Journal of Critical Care 67 (2022) 44–56

Contents lists available at ScienceDirect

Journal of Critical Care

journal homepage: www.journals.elsevier.com/journal-of-critical-care

Diagnostic and prognostic prediction models in ventilator-associated pneumonia: Systematic review and meta-analysis of prediction modelling studies

Tuomas Frondelius [a], Irina Atkova [b], Jouko Miettunen [c,e], Jordi Rello [d,f,g], Miia M. Jansson [a,*]

[a] Research Unit of Medical Imaging, Physics and Technology, University of Oulu, Oulu, Finland
[b] University of Oulu, Oulu, Finland
[c] Center for Life Course Health Research, University of Oulu, Oulu, Finland
[d] CIBER de Enfermedades Respiratorias, CIBERES, Instituto de Salud Carlos III, Barcelona, Spain
[e] Medical Research Center Oulu, Oulu University Hospital, University of Oulu, Oulu, Finland
[f] Clinical Research/Epidemiology In Pneumonia & Sepsis (CRIPS), Vall d'Hebron Institut of Research (VHIR), Barcelona, Spain
[g] Clinical Research, CHU Caremeau, Nimes, France

ARTICLE INFO

Keywords:
Mechanical ventilation
Exhaled human breath
Machine learning
Predictive analytics
Prognostic model
Ventilator-associated pneumonia

ABSTRACT

Purpose: Existing expert systems have not improved the diagnostic accuracy of ventilator-associated pneumonia (VAP). The aim of this systematic literature review was to review and summarize state-of-the-art prediction models detecting or predicting VAP from exhaled breath, patient reports and demographic and clinical characteristics.

Methods: Both diagnostic and prognostic prediction models were searched from a representative list of multidisciplinary databases. An extensive list of validated search terms was added to the search to cover papers failing to mention predictive research in their title or abstract. Two authors independently selected studies, while three authors extracted data using predefined criteria and data extraction forms. The Prediction Model Risk of Bias Assessment Tool was used to assess both the risk of bias and the applicability of the prediction modelling studies. Technology readiness was also assessed.

Results: Out of 2052 identified studies, 20 were included. Fourteen (70%) studies reported the predictive performance of diagnostic models to detect VAP from exhaled human breath with a high degree of sensitivity and a moderate specificity. In addition, the majority of them were validated on a realistic dataset. The rest of the studies reported the predictive performance of diagnostic and prognostic prediction models to detect VAP from unstructured narratives [2 (10%)] as well as baseline demographics and clinical characteristics [4 (20%)]. All studies, however, had either a high or unclear risk of bias without significant improvements in applicability.

Conclusions: The development and deployment of prediction modelling studies are limited in VAP and related outcomes. More computational, translational, and clinical research is needed to bring these tools from the bench to the bedside.

Registration: PROSPERO CRD42020180218, registered on 05-07-2020.

1. Introduction

Lifesaving, mechanical ventilation (MV) also predisposes patients to numerous resource-intensive, morbid, and lethal complications such as ventilator-associated pneumonia (VAP), acute respiratory distress syndrome, sepsis, and atelectasis. Infection-related complications increase the number of ventilator-days by 5, the Intensive Care Unit (ICU) length of stay (LOS) by 11 days, and hospital LOS by 12 days [1,2]. Part of the costs, however, have arisen from the current inability to systematically diagnose the disease at an early stage, potentially leading to delayed treatment or overuse of broad-spectrum antimicrobials.

Current diagnostic approaches (e.g., signs and symptoms, microbiological cultures, and visual output of plain radiographs) and conventional surveillance methods (e.g., manual chart reviews, prevalence surveys, discharge codes, electronic surveillance algorithms) are insensitive [1,3], time-consuming [4,5], and frequently expensive [5]. In

* Corresponding author.
E-mail addresses: juha.frondelius@oulu.fi (T. Frondelius), irina.atkova@oulu.fi (I. Atkova), jouko.miettunen@oulu.fi (J. Miettunen), jrello@crips.es (J. Rello), miia.jansson@oulu.fi (M.M. Jansson).

https://doi.org/10.1016/j.jcrc.2021.10.001

addition, they are unable to identify an individual who would benefit from a certain type of treatment based on the predicted disease course.

Artificial Intelligence (AI) is transforming medical practice and precision medicine in ICUs [6,7]. Emerging technologies (e.g., image-based and biochemical approaches, sequencing technologies) using machine learning (ML) techniques will overcome the constraints of current diagnostic approaches [4,5,8]; these data-intensive system will integrate and analyze various structured and unstructured data from different sources to monitor temporal trends, identify risk factors, and predict morbidity and mortality, and thus, facilitate fast clinical decision making for infectious diseases [9] in a timely manner [4].

Since 2014, researchers have developed automated methods for objective VAP detection [10-16]. These expert systems, however, have not improved the diagnostic accuracy of VAP [17]. In addition, automation without AI techniques does not predict the probability or risk of the future occurrence of complications in individuals at risk. For that reason, we reviewed and summarized state-of-the-art prediction models (e.g., local ecology and/or respiratory surveillance culture based AI or ML algorithms, defined as computational models able to learn from data gathered in the ICU) to generate patient-specific predictions of VAP [18]. Due to the lack of a consensus on a "*gold standard*" definition, all diagnostic criteria were taken into account. We hypothesized that the probability or risk of VAP and related outcomes could be predicted. Our primary objective was to evaluate the predictive performance of existing diagnostic and prognostic models. Our secondary objective was to assess both the risk of bias (ROB) and the applicability of the prediction models, and the level of clinical readiness as part of the maturity definition.

2. Material and methods

2.1. Study design

This systematic review and meta-analysis was conducted in accordance with the guide to systematic reviews and meta-analysis of prediction model performance [19] and the statement of Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies [20]. The study protocol was registered in the international prospective register of systematic reviews (PROSPERO) before the start of the study (CRD42020180218).

2.2. Information sources

A preliminary search for existing reviews on the topic was conducted in the JBI Database of Systematic Reviews and Implementation Reports, Cochrane Database of Systematic Reviews, and the PROSPERO Database of Systematic Review Protocols by two reviewers (IA, MMJ) in April 2020 to assess the volume of relevant studies and to identify existing reviews. An actual search was conducted in several multidisciplinary databases (ACM Digital Library/ ACM Guide to Computing Literature, Astrophysics Data System, arXiV, IEEE Xplore Digital Library, Academic Search Ultimate, Cumulative Index to Nursing and Allied Health Literature [CINAHL], CT.gov, PubMed [Medline], Scopus, Web of Science) with the assistance of an information specialist in June 2020. The controlled (MesH in Medline Ovid and PubMed) and free-text terms were used to build sensitive search strategy (Appendix A). In addition, an extensive list of validated search terms was added to the search to cover papers failing to mention predictive research in their title or abstract [21].

