European Journal of Internal Medicine 121 (2024) 76–87

Contents lists available at ScienceDirect

European Journal of Internal Medicine

journal homepage: www.elsevier.com/locate/ejim

Original Article

Early prediction of ventilator-associated pneumonia with machine learning models: A systematic review and meta-analysis of prediction model performance☆

Tuomas Frondelius[a], Irina Atkova[b], Jouko Miettunen[c,d], Jordi Rello[e,f,g], Gillian Vesty[h], Han Shi Jocelyn Chew[i], Miia Jansson[a,j,*]

[a] Research Unit of Health Sciences and Technology, University of Oulu, Oulu, Finland
[b] University of Oulu, Oulu, Finland
[c] Research Unit of Population Health, University of Oulu, Oulu, Finland
[d] Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland
[e] Global Health eCore, Vall d'Hebron Institute of Research (VHIR), Barcelona, Spain
[f] Centro de Investigacion Biomedica en Red de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III, Madrid, Spain
[g] Unité de Recherche FOVERA, Réanimation Douleur Urgences, Centre Hospitalier Universitaire de Nîmes, Nîmes, France
[h] School of Accounting, RMIT University, Melbourne, Australia
[i] Alice Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[j] Research Unit of Health Sciences and Technology, University of Oulu, Oulu, Finland, RMIT University, Melbourne, Australia

ARTICLE INFO

Keywords:
Artificial intelligence
Artificial ventilation
Machine learning
Meta-analysis
Predictive analytics
Ventilator-associated pneumonia

ABSTRACT

Background: Machine learning-based prediction models can catalog, classify, and correlate large amounts of multimodal data to aid clinicians at diagnostic, prognostic, and therapeutic levels. Early prediction of ventilator-associated pneumonia (VAP) may accelerate the diagnosis and guide preventive interventions. The performance of a variety of machine learning-based prediction models were analyzed among adults undergoing invasive mechanical ventilation.

Methods: This systematic review and meta-analysis was conducted in accordance with the Cochrane Collaboration. Machine learning-based prediction models were identified from a search of nine multi-disciplinary databases. Two authors independently selected and extracted data using predefined criteria and data extraction forms. The predictive performance, the interpretability, the technological readiness level, and the risk of bias of the included studies were evaluated.

Results: Final analysis included 10 static prediction models using supervised learning. The pooled area under the receiver operating characteristics curve, sensitivity, and specificity for VAP were 0.88 (95 % CI 0.82–0.94, I^2 98.4 %), 0.72 (95 % CI 0.45–0.98, I^2 97.4 %) and 0.90 (95 % CI 0.85–0.94, I^2 97.9 %), respectively. All included studies had either a high or unclear risk of bias without significant improvements in applicability. The care-related risk factors for the best performing models were the duration of mechanical ventilation, the length of ICU stay, blood transfusion, nutrition strategy, and the presence of antibiotics.

Conclusion: A variety of the prediction models, prediction intervals, and prediction windows were identified to facilitate timely diagnosis. In addition, care-related risk factors susceptible for preventive interventions were identified. In future, there is a need for dynamic machine learning models using time-depended predictors in conjunction with feature importance of the models to predict real-time risk of VAP and related outcomes to optimize bundled care.

Registration: PROSPERO CRD42022367014, registered on 24-10-2022.

* Corresponding author.
E-mail address: miia.jansson@oulu.fi (M. Jansson).

https://doi.org/10.1016/j.ejim.2023.11.009
Received 15 May 2023; Received in revised form 16 October 2023; Accepted 6 November 2023
Available online 18 November 2023

1. Introduction

Although lifesaving, mechanical ventilation predisposes patients to ventilator-associated pneumonia (VAP), which is the most serious respiratory infection, occurring in 23.5–39.3 % (12.2 to 22.4 cases per 1000 ventilator days) of intubated patients [1–2]. VAP accounts for more than half of all intensive care unit (ICU) antibiotic prescriptions being associated with multidrug-resistant organisms and overuse of broad-spectrum antibiotics [3] and significant mortality (22–43 %) rates [1]. VAP is also an important contributing factor to other causes of death, particularly sepsis and multiple-organ failure.

Since 2014, researchers have started to develop near real-time automation for the detection of VAP [4–10]. These expert systems are efficient and useful as a quality indicator [11], but they have not improved the diagnostic accuracy of VAP [8,12–13]. In addition, they are resource-intensive and unable to identify an individual who would benefit from a certain type of treatment based on the predicted disease course. Full compliance with preventive interventions has decreased morbidity and halved the daily bed occupancy and use of ventilators [14].

ICU is a high-tech environment for intensive and invasive monitoring. ICU clinical information systems collect, store, process, and present high-volume, high-velocity, and high-variety information (e.g., structured and unstructured data, waveforms) from multiple systems (e.g., laboratory, pathology, and radiology systems) and solutions (e.g., monitors, ventilators, infusion pumps) simultaneously, continuously, and synchronously to support clinical decision-making. Machine learning-based prediction models can continuously catalog, classify, and correlate large amounts of multimodal data to aid clinicians at diagnostic, prognostic, and therapeutic levels (e.g., risk assessment, patient profiling, resource allocation) [15]. Despite its clinical potential, medical artificial intelligence (AI) is not yet a universal solution due to uncertainty and distrust of AI predictions.

Diagnostic and prognostic prediction models may accelerate the diagnosis (e.g., stratify patients into high- and low-risk groups), guide preventive interventions (e.g., identify risk factors, predict the disease course), and determine the best treatment options (e.g., predict weaning failure) decreasing the risk of VAP and its consequences (e.g., prolonged mechanical ventilation, antimicrobial resistance, mortality). Compared to one-time activity (e.g., manual diagnosis) and static baseline information (e.g., clinical scoring tools), various prediction intervals and prediction windows can facilitate timelier diagnosis to guide preventive interventions and additional monitoring in VAP. In addition, dynamic clinical prediction models can increase predictive power across a certain time by computing and updating candidate predictors on a continuous basis over time [16]. In prior literature, however, the majority of prediction modeling studies have been retrospective with a high risk of bias due to insufficiently reported eligibility and recruitment methods as well as unvalidated and unadjusted electronic health record (EHR) data [16]. Correspondingly, the level of technology readiness has been low due to lack of clinical validation and workflow integration. In addition, none of the previous reviews have focused on the interpretability of machine learning-based prediction models [16–17] to foster explainability and trust in AI predictions.

For that reason, we reviewed and summarized the state of existing multivariable diagnostic and prognostic prediction models in VAP and related outcomes in adult patients undergoing invasive mechanical ventilation (IMV). Due to the lack of consensus on a "*gold standard*" definition, all diagnostic criteria were considered [2]. The primary objective was to compare the predictive performance of competing machine learning-based prediction models. Our secondary objectives were to identify risk factors for VAP and assess the interpretability, the technological readiness level and the risk of bias of the included studies.

This systematic review and meta-analysis was conducted in accordance with the Cochrane Collaboration's tool [18] and the PRISMA-DTA (Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies) statement [19]. The formal research question was as follows: *to what extent are existing machine learning-based prediction models able to predict VAP and related outcomes in adults undergoing invasive mechanical ventilation*? The study protocol was registered in the international prospective register of systematic reviews (PROSPERO) before the start of the study (CRD42022367014).

2. Material and methods

2.1. Literature search and inclusion criteria

An actual search was conducted in nine multi-disciplinary databases (ACM Digital Library/ACM Guide to Computing Literature, Astrophysics Data System, arXiV, Academic Search Ultimate, Cumulative Index to Nursing and Allied Health Literature [CINAHL], IEEE Xplore Digital Library, PubMed [Medline], Scopus, Web of Science) with the assistance of information specialists in October 2022, and the saved searches were then automatically updated until October 2023. The controlled (MesH in Medline Ovid and PubMed) and free-text terms were used through the Boolean operators (Table 1). In addition, manual searches of the reference lists, citations, and related articles (PubMed function) of the included studies were undertaken to identify additional studies missed from the original electronic searches. All original prediction modeling studies written in English were included if they met the predefined inclusion criteria (PICOTS):

- Population: Adults undergoing IMV.
- Index: Machine learning-based diagnostic and prognostic prediction models using regression (e.g., logistic regression) or non-regression (e.g., random forests, neural networks, and support vector machines) modeling techniques.
- Comparator: Manual diagnosis with and without clinical scoring tools.
- Outcomes: VAP with and without related outcomes.
- Timing: Models to be used prior to VAP and at the moment of diagnosis.
- Setting: Models to inform clinical decision making in ICU setting.

