Table S6 Patients' characteristics and division into development and validation cohort. Table S7 Average Odds Difference computed for each demographic level. CCU: Coronary Care Unit / Cardiac Care Unit, CSRU: Cardiac Surgery Recovery Unit, MICU: Medical Intensive Care Unit, SICU: Surgical Intensive Care Unit, TSICU: Trauma Surgical Intensive Care Unit. Table S8 Performances of the regression model in subpopulations. CCU: Coronary Care Unit / Cardiac Care Unit, CSRU: Cardiac Surgery Recovery Unit, MICU: Medical Intensive Care Unit, SICU: Surgical Intensive Care Unit, TSICU: Trauma Surgical Intensive Care Unit. Supplementary figures. Figure S1 CIFs for VAP stratified by age group (18–35, 35–50, 50–70, 70–80, 80+ years). Figure S2 CIFs for VAP stratified by biological sex (male and female). Figure S3 CIFs for VAP stratified by ethnicity (Black and White). Analysis includes only patients identified as Black or White. Figure S4 CIFs for VAP with reason for admission as a covariate. Each curve represents the incidence for patients admitted for a specific reason. Figure S5 CIFs for VAP estimated using a Fine-Gray model with comorbidities as covariates. Each curve represents the incidence for patients with a single comorbidity present. COPD: Chronic Obstructive Pulmonary Disease. Figure S6 Bar chart displaying the relative importance of the top 40 features in predicting ventilator-associated pneumonia onset. Figure S7 Graphical representation of Average Odds Differences. CCU: Coronary Care Unit / Cardiac Care Unit, CSRU: Cardiac Surgery Recovery Unit, MICU: Medical Intensive Care Unit, SICU: Surgical Intensive Care Unit, TSICU: Trauma Surgical Intensive Care Unit. Figure S8 Performance evaluation of the VAP prediction model at the 24-hour prediction point. The figure includes receiver operating characteristic (ROC) curves, precision-recall (PR) curves, and histograms of lead times for both the training and test datasets. Figure S9 Performance evaluation of the VAP prediction model at the 48-hour prediction point. The figure includes receiver operating characteristic (ROC) curves, precision-recall (PR) curves, and histograms of lead times for both the training and test datasets.

Clinical trial number

Not applicable.

Authors' contributions

SZ: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation. LR: Writing – review & editing, Writing – original draft, Methodology, Investigation. FA: Writing – original draft, Methodology, Conceptualization. AA: Writing – review & editing, Writing – original draft, Resources, Project administration, Funding acquisition, Conceptualization. AG: Writing – review & editing, Writing – original draft, Methodology, Visualization, Investigation. GG: Writing – review & editing, Writing – original draft, Supervision. VS: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation.

Funding

The study was partially funded by the Italian Ministry of Health — Current Research IRCCS.

Data availability

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

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

VS provided consulting services to U-Care Medical s.r.l.,. LR has received consulting fees from U-Care Medical s.r.l. AA is CEO of U-Care Medical. FA and AA are shareholders of U-Care Medical. FA and SZ are employees of U-Care Medical. AA, FA and SZ filed for the European Patent Application No. 24202572.4 "System and method for management and prediction of invasive mechanical ventilation necessity". GG is president-elect of Società Italiana di Anestesia, Analgesia, Rianimazione e Terapia Intensiva (S.I.A.A.R.T.I.).

Received: 23 June 2025 Accepted: 3 November 2025
Published online: 13 January 2026

References

1. Rosenthal VD, Memish ZA, Bearman G (2025) Preventing ventilator-associated pneumonia: a position paper of the International Society for Infectious Diseases, 2024 update. Int J Infect Dis 151:107305. https://doi.org/10.1016/j.ijid.2024.107305
2. Zhang J, Yang P, Zeng L, Li S, Zhou J (2024) Ventilator-associated pneumonia prediction models based on AI: scoping review. JMIR Med Inform 12:e57026–e57026. https://doi.org/10.2196/57026
3. Giang C et al (2021) Predicting ventilator-associated pneumonia with machine learning. Medicine (Baltimore) 100(23):e26246. https://doi.org/10.1097/MD.0000000000026246
4. Liang Y et al (2022) Early prediction of ventilator-associated pneumonia in critical care patients: a machine learning model. BMC Pulm Med 22(1):250. https://doi.org/10.1186/s12890-022-02031-w
5. Faucher M et al (2022) Early prediction of ventilator-associated pneumonia in ICU patients using an interpretable machine learning algorithm. https://doi.org/10.20944/preprints202206.0149.v1
6. Johnson AEW et al (2016) MIMIC-III, a freely accessible critical care database. Sci Data 3(1):160035. https://doi.org/10.1038/sdata.2016.35
7. Johnson A, Pollard T, Mark R (2016) MIMIC-III Clinical Database (version 1.4). PhysioNet 10(C2XW26):2. https://doi.org/10.13026/C2XW26.
8. Johnson AEW, Wac M, Yarnell CJ, Gow B Ventilation.sql script in MIT-LCP/mimic-code. Available: https://github.com/MIT-LCP/mimic-code/blob/e39825259beaa9d6bc9b99160049a5d251852aae/mimic-iv/concepts/treatment/ventilation.sql. Accessed: Oct. 25, 2024