2.3. Study selection

The study selection was carried out independently by two researchers (IA, MMJ) on titles and abstracts and then on the full text. All prediction modelling studies with and without external validation as well as external validation studies with and without model updating were included if they met the predefined inclusion criteria (PICOTS):

- Population: adult patients undergoing MV.
- Index: diagnostic and prognostic prediction models using classical methodologies (e.g., logistic regression or survival models) and deep learning (e.g., random forests, neural networks, and support vector machines).
- Comparator: no comparator.
- Outcome(s): VAP with and without clinical outcomes (e.g., ICU mortality, duration of MV, ICU length of stay).
- Timing: Models to be used prior to VAP and at the moment of diagnosis.
- Setting: ICUs in high-income countries.

We included all original and peer-reviewed development and validation studies written in English. In addition, preprints, scientific reports, and studies included in the previous systematic reviews were included if they reported psychometric properties of the primary prediction modelling studies. We did not apply any restrictions on the date or publication status to the searches. We excluded studies related to animals, non-human samples, and expert systems. 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 (Appendix B).

2.4. Assessment of methodological quality

The Prediction Model Risk of Bias Assessment Tool (PROBAST) was used to assess both the ROB and the applicability of the prediction modelling studies [22]. The PROBAST includes 20 signalling questions across four key domains (e.g., participants, predictors, outcome, analysis), while each domain is judged for a risk of bias (e.g., low, high, or unclear). Three reviewers (TF, IA, MMJ) assessed the quality and disagreements were resolved by consensus. We also assessed the overall certainty of evidence (e.g., inconsistency, imprecision, indirectness, and publication bias) in the prognostic studies using the GRADE framework [23].

2.5. Data extraction

The Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies checklist was used to guide the data extraction [24]. Three authors (TF, IA, MMJ) extracted the data (e.g., source of data, participants, outcomes, predictors, sample size, missing data, model development, model performance, model evaluation, results) from the included studies. AI level of readiness (range 1–9) was assessed by applying the general concept of technology readiness levels [25].

2.6. Statistics

The forest plot includes statistics for accuracy, the area under curve (AUC), sensitivity, specificity, positive predictive value (PPV), the negative predictive value (NPV), and their 95% confidence intervals (CIs). A meta-analysis was conducted to estimate the effect of model performance. In these analyses only the best model from each study was included, resulting in pooled estimates of two or three models in each set of statistics.

3. Results

3.1. Study selection process

We identified 3688 records. After exclusion of duplicates, 2052 records were identified for screening. Following screening, 67 records were considered eligible for full-text evaluation. Based on the full-text

evaluation, 46 records were excluded, resulting in 20 unique studies (18 full and 2 conference abstracts) that met our inclusion criteria for the systematic review. The PRISMA flow diagram of the study selection is presented in Fig. 1.

3.2. General characteristics of included studies

Seventy percent ($n = 14$) of the included studies used a sensor-based electronic (e) nose [26-36], an acoustic wave based e-nose [37], and gas chromatography–mass spectrometry (GC–MS) based [38,39] assays to predict VAP and/or a causative pathogen from exhaled human breath (Appendix C). Two (10%) studies [40,41] used Natural Language Processing (NLP)-based techniques to detect VAP from unstructured narratives (Appendix D). The rest [4 (20%)] of the included studies predicted VAP with [42] and without related outcomes [43-45] from baseline demographics and clinical characteristics (Appendix E).

The definition of VAP was not uniform [18]; eight of included studies used clinical criteria with microbiological confirmation [30,33,35,38,39,41,42,41]. In four studies, clinical criteria were used in conjunction with [30,38] and without a Clinical Pulmonary Infection Score [27,28]. In addition, nine studies used solely microbiological [29,31,32,34,36,37,45] and CT-based diagnoses [26], and Ventilator Associated Event (VAE) surveillance criteria [40].

3.3. Predictive performance of diagnostic prediction models

The sensitivity of the prediction models varied from 56% to 100% for the e-nose based assays [29,30,33-36], from 67% to 92% for the NLP-based techniques [40,41], and from 78% to 90% for the risk models [44,45]. The pooled sensitivity of the e-noses in the prediction of clinically defined VAP (Table 1) and *P. aeruginosa* infections (Table 2) were 0.90 (95% CI 0.85–0.94; $I = 17.2\%$; $p = 0.299$) and 0.91 (95% CI 0.79–1.04; $I = 72.9\%$, $p = 0.055$), respectively (Figs. 2 and 3).

The specificity of the prediction models varied from 56% to 100% for the e-nose based assays [29,30,33-35], from 97% to 100% for the NLP-based techniques [40,41], and from 0.06% to 96% for the risk models [44,45]. The pooled specificity of the e-noses in the prediction of clinically defined VAP was 0.75 (95% CI 0.58–0.91; $I = 83.1\%$; $p = 0.015$).

The accuracy of the prediction models varied from 68% to 100% for the e-nose and GC–MS based assays [26,29,31,32,34-36,39]. The pooled accuracy of the e-noses to predict *P. aeruginosa* infections was 0.93 (95% CI 0.89–0.98; $I = 0.0\%$; $p = 0.889$). The accuracy of NLP-based techniques was 98% [40].

The PPV of the prediction models varied from 64% to 93% for the e-nose based assays [30,34-36], from 90% to 100% for the NLP-based techniques [40,41], and from 6% to 86% for the risk models [44,45]. The pooled PPV of e-noses in the prediction of clinically defined VAP and *P. aeruginosa* infections were 0.89 (95% CI 0.83–0.95; $I = 0.0\%$; $p = 0.415$) and 0.87 (95% CI 0.76–0.99; $I = 54.4\%$; $p = 0.138$), respectively.

The NPV of the prediction models varied from 77% to 92% for the e-nose based assays [30,35] and from 66% to 100% for the risk models [45]. The pooled NPV of the e-noses in the prediction of clinically defined VAP was 0.84 (95% CI 0.77–0.91; $I = 0.0\%$; $p = 0.780$). The NPV for NLP was 98% [40,41].

The AUC of the prediction models varied from 0.62 to 0.98 for the e-nose based assays [30,33,35,36] and from 0.51 to 0.86 for the risk models [44,45]. The pooled AUC of the e-noses in the prediction of clinically defined VAP and *P. aeruginosa* infections were 0.83 (95% CI 0.77–0.89; I = 0.0%; $p = 0.625$) and 0.96 (95% CI 0.91–1.02; $I = 30.3\%$; $p = 0.231$), respectively.

3.4. Predictive performance of prognostic prediction models

Only one study predicted the risk of VAP and death in patients with and without VAP demonstrating moderate accuracy [42]. Further analysis was not possible due to lack of studies.

