ISSN: 2581 – 3633

International Journal of Medical Science and Advanced Clinical Research (IJMACR)

Available Online at: *https://www.ijmacr.com*

Volume – 9, Issue – 5, September – 2026, Page No.: 552 – 567

From Algorithm to bedside: A narrative review of machine learning and deep learning in early prediction of ventilator associated pneumonia and a roadmap for implementation in developing countries

[1]Queeny Baruah, Post Graduate Student, Department of Microbiology, Sharda School of Medical Sciences and Research, Sharda Hospital, Greater Noida, India,

[2]V.S. Randhawa, Professor, Department of Microbiology, Sharda School of Medical Sciences and Research, Greater Noida, India

[3]Ayushi Mishra, Post Graduate Student, Sharda School of Medical Sciences and Research, Sharda Hospital, Greater Noida, India

[4]Ram Murti Sharma, Medical Superintendent, Sharda School of Medical Sciences and Research and associated hospitals, Greater Noida, India.

[5]Sonisha Gupta, Professor, Department of Respiratory Medicine, Sharda School of Medical Sciences and Research, Greater Noida, India

[6]Shefali Singh, Post Graduate Student, Sharda School of Medical Sciences and Research, Sharda Hospital, Greater Noida, India

[7]Ankit, Consultant, Department of Anesthesia, Sharda School of Medical Sciences and Research and associated hospitals, Greater Noida, India

[8]Tarana Sarwat, Professor, Department of Microbiology, Sharda School of Medical Sciences and Research, Greater Noida, India

Corresponding Author: V.S. Randhawa, Professor, Department of Microbiology, Sharda School of Medical Sciences and Research, Greater Noida, India.

How to citation this article: Queeny Baruah, V.S. Randhawa, Ayushi Mishra, Ram Murti Sharma, Sonisha Gupta, Shefali Singh6, Ankit, Tarana Sarwat, "From Algorithm to bedside: A narrative review of machine learning and deep learning in early prediction of ventilator associated pneumonia and a roadmap for implementation in developing countries", IJMACR– September–2026, Volume – 9, Issue – 5, P. No. 552 – 567.

Type of Publication: Review Article

Conflicts of Interest: Nil

Abstract

Ventilator-associated pneumonia (VAP) is one of the most common healthcare-associated infections among mechanically ventilated intensive care unit (ICU) patients. It is associated with increased mortality, prolonged hospitalization, inappropriate broad-spectrum antimicrobial use, and higher healthcare costs. Conventional diagnostic approaches, including the Clinical Pulmonary Infection Score (CPIS) and clinical assessment, have limited predictive accuracy and often identify VAP only after its onset. This narrative review evaluates the emerging role of machine learning (ML) and deep learning (DL) algorithms for early VAP prediction. Unlike previous inconclusive meta-analyses, this review highlights consistent evidence from studies published

between 2016 and 2025 demonstrating the superior performance of ML and DL models over traditional methods. A meta-analysis of ten supervised ML models reported an overall area under the receiver operating characteristic curve (AUROC) of 0.88, with a sensitivity of 0.72 and specificity of 0.90. Among individual models, Random Forest achieved an AUROC of 0.84 using 42 clinical variables, while XGBoost showed the highest predictive performance (AUROC 0.831–0.94). The PREDICT-LSTM framework achieved an AUROC above 0.85 and enabled prediction 6–24 hours before VAP onset using continuous physiological data. Other DL models incorporating electronic health records and CT imaging predicted VAP more than 48 hours before conventional diagnosis, with AUROC values exceeding 0.87. Explainable AI techniques, including SHAP analysis, further enhanced model interpretability. Overall, ML and DL models enable earlier and more accurate VAP prediction, supporting timely intervention, improved antimicrobial stewardship, and data-driven critical care.

Keywords: Ventilator-Associated Pneumonia, Machine Learning, Deep Learning, Xgboost, Intensive Care Unit, Predictive Analytics, Long Short-Term Memory.

Introduction

Ventilator-associated pneumonia (VAP) refers to a hospital-acquired infection, which occurs within two or more days after the initiation of mechanical ventilation. Mechanical ventilation is frequently associated with ventilator-associated pneumonia (VAP), the most severe form of hospital-acquired pneumonia. This infection is the source of more than 50% of ICU cases. The use of antibiotics is closely associated with both multidrug resistance and the use of broad-spectrum antimicrobial agents and leads to sepsis and multiple organ failure with a mortality rate of 22–43% [1.]

The Clinical Pulmonary Infection Score (CPIS) that comprises a total of 6 items which include temperature, leukocyte count, qualitative assessment of tracheal secretions, oxygenation, chest radiograph, and tracheal aspirate bacterial culture can be used[2,3]. The score has been found to have moderate to good accuracy for predicting VAP, is simple to implement, and widely used in the diagnosis of VAP[2,3].

Although the methods for the prevention and treatment of ventilator-associated pneumonia (VAP) are well established, prediction of the disease occurrence remains a matter of concern. The problem is complicated by the fact that there is no generally accepted clinical definition of VAP, resulting in frequent diagnostic errors, including false-negative and delayed diagnoses, which are among the most common healthcare-related adverse events.[4,5]

Now, machine learning (ML) and Deep Learning (DL) models have been proposed to predict ventilator-associated pneumonia (VAP) hours to days before its diagnosis and to stratify the risk of developing VAP using routinely collected intensive care unit (ICU) data.[6,7,8]

The modern development of machine learning (ML), and more specifically, the recent advances in deep learning (DL) offer an entirely new approach to the conventional, intuitive diagnosis of ventilator-associated pneumonia (VAP). By enabling the transition from static, rule-based reasoning to dynamic, data-driven prediction, these methods allow for improved early detection, surveillance, and treatment of VAP in mechanically ventilated intensive care unit (ICU) patients.