9. Boles J-M et al (2007) Weaning from mechanical ventilation. Eur Respir J 29(5):1033–1056. https://doi.org/10.1183/09031936.00010206
10. Singer M et al (2016) The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA 315(8):801. https://doi.org/10.1001/jama.2016.0287
11. Mumtaz H et al (2023) Ventilator associated pneumonia in intensive care unit patients: a systematic review. Ann Med Surg 85(6):2932–2939. https://doi.org/10.1097/MS9.0000000000000836
12. Garnier M et al (2023) Epidemiology, risk factors and prognosis of ventilator-associated pneumonia during severe COVID-19: multicenter observational study across 149 European intensive care units. Anaesth Crit Care Pain Med 42(1):101184. https://doi.org/10.1016/j.accpm.2022.101184
13. Motowski H, Ilges D, Hampton N, Kollef MH, Micek ST (2023) Determinants of mortality for ventilated hospital-acquired pneumonia and ventilator-associated pneumonia. Crit Care Explor 5(3):e0867. https://doi.org/10.1097/CCE.0000000000000867
14. Sano M et al (2022) Risk factors for antibiotic resistance in hospital-acquired and ventilator-associated pneumonia. J Infect Chemother 28(6):745–752. https://doi.org/10.1016/j.jiac.2022.02.012
15. Lisboa T et al (2008) The ventilator-associated pneumonia PIRO score. Chest 134(6):1208–1216. https://doi.org/10.1378/chest.08-1106
16. Schurink CAM et al (2004) Clinical pulmonary infection score for ventilator-associated pneumonia: accuracy and inter-observer variability. Intensive Care Med 30(2):217–224. https://doi.org/10.1007/s00134-003-2018-2
17. Rambaud J et al (2023) Clinical decision support system to detect the occurrence of ventilator-associated pneumonia in pediatric intensive care. Diagnostics 13(18):2983. https://doi.org/10.3390/diagnostics13182983
18. Schurink CAM et al (2007) A bayesian decision-support system for diagnosing ventilator-associated pneumonia. Intensive Care Med 33(8):1379–1386. https://doi.org/10.1007/s00134-007-0728-6
19. Meng H et al (2025) Prediction model, risk factor score and ventilator-associated pneumonia: a two-stage case-control study. J Microbiol Immunol Infect 58(1):94–102. https://doi.org/10.1016/j.jmii.2024.11.005
20. De Souza Dos Santos G, Alves de Carvalho França de Macedo V, Oliniski Reikdal S, Graf ME, Mario Martin B, Joaquim Meier M (2025) Ventilator-associated pneumonia risk factors in patients with severe COVID-19 in Southern Brazil: a retrospective observational study. Infect Dis Health 30(1):38–49. https://doi.org/10.1016/j.idh.2024.07.004
21. Mangioni D et al (2023) Incidence, microbiological and immunological characteristics of ventilator-associated pneumonia assessed by bronchoalveolar lavage and endotracheal aspirate in a prospective cohort of COVID-19 patients: CoV-AP study. Crit Care 27(1):369. https://doi.org/10.1186/s13054-023-04658-5
22. Grasselli G et al (2021) Hospital-acquired infections in critically ill patients with COVID-19. Chest 160(2):454–465. https://doi.org/10.1016/j.chest.2021.04.002
23. Wałaszek M et al (2023) Ventilator-associated pneumonia in Polish intensive care unit dedicated to COVID-19 patients. BMC Pulm Med 23(1):443. https://doi.org/10.1186/s12890-023-02743-7
24. Rosenthal VD et al (2024) Assessing the impact of a multidimensional approach and an 8-component bundle in reducing incidences of ventilator-associated pneumonia across 35 countries in Latin America, Asia, the Middle East, and Eastern Europe. J Crit Care 80:154500. https://doi.org/10.1016/j.jcrc.2023.154500
25. Samadani A, Wang T, van Zon K, Celi LA (2023) VAP risk index: early prediction and hospital phenotyping of ventilator-associated pneumonia using machine learning. Artif Intell Med 146:102715. https://doi.org/10.1016/j.artmed.2023.102715
26. Cabrera-Tejada GG et al (2023) Estimation of additional costs in patients with ventilator-associated pneumonia. Antibiotics 13(1):2. https://doi.org/10.3390/antibiotics13010002

Publisher's Note