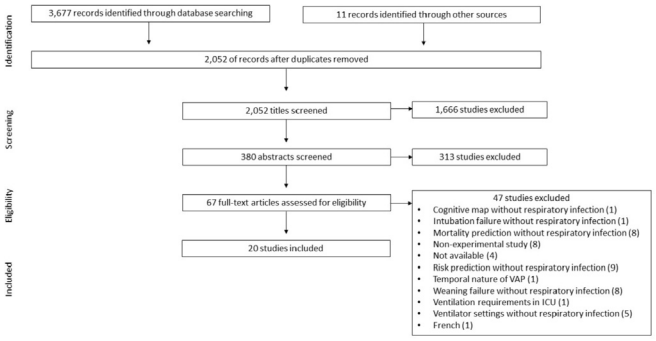

Fig. 1. Flow diagram showing the literature search and results. The flow of information through the different phases of our systematic review was recorded according to the PRISMA reporting guidelines. We identified 1496 records by searching the Scopus, Web of Science and Academic Search Ultimate databases, 688 records by searching the PubMed database, and 340 records by searching the CINAHL database. We identified a further 1059 records by searching the ACM Guide to Computing Literature, arXiV, Astrophysics Data System, CT.gov, and IEEE Xplore Digital Library databases. This literature search resulted in a total of 2052 records for our systematic review. In addition, manual searches of the reference lists, citations, and related articles of included studies revealed eleven references.

3.5. AI level of readiness

Eighty percent ($n = 16$) of the included studies were validated on realistic dataset (level 5) other than the original training and testing population (Appendix C–E). Studies reporting real-time testing (level 6), workflow integration (level 7), clinical outcome evaluation (level 8), and model integration (level 9) were not identified.

3.6. Methodological quality of included studies

The majority of the included studies were prospective cohort [26-28,30,33,39,44,45] or non-nested case-control studies [29,31,34-36,38]; while only one clinical trial [32] and three register studies [40,42,43] were identified. All studies had either a high [26-30,33,35,38,39,42,44,45] or unclear [31,32,34,36,37,40,41,43] risk of bias without significant improvements in applicability (Table 3). High ROB most often originated in the domain "participants" and "analysis". In addition, it was unclear which variables were eventually used by the AI model.

3.7. Overall certainty of evidence

In general, the body of evidence is low due to the lack of randomized controlled trials. The overall certainty of evidence was difficult to estimate due to the low number of prognostic studies and heterogenous results [42,43].

4. Discussion

Our study reports the development of diagnostic and prognostic prediction models in this at-risk population of VAP. The majority of the included studies reported the predictive performance of diagnostic models to detect VAP from various sources. However, only one study predicted a clinical prognosis. Although the majority of them were validated on a realistic dataset, there was a high or unclear risk of bias without significant improvements in applicability.

The studies most often reported the predictive performance of diagnostic models to detect VAP and/or its causative pathogen from exhaled human breath with a high degree of sensitivity and a moderate specificity. Overall, *Pseudomonas aeruginosa* was found to be the most studied pathogen, the management of which requires prompt and adequate antimicrobial exposure [46] due to increased mortality [47] and duration of mechanical ventilation. According to a recent report from EU/EEA

Table 1
Meta-analysis results of the e-nose predictions of clinically defined ventilator-associated pneumonia from exhaled human breath.

Outcome	Model	ES (95% CI)	Weight (%)
Sensitivity			
Bos et al., 2014 [30]	In-set analysis (SPLS)	0.94 (0.87–1.01)	36.91
Schnabel et al., 2015 [33]	In-set analysis (RF with PCA)	0.88 (0.81–0.95)	37.53
Chen et al., 2020 [35]	In-set analysis (Decision tree)	0.86 (0.77–0.95)	25.56
Subtotal (I-squared = 17.2%, $p = 0.299$)		0.90 (0.85–0.94)	100.0
Specificity			
Schnabel et al., 2015 [33]	In-set analysis (FR with PCA)	0.66 (0.56–0.76)	49.42
Chen et al., 2020 [35]	In-set analysis (Decision tree)	0.83 (0.74–0.92)	50.58
Subtotal (I-squared = 83.1%, $p = 0.015$)		0.75 (0.58–0.91)	100.0
Positive predictive value			
Bos et al., 2014 [30]	In-set analysis (SPLS)	0.91 (0.83–0.99)	51.62
Chen et al., 2020 [35]	In-set analysis (Decision tree)	0.86 (0.77–0.95)	48.38
Subtotal (I-squared = 0.0%, $p = 0.415$)		0.89 (0.83–0.95)	100.0
Negative predictive value			
Bos et al., 2014 [30]	In-set analysis (SPLS)	0.85 (0.75–0.95)	44.54
Chen et al., 2020 [35]	In-set analysis (Decision tree)	0.83 (0.74–0.92)	55.46
Subtotal (I-squared = 0.0%, $p = 0.780$)		0.84 (0.77–0.91)	100.0
Area under curve			
Schnabel et al., 2015 [33]	In-set analysis (RF with PCA)	0.82 (0.74–0.90)	54.51
Chen et al., 2020 [35]	In-set analysis (Decision tree)	0.85 (0.76–0.94)	45.49
Subtotal (I-squared = 0.0%, $p = 0.625$)		0.83 (0.77–0.89)	100.0

CI: Confidence Intervals; ES = Effect Size; I-squared = Heterogeneity; PCA = Principal Component Analysis; RF = Random Forest; SPLS = Sparse Partial least square.
NOTE: Weights are from a random effects analysis.

countries, 32.1% of *P. aeruginosa* isolates have been resistant to at least one of the antimicrobial groups whereas 19.2% have been resistant to two or more antimicrobial groups [48] highlighting the potential of surveillance culture based algorithms to restrict the use of broad-spectrum antimicrobials [18].

The analysis of volatile organic compounds (VOCs) in exhaled human breath appears to be a promising means for noninvasive detection and monitoring of infectious diseases [49]. In this study, the pooled sensitivity of e-noses to predict clinically defined VAP and/or its causative pathogen demonstrated a high level of sensitivity. The pooled specificity of the e-noses, however, demonstrated only moderate specificity. Some of these technologies (e.g., GC–MS) require bulky instruments, complex sampling methods, and qualified personnel, which have limited their application in on-site testing [38,39]. For this reason, commercial e-noses (e.g., Cyranose® 320, DiagNose, BreathSpec®, ChemPro100i) with analytical platforms have been applied for on-site testing [50], but they have suffered from laborious pre-treatment, relatively low sensitivity, and incapability to detect unknown targets. In addition, elaborated techniques have been limited in molecular selectivity, and have suffered from high moisture sensitivity and power demands [50].