2.2. Selection of relevant studies

Two reviewers (MMJ and IA) screened the data independently. In case of disagreement, conflict was solved by discussion with each other. The study selection was carefully documented using the Covidence Systematic Review Software tool and Microsoft Excel spreadsheet to ensure its repeatability.

2.3. Data extraction and assessment of methodological quality

Two reviewers (MMJ and TF) assessed the risk of bias using the Prediction Model Risk of Bias Assessment Tool (PROBAST) [20]. In case of disagreement, conflict was solved by discussion with each other. The PROBAST includes 20 signaling questions across four key domains (participants, predictors, outcome, analysis), while each domain is assessed for a low, high, or unclear risk of bias (Table 2). Two reviewers (MMJ and IA) extracted the data (e.g., source of data, participants, outcomes, predictors, sample size, model development, model performance, model evaluation, results) using the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies (CHARMS) checklist [21] in conjunction with the PROBAST tool [20]. Discrimination (model's ability to differentiate between patients with and without VAP) was extracted to estimate prediction models' ability to distinguish patients with and without VAP [range from 0.5 (no discriminative ability) to 1 (perfect discriminative ability)]. Due to lack of calibration blots, summarization of calibration (the agreement between the frequency of observed events with the predicted probabilities) was impossible. The technological readiness level (range 1–9) were

Table 1
Search strategy.

Databases	PI(COT)S	MesH in Medline Ovid and PubMed	No of references
Scopus, Web of Science, and Academic Search Ultimate	Population	"respiration, artificial" OR "artificial respiration" OR "ventilator-associated" OR "ventilator-induced" OR "mechanical ventilation"	Scopus 296 Web of Science 180 Academic Search Premier 111
	Intervention	(stratification OR "ROC curve" OR discriminat* OR "c-statistic" OR "area under the curve" OR auc OR calibration* OR algorithm* OR multivariable OR "multi-variable" OR diagnos* OR prognos*) AND ("Artificial Intelligence" OR "Machine Learning" OR "Artificial Learning" OR "Bayesian Learning" OR "Deep Learning" OR "Knowledge Representation" OR "Neural Network*" OR "Probabilistic Network*" OR "Statistical Learning" OR "Support vector machine*" OR "Generalized linear model*" OR "Naive bayes*" OR "Ensemble method*" OR "Neural network model*" OR "Decision tree*")	
	Setting	hospital* OR "operation theater*" OR "emergency department*" OR "recovery room*" OR "tertiary care center*" OR "intensive care unit*" OR "operation room*" OR "Clinical Decision Unit*" OR "Clinical Observation Unit*" OR "respiratory care unit*" OR "Trauma Center*"	
PubMed (Medline)	Population	((("Respiration, Artificial"[Mesh]) OR "Pneumonia, Ventilator-Associated"[Mesh]) OR ("ventilator-associated"[Text Word] OR "ventilator-induced"[Text Word] OR "mechanical ventilation"[Text Word] OR "artificial respiration"[Text Word]))	270
	Intervention	((((((("Artificial Intelligence"[Mesh]) OR ("Artificial Intelligence"[Text Word] OR "Machine Learning"[Text Word] OR "Artificial Learning"[Text Word] OR "Bayesian Learning"[Text Word] OR "Deep Learning"[Text Word] OR "Knowledge Representation"[Text Word] OR Neural Network*[Text Word] OR Probabilistic Network*[Text Word] OR "Statistical Learning"[Text Word] OR Support vector machine*[Text Word] OR Generalized linear model*[Text Word] OR Naive bayes*[Text Word] OR Ensemble method*[Text Word] OR Neural network model*[Text Word] OR Decision tree*[Text Word] AND ((("ROC Curve"[Mesh]) OR (stratification[Text Word] OR "ROC curve"[Text Word] OR discriminat*[Text Word] OR "c-statistic"[Text Word] OR "c statistic"[Text Word] OR "area under the curve"[Text Word] OR AUC[Text Word] OR calibration*[Text Word] OR algorithm*[Text Word] OR multivariable [Text Word] OR "multi-variable"[Text Word] OR diagnos*[Text Word] OR prognos*[Text Word]))))	
	Setting	(((((((("Hospitals"[*Mesh*]) OR "Clinical Observation Units"[*Mesh*]) OR "Intensive Care Units"[*Mesh*]) OR "Operating Rooms"[*Mesh*]) OR "Emergency Service, Hospital"[*Mesh*])) OR (hospital*[Text Word] OR operation theater*[Text Word] OR emergency department*[Text Word] OR recovery room*OR tertiary care center*[Text Word] OR intensive care unit*[Text Word] OR operation room*[Text Word] OR Clinical Decision Unit*[Text Word] OR Clinical Observation Unit*[Text Word] OR respiratory care unit*[Text Word] OR Trauma Center*[*Text Word*]))	
CINAHL	Population	((MH "Respiration, Artificial+" OR MH "Pneumonia, Ventilator-Associated") OR ("artificial respiration" OR "ventilator-associated" OR "ventilator-induced" OR "mechanical ventilation"))	65 (peer reviewed)
	Intervention	(MH "ROC Curve" OR (stratification OR "ROC curve" OR discriminat* OR "c-statistic" OR "area under the curve" OR AUC OR calibration* OR algorithm* OR multivariable OR "multi-variable"OR diagnos* OR prognos*)) AND (MH "Artificial Intelligence+" OR ("Artificial Intelligence" OR "Machine Learning" OR "Artificial Learning" OR "Bayesian Learning" OR "Deep Learning" OR "Knowledge Representation" OR "Neural Network*" OR "Probabilistic Network*" OR "Statistical Learning" OR "Support vector machine*" OR "Generalized linear model*" OR "Naive bayes*" OR "Ensemble method*" OR "Neural network model*" OR "Decision tree*"))	
	Setting	(((MH "Hospitals+") OR (MH "Observation Units") OR (MH "Intensive Care Units+") OR (MH "Operating Rooms") OR (MH "Emergency Service+")) OR (hospital* OR "operation theater*" OR "emergency department*" OR "recovery room*" OR "tertiary care center*" OR "intensive care unit*" OR "operation room*"	

(*continued on next page*)

Table 1 (*continued*)