This narrative review assesses the role of ML and DL in early prediction and risk stratification of VAP, analyzes the existing studies, and highlights the opportunities and limitations for the clinical application of these technologies.

Materials and Methods

The purpose of this narrative review is to assess the literature for studies evaluating the performance of ML and DL models for the early prediction of VAP in mechanically ventilated ICU patients. The review will pursue four additional objectives. First, by analyzing the cited literature, this study will identify the most common and effective ML models for VAP prediction. Second, the review will determine the key predictors of VAP. Third, it will evaluate the interpretability of the predictive models and their potential application in ICUs, including an assessment of the performance of various interpretability techniques, such as SHAP. Lastly, the review will highlight the limitations of the current literature and suggest areas of research for the future study.

The narrative review methodology was conducted to prepare this article. The literature search strategy included databases such as PubMed, Scopus, Embase, MEDLINE, Web of Science, Cochrane Library, and CINAHL. The inclusion criteria comprised systematic reviews, meta-analyses, scoping reviews, and research articles published in peer-reviewed journals between 2016 and 2025. The studies focusing only on pneumonia without ventilator-associated pneumonia (VAP) or those concentrating on other infections in intensive care units (ICUs) and not addressing VAP were excluded.

This study has involved a database Multi-Institutional Input/output Intensive Care Units (MIMIC-IV) of critically ill patients maintained by the Massachusetts Institute of Technology (MIT) and Beth Israel Deaconess Medical Center (BIDMC) with the participation of hospitals' wards in 2008-2019, and another database named MIMIC-III of de-identified health information [8,9,10,11,3,12]. age, sex, admission source (medical intensive care unit vs. coronary care unit, surgical intensive care unit, cardiac surgery recovery unit, and trauma surgical intensive care unit), admission type to enable early recognition of first-occurrence ventilator-associated pneumonia (VAP), the total number of 42 variables were selected from the MIMIC-III database according to the related studies and literature to train the predictive model. The included features were emergency or elective admission, reintubation, pre-existent comorbidities, worst PaO 2 /FiO 2, white blood cell count, and temperature during the first 24 hours of ventilation, worst APACHE III score and its components, as well as the lowest Simplified Acute Physiology Score (SAPS) II and Sequential Organ Failure Assessment (SOFA) score during the first 24 hours in the intensive care unit (ICU) [3,13,14,15,16].

Amongst machine learning algorithms, Random Forest has become one of the standard procedures for predictive analytics in the VAP research field. A random forest model, which is an ensemble of decision trees, was constructed using a five-fold cross-validation from the MIMIC-III database to devise an early VAP predictor utilizing 42 risk factors including admission vital signs, laboratory investigations, and duration of mechanical ventilation [3,8,17,18]. Gradient boosting algorithms have proven to be highly efficient in ICU prediction models. XGBoost (eXtreme Gradient Boosting) is an ensemble algorithm that makes use of iterative improvements of prior models to correct errors. This algorithm is reportedly more computationally efficient than traditional gradient boosting methods while also allowing greater flexibility in terms of handling missing data and non-linearity in the response. The hyperparameters of the leaf weights and depths were adjusted in this study, followed by a ten-fold cross-validation procedure, wherein one

tenth of the data was used as a validating set. The algorithm has since been used to predict multi-drug resistant organism associated VAP using MIMIC-IV data of patients diagnosed with VAP[3,8,10,19,20]. To further expand upon prior comparative analyses, we tested fourteen machine learning algorithms AdaBoost, CatBoost, decision trees, Extra Trees, GBDT, KNN, LDA, LightGBM, logistic regression, MLP, Naive Bayes, Random Forest, SVM, and XGBoost on the MIMIC-IV training set and an external validation set [3,8,10,17,18,19,20,21,27,28,29,30]. Deep learning uses multilayer neural networks to learn features from the data automatically. With that said, compared to traditional machine learning algorithms, deep learning does not require much manual feature engineering. Instead, it can automatically learn temporal and nonlinear dependencies in the data. For example, the PREDICT framework with LSTM networks (Long Short-Term Memory): This framework uses recurrent networks to analyze time series data based on MIMIC-IV. It uses a 24-hour observation window that includes respiratory rate, SpO2, heart rate, temperature, and mean arterial pressure (MAP). LSTM networks can learn long-term temporal dependencies by storing past information in memory cells. As a result, they are convenient for use in the ICU to assess the risk dynamically and continuously; however, their "black box" nature makes them difficult to interpret.[11] A major advantage of the study was the inclusion of CT scans, electronic medical records and other clinical sources of information. Unlike the conventional use of tabular clinical data in machine learning algorithms, multisource ensemble models have the potential to offer greater robustness and lead time for prediction. The prediction window identified by the model is over 48 hours, suggesting ample opportunity for intervention and antibiotic stewardship.[8] Another important trend is the rise of so-called explainable artificial intelligence (XAI), which aims to solve the problem of the black-box nature of the majority of modern algorithms. In particular, several recent VAP studies have used the approach of SHAP (SHapley Additive exPlanations), which explains the output of a given model by estimating the SHapley values from each predictor; thus, making the inner workings of the algorithm more interpretable and understandable for clinicians.[10,21,22] All employed SHAP-based interpretability frameworks, and some additionally adopted TRIPOD-AI compliant reporting to improve methodological transparency and reproducibility.[10]

In VAP prediction studies, the AUROC is the most reported metric for evaluating discriminative performance across scoring instruments and ML models. Comparisons of traditional scores (e.g., Clinical Pulmonary Infection Score, CPIS) against ML models (XGBoost, Random Forest, logistic regression with LASSO) have found that ML models outperform traditional scores, a difference partly explained by variations in VAP definition and prediction lead time. Notably, XGBoost achieved the highest AUROC (0.91) when dynamic respiratory mechanics and antibiotic exposure history were included, whereas scores relying solely on static clinical variables showed AUROC values rarely exceeding 0.78, underscoring that ML captures complex, nonlinear interactions that improve early VAP risk stratification.