NLP-based techniques have been found to be a more timely, cost-effective, and accurate alternative to manual surveillance [41]. In this study, NLP-based techniques were used to extract unstructured data from electronic health records in order to classify patients with and without clinically defined VAP [41]. In addition, these techniques were used to detect sepsis and ventilator-associated events from the MIMIC II dataset [40]. Both included studies demonstrated moderate to high

Table 2
Meta-analysis results of e-nose predictions of *P. Aeruginosa* infections from exhaled human breath.

Outcome	Model	ES (95% CI)	Weight (%)
Accuracy			
Chiu et al., 2014 [31]	In-set analysis (RAW)	0.93 (0.87–0.99)	50.59
Liao et al., 2019 [34]	In-set analysis (ENN)	0.95 (0.86–1.04)	23.81
Liao et al., 2020 [36]	In-set analysis (SVM)	0.92 (0.84–1.0)	25.61
Subtotal (I-squared = 0.0%, $p = 0.889$)		0.93 (0.89–0.98)	100.0
Sensitivity			
Liao et al., 2019 [34]	In-set analysis (ENN)	0.97 (0.90–1.04)	56.35
Liao et al., 2020 [36]	In-set analysis (SVM)	0.84 (0.73–0.95)	43.65
Subtotal (I-squared = 72.9%, $p = 0.055$)		0.91 (0.79–1.04)	100.0
Positive predictive value			
Liao et al., 2019 [34]	In-set analysis (ENN)	0.93 (0.83–1.03)	53.94
Liao et al., 2020 [36]	In-set analysis (SVM)	0.81 (0.69–0.93)	46.06
Subtotal (I-squared = 54.4%, $p = 0.138$)		0.87 (0.76–0.99)	100.0
Area under curve			
Liao et al., 2019 [34]	In-set analysis (ENN)	0.98 (0.93–1.03)	65.72
Liao et al., 2020 [36]	In-set analysis (SVM)	0.93 (0.84–1.01)	34.28
Subtotal (I-squared = 30.3%, $p = 0.231$)		0.96 (0.91–1.02)	100.0

CI: Confidence Intervals; ENN = Ensemble Neural Network; ES = Effect Size; I-squared = Heterogeneity; SVM = Support Vector Machine.
NOTE: Weights are from a random effects analysis.

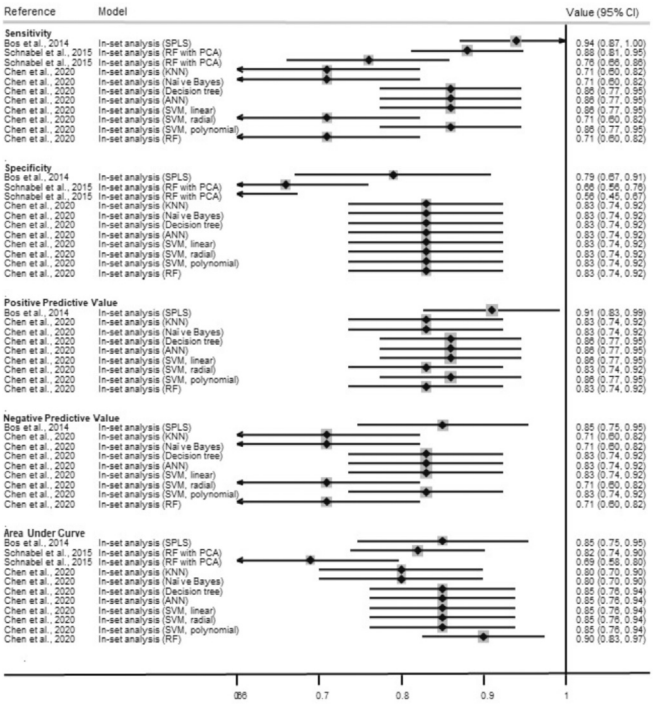

Fig. 2. Forest plot of sensitivity, specificity, positive and negative predictive values, and area under the ROC curve of e-nose analyses to predict clinically defined Ventilator-Associated Pneumonia from exhaled human breath.

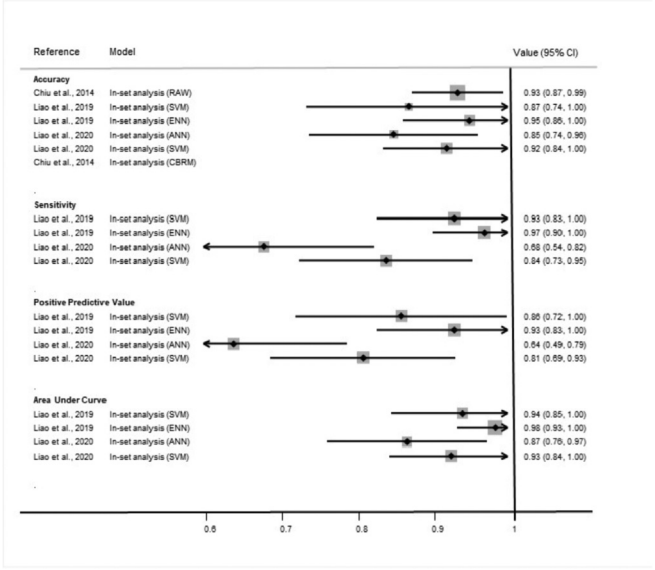

Fig. 3. Forest plot of accuracy, sensitivity, positive predictive value, and area under the ROC curve of e-nose analyses to predict *P. aeruginosa* infection from exhaled human breath.

sensitivity (67–92%) and specificity (97–100%) but failed in the external validation. The relatively low sensitivity in the study by Daza et al. [40] might be due to the low incidence of VAP patients in their dataset and relatively small number of patients.

A moderate predictive performance was observed within the existing diagnostic and prognostic prediction models, possibly due to the limited amount of data used to train the models. The development of breath analysis tools and models requires manual data collection from patients, which is time consuming and expensive, which severely limits the availability of training data for the models. The limited amount of data may lead to biases in the models and the patients may also suffer from comorbidities, which can also lead to errors and reduce the model performance. Datasets containing electronic healthcare records and physiological signals are more readily available, but they may lack appropriate annotations. Because of this, the data needs to be annotated manually which is time consuming and limits the amount of usable data for model development. In the literature, the use of multimodal data has improved the predictive performance. Utilizing multimodal data to train automatic diagnostic and prognostic ML models allows the models to learn more contextually relevant features and significantly improves the model performance.

According to our findings, the development and deployment of AI is limited in critical care settings although more than half of healthcare leaders expect that the widespread adoption of AI will take less than five years [51]. The limited clinical relevance, lack of suitable data and infrastructures, costs, as well as legal and ethical considerations have hampered the clinical integration and implementation of AI-solutions [52,53]. While AI may enable the development of accurate tools, their introduction must follow careful consideration of real-world clinical utility, efficiency, and existing workflows [54]. Future studies using AI should focus on the benefits of preventive interventions of infectious and non-infectious complications [55]. It is still unclear, however, what kind of evidence will be needed to recommend the widespread adoption of new AI systems [56].