Databases	PI(COT)S	MesH in Medline Ovid and PubMed	No of references
		OR "Clinical Decision Unit*" OR "Clinical Observation Unit*" OR "respiratory care unit*" OR "Trauma Center*"))	
ACM Guide to Computing Literature	Population	[[All: "respiration, artificial"] OR [All: "artificial respiration"] OR [All: "ventilator-associated"] OR [All: "ventilator-induced"] OR [All: "mechanical ventilation"]]	275
	Intervention	[[All: "artificial intelligence"] OR [All: "machine learning"] OR [All: "artificial learning"] OR [All: "bayesian learning"] OR [All: "deep learning"] OR [All: "knowledge representation"] OR [All: "neural network"] OR [All: "probabilistic network"] OR [All: "statistical learning"] OR [All: "support vector machine"] OR [All: "generalized linear model"] OR [All: "naive bayesian"] OR [All: "ensemble method"] OR [All: "neural network model"] OR [All: "decision tree"] OR [All: "proportional hazards model"]] AND [[All: "roc curve"] OR [All: discriminat*] OR [All: "c-statistic"] OR [All: "area under the curve"] OR [All: auc] OR [All: calibration*] OR [All: algorithm*] OR [All: multivariable] OR [All: "multi-variable"] OR [All: diagnos*] OR [All: prognos*]]	
	Setting	[[All: hospital*] OR [All: "operation theater"] OR [All: "emergency department"] OR [All: "recovery room"] OR [All: "tertiary care center"] OR [All: "intensive care unit"] OR [All: "operation room*"] OR [All: "clinical decision unit"] OR [All: "clinical observation unit"] OR [All: "respiratory care unit"] OR [All: "trauma center"]]	
arXiV	Population	"respiration, artificial" OR "artificial respiration" OR "ventilator-associated" OR "ventilator-induced" OR "mechanical ventilation"	0
	Intervention	"multivariable prediction model*" OR "multi-variable prediction model*"	
Astrophysics Data System	Population	("respiration, artificial" OR "artificial respiration" OR "ventilator-associated" OR "ventilator-induced" OR "mechanical ventilation")	0
	Intervention	("Artificial Intelligence" OR "Machine Learning" OR "Artificial Learning" OR "Bayesian Learning" OR "Deep Learning" OR "Knowledge Representation" OR "Neural Network" OR "Probabilistic Network" OR "Statistical Learning" OR "Support vector machine" OR "Generalized linear model" OR "Naive bayes" OR "Ensemble method" OR "Neural network model" OR "Decision tree") AND (stratification OR "ROC curve" OR discriminat* OR "c-statistic" OR "area under the curve" OR AUC OR calibration* OR algorithm* OR multivariable OR "multi-variable" OR diagnos* OR prognos*)	
	Setting	(hospital* OR "operation theater" OR "emergency department" OR "recovery room" OR "tertiary care center" OR "intensive care unit" OR "operation room" OR "Clinical Decision Unit" OR "Clinical Observation Unit" OR "respiratory care unit" OR "Trauma Center")	
IEEE Xplore Digital Library	Population	"respiration, artificial" OR "artificial respiration" OR "ventilator-associated" OR "ventilator-induced" OR "mechanical ventilation" AND multivariable prediction model* OR multi-variable prediction model*	6
	Intervention	("Artificial Intelligence" OR "Machine Learning" OR "Artificial Learning" OR "Bayesian Learning" OR "Deep Learning" OR "Neural Network") AND (stratification OR "ROC curve" OR discriminat* OR "c-statistic" OR "area under the curve" OR AUC OR calibration* OR algorithm* OR multivariable OR "multi-variable" OR diagnos* OR prognos*)	
	Setting	"intensive care unit"	

assessed according to Fleuren et al. [22]. The interpretability of the included studies was assessed by whether the authors calculated the feature importance (e.g., Shapley values) for all the input features for a given model (a positive score indicates larger effect to predict a certain variable). In addition, the number of candidate predictors was considered for model interpretability.

2.4. Data analysis

The retrieved discrimination [e.g., area under the receiver operating characteristics (ROC) curve, AUROC] was summarized into a weighted average. In meta-analysis, we pooled prognostic prediction models with effect sizes (AUROCs) that used same measurement times for the same outcome. Standard errors were estimated based on normal distribution assumption. Due to heterogeneity of prediction modeling studies, random effects meta-analysis was performed [18]. The forest plot includes statistics for AUROC, sensitivity, and specificity with 95 % confidence intervals. In addition, the Higgins I^2 test was used to evaluate heterogeneity between the included studies ($I^2 \leq 25$ % for low, $I^2 < 50$ % for moderate, $I^2 \geq 50$ % for high) [23].

3. Results

3.1. Study selection and characteristics

Study selection (Fig. 1) was performed in three stages: At the first stage ($N = 1205$), the studies were screened for duplicate hits. At the

Table 2
Open data sources in retrospective cohort studies.

Database	Sample	Data sources	Research data sets
eICU Collaborative Research Database (eICU-CRD)	Over 200,000 ICU admission across the United States between 2014 and 2015	Clinical records	Patient demographics Clinical characteristics (e.g., outcome information) Laboratory test results Physiological measurements (e.g., vital signs) Medications and fluid balances Respiratory care (e.g., ventilator settings) Billing-related information (e.g., ICD-9 codes)
Medical Information Mart for Intensive Care (MIMIC)	Over 60,000 ICU admission in the Beth Israel Deaconess Medical Center between 2001 and 2012	Hospital EHR; ICU CIS; Social Security Administration Death Master File	Patient demographics Clinical characteristics (e.g., outcome information) Laboratory test results Reports of electrocardiogram and imaging studies Physiological measurements (e.g., vital signs) Medications and fluid balances Respiratory care (e.g., ventilator settings) Billing-related information (e.g., ICD-9, DRG, and CPT codes)
National Trauma Data Bank	Pediatric and adult patients admitted to Level I, II, III, IV, V or undesignated trauma centers between 2007 – 2020	Clinical records	Patient demographics Clinical characteristics (e.g., outcome information) Physiological measurements (e.g., vital signs) Pre-hospital information Emergency department information Hospital events Medications and fluid balances Respiratory care Billing-related information (e.g., ICD-10 and AIS codes)
OUTCOMEREA	ICU admissions across France starting from 1997 (–ongoing)	Clinical records	Patient demographics (e.g., admission features and diagnoses) Clinical characteristics (e.g., outcome information) Physiological measurements (e.g., vital signs)

second stage, potentially relevant studies were assessed by comparing the titles and abstracts ($n = 702$) against the predetermined inclusion criteria. At the third stage, the studies ($n = 21$) that appeared to meet the inclusion criteria were obtained for detailed assessment against the inclusion criteria. Eventually, 10 original prediction modeling studies were identified to meet the inclusion criteria (Fig. 2).

Among the 20 prediction modeling studies identified and assessed for eligibility, 12 were excluded due to wrong article type (conference abstract), design (lack of comparator) or outcome (lack of psychometrics). Majority (90 %) of included studies (Appendix 1) were retrospective cohort studies using either EHR [24–26] or open [27–32] data sources (Table 2). All included studies used the ICD-9 code for VAP (997.31) [24–32]. The most common study aims were predicting VAP without its consequences. The timeline for VAP diagnosis and VAP variable extraction varied from the first hour after ICU admission [25] to 24–48 h after initiation of IMV [27–30], and beyond.

3.2. Risk factors

Table 3 summarizes the identified risk factors for VAP. The identified patient-specific risk factors were: age [30], Glasgow coma scale (GCS) [27–29], comorbidities (e.g., diabetes, hypertension, coronary artery disease) [24], pneumothorax [24] as well as APACHE III, SOFA, and SAPS II-scores in mixed medical-surgical patients [30–31] and the Injury Severity Score in trauma patients [24,32]. The identified clinical criteria-related risk factors were: white blood cell (WBC) count [27–28, 30,33], PaO2/FiO_2-ratio [30,33], body temperature [30,33], and sputum production [29,33] and color [33]. The care-related risk factors for the best performing models were the time from injury to emergency department [24], intubation [32], the duration of mechanical ventilation [27,29,33], the length of ICU stay [32], blood transfusion [24,31], nutrition strategy [31], and the presence of antibiotics [29].

3.3. Predictive performance

Most studies used more than one evaluation metric related to discrimination (e.g., AUROC, sensitivity, and specificity) whereas calibration performance was not evaluated due to lack of studies (Appendix 1). The pooled AUROC for VAP (Fig. 3) and early VAP (Fig. 4) were 0.88 (95 % CI 0.82–0.94, I^2 98.4 %) and 0.84 (95 % CI 0.76–0.91, I^2 98.7 %), respectively. Effect sizes for model development are described in Fig. 5. The pooled sensitivity and specificity for VAP were 0.72 (95 % CI 0.45–0.98, I^2 97.4 %) and 0.90 (95 % CI 0.85–0.94, I^2 97.9 %), respectively. Compared to clinical scoring tools, the ML models outperformed the PIRO (predisposition, insult, response, organ dysfunction) and CPIS (clinical pulmonary infection score) scoring tools at all prediction times [29].

3.4. Interpretability

All of the included studies used supervised learning where conventional classification models were trained end-to-end. A set of candidate predictors varied from three [27–28] to 42 [30]. All predictors were available at the time the model was intended for use. Some of them (e.g., WBC count, body temperature, sputum production), however, were not excluded from the outcome definition [27–28,30,33]. Feature importance was evaluated in 70 % studies [24–25, 27–31) and reported in half of the studies [24,28–32]. Some studies utilized feature importance to optimize the model performance but didn't report the feature importance of the tested models [25,27]. Additionally, feature importance was reported and used to develop minimal input model in some studies, but the feature importance was not reported for the resulting models [28–29].