Results

Machine learning and deep learning methods have proven to be effective in the early prediction of ventilator-associated pneumonia. The results of the study revealed that most of the reviewed research demonstrated better performance than traditional approaches. According to

Table 1, the ten studies that employed supervised machine learning algorithms to predict VAP revealed a pooled AUROC of 0.88. Mechanical ventilation, length of ICU stay, blood transfusion, nutritional support, and antibiotics were the key predictors. Random Forest, which is one of the most commonly used algorithms, had an AUROC of 0.84 when 42 clinical and laboratory variables were used to predict VAP in the MIMIC-III database. The study also found that XGBoost is a powerful model in VAP prediction, and its AUROC ranged from 0.831 to 0.94 in different applications, including prediction of multidrug-resistant organisms, and mortality. Serum potassium, blood urea nitrogen, red cell distribution width, anion gap, duration of mechanical ventilation, and comorbidities were the key variables used in the prediction model.

The study also revealed that deep learning algorithms have improved the prediction of VAP. For instance, the PREDICT LSTM framework achieved an AUROC of over 0.85 and could use respiratory rate, oxygen saturation, heart rate, temperature, and mean arterial pressure to predict VAP. Table 1 also shows that there is a growing interest in the development of early warning systems for VAP. Some of the models can predict VAP several days before the clinical diagnosis, while others that use electronic health records and CT scans can predict it more than 48 hours earlier. The use of explainability techniques such as SHAP has also improved the understanding of the contribution of different variables to model predictions. This has improved the trust and transparency of these models, which has improved their adoption in the clinical field.

The usual ways to identify VAP were based on clinical evaluation, bedside scoring schemes, and surveillance algorithms. The Clinical Pulmonary Infection Score (CPIS) showed poor discriminatory power (area under the receiver operating characteristic curve [AUROC] 0.60-0.72) and included temperature, white blood cell count, tracheal secretions consistency, oxygenation, chest radiograph findings, and microbiologic results. Physician assessment was similarly inconsistent and subjective, and the CDC/NHSN surveillance criteria had little predictive value for VAP. In contrast, the ML algorithms in Table 2 demonstrated substantially better predictive power and were able to identify patients at risk of developing VAP at an earlier time point than conventional methods. The Random Forest algorithm demonstrated an AUROC of 0.84 using 42 variables related to the patients' clinical characteristics extracted from the MIMIC-III database. In turn, the XGBoost consistently showed the best performance (AUROC 0.831-0.94). Predictive models highlighted the duration of invasive mechanical ventilation, length of ICU stay, inflammatory markers such as C-reactive protein, potassium levels, blood urea nitrogen, red cell distribution, and preexisting comorbidities as important predictors of VAP. A meta-analysis of 10 different ML models revealed an increase in pooled AUROC of 0.88, sensitivity of 0.72, and specificity of 0.9. Some studies utilized deep learning approaches such as PREDICT LSTM, and showed the AUROC of over 0.85, with prediction 6, 12, and 24 hours before VAP diagnosis based on continuous physiologic monitoring. Similarly, Table 2 highlights the development of multi-source prediction models that utilize both electronic health records and CT scans to predict VAP with over 48 hours of lead time before diagnosis. Another approach using ensemble models can predict hospital-acquired VAP seven days earlier than the conventional methods.

Table 1: Characteristics and Performance of ML Models Applied to Ventilator-Associated Pneumonia Prediction and Prognostication

Journal [Ref]	Study Type / Algorithm	Key Findings	AUROC
J Crit Care / Eur J Intern Med[1]	Systematic Review & Meta-Analysis — 10 supervised static ML models	Sensitivity 0.72; specificity 0.90 (pooled). Key features: MV duration, ICU LOS, blood transfusion, nutrition, antibiotics. PROSPERO: CRD42022367014.	0.88 (pooled)
JMIR Medical Informatics[8]	Scoping Review (PRISMA-ScR) — Random Forest (45%), Neural Networks (36%), XGBoost (36%)	11 studies included. Random forest most common. MIMIC-III dominant dataset. VAP incidence 5–40%; attributable mortality ~10%.	NR (scoping)
BMC Pulmonary Medicine[3]	ML Model Development (Primary Study) — Random Forest (5-fold CV)	Sensitivity 0.74; specificity 0.71 at 24 h post-intubation. 42 VAP risk factors identified from admission vitals and labs.	0.84
PLOS ONE[10]	ML Model Development (Mortality Prediction) — 14 models incl. AdaBoost, CatBoost, XGBoost, RF, SVM, LightGBM, MLP, etc.	TRIPOD-AI compliant. XGBoost best performer. SHAP used for interpretability. Predicts in-hospital mortality in VAP patients.	0.88 (XGBoost)
Computational Biology and Medicine[25]	ML Model — MDRO-VAP Prediction — Logistic Regression, XGBoost, Random Forest, Gradient Boosting	XGBoost best: AUC 0.831 (95% CI 0.785–0.877). SHAP: RDW, MV duration, anion gap, basophil%, neutrophil%, kidney disease, hypertension significant.	0.831 (XGBoost)
PMC[26]	ML Development + External Validation — Best-performing ensemble model (unspecified)	AUROC 0.866 (internal), 0.817 (external). Predicts VAP up to 1.5 days before clinical diagnosis. Key: platelet count, PEEP, MV duration, inflammatory markers.	0.86 / 0.81
J Clin Med[11]	Deep Learning (LSTM) — PREDICT: LSTM neural network (24 h temporal	Best AUPRC: 96.0%, 94.1%, 94.7% at 6, 12, 24 h. Outperformed XGBoost (94.1%) and RF	0.85+ (LSTM)

	windows)	(91.7%) at 6 h. Key signs: RR, SpO_2, HR, Temp, MAP.	
Clinical Respiratory Journal[27]	AI Framework (Multisource) — AI-enhanced ensemble, multisource analytics (CT + EHR)	Longest lead time >48 h. Addresses high-risk early identification using multisource data including CT imaging.	0.87+

Abbreviations: AUROC = Area Under the Receiver Operating Characteristic Curve; RF = Random Forest; XGBoost = Extreme Gradient Boosting; LSTM = Long Short-Term Memory; SHAP = SHapley Additive exPlanations; AUPRC = Area Under Precision-Recall Curve; NR = Not Reported.