According to our best knowledge, this is the first review reporting the development of diagnostic and prognostic prediction models in patients at risk of VAP. This review comes at an important time in the pandemic, highlighting the risk-based prevention of complications occurring in individuals at risk. This systematic review has several limitations. First, explorative prediction research is difficult to find in Medline, using any of the currently available search filters [21]. The lack of studies in our systematic review may also be related to the search strategy and screening criteria that focused solely on VAP; we did not systematically examine other complications occurring during MV which could potentially have provided additional information on risk factors. Second, we did not extract the predictive performance of other clinical outcomes in patients with VAP due to the lack of studies [42]. Third, the definition of respiratory infections was not uniform in the included studies. Moreover, quantitative cultures and/or CT-based confirmation alone are insufficient to distinguish patients with and without VAP. In addition, the inter-observer agreement in calculating the Clinical Pulmonary Infection Score (CPIS) has been shown to be poor [56]. Fourth, e-noses show promising results in VAP diagnosis, but so far, these devices have been trained using small breath sample datasets collected from limited number of patients. For this reason, the ML algorithms used to analyze the sensor array data may not be generalized to other datasets of patients. In future, the predictive performance of these devices should be validated with multiple datasets before clinical applications. Fifth, there is a lack of consistency between different statistical measures. In addition, the lack of studies reduces the reliability of pooling. Lastly, prognostic models are rarely developed although the knowledge of prognosis is critical in the planning of diagnostic interventions and predicting the likely effect of treatment. In addition, a clear definition of risk factors is highlighted.

An ideal diagnostic test should have a high level of accuracy, and be low cost, noninvasive, easily repeatable at specific intervals, non-technical, and clinically available at the bedside. Advanced methods for early and accurate diagnoses are urgently needed to distinguish patients with VAP and ventilator-associated tracheobronchitis from airway colonization. Infection-related VOC profiles require further research, however; there is a lack of clinically approved VOC biomarkers.

Table 3
The Prediction Model Risk of Bias Assessment Tool was used to assess both the risk of bias and the applicability of prediction modelling studies (Wolff et al., 2019).

Study	ROB				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applicability
Bos et al., 2014 [30]	+	+	+	−	−	−	+	−	−
Chen et al., 2020 [35]		+	+	−	?	+	+	−	?
Chiu et al., 2014 [31]	−	+	?	−	?	−	?	?	?
Gao et al., 2016 [38]	−	+	+	−	+	+	+	−	+
Hanson et al., 2005 [27]	+	+	−	−	−	+	+	−	−
Hockstein et al., 2004 [26]	+	+	−	−	−	+	−	−	−
Hockstein et al., 2005 [28]	+	+	−	−	−	+	+	−	−
Humphreys et al., 2011 [29]	−	+	–	−	−	+	+	−	−
Liao et al., 2019 [34]	−	+	?	−	−	+	−	?	−
Liao et al., 2020 [36]	−	+	?	−	−	+	?	?	?
Schnabel et al., 2015 [33]	+	+	+	−	+	+	+	−	+
Shih et al., 2010 [37]	?	?	?	−	+	?	?	?	?
Tang et al., 2014 [32]	−	+	?	?	−	?	?	?	?
van Oort et al. 2017 [39]	+	+	+	−	+	+	+	−	+
Daza et al., 2016 [40]	−	?	−	−	−	?	+	?	?
Ding et al., 2019 [33]	?	?	+	?	?	?	+	?	?
Liquet et al., 2012 [42]	−	+	+	+	+	+	+	−	+
Pearl et al., 2012 [43]	−	+	?	+	−	+	?	?	?
Schurink et al., 2007 [44]	+	+	+	−	+	+	+	−	+
Visscher et al., 2008 [45]	+	+	+	−	+	+	+	−	+

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.

5. Conclusion

The development and deployment of prediction modelling studies are limited concerning VAP and related outcomes. Our findings suggest that computational, translational, and clinical research to bring these tools from the bench to the bedside is an unmet clinical need. In addition, advanced methods for early and accurate diagnosis are urgently needed to distinguish ventilated patients with and without lower respiratory infections. Our findings may inform the development of diagnostic and prognostic prediction models in this at-risk population of lower respiratory infections in the future and suggest the need to improve the specificity of exhaled breath tests for intubated patients.

Author contributions

MMJ conceived the study. IA and MMJ developed the search strategy and performed the literature search. IA and MMJ did the study selection. TF, IA, and MMJ assessed the quality of the included studies. In addition, TF, IA, and MMJ carried out the data extraction for the systematic review. JM did the statistical analyses for the meta-analysis. MMJ wrote the first draft of the manuscript. All authors contributed to the interpretation of the data and critical revision of the manuscript and approved the final manuscript. All authors confirm the accuracy and integrity of the work.

Compliance with ethical standards

Not applicable.

Declaration of competing interest

None.

Acknowledgements

We would like to acknowledge Informational Specialists Sirpa Grekula and Pertti Martinmäki, respectively. 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

Appendix A

An extensive list of validated search terms was used to the search to cover papers failing to mention predictive research in their title or abstract [20] (Geersing et al., 2012).

Databases	PI(*COT*)S	MesH in Medline Ovid and PubMed	No. of references
Scopus, Web of Science and Academic Search Ultimate (17.6.2020)	Population	((TITLE-ABS-KEY ("respiration, artificial" OR "artificial respiration" OR "ventilator-associated" OR "ventilator-induced" OR "mechanical ventilation"))	1496
	Intervention	(TITLE-ABS-KEY (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 ((TITLE-ABS-KEY ("Artificial Intelligence" OR "Machine Learning" OR "Artificial Learning" OR "Bayesian Learning" OR "Deep Learning" OR "Knowledge Representation" OR "Neural Network*") OR TITLE-ABS-KEY ("Probabilistic Network*" OR "Statistical Learning" OR "Support vector machine*" OR "Generalized linear model*") OR TITLE-ABS-KEY ("Naive bayes*" OR "Ensemble method*" OR "Neural network model*" OR "Decision tree*") OR TITLE-ABS-KEY ("Proportional hazards model*" OR "Long short term memory" OR "Natural language processing" OR "Speech recognition" OR robotics OR sensor* OR gamification OR "Automated planning"))))	
		Separate search: (TITLE-ABS-KEY ("multivariable prediction model*" OR "multi-variable prediction model*"))	
	Setting	hospital* OR "operation theatre*" OR "emergency department*" OR "recovery room*" OR"tertiary care center*" OR"intensive care unit*" OR"operation room*" OR"Clinical Decision Unit*" OR"Clinical Observation	

(continued on next page)

(*continued*)