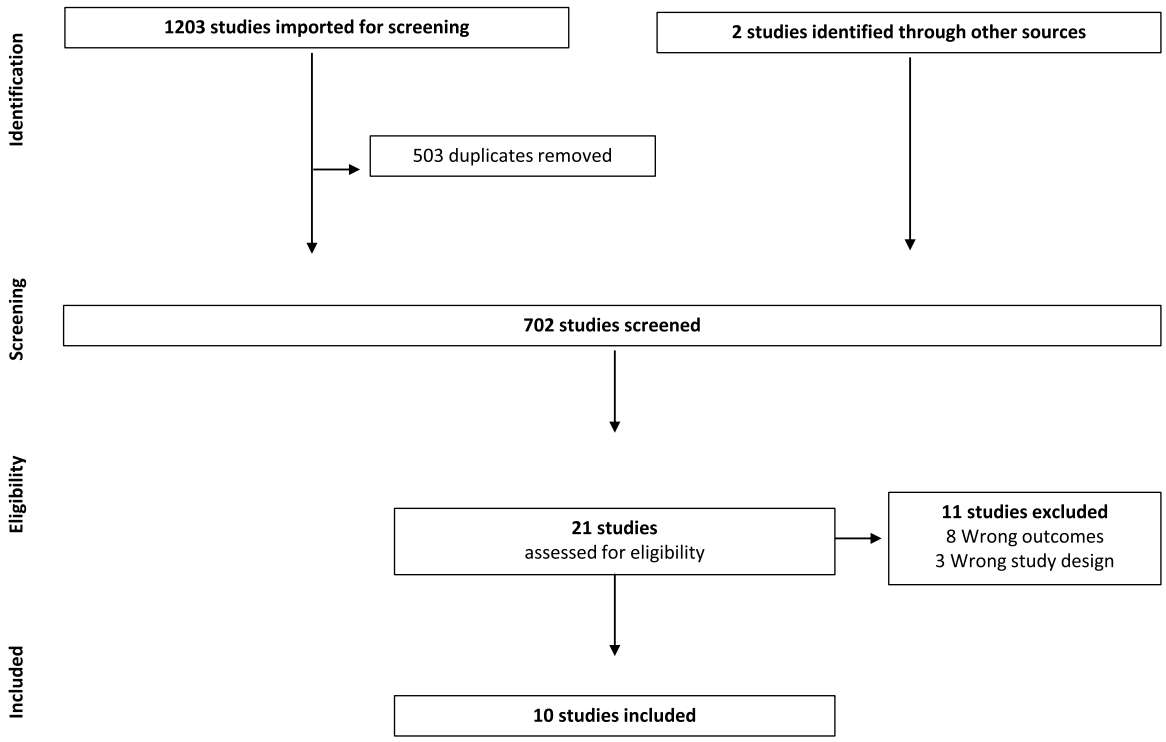

Fig. 1. Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) flow diagram of the study selection. Among the 20 prediction modeling studies identified and assessed for eligibility, 12 were excluded due to wrong article type (conference abstract), design (lack of comparator) or outcome (lack of psychometrics).

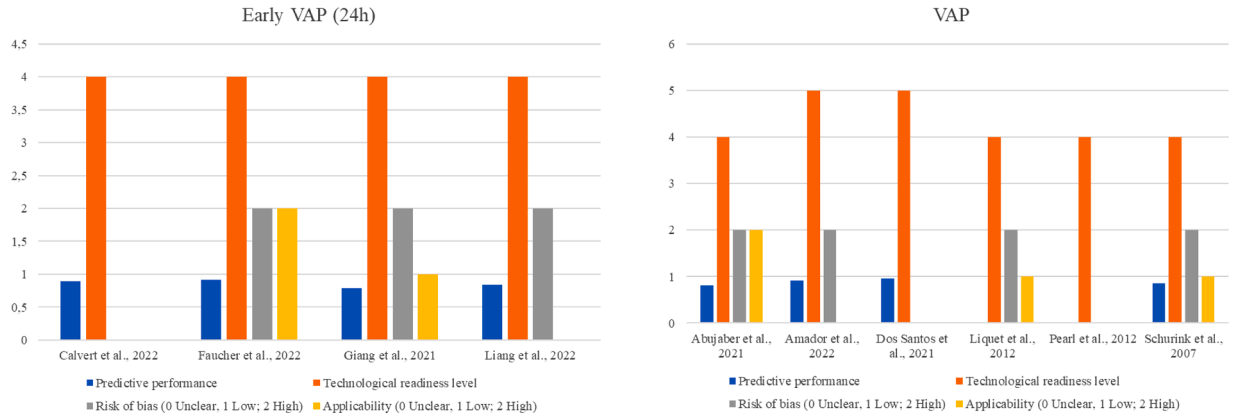

Fig. 2. Summary of included studies by the definition of VAP, predictive performance (AUROC), and the overall risk of bias and applicability.

3.5. *Technological readiness level*

The majority (80 %) of included studies (Fig. 2) were in the development phase (level 4). Two (20 %) of the included studies [25–26] were validated on an external dataset (level 5). Real-time data (level 6), workflow integration (level 7), clinical outcome evaluation (level 8), or model integration (level 9) were not considered (Appendix 1).

3.6. *Risk of bias*

In general, the current body of evidence was low due to the lack of clinical trials. All included studies had either a high or unclear risk of bias without significant improvements in applicability (Table 4). High risk of bias was most often originated in the domain "*participants*" due to lack of adjustment of the baseline risk/hazard in the analysis, "*outcome*" due to suboptimal method of outcome determination, and "*analysis*" due to low number of participants with the outcome (events per variable). Low to high concerns related to applicability were detected (Fig. 2).

4. Discussion

A variety of prediction models, prediction intervals, and prediction windows (e.g., 6, 12, 24, and 48 h) were identified. Compared to manual diagnosis and clinical scoring tools, the included machine learning-based prediction models demonstrated sufficient discrimination ability. Given the costs and complications associated with VAP, once implemented and prospectively validated, these models may diagnose VAP faster and more accurately than clinicians.

Contrary to diagnostic criteria, early (24 h) VAP was predicted to facilitate timely diagnosis to guide preventive interventions and additional monitoring [27–31]. The predictors of the best performing models included care-related risk factors, which are susceptible for preventive interventions (e.g., sedation and antibiotic strategies) [34,35]. Additional work is needed to determine the ideal data collection and

Table 3
Identified risk factors for ventilator-associated pneumonia (VAP).

Author(s), year	Source of data with participants	Outcome to be predicted	Risk factors
Abujaber et al., 2021	Patients with traumatic brain injury in Qatar, 2014–2019 ($n = 772$)	VAP ($n = 169$)	Time to emergency department, blood transfusion, comorbidity, Injury Severity Score, pneumothorax
Amador et al., 2022	Mixed medical surgical patients in 3 ICUs in Brazil, 2016–2018 ($n = 5474$)	VAP ($n = 39$)	Administrative data, vital signs, lab results
Calvert et al., 2022	MIMIC-III (v1.3) database, 2001–2012 ($n = 20,487$)	VAP at 24 ($n = 469$) hour-window	White blood cell count, Glasgow Coma Scale, duration of mechanical ventilation
Dos Santos et al., 2021	Mixed medical surgical patients in Southern Brazil, 2017 ($n = 5105$)	VAP ($n = 9$)	Vital signs, lab results, free text
Faucher et al., 2022	MIMIC-III (v1.3) database, 2001–2012 ($n = 19,141$)	VAP at 12, 24, 36, and 48 ($n = 470$) hour-window	White blood cell count, Glasgow Coma Scale, duration of mechanical ventilation
Giang et al., 2021	MIMIC-III (v1.3) database ($n = 6126$)	VAP at 6, 12, 24, and 48 h-window ($n = 524$)	Duration of mechanical ventilation, the presence of antibiotics, sputum test frequency, Glasgow Coma Scale
Liang et al., 2022	MIMIC-III database, 2001–2012 ($n = 38,515$)	VAP at 24 ($n = 212$) hour-window	Admission source, APACHE III, SOFA, age, body temperature, PaO_2/FiO_2 ratio, white blood cell count
Liquet et al., 2012	OUTCOMEREA database, 1996–2007 ($n = 2871$)	VAP ($n = 433$)	Age, SAPS II, parenteral nutrition
Pearl et al., 2012	National Trauma Data Bank, 2001–2005 ($n = 1438,035$)	VAP (n=NR)	Intubation, Injury Severity Score, intensive care unit length of stay >2 days
Schurink et al., 2007	Neurosurgical patients, University Medical center Utrecht, 2000–2003 ($n = 872$)	VAP ($n = 58$) Possible VAP ($n = 78$) Probable VAP ($n = 21$)	Duration of mechanical ventilation, white blood cell count, body temperature, antipyretic drugs, sputum production and color, PaO_2/FiO_2 ratio, X-ray

prediction windows and distinguish ventilator-associated tracheobronchitis (VAT) from VAP [1,36]. Decreasing the progress (e.g., dysbiosis) from VAT to VAP would improve patient outcomes [37].