Table 2: Comparative Analysis of Traditional Scoring Systems and Machine Learning Algorithms for Ventilator-Associated Pneumonia

Method / Approach	Diagnostic Performance	Prediction Timing & Data Source	Key Features, Interpretability & Notes
Traditional Methods			
Clinical Pulmonary Infection Score (CPIS) Traditional scoring	AUROC 0.60–0.72 Sensitivity ~0.50–0.65 Specificity ~0.55–0.70	At time of assessment (no early warning) Multi-centre ICU bedside scoring	High interpretability (clinician-interpreted). Uses temperature, WBC, tracheal secretions, oxygenation (PaO2/FiO2), CXR, tracheal aspirate culture. Easy bedside use but poor individual-level discrimination.
Conventional Clinical Judgement (physician assessment) — Traditional clinical	AUROC 0.60–0.70 Sensitivity/specificity variable	No advance warning Routine ICU practice	High interpretability (direct physician reasoning). Depends on clinical experience; subject to cognitive bias; no standardised algorithm.
CDC/NHSN VAE/VAP Surveillance Criteria Traditional surveillance	AUROC NR Sensitivity/specificity moderate	Post-event (retrospective flag) Hospital infection surveillance	Moderate interpretability (rule-based). SpO2, FiO2, PEEP changes; microbiological culture; designed for surveillance, not real-time prediction.

Machine Learning Models			
Random Forest Liang et al., 2022 — ML ensemble	AUROC 0.84 Sensitivity 0.74 Specificity 0.71	24h post-intubation MIMIC-III (retrospective)	Moderate interpretability (feature importance). 42 risk factors; MV duration, ICU LOS, admission vitals, laboratory indices. Five-fold CV. Most commonly used algorithm (~45% of VAP studies).
XGBoost – TBI-VAP Ashrafi et al., 2025 — ML gradient boosting	AUROC 0.94 Sensitivity/specificity NR	Early prediction (timing NR) MIMIC-III (TBI patients)	High interpretability (SHAP analysis). ICU LOS, hospital LOS, serum potassium, BUN. Best performer across 6 models; accuracy 0.875. Outperforms benchmarks by 23.4%.
XGBoost – MDRO-VAP Cui et al., 2025 — ML gradient boosting	AUROC 0.831 (95% CI 0.785–0.877)	Prediction before MDRO isolation MIMIC-IV (n=972)	High interpretability (SHAP analysis). RDW, MV duration, anion gap, basophil%, neutrophil%. Comorbidities: kidney disease (p=0.014), hypertension (p=0.006).
XGBoost – In-hospital Mortality Wei et al., 2025 — ML gradient boosting	AUROC 0.882 / 0.871 (training / external)	Admission to ICU MIMIC-IV (training + external)	High interpretability (SHAP, TRIPOD-AI). 14 models compared; best was XGBoost. SHAP dashboard for clinical transparency. Predicts VAP-associated mortality, not VAP onset.
ML Ensemble PMC12870606, 2025 — ML ensemble	AUROC 0.866 (internal) 0.817 (external)	Up to 7 days before clinical diagnosis Prospective ICU + MIMIC-IV external	Moderate interpretability (unspecified). Platelet count, PEEP, MV duration, inflammatory markers. Only study with true prospective

			design and 7-day advance warning.
Pooled ML – Meta-Analysis Frondelius, 2024 — 10 ML models	AUROC 0.88 (0.82–0.94) Sensitivity 0.72 (0.45–0.98) Specificity 0.90 (0.85–0.94)	Variable across studies Multiple ICUs; PubMed/Cochrane/WoS systematic search	Variable interpretability (I^2>97%). Care-related factors dominate: MV duration, ICU LOS, blood transfusion, nutrition strategy, antibiotics. High heterogeneity limits pooled conclusions.
PREDICT (LSTM) Agard et al., 2025 — Deep learning	AUROC 0.85+ (prospective) AUPRC 96.0% at 6h	6, 12, 24h pre-VAP MIMIC-IV (retrospective + prospective)	Low interpretability (black-box DL). 5 vital signs: RR, SpO2, HR, Temp, MAP; 24h temporal windows. Best for high-acuity continuous monitoring. Outperforms RF at 6h (AUPRC 96.0% vs 91.7%).
AI Multisource Framework Zhang et al., 2025 — AI ensemble (multisource)	AUROC 0.87+	>48h (longest lead time) Prospective ICU (clinical deployment)	Moderate interpretability (ensemble). CT imaging + EHR + multisource analytics. Longest early warning window. Requires advanced digital ICU infrastructure.

NR = Not Reported. AUROC = Area Under the Receiver Operating Characteristic Curve. MV = Mechanical Ventilation. ICU = Intensive Care Unit. LOS = Length of Stay. SHAP = SHapley Additive explanations. DL = Deep Learning. MDRO = Multi-Drug-Resistant Organism. TBI = Traumatic Brain Injury. VAP = Ventilator-Associated Pneumonia. LSTM = Long Short-Term Memory. EHR = Electronic Health Records.

Discussion

The current review proves that ML and DL models demonstrate better performance than conventional clinical predictors in terms of early identification of ventilator-associated pneumonia. The results highlight a limited ability of both CPIS (the clinical pulmonary infection score) and the physician's assessment to predict VAP accurately. This limitation is primarily attributed to the delayed appearance of relevant clinical signs. In contrast, AI models can track changes in multiple variables and better detect subtle clinical deterioration that may be missed by healthcare providers.