Databases	PI(*COT*)S	MesH in Medline Ovid and PubMed	No. of references
		Unit*" OR"respiratory care unit*" OR"Trauma Center*"	
PubMed (Medline) (17.6.2020)	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]))	688
	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] OR Proportional hazards model*[Text Word] OR "Long short term memory"[Text Word] OR "Natural language processing"[Text Word] OR "Speech recognition"[Text Word] OR Robotics[Text Word] OR Sensor*[Text Word] OR Gamification[Text Word] OR "Automated planning"[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]))))	
		Separate search: multivariable prediction model* OR multi-variable prediction model*	
	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 theatre*[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 (17.6.2020)	Population	((MH "Respiration, Artificial+") OR (MH "Pneumonia, Ventilator-Associated")) OR ("respiration, artificial" OR"artificial respiration" OR "ventilator-associated" OR "ventilator-induced" OR "mechanical ventilation")	340
	Intervention	(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*" OR"Proportional hazards model*" OR"Long short term memory" OR "Natural language processing" OR "Speech recognition" OR Robotics OR Sensor* OR Gamification OR "Automated planning") AND (MH "ROC Curve") 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*)	
		Separate search: multivariable prediction model* OR multi-variable prediction model*	
	Setting	((TITLE-ABS-KEY (hospital* OR "operation theatre*" 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 TITLE-ABS-KEY ("Trauma Center*")))	
ACM Guide to Computing Literature (26.6.2020)	Population	[[All: "respiration, artificial"] OR [All: "artificial respiration"] OR [All: "ventilator-associated"] OR [All: "ventilator-induced"] OR [All: "mechanical ventilation"]]	193
	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 bayes*"] OR [All: "ensemble method*"] OR [All: "neural network model*"] OR [All: "decision tree*"] OR [All: "proportional hazards model*"] OR [All: "long short term memory"] OR [All: "natural language processing"] OR [All: "speech recognition"] OR [All: robotics] OR [All: sensor*] OR [All: gamification] OR [All: "automated planning"]] [[All: stratification] OR [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*]]	
		Separate search: multivariable prediction model* OR multi-variable prediction model*	
	Setting	[[All: hospital*] OR [All: "operation theatre*"] 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 (20.8.2020)	Population	"respiration, artificial" OR "artificial respiration" OR "ventilator-associated" OR "ventilator-induced" OR "mechanical ventilation"	43
	Intervention	"multivariable prediction model*" OR "multi-variable prediction model*"	
Astrophysics Data System (20.8.2020)	Population	("respiration, artificial" OR "artificial respiration" OR "ventilator-associated" OR "ventilator-induced" OR "mechanical ventilation")	190
	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*" OR "Proportional hazards model*" OR "Long short term memory" OR "Natural language processing" OR "Speech recognition" OR Robotics OR Sensor* OR Gamification OR "Automated planning") 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 theatre*" 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*")	
CT.gov (20.8.2020)	Population	("respiration, artificial" OR "artificial respiration" OR "ventilator-associated" OR "ventilator-induced" OR "mechanical ventilation")	277
	Setting	"intensive care unit"	

(*continued*)

Databases	PI(*COT*)S	MesH in Medline Ovid and PubMed	No. of references
IEEE Xplore Digital Library (20.8.2020)	Population	"respiration, artificial" OR "artificial respiration" OR "ventilator-associated" OR "ventilator-induced" OR "mechanical ventilation" AND multivariable prediction model* OR multi-variable prediction model*	356
	Intervention	"Artificial Intelligence" OR "Machine Learning" OR "Artificial Learning" OR "Bayesian Learning" OR "Deep Learning" OR "Neural Network" 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*	
	Setting	"intensive care unit"	

Appendix B

The list of sources of "grey literature".

- The System for Information on Grey Literature in Europe (www.opengrey.eu)
- The DART-Europe E-theses Portal (www.dart-europe.eu)
- The EThOS e-theses online service (https://ethos.bl.uk)
- WorldCat (https://www.worldcat.org/)
- MedNar (http://allcatsrgrey.org.uk/wp/knowledgebase/mednar/)
- The Netherlands Trial Register (https://www.trialregister.nl/)
- The International Clinical Trials Registry Platform (https://www.who.int/ictrp/search/searchtips/en/)
- The GSK Study Register (gsk-clinicalstudyregister.com)
- Regulatory agencies: FDA/EMEA

Appendix C

Data extraction of diagnostic prediction models to predict VAP and its causative pathogen from exhaled human breath.

Author(s)	Source of data	Participants	Outcome(s) to be predicted	Candidate predictors (or index tests)	Sample size	Missing data	Model development	Model performance	Model evaluation	Result(s)	Level of readiness
Bos et al., 2014	Cohort study	More than 7 days ventilated patients	Clinical criteria of VAP with microbiological confirmation, CPIS for VAP	The sensor array response data	28 cases and 17 controls	Patients with missing data were excluded	SPLS	ROC	Internal validation (cross--validation)	eNose: AUC for non-colonized controls 0.84 (0.68–1.0); SEN 94%; SPE 79%; PPV 84%; NPV 92%. eNose: AUC for colonized-controls 0.85; SEN 92%; SPE 79%; PPV 81%; NPV 92% CPIS: AUC 0.89 (0.80–0.99); SEN 87%; SPE 86%; PPV 72%; NPV 92% CPIS + eNose: AUC 0.95 (95% CI 0.88–1.0); SEN 94%; SPE 86%; PPV 81%, NPV 92%	5
Chen et al., 2020	Case-control study	Mechanically ventilated patients, National Taiwan University Hospital 02/2017–06/2019	Clinical criteria of VAP with microbiological confirmation (*P. aeruginosa*)	The sensor array response data	33 cases and 26 controls	NR	KNN Naive Bayes, Decision Tree, ANN, SVM, RF	Confusion matrix, ROC	Internal validation (bootstrapping)	Mean ACC 81%; SEN 79%; SPE 83%; PPV 85%; NPV 77%; Kappa 0.85; AUC 0.62 (0.08)	5
Chiu et al., 2014	Case-control study	Mechanically ventilated patients, Taipei Medical University.	Pathogens causing VAP (*P. aeruginosa. Klebsiella pneumoniae*, *S. aureus*)	The sensor array response data	76 cases and 41 controls	NR	CRBM	Classification accuracy		Normal vs infected: RAW ACC 87%; CFI 27.80, CRBM ACC 92%; CFI 15.30 *P. aeruginosa*: RAW ACC 93%; CFI 8.53, CRBM ACC 100%; CFI 4.81 *K. pneumoniae*: RAW ACC 95%; CFI 6.71, CRBM ACC	5

(*continued on next page*)

(*continued*)