It must be noted, however, that predictive power may be lost after a certain time [29], which reduces the overall relevance of the candidate predictors (e.g., patient-related risk factors). For that reason, case-specific variables (e.g., the components of the ventilator bundle) should be considered [38] to develop dynamic clinical prediction models [39]. In addition, other ventilator-associated events should also be taken into account to limit complications and improve outcomes [40–41].

Intelligible machine learning models were increasingly used to understand the prediction mechanism [25,29]. However, only half of the included studies reported feature importance for the candidate predictors. Typically, the feature importance scores were computed for random forest or boosting models yielding a numeric estimate for predictors (risk factor) contribution to the final predictions (risk of VAP), as well as allowing the features to be ordered and visualized according to their contribution to the positive prediction. It should be noted, however, that the feature importance scores are model specific and focus on correlation between the features and outcome, rather than causal relationships.

In line with previous literature [16–17], retrospective cohort design seems to be the most frequent non-clinical study design, since treatment of patients has not been influenced by the use of AI. In addition, the development of the technological readiness level seems to be horizontal instead of diagonal. According to our findings, the majority of included studies were in the development phase, where the models are tested and optimized without external validation. Thus, a significant amount of development is still required to improve the maturity of technologies during their conceptualization, development, and application stages. In addition, the models need to be tested using real-time data. Additional development is also needed to introduce the models to clinical workflow, evaluate clinical outcomes, and integrate the models in the hospital environment.

Conventional machine learning models used in the included studies had minimal issues regarding scalability and computational demands due to the relatively low number of predictors and low model complexity. This leaves patient privacy and data security as more important concerns when bringing the models to clinical environment. Data-acquisition and sampling methods used in clinical setting should extract relevant predictors in real-time (either using bedside measurements or electronic health records) without compromising patient security or privacy. Feature extraction should also be done in accordance with the relevant regulations and laws, which need to be considered when bringing the models to the next level of readiness.

In future, unsupervised (e.g., clustering based models) and semi-supervised (e.g., generative adversarial neural networks) machine learning algorithms can potentially reveal novel features for prediction of VAP from patient data. The models might be able to learn features or combinations of features which go unseen by human observer and conventional statistical analyses. However, it is important to keep in mind that the models might learn to identify correlations rather than causal features and the results may be difficult to interpret. Additionally, validating the performance of unsupervised models can be difficult due to the lack of clear target or ground truth in unsupervised learning. This issue can be alleviated by using labeled data during model development and training.

The included studies demonstrated sufficient discrimination ability, which is a prerequisite for clinical acceptance [42–43]. Prior to that, however, external validation is warranted in a clinical workflow. In future, it is important to determine whether the model addresses treated and/or non-treated patients and how the treatment effects were handled in the models. Evaluation of the feature importance would improve the robustness, interpretability, and stability of explanations, and reveal important predictors and potential biases in the datasets. In addition, real-time data infrastructures should be developed to respond to unknown unknowns.

This systematic review has several limitations. First of all, gray literature was not included. In addition, meta-regression was not conducted due to low number of studies. Second, the high risk of bias was most often originated in the domain "*participants*" due to insufficiently reported eligibility and recruitment methods as well as un-validated and unadjusted EHR data, "*outcome*" due to suboptimal method of outcome determination (e.g., clinical criteria), and "*analysis*" due to unreasonable events per variable. In addition, differences in predictors, prediction windows, study characteristics, case-mix, statistical analysis, and selective reporting were potential sources of heterogeneity across the included studies. Finally, calibration performance was not evaluated due to lack of studies. In addition, the impact of bundled care in the risk of VAP is unknown due to lack of discrete time-to-event data.

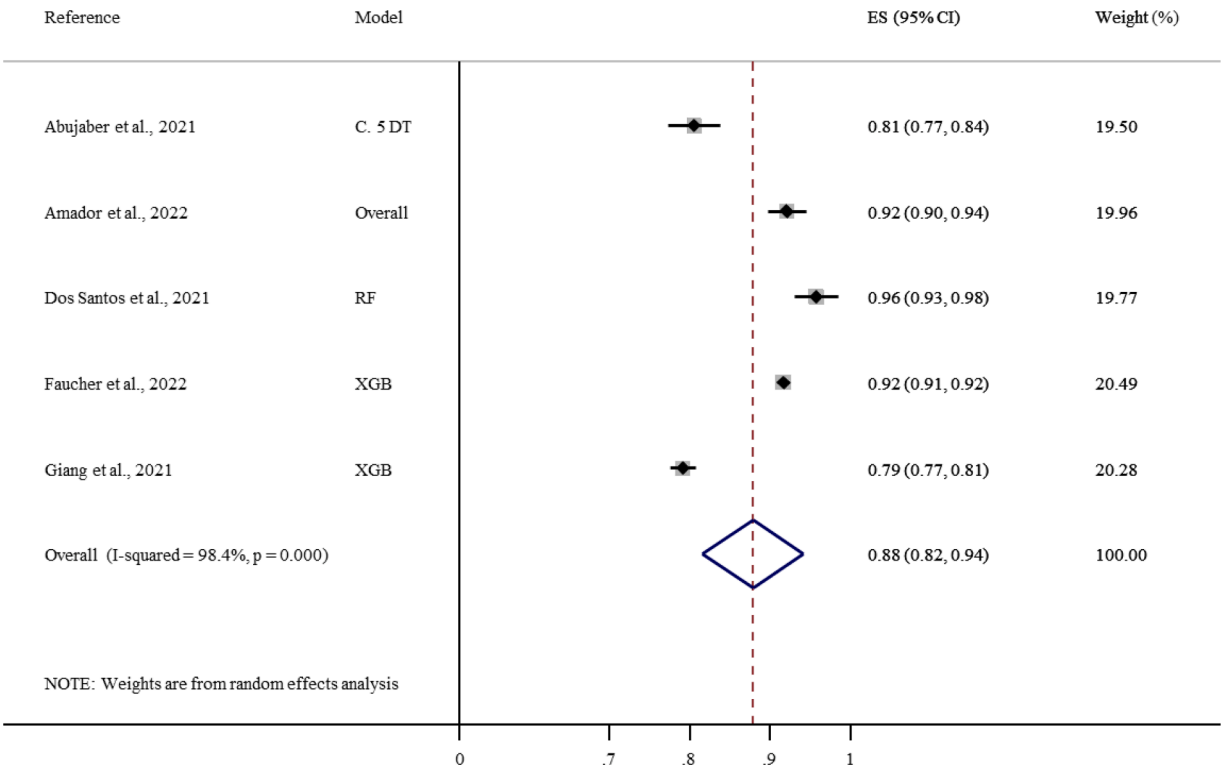

Fig. 3. Forest plot of area under the receiver operating characteristics curve of machine learning models to predict VAP.