Prolonged mechanical ventilation suppresses host airway host defense, and the concomitant invasive procedures plus broad-spectrum antibiotic use increase the likelihood

of resistance and acquisition of multidrug resistant organisms and ventilator associated pneumonia.[23]

The findings of Liang et al. (2022) [3] support the growing importance of ensemble machine learning approaches in critical care medicine. Their Random Forest model was able to produce clinically useful early predictions based on routinely collected intensive care variables. Similar conclusions were reached by Frondelius et al. (2024)[1], whose meta-analysis showed pooled AUROC values approaching 0.88 across ten machine learning models. Both studies identified care-related variables such as mechanical ventilation duration, ICU length of stay, antibiotic exposure, and laboratory markers as major predictors of VAP development.

Among gradient boosting algorithms, XGBoost consistently emerged as the most robust predictive model across studies. Such as Ashrafi et al. (2025)[21], and Wei et al. (2025) [8] independently demonstrated that XGBoost showed better performance than traditional machine learning algorithms such as Logistic Regression, Support Vector Machine, Random Forest, AdaBoost, and Artificial Neural Networks. This indicates that gradient boosting approaches could be more efficient in the analysis of non-linear ICU data with multiple variables.

Deep learning approaches appear particularly promising for continuous ICU monitoring because of their ability to process sequential temporal data. Agard et al. (2025) [11] demonstrated that the PREDICT LSTM was able to predict the development of VAP 6, 12, and 24 hours before official diagnosis. The ability to use continuously measured variables such as respiratory rate, oxygen saturation, temperature, heart rate, and mean arterial pressure reflects the shift towards predictive surveillance in critical care medicine.

The multisource AI framework proposed by Zhang et al. (2025)[8] represents another significant step by incorporating CT scans, electronic health records, and various other clinical data sources. The model's ability to utilize multiple data types, as opposed to the tabular data used by conventional ML algorithms, enhances both the robustness and predictive window. The prediction window of over 48 hours emphasizes the potential for implementing preventive measures and optimizing antibiotic use.

Although there are positive results in the studies conducted, there are several limitations to the research. The majority of the studies reviewed were conducted using data sets from a single centre, including MIMIC-III and MIMIC-IV databases. This limitation indicates that the results of the research cannot be applied to populations in different geographic locations. Frondelius et al. (2024)[1] additionally reported substantial heterogeni city (I^2 >97%) among pooled machine learning studies, emphasizing variability in dataset composition, diagnostic definitions, predictor selection, and validation methodologies.

In their study, the authors trained a machine learning model to predict a patient's survival or death based on their clinical presentation when admitted to the ICU connected to a mechanical ventilator. In the experiment, they used a cohort of 660 patients hospitalized in three intensive care units at AIIMS, Delhi. The researchers reduced 98 different variables to 39 most important features using the Shapley-Additive-explanations (SHAP) and a Random Forest algorithm. They found that the three most predictive features for mortality were the Sequential Organ Failure Assessment (SOFA) score, International Normalized Ratio (INR), and respiratory rate. The researchers evaluated many binary classification

models and report that the K-nearest-neighbor (KNN) model with the SHAP features that were selected using the Random Forest, had the best performance on the testing set with an accuracy of 0.80 and an area under the receiver operating characteristic (AUROC) of 0.84. It significantly outperformed the similar model that only used the SOFA score with an accuracy of 0.73 and AUROC of 0.73. The authors concluded that automated machine learning methods could help with patients' early triage and treatment in ICUs by assisting doctors in clinical decision-making.[24]

Overall, based on the information provided, artificial intelligence in medicine appears to be best utilized as an ancillary tool to physicians rather than a replacement for them. Work towards validating the technology across multiple medical centers, ensuring its transparency through explainable artificial intelligence, testing it in real-world settings such as intensive care units, and publishing standardized results for review will all be important steps in the technology's potential future use.