Author(s)	Source of data	Participants	Outcome(s) to be predicted	Candidate predictors (or index tests)	Sample size	Missing data	Model development	Model performance	Model evaluation	Result(s)	Level of readiness
										100%; CFI 2.89 *S. aureus*: RAW ACC 92%; CFI 9.92, CRBM ACC 100%; CFI 5.43 *Candida*: RAW ACC 96%; CFI 5.09, CRBM ACC 100%; CFI 3.07	
Gao et al., 2016	Case-control study	Mechanically ventilated patients, the Sir Run Shaw Hospital China, 2014–2016	Clinical criteria of VAP with microbiological confirmation (*Acinetobacter baumannii)*, CPIS for VAP	GC–MS data	40 cases (infection and colonization groups) and 20 controls	NR	PCA, PLS discriminant analysis	ROC	Internal validation	0.89 and 0.88	5
Hanson et al., 2005	Cohort study	Mechanically ventilated patients with surgery, Hospital of the University of Pennsylvania	CPIS for VAP	The sensor array response data	19 cases and 19 controls	NR	Linear and nonlinear PLS regression	Classification measures	Internal validation (cross--validation)	r^2 0.81 ($p = 0.0001$); mean bias 0.0 (limits ± 2.6)	5
Hockstein et al., 2004	Cohort study	Mechanically ventilated patients with surgery, University of Pennsylvania Hospital, 08–10/2003	CT confirmed VAP	The sensor array response data	13 cases and 12 controls	NR	SVM	PCA	Internal validation (cross-validation) External validation (cross-validation)	ACC 91.6–100.0% ACC >80%	5
Hockstein et al., 2005	Cohort study	Mechanically ventilated patients with surgery, University of Pennsylvania Hospital, 08–12/2003	CPIS for VAP	The sensor array response data	15 cases and 29 controls	Records with missing data were excluded	KNN	Classification accuracy	Internal validation (cross-validation)	ACC 70%	5
Humphreys et al., 2011	Case-control study	Mechanically ventilated patients, Cheltenham General and Gloucestershire Royal Hospitals	Bronchoalveolar lavage	The sensor array response data	44 cases and 6 controls	NR	PCA, LDA	Classification accuracy	Internal validation (LOO	cross-validation)	Group 1: ACC 83%; SEN 74–95%; SPE 77–100% Group 2: ACC 68%; SEN 67–69%; SPE 67–69% Group 3: ACC 77%; SEN 56–84%; SPE 81–97% Group 4: ACC 76%; SEN 60–100%; SPE 81–100%
	5										
Liao et al., 2019	Case-control study	Mechanically ventilated patients with neurosurgery, Taipei Medical University Hospital	Pathogen causing VAP (*P. aeruginosa*)	The sensor array response data	12 cases and 12 controls	NR	SVM ENN	Classification measures	Internal validation (cross-validation)	ACC 0.87; SEN 93%; PPV 86%; AUC 0.94 (0.0301) ACC 0.947 (0.0135); SEN 97%; PPV 93%; AUC 0.98 (0.0058)	5

(*continued*)

Author(s)	Source of data	Participants	Outcome(s) to be predicted	Candidate predictors (or index tests)	Sample size	Missing data	Model development	Model performance	Model evaluation	Result(s)	Level of readiness
Liao et al., 2020	Case-control study	Mechanically ventilated patients with cardiopulmonary surgery, Taipei Medical University Hospital	Pathogen causing VAP (*P. aeruginosa*)	The sensor array response data	20 cases and 20 controls	NR	ANN SVM	ROC	Internal validation (cross-validation)	ACC 0.86 (0.02); SEN 68%; PPV 64%; AUC 0.87 (0.05) ACC 92%; SEN 84%; PPV 82%; AUC 0.93 (0.017)	5
Schnabel et al., 2015	Cohort study	Mechanically ventilated patients, Maastricht University Medical Centre, 2009–2011	Clinical criteria of VAP with microbiological confirmation	The sensor array response data	72 cases and 53 controls	NR	RF with PCA	ROC	Internal validation (out-of-bag-error)	Group 1: AUC 0.82 (0.73–0.91); SEN 88%; SPE 66% Group 2: AUC 0.69 (0.57–0.81); SEN 76%; SPE 56%	5
Shih et al., 2010	Unclear	ICU pneumonia patients	Most common pathogens causing respiratory infection (*P. aeruginosa, A. baumannii, K. pneumoniae, S. aureus, Acinetobacter lwoffi*)	Acoustic wave based electronic nose system	96 patients, 128 breath samples; 106 single bacteria; 11 multiple bacterial infections; 11 normal samples	NR	MDA	χ^2	Internal validation	ACC: 98%; χ^2: 166 (p-value 0.0034), χ^2: 83.4 (p-value 0.7269)	5
Tang et al., 2014	Clinical trial	NR	Pathogens causing VAP (*K. pneumoniae*, *P. aeruginosa*)	The sensor array response data	74 cases and 43 controls	NR	KNN	Classification accuracy, CFI	Internal validation	ACC 100%; CFI 0.73 for infected patients	5
van Oort et al. 2017	Cohort study		Clinical criteria of VAP with microbiological confirmation	VOCs identified from GC–MS data	25 cases (infection and colonization groups) and 68 controls	NR	PCA, PLS discriminant analysis	ROC	Internal validation (LOO	cross-validation)	In-set analysis: ACC 87% LOO: ACC 73%

5

ACC = Accuracy; ANN = Artificial Neural Network; AUC = Area Under Curve; BN = Bayesian Network; CFI = Clustering Fisher Index; CRBM = Continuous Restricted Boltzmann Machine; CI = Confidence Intervals; CPIS = Clinical Pulmonary Infection Score; CT = computed tomography; ENN = Ensemble Neural Network; GC–MS = gas chromatography–mass spectrometry; KNN = K-Nearest Neighbors; LOO = Leave one out cross-validation; MDA = Multiple discriminant analysis; NLP = Natural Language Processing; NPV = Negative Predictive Value; NR = Not Reported; LDA = Linear discriminant analysis; PCA = Principal Component Analysis; PLS = Partial Least Squares; PPV = Positive Predictive Value; ROC = Receiver Operating Characteristics; SEN = Sensitivity, SPE = Specificity; SPLS = Sparse partial least square; SVM = Support Vector Machine; RF = Random Forest; VAP = Ventilator-Associated Pneumonia, VOC = Volatile Organic Compound; χ^2 = Chi-squared test.

Appendix D

Data extraction of the diagnostic prediction models to predict VAP from unstructured narratives.