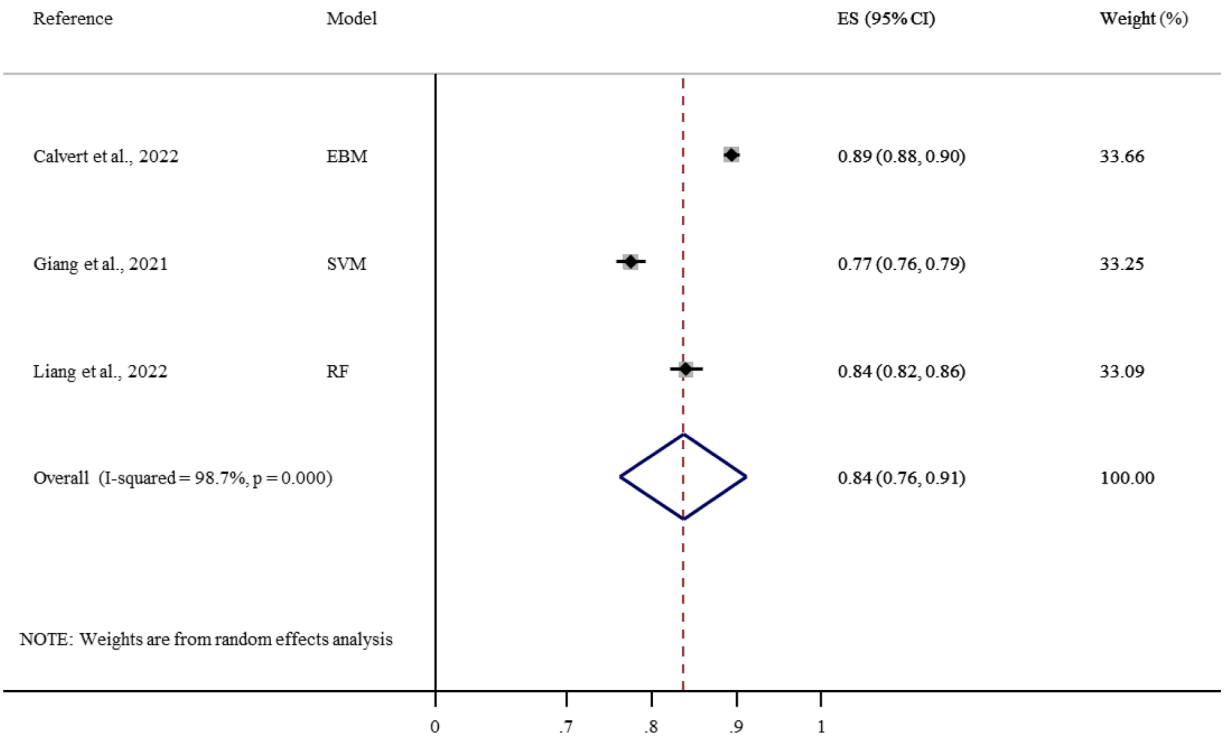

Fig. 4. Forest plot of area under the receiver operating characteristics curve of machine learning to predict early VAP.

5. Conclusion

A variety of the prediction models, prediction intervals, and prediction windows were identified to facilitate timely diagnosis. In addition, care-related risk factors susceptible for preventive interventions were identified. In future, there is a need for dynamic machine learning models using time-depended predictors in conjunction with feature importance of the models to predict real-time risk of VAP and related outcomes to optimize bundled care in adults undergoing IMV.

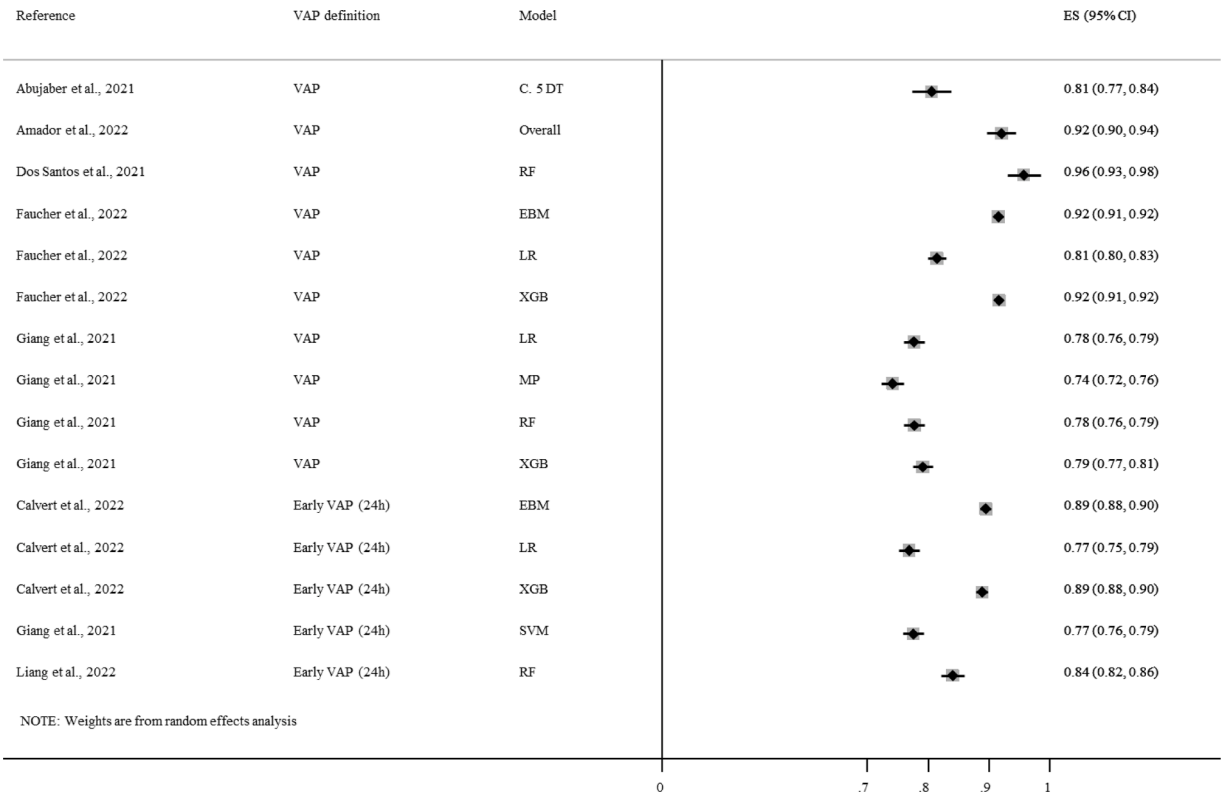

Fig. 5. Forest plot of area under the receiver operating characteristics curve of machine learning for model development.

Table 4
Risk of bias assessment.

	ROB				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applica-bility
Abujaber et al., 2021	–	+	?	–	–	+	?	–	–
Amador et al., 2022	–	+	?	?	+	+	?	–	?
Calvert et al., 2022	–	?	?	?	+	?	?	?	?
Dos Santos et al., 2021	–	?	?	?	+	?	?	?	?
Faucher et al., 2022	–	+	–	–	+	+	+	–	–
Giang et al., 2021	–	+	–	+	+	+	+	–	+
Liang et al., 2022	–	+	?	+	+	+	?	–	?
Liquet et al., 2012	–	+	+	+	+	+	+	–	+
Pearl et al., 2012	–	+	?	+	–	+	?	?	?
Schurink et al., 2007	+	+	+	–	–	+	+	–	+

ROB = risk of bias.
+ indicates low ROB/low concern regarding applicability.
- indicates high ROB/high concern regarding applicability.
? indicates unclear ROB/unclear concern regarding applicability.

Statements and declarations

Funding

No funding was received for the conduct of the review.

Institutional review board statement

Not applicable.

Informed consent statement

Not applicable.

Data availability

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

CRediT authorship contribution statement

Tuomas Frondelius: Conceptualization, Formal analysis, Writing – review & editing. **Irina Atkova:** Conceptualization, Data curation, Writing – review & editing. **Jouko Miettunen:** Formal analysis, Visualization, Writing – review & editing. **Jordi Rello:** Writing – review & editing, Supervision. **Gillian Vesty:** Writing – review & editing. **Han Shi Jocelyn Chew:** Writing – review & editing. **Miia Jansson:** Conceptualization, Data curation, Methodology, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research is connected to the DigiHealth-project, a strategic profiling project at the University of Oulu. The project is supported by the Academy of Finland (project number 326291) and the University of Oulu. In addition, we would like to acknowledge Informational Specialists Sirpa Grekula and Pertti Martinmäki, respectively.

Appendix 1

Table A1
Data extraction from predictive modeling studies.