A Roadmap for Implementation of Ventilator-Associated Pneumonia Prediction

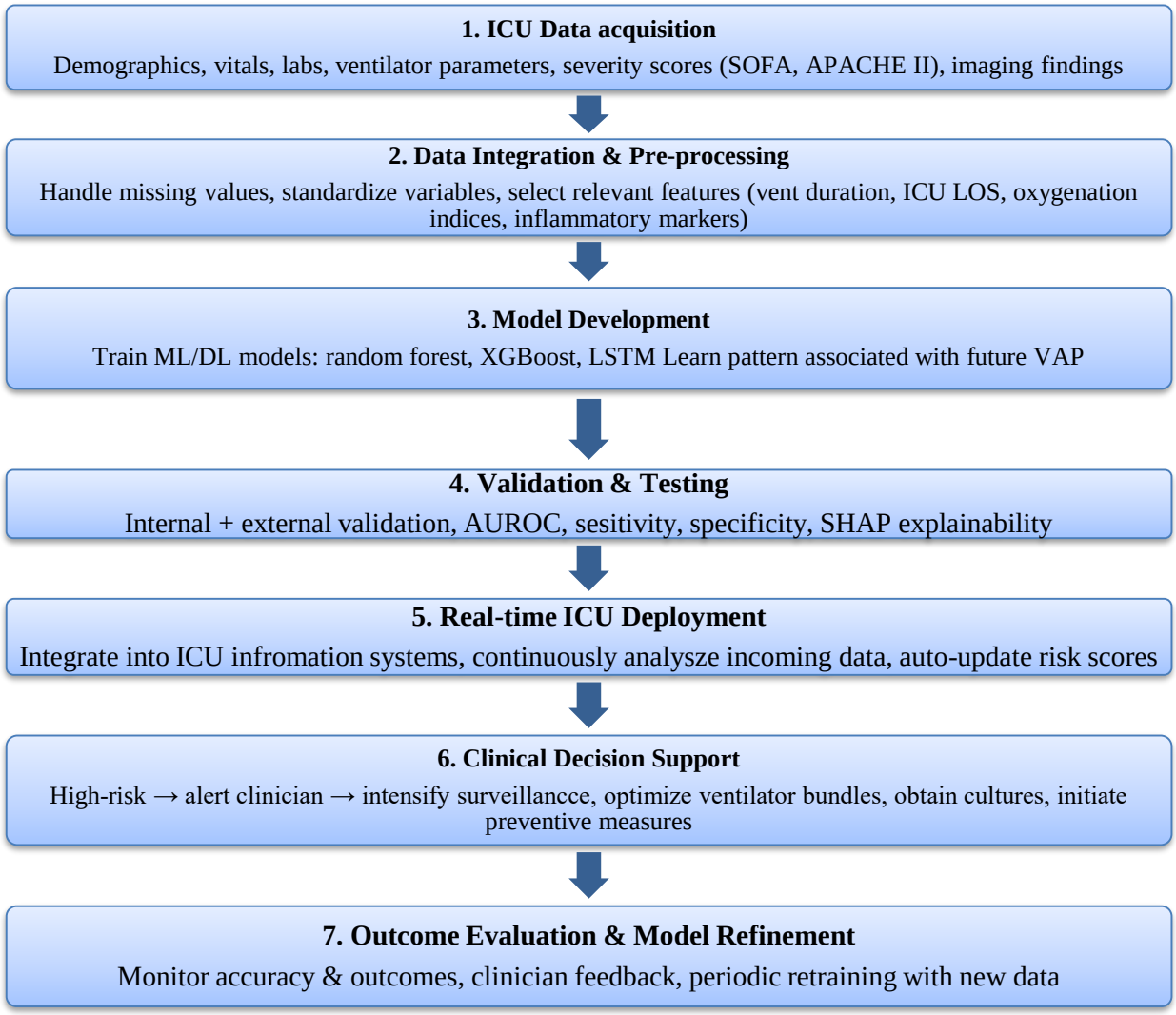

Figure 1. Proposed workflow for an artificial intelligence (AI)-based predictive model for ventilator-associated pneumonia (VAP) in intensive care unit (ICU) patients

A Roadmap for Implementation of Ventilator-Associated Pneumonia Prediction as demonstrated in figure 1 gives the clinical implementation framework for ventilator-associated pneumonia prediction consists of seven steps, beginning with ICU Data Acquisition. In the first stage, the required data is collected from various sources, including patient demographics, vital signs, laboratory tests, radiographic findings, and others. The data acquired from different sources in the ICU consists of a patient's age, gender, heart rate, respiratory rate, peripheral oxygen saturation, temperature, invasive and non-invasive blood pressure, laboratory tests results such as white blood cell count, arterial blood gases, and inflammatory biomarkers, ventilator parameters such as tidal volume, positive end-expiratory pressure, and the fraction of inspired oxygen, severity of illness indicators such as Sequential Organ Failure Assessment and Acute Physiology and Chronic Health Evaluation III, and radiographic findings. The next step is Data Integration and Preprocessing, which addresses the issue of missing and irrelevant data in the integrated database. This step also includes the elimination of noise and the transformation of data into a suitable format for the chosen data mining techniques. For instance, the data may be standardized using the z-score standardization method if continuous variables are used in the analysis. In addition, the variables that are not relevant to the prediction model or that have no significant impact on the outcome are removed. Some of the variables used in the analysis include the duration of mechanical ventilation and ICU stay, oxygenation, and inflammatory biomarkers such as the PaO2/FiO2 ratio. The data is then used to train the machine learning model in the Model Development stage. The most common data mining techniques used in model development are random forests, gradient boosting, and long short-term memory. The developed model is then subjected to the Validation and Testing Process. This stage determines the accuracy of the developed model. The most appropriate statistical evaluation measures should be used to report the accuracy of the model. These assessment metrics include area under the receiver operating curve, sensitivity, specificity, positive predictive value, negative predictive value, and calibration slope. In addition, the model can be validated using an independent data set that was not used to train the model. In the next step, the model is deployed in real-time ICU surveillance, where it is used to monitor patients' vital signs and predict the likelihood of developing VAP. This involves integrating the model into the existing ICU surveillance system or electronic health records. The model then continuously analyzes the data stream and updates the patient's risk score in real-time without requiring any input from clinicians. The risk score is then used to trigger alerts in the Clinical Decision Support component. This component uses predefined evidence-based standard operating procedures and critical thresholds to determine when to notify clinicians. If a patient is determined to be at risk of developing VAP, alerts are sent to clinicians, and evidence-based measures are recommended to reduce the likelihood of VAP developing. The recommended actions include enhanced surveillance and rapid response, the implementation of bundles such as elevating the head of the bed, subglottic suctioning, and chlorhexidine mouth care, and obtaining lower respiratory tract cultures to identify causative organisms. The final stage of the framework is Outcome Evaluation and Model Refinement. This stage ensures that the model is continuously updated and improved based on new data and clinician feedback. In addition, the stage monitors the model's performance in terms of its

ability to predict patient outcomes and the impact that the model has on patient outcomes. The model's performance is then used to refine it and enhance its accuracy and applicability in the ICU environment. The process also involves clinicians reviewing the model's recommendations and providing feedback to the developers. This feedback can then be used to improve the model and ensure that it is appropriately integrated into the ICU environment. Finally, the model's performance is evaluated on a regular basis to ensure that it is up-to-date and relevant.