Author(s), year	Source of data with participants	Type of prediction model study Development /Validation	Outcome to be predicted Diagnostic /Prognostic	Predictors	Model development	Model evaluation (AI level of readiness)	Result(s)	
Abujaber et al., 2021	Patients with TBI in Qatar, 2014–2019 (n = 772)	Development	Prognostic	VAP (n = 169)	Time to ED, blood transfusion, comorbidity, ISS, pneumothorax	C.5 DT model	Internal validation (Level 4)	AUC 80.5 %; SEN 43 %; SPE 95 %; ACC 83.5 %; PRE 71 %; NPV 86 %; F-score 54 %
Amador et al., 2022	Mixed medical surgical patients in 3 ICUs in Brazil, 2016–2018 (n = 5474)	Development + validation	Prognostic	VAP (n = 39)	Administrative data, vital signs, lab results	XGB	External validation in 2019 (n = 1069) (Level 5)	Robustness 4.9 %; stability 7.4 %
						RF		Robustness 5.0 %; stability 7.2 %
						LR		Robustness 5.5 %; stability 7.6 %
Calvert et al., 2022	MIMIC-III (v1.3) database, 2001–2012 (n = 20,487)	Development	Prognostic	VAP at 24 (n = 469) hour-window	WBC, GCS, duration of MV	XGB	Internal validation (Level 4)	AUC 0.888
						EBM		AUC 0.894
						LR		AUC 0.768
Dos Santos et al., 2021	Mixed medical surgical patients in Southern Brazil, 2017 (n = 5105)	Development + validation	Prognostic	VAP (n = 9)	Vital signs, lab results, free text	RF	External validation in 2018 (Level 5)	AUROC 95.67 % (SD 0.15); SEN 95.69 % (SD 0.48); SPE 89.02 % (SD 0.05); ACC 89.0 % (SD 0.05); NPV 99.99 % (SD 0.0); PPV 2.62 % (SD 0.02)
						LR		NR
						CNN		NR
Faucher et al., 2022	MIMIC-III (v1.3) database, 2001–2012 (n = 19,141)	Development	Prognostic	VAP at 12, 24, 36, and 48 (n = 470) hour-window	WBC, GCS, duration of MV	XGB	NR (Level 4)	AUC 0.916
						EBM		AUC 0.915
						LR		AUC 0.814
Giang et al., 2021	MIMIC-III (v1.3) database (n = 6126)	Development	Prognostic	VAP at 6, 12, 24, and 48 h-window (n = 524)	Duration of MV, the presence of antibiotics, sputum test frequency, GCS	XGB	Internal validation (Level 4)	AUROC 0.791
						RF		AUROC 0.777
						LR		AUROC 0.776
						SVM		AUROC 0.775
						MP		AUROC 0.741
Liang et al., 2022	MIMIC-III database, 2001–2012 (n = 38,515)	Development	Prognostic	VAP at 24 (n = 212) hour-window	Admission source, APACHE III, SOFA, age, body temperature, PaO_2/FiO_2 ratio, WBC	RF	Internal validation (Level 4)	AUC 84 %; SEN 74 %; SPE 71 %
Liquet et al., 2012	OUTCOMEREA database, 1996–2007 (n = 2871)	Development	Prognostic	VAP (n = 433)	Age, SAPS II, parenteral nutrition	Markov models	Internal validation (Level 4)	Semi-Parametric LCV 4.04
								Non-homogenous LCV 4.06
								Parametric LCV 4.46

(*continued on next page*)

Table A1 (*continued*)

Author(s), year	Source of data with participants	Type of prediction model study Development /Validation	Outcome to be predicted Diagnostic /Prognostic	Predictors	Model development	Model evaluation (AI level of readiness)	Result(s)	
Pearl et al., 2012	National Trauma Data Bank, 2001–2005 (n = 1438,035)	Development	Prognostic	VAP (n=NR)	Intubation, ISS, ICU LOS >2 days	ANN	Internal validation (Level 4)	True 85 %; False 87 %; Gini 0.80
Schurink et al., 2007	Neurosurgical patients, University Medical center Utrecht, 2000–2003 (n = 872)	Development	Diagnostic	VAP (n = 58)	Duration of MV, WBC, body temperature, antipyretic drugs, sputum production and color, PaO_2/FiO_2 ratio, X-ray	BN	Internal validation (Level 4)	AUC 0.857 (95% CI 0.827–0.888), SEN 80%; SPE 80%; PPV 6.1%; NPV 99.6%
				Possible VAP (n = 78) Probable VAP (n = 21)				AUC 0.884 (95 % CI 0.842–0.925) AUC 0.875 (95 % CI 0.804–0.945)

ACC = Accuracy; ANN = Artificial Neural Network; AUC = Area Under Curve; AUROC = Area Under the Receiver Operating Curve; BN = Bayesian Network; CI = Confidence Intervals; EBM = Explained Boosting Machine; ED = Emergency Department; GCS Glasgow Coma Scale; ICU = Intensive care unit; ISS = Injury Severity Score; LCV = Likelihood cross-validation; LOS = length of stay; LR = Logistic regression; MIMIC = Multiparameter Intelligent Monitoring in Intensive Care; MP = Multilayer Perceptron; MV = Mechanical Ventilation; NPV = Negative Predictive Value; NPA VAP = VAP attributed to *P. Aeruginosa*; NR = not reported; PPV = Positive Predictive Value; RF = random forest; SEN = Sensitivity, SPE = Specificity; TBI = traumatic brain injury; VAP = Ventilator-Associated Pneumonia; WBC = White Blood Cell count; XGB = eXtreme Gradient Boosting.

References

[1] Ramírez-Estrada S, Lagunes L, Peña-López Y, Vahedian-Azimi A, Nseir S, Arvaniti K, et al. The EU-VAE study investigators group. assessing predictive accuracy for outcomes of ventilator-associated events in an international cohort: the EUVAE study. Intensive Care Med 2018:1212–20. https://doi.org/10.1007/s00134-018-5269-7.

[2] Jansson M, Ala-Kokko T, Ahvenjärvi L, Karhu J, Ohtonen P, Syrjälä H. What is the applicability of a novel surveillance concept of ventilator-associated events? Infect Control Hosp Epidemiol 2017;38:983–8. https://doi.org/10.1017/ice.2017.106.

[3] Tedja R, Nowacki A, Fraser T, Fatica C, Griffiths L, Gordon S, et al. The impact of multidrug resistance on outcomes in ventilator-associated pneumonia. Am J Infect Control 2014;42:542–5. https://doi.org/10.1016/j.ajic.2013.12.009.

[4] Hebert C, Flaherty J, Smyer J, Ding J, Mangino JE. Development and validation of an automated ventilator-associated event electronic surveillance system: a report of a successful implementation. Am J Infect Control 2018;46:316–21. https://doi.org/10.1016/j.ajic.2017.09.006.

[5] Shenoy ES, Rosenthal ES, Shao YP, Biswal S, Ghanta M, Ryan EE, et al. Real-time, automated detection of ventilator-associated events: avoiding missed detections, misclassifications, and false detections due to human error. Infect Control Hosp Epidemiol 2018;39:826–33. https://doi.org/10.1017/ice.2018.97.

[6] Mann T, Ellsworth J, Huda N, Neelakanta A, Chevalier T, Sims KL, et al. Building and Validating a computerized algorithm for surveillance of ventilator-associated events. Infect Control Hosp Epidemiol 2015;36. https://doi.org/10.1017/ice.2015.127. 999–03.

[7] Nuckchady D, Heckman MG, Diehl NN, Creech T, Carey D, Domnick R, et al. Assessment of an automated surveillance system for detection of initial ventilator-associated events. Am J Infect Control 2015;43:1119–21. https://doi.org/10.1016/j.ajic.2015.05.040.

[8] Klein Klouwenberg PM, van Mourik MS, Ong DS, Horn J, Schultz MJ, Cremer OL, et al. MARS consortium. electronic implementation of a novel surveillance paradigm for ventilator-associated events. feasibility and validation. Am J Respir Crit Care Med 2014;189:947–55. https://doi.org/10.1164/rccm.201307-1376OC.

[9] Resetar E, McMullen KM, Russo AJ, Doherty JA, Gase KA, Woeltje KF. Development, implementation and use of electronic surveillance for ventilator-associated events (VAE) in adults. AMIA Annu Symp Proc 2014:1010–7.