The proposed workflow for developing and implementing an AI-based model for predicting ventilator-associated pneumonia in ICU patients is illustrated in Figure 1. First of all, collecting the required data falls under the responsibility of the intensive care medical staff. It should include the patient's demographical information, vital signs, laboratory tests, ventilator parameters, evidence of disease severity (SOFA and APACHE II scores), and diagnostic imaging. The next step is to prepare a data set for analysis by combining and structuring the input information, removing irrelevant or missing parts, and choosing significant indicators, such as the duration of treatment under a ventilator, ICU stay, oxygen saturation levels, and inflammatory factors.

The pre-processed database was used to develop the model, which was trained using machine learning and deep learning algorithms, such as Random Forest, XGBoost, and long short-term memory networks, for the recognition of future VAP patterns. The performance of the developed algorithm was evaluated during the validation and test stages by conducting internal and external validations and comparing the results using the area under the receiver operating characteristic curve, sensitivity, specificity, and explainability tools, such as SHAP analysis.

After the prediction model has been validated, it can be deployed and implemented in ICU information systems. The model should analyze the constant input data stream about the patients' current status and update the VAP risk scores accordingly. For patients at high risk, clinicians could get alerts and respond by performing the recommended evidence-based procedures, such as surveillance for ventilator-associated pneumonia, implementing bundles for the prevention of VAP, obtaining the necessary cultures, and taking other actions. Lastly, the model should be evaluated and updated constantly based on treatment outcome assessments and feedback received to ensure that it remains relevant and accurate.

Artificial intelligence, machine learning, and deep learning have great potential in predicting ventilator-associated pneumonia (VAP) among mechanically ventilated patients. Advanced algorithms offer more accurate and earlier predictions than classical score systems since they can utilize a broader range of variables to detect VAP development. For instance, the use of ensemble methods and XGBoost is considered highly efficient for VAP prediction, whereas long short-term memory architectures enable continuous surveillance of the patient's state in real-time. Thus, the application of these technologies can prolong the time of VAP prediction and improve the accuracy of detection.

Declarations

Acknowledgements

The authors sincerely acknowledge the Departments of Medical Microbiology, Critical Care and Respiratory Medicine, School of Medical Sciences and Research, Sharda University, Greater Noida, for providing the

academic environment and resources that facilitated the preparation of this review article.

Authors' Contribution

Queeny Baruah, Ayushi Mishra, and Shefali Singh were involved in the conceptualization of the review, comprehensive literature search, critical analysis and interpretation of the literature, manuscript drafting, preparation of figures and tables, editing of the manuscript.

Dr. V.S. Randhawa provided conceptual guidance, supervised the preparation of the manuscript, critically reviewed the intellectual content, and approved the final version of the manuscript for publication. Dr. Ram Murti Sharma and Dr. Ankit provided the feedback from the critical care point of view and edited the manuscript. Dr. Sonisha Gupta provided the feedback from the pulmonary point of view and edited the manuscript.

Data Availability Statement

No new data were generated or analysed during the preparation of this review article. All information presented is based on previously published literature, which has been appropriately cited.

Ethical Approval Statement

Ethical approval was not required for this review article because it does not involve human participants, animals, clinical specimens, or primary experimental data.

Informed Consent Statement

Informed consent was not required because this review article does not involve human participants or identifiable personal data.

References

1. Frondelius T, Atkova I, Miettunen J, Rello J, Vesty G, Chew HS, Jansson M. Early prediction of ventilator-associated pneumonia with machine learning models: A systematic review and meta-analysis of prediction model performance☆. European Journal of Internal Medicine. 2024 Mar 1;121:76-87.
2. Gaudet A, Martin-Loeches I, Povoa P, Rodriguez A, Salluh J, Duhamel A, Nseir S, TAVeM Study Group. Accuracy of the clinical pulmonary infection score to differentiate ventilator-associated tracheobronchitis from ventilator-associated pneumonia. Annals of Intensive Care. 2020 Jan 1;10(1):101.
3. Liang Y, Zhu C, Tian C, Lin Q, Li Z, Li Z, Ni D, Ma X. Early prediction of ventilator-associated pneumonia in critical care patients: a machine learning model. BMC Pulmonary Medicine. 2022 Jun 25;22(1):250.
4. Samadani A, Wang T, van Zon K, Celi LA. VAP risk index: early prediction and hospital phenotyping of ventilator-associated pneumonia using machine learning. Artificial Intelligence in Medicine. 2023 Dec 1;146:102715.
5. Bates DW, Levine D, Syrowatka A, Kuznetsova M, Craig KJ, Rui A, Jackson GP, Rhee K. The potential of artificial intelligence to improve patient safety: a scoping review. NPJ digital medicine. 2021 Mar 19;4(1):54.
6. Hendriksen JM, Geersing GJ, Moons KG, de Groot JA. Diagnostic and prognostic prediction models. Journal of Thrombosis and Haemostasis. 2013 Jun;11:129-41.
7. Pirracchio R, Cohen MJ, Malenica I, Cohen J, Chambaz A, Cannesson M, Lee C, Resche-Rigon M, Hubbard A, ACTERREA Research Group. Big data and targeted machine learning in action to assist medical decision in the ICU. Anaesthesia Critical Care & Pain Medicine. 2019 Aug 1;38(4):377-84.