[10] Stevens JP, Silva G, Gillis J, Novack V, Talmor D, Klompas M, et al. Automated surveillance for ventilator-associated events. Chest 2014;146:1612–8. https://doi.org/10.1378/chest.13-2255.

[11] Ramirez-Estrada S, Peña-Lopez Y, Kalwaje Eshwara V, Rello J. Ventilator-associated events versus ventilator-associated respiratory infections-moving into a new paradigm or merging both concepts, instead? Ann Transl Med 2018;6:425. https://doi.org/10.21037/atm.2018.10.54.

[12] Pouly O, Lecailtel S, Six S, Préau S, Wallet F, Nseir S, et al. Accuracy of ventilator-associated events for the diagnosis of ventilator-associated lower respiratory tract infections. Ann Intensive Care 2020;10(6). https://doi.org/10.1186/s13613-020-0624-6.

[13] Fan Y, Gao F, Wu Y, Zhang J, Zhu M, Xiong L. Does ventilator-associated event surveillance detect ventilator-associated pneumonia in intensive care units? A systematic review and meta-analysis. Crit Care 2016;20:338. https://doi.org/10.1186/s13054-016-1506-z.

[14] Rello J, Afonso E, Lisboa T, Ricart M, Balsera B, Rovira A, et al. FADO project investigators. a care bundle approach for prevention of ventilator-associated pneumonia. Clin Microbiol Infect 2013;19:363–9. https://doi.org/10.1111/j.1469-0691.2012.03808.x.

[15] Jansson MM, Ohtonen P, Alalääkkölä T, Heikkinen J, Mäkiniemi M, Lahtinen S, et al. Artificial Intelligence-based technologies in surgical scheduling: content validity assessment of required functionalities. BMC Health Serv Res 2022;22:1513. https://doi.org/10.1186/s12913-022-08780-y.

[16] van de Sande D, van Genderen ME, Huiskens J, Gommers D, van Bommel J. Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. Intensive Care Med 2021;47:750–60. https://doi.org/10.1007/s00134-021-06446-7.

[17] Frondelius T, Atkova I, Miettunen J, Rello J, Jansson MM. Diagnostic and prognostic prediction models in ventilator-associated pneumonia: systematic review and meta-analysis of prediction modelling studies. J Crit Care 2022;67:44–56. https://doi.org/10.1016/j.jcrc.2021.10.001.

[18] Debray TPA, Damen JAAG, Snell KIE, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ 2017;356. https://doi.org/10.1136/bmj.i6460.

[19] McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM. The PRISMA-DTA group. preferred reporting items for systematic reviews and meta-analysis of diagnostic test accuracy studies the prisma-dta statement. JAMA J Am Med Assoc 2018;319:388–96. https://doi.org/10.1001/jama.2017.19163.

[20] Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170:51–8. https://doi.org/10.7326/M18-1376.

[21] Fleuren LM, Thoral P, Shillan D, Ercole A, Elbers PWG. Right data right now collaborators. machine learning in intensive care medicine: ready for take-off? Intensive Care Med 2020;46:1486–8. https://doi.org/10.1007/s00134-020-06045.

[22] Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Reitsma JB, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PlosMed 2014;11. https://doi.org/10.1371/journal.

[23] Higgins JPT, Thompson SG, Deeks J, Altman DG. Measuring inconsistency in meta-analyses. Br Med J 2013;327:557–60.

[24] Abujaber A, Fadlalla A, Gammoh D, Al-Thani H, El-Menyar A. Machine learning model to predict ventilator associated pneumonia in patients with traumatic brain injury: the C.5 decision tree approach. Brain Inj 2021;6:1–8. https://doi.org/10.1080/02699052.2021.1959060.

[25] Amador T, SaturinoA Veloso A, Zivani N. Early identification of ICU patients at risk of complications: regularization based on robustness and stability of explanations. Artif Intell Med 2022;128:102283. https://doi.org/10.1016/j.artmed.2022.102283.

[26] Dos Santos RP, Menezes A, Lukasewicz S, et al. Automated healthcare-associated infection surveillance using an artificial intelligence algorithm. Infect. Prev. Pract. 2021;3:100167. https://doi.org/10.1016/j.infpip.2021.100167.

[27] Calvert J, Faucher M, Casie Chetty C, Shokouhi S, Evans D, Barnes G, Mao Q. Early prediction of ventilator-associated pneumonia in intensive care unit patients using an intelligible machine learning algorithm. Am J Respir Crit Care Med 2022;205:A3451.

[28] Faucher M, Casie Chetty S, Shokouhi S, Barnes G, Rahmani K, Calvert J, et al. Early Prediction of ventilator-associated pneumonia in ICU patients using an

interpretable machine learning algorithm. Preprints 2022:2022060149. https://doi.org/10.20944/preprints202206.0149.v1.
[29] Giang C, Calvert J, Rahmani K, Barnes G, Siefkas A, Green-Saxena A, et al. Predicting ventilator-associated pneumonia with machine learning. Medicine (Baltimore) 2021;100:e26246. https://doi.org/10.1097/MD.0000000000026246.
[30] Liang Y, Zhu C, Tian C, Lin Q, Li Z, Li Z, et al. Early prediction of ventilator-associated pneumonia in critical care patients: a machine learning model. BMC Pulm Med 2022;22:250. https://doi.org/10.1186/s12890-022-02031-w.
[31] Liquet B, Timsit J, Rondeau V. Investigating hospital heterogeneity with a multi-state frailty model: application to nosocomial pneumonia disease in intensive care units. BMC Med Res Methodol 2021;12:79.
[32] Pearl A, Bar-Or D. Decision support in trauma management: predicting potential cases of Ventilator Associated Pneumonia. Stud Health Technol Inform 2012;180:305–9.
[33] Schurink CA, Visscher S, Lucas PJ, van Leeuwen HJ, Buskens E, Hoff RG, et al. A bayesian decision-support system for diagnosing ventilator-associated pneumonia. Int Care Med 2007;33:1379–86.
[34] Ramirez-Estrada S, Peña-Lopez Y, Rello J. The effects of sedatives, neuromuscular blocking agents and opioids on ventilator-associated events. Eur J Anaesthesiol 2020;37:67–9.
[35] Peña-López Y, Ramírez-Estrada S, Serrano-Megías M, Lagunes L, Rello J. E.U.V.A.E. study group. short-acting sedative-analgesic drugs protect against development of ventilator-associated events in children: secondary analysis of the EUVAE Study. Respir Care 2021;66(5):798–805. https://doi.org/10.4187/respcare.08597.
[36] Keane S, Vallecoccia MS, Nseir S, Martin-Loeches I. How can we distinguish ventilator-associated tracheobronchitis from pneumonia? Clin Chest Med 2018;39:785–96.
[37] Rello J, Schrenzel J, Tejo AM. New insights into pneumonia in patients on prolonged mechanical ventilation: need for a new paradigm addressing dysbiosis. J Bras Pneumol 2021;47:e20210198.
[38] Kalimouttou A, Lerner I, Cheurfa C, Jannot AS, Pirracchio R. Machine-learning-derived sepsis bundle of care. Intensive Care Med 2023;49:26–36. https://doi.org/10.1007/s00134-022-06928-2.
[39] Heyard R, Timsit JF, Essaied W, Held L, consortium COMBACTE-MAGNET. Dynamic clinical prediction models for discrete time-to-event data with competing risks-A case study on the OUTCOMEREA database. Biom J 2019;61:514–34.
[40] Peña-López Y, Ramirez-Estrada S, Eshwara VK, Rello J. Limiting ventilator-associated complications in ICU intubated subjects: strategies to prevent ventilator-associated events and improve outcomes. Expert Rev Respir Med 2018;12(12):1037–50. https://doi.org/10.1080/17476348.2018.1549492.
[41] Sergio Ramírez-Estrada, Peña-Lopez Y, Vieceli T, Rello J. Ventilator-associated events: from surveillance to optimizing management. J Int Med 2022. https://doi.org/10.1016/j.jointm.2022.09.004.
[42] Šimundić AM. Measures of diagnostic accuracy: basic definitions. EJIFCC 2009;19:203–11.
[43] Jones CM, Athanasiou T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. Ann Thorac Surg 2005;79:16–20. https://doi.org/10.1016/j.athoracsur.2004.09.040.