8. Zhang J, Yang P, Zeng L, Li S, Zhou J. Ventilator-associated pneumonia prediction models based on AI: scoping review. JMIR Medical Informatics. 2024 May 14;12(1):e57026.
9. Wu WT, Li YJ, Feng AZ, Li L, Huang T, Xu AD, Lyu J. Data mining in clinical big data: the frequently used databases, steps, and methodological models. Military Medical Research. 2021 Aug 11;8(1):44.
10. Agard G, Roman C, Guervilly C, Forel JM, Orléans V, Barrau D, Auquier P, Ouladsine M, Boyer L, Hraiech S. An innovative deep learning approach for ventilator-associated pneumonia (vap) prediction in intensive care units—pneumonia risk evaluation and diagnostic intelligence via computational technology (predict). Journal of Clinical Medicine. 2025 May 13;14(10):3380.
11. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, Moody B, Szolovits P, Anthony Celi L, Mark RG. MIMIC-III, a freely accessible critical care database. Scientific data. 2016 May 24;3(1):1-9.
12. Akça O, Koltka K, Uzel S, Çakar N, Pembeci K, Sayan MA, Tütüncü AS, Karakas SE, Çalangu S, Özkan T, Esen F. Risk Factors for Early-onset, Ventilator-associated Pneumonia in Critical Care Patients: Selected Multiresistantversus Nonresistant Bacteria. Anesthesiology. 2000 Sep 1;93(3):638-45.
13. Galal YS, Youssef MR, Ibrahiem SK. Ventilator-Associated Pneumonia: Incidence, Risk Factors and Outcome in Paediatric Intensive Care Units at Cairo University Hospital. J Clin Diagn Res. 2016;10(6):SC06-SC11. doi:10.7860/JCDR/2016/18570.7920
14. Tejerina E, Frutos-Vivar F, Restrepo MI, Anzueto A, Abroug F, Palizas F, González M, D'Empaire G, Apezteguía C, Esteban A, Internacional Mechanical Ventilation Study Group. Incidence, risk factors, and outcome of ventilator-associated pneumonia. Journal of critical care. 2006 Mar 1;21(1):56-65.
15. Liang YJ, Li ZL, Wang L, Liu BY, Ding RY, Ma XC. Comparision of risk factors and pathogens in patients with early-and late-onset ventilator-associated pneumonia in intensive care unit. Zhonghua nei ke za zhi. 2017 Oct 1;56(10):743-6.
16. Amador T, Saturnino S, Veloso A, Ziviani N. Early identification of ICU patients at risk of complications: regularization based on robustness and stability of explanations. Artificial Intelligence in Medicine. 2022 Jun 1;128:102283.
17. Giang C, Calvert J, Rahmani K, Barnes G, Siefkas A, Green-Saxena A, Hoffman J, Mao Q, Das R. Predicting ventilator-associated pneumonia with machine learning. Medicine. 2021 Jun 11;100(23):e26246.
18. Ding C, Guo Y, Mo Q, Ma J. Prediction model of postoperative severe hypocalcemia in patients with secondary hyperparathyroidism based on logistic regression and XGBoost algorithm. Computational and mathematical methods in medicine. 2022;2022(1):8752826.
19. Yang S, Cao L, Zhou Y, Hu C. A retrospective cohort study: predicting 90-day mortality for ICU trauma patients with a machine learning algorithm using XGBoost using MIMIC-III database. Journal of Multidisciplinary Healthcare. 2023 Dec 31:2625-40.
20. Cotia AL, Scorsato AP, da Silva Victor E, Prado M, Gagliardi G, de Barros JE, Generoso Jr JR, de Menezes FG, Hsieh MK, Lopes GO, Edmond MB. Integration of an electronic hand hygiene auditing system with electronic health records using machine learning to predict hospital-acquired infection in a

health care setting. American journal of infection control. 2025 Jan 1;53(1):58-64.

21. Hsu JF, Lin YC, Lin CY, Chu SM, Cheng HJ, Xu FW, Huang HR, Liao CC, Fu RH, Tsai MH. Deep learning models for early and accurate diagnosis of ventilator-associated pneumonia in mechanically ventilated neonates. Computers in Biology and Medicine. 2025 May 1;189:109942.
22. Torres A, Niederman MS, Chastre J, Ewig S, Fernandez-Vandellos P, Hanberger H, Kollef M, Bassi GL, Luna CM, Martin-Loeches I, Paiva JA. Summary of the international clinical guidelines for the management of hospital-acquired and ventilator-acquired pneumonia. ERJ open research. 2018 Jun 26;4(2).
23. Cui Z, Dong Y, Yang H, Li K, Li X, Ding R, Yin Z. Machine learning prediction models for multidrug-resistant organism infections in ICU ventilator-associated pneumonia patients: analysis using the MIMIC-IV database. Computers in Biology and Medicine. 2025 May 1; 190:110028.
24. Peltekian AK, Liao WT, Guggilla V, et al. Developing and externally validating machine learning models to forecast short-term risk of ventilator-associated pneumonia. Preprint. medRxiv. 2026;2026.01.28.26344858. Published 2026 Jan 30. doi:10.64898/2026.01.28.26344858
25. Zhang J, Wang Y, Cao Y, Li J, Dai S, Li Y, Zhang Z, Zhang X, Yang R, Zhang X, Chen J. Risk Factor Assessment and Predictive Modeling for Ventilator-Associated Pneumonia: Design and Clinical Implementation of an Artificial Intelligence-Enhanced Early Detection Framework Using Multisource Data Analytics. The Clinical Respiratory Journal. 2025 Dec;19(12):e70144.
26. Faucher M, Chetty SC, Shokouhi S, Barnes G, Rahmani K, Calvert J, Mao Q. Early prediction of ventilator-associated pneumonia in ICU patients using an interpretable machine learning algorithm.
27. Dos Santos RP, Silva D, Menezes A, Lukasewicz S, Dalmora CH, Carvalho O, Giacomazzi J, Golin N, Pozza R, Vaz TA. Automated healthcare-associated infection surveillance using an artificial intelligence algorithm. Infection Prevention in Practice. 2021 Sep 1;3(3):100167.
28. Chen CY, Lin WC, Yang HY. Diagnosis of ventilator-associated pneumonia using electronic nose sensor array signals: solutions to improve the application of machine learning in respiratory research. Respiratory research. 2020 Feb 7;21(1):45.
29. Tsai HC, Hsieh CY, Sung SF. Application of machine learning and natural language processing for predicting stroke-associated pneumonia. Frontiers in Public Health. 2022 Sep 29; 10:1009164.