Author (s)	Source of data	Participants	Outcome(s) to be predicted	Candidate predictors (or index tests)	Sample size	Missing data	Model development	Model performance	Model evaluation	Result(s)	Level of readiness
Daza et al., 2016	Register study	MIMIC II database	Sepsis, VAE surveillance criteria	NR	60	NR	A framework for knowledge-based temporal abstraction	Classification accuracy, F-measure	Internal validation	ACC 98%; SEN 67%; SPE 100%; PPV 100%; NPV 98%; F-measure 0.8; Kappa 0.93 One VAP patient missed by ClinicalTime.	4
Ding et al., 2019	Unclear	NR	Clinical criteria of VAP with microbiological confirmation	NR	223	NR	NR	Classification accuracy	Internal validation	SEN 92%; SPE 97%; PPV 90%; NPV 98%	4

ACC = Accuracy; ANN = Artificial Neural Network; AUC = Area Under Curve; NPV = Negative Predictive Value; CI = Confidence Intervals; PPV = Positive Predictive Value; SEN = Sensitivity, SPE = Specificity; VAE = Ventilator-Associated Event; VAP = Ventilator-Associated Pneumonia.

Appendix E

Data extraction of the diagnostic and prognostic prediction models to predict VAP and/or a causative pathogen from the baseline demographics and clinical characteristics.

Author (s)	Source of data	Participants	Outcome(s) to be predicted	Candidate predictors (or index tests)	Sample size	Missing data	Model development	Model performance	Model evaluation	Result(s)	Level of readiness
Liquet et al., 2012	Register study	Outcomera database, 1996–2007	Clinical criteria of VAP with microbiological confirmation, discharge, and death in the ICU	13 variables	2871	NR	Markov models	LCV	Internal validation (cross-validation)	Semi-Parametric LCV 4.04; Non-homogenous LCV 4.06; Parametric LCV 4.46	4
Pearl et al., 2012	Register study	National Trauma Data Bank, 2001–2005	VAP (undefined criteria)	9 variables	1,438,035	Records with missing data were excluded	ANN	Gini values	Internal validation (cross-validation)	True 85%; false 87%; gini 0.80	4
Schurink et al., 2007	Cohort study	Patients with neurosurgery, University Medical center Utrecht 2000–2003	Clinical criteria of VAP with microbiological confirmation	7 variables	872	NR	BN	Classification measures	Internal validation	Approach 1: VAP: AUC 0.86 (0.83–0.89), optimal cut-off point 46% with a SEN and SPE of 80% (PPV 6%, NPV 100%) Possible VAP: AUC 0.89 (0.84–0.93), optimal cut-off point 53% Probable VAP: AUC 0.88 (0.80–0.95), optimal cut-off point 53% Approach B: VAP: AUC 0.85 (0.79–0.90), optimal cut-off point 78% with a SEN and SPE of 79% (PPV 86%, NPV 66%) Possible VAP: AUC 0.85 (0.79–0.92), optimal cut-off point 78% Probable VAP: AUC 0.88 (0.80–0.95), optimal cut-off point 53%	5
Visscher et al., 2008	Cohort study	Patients with neurosurgery, University Medical center Utrecht 2000–2003	Pathogens causing VAP	NR	157	NR	BN	AUC	Internal validation	Analysis 1: AUC 0.51 (0.39–0.63) to 0.77 (0.64–0.91); SEN 85–90%; SPE 0.06–66%; PPV 13–33%; NPV 70–97% Analysis 2: AUC 0.83 (0.68–0.98) to 0.92 (0.85–0.982); SEN 78–86%; SPE 78–84%; PPV 38–61%; NPV 90–97% Analysis 3: AUC 0.6 (0.74–0.98) to 0.93 (0.88–0.98); SEN 78–86%; SPE 77–96%; PPV 22–64%; NPV 86–97%	5

ACC = Accuracy; ANN = Artificial Neural Network; AUC = Area Under Curve; BN = Bayesian Network; NPV = Negative Predictive Value; CI = Confidence Intervals; ICU = Intensive care unit; LCV = Likelihood cross-validation; PPV = Positive Predictive Value; SEN = Sensitivity, SPE = Specificity; VAP = Ventilator-Associated Pneumonia.

References

[1] 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.

[2] Ramírez-Estrada S, Lagunes L, Peña-López Y, Vahedian-Azimi A, Nseir S, Arvaniti K, et al. Assessing predictive accuracy for outcomes of ventilator-associated events in an international cohort: the EUVAE study. Intensive Care Med. 2018;44:1212–20. https://doi.org/10.1007/s00134-018-5269-7.

[3] Fernando SM, Tran A, Cheng W, Klompas M, Kyeremanteng K, Mehta S, et al. Diagnosis of ventilator-associated pneumonia in critically ill adult patients-a systematic

review and meta-analysis. Intensive Care Med. 2020;46:1170–9. https://doi.org/10.1007/s00134-020-06036-z.
[4] Millot G, Voisin B, Loiez C, Wallet F, Nseir S. The next generation of rapid point-of-care testing identification tools for ventilator-associated pneumonia. Ann Transl Med. 2017;5:451. https://doi.org/10.21037/atm.2017.11.05.
[5] Maugeri G, Lychko I, Sobral R, Roque ACA. Identification and antibiotic-susceptibility profiling of infectious bacterial agents: a review of current and future trends. Biotechnol J. 2019(4):e1700750. https://doi.org/10.1002/biot.201700750.
[6] Pirracchio R, Cohen MJ, Malenica I, Cohen J, Chambaz A, Cannesson M, et al. Big data and targeted machine learning in action to assist medical decision in the ICU. Anaesth Crit Care Pain Med. 2019;38:377–84. https://doi.org/10.1016/j.accpm.2018.09.008.
[7] Jansson M, Rubio J, Gavaldà R, Rello J. Artificial intelligence for clinical decision support in critical care, required and accelerated by COVID-19. Anaesth Crit Care Pain Med. 2020;39:691–3. https://doi.org/10.1016/j.accpm.2020.09.010.
[8] Douglas IS. New diagnostic methods for pneumonia in the ICU. Curr Opin Infect Dis. 2016;29:197–204. https://doi.org/10.1097/QCO.0000000000000249.
[9] Roth JA, Battegay M, Juchler F, Vogt JE, Widmer AF. Introduction to machine learning in digital healthcare epidemiology. Infect Control Hosp Epidemiol. 2018;39:1457–62. https://doi.org/10.1017/ice.2018.265.
[10] Klein Klouwenberg PM, van Mourik MS, Ong DS, Horn J, Schultz MJ, Cremer OL, et al. 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.
[11] 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;2014:1010–7.
[12] 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.
[13] Mann T, Ellsworth J, Huda N, Neelakanra A, Chevalier T, Sims KL, et al. Building and validating a computerized algorithm for surveillance of ventilator-associated events. Infect Control Hosp Epidemiol. 2015;36:999–1003. https://doi.org/10.1017/ice.2015.127.
[14] 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.
[15] 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.
[16] 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.
[17] 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.
[18] De Bus L, Saerens L, Gadeyne B, Boelens J, Claeys G, De Waele JJ, et al. Development of antibiotic treatment algorithms based on local ecology and respiratory surveillance cultures to restrict the use of broad-spectrum antimicrobial drugs in the treatment of hospital-acquired pneumonia in the intensive care unit: a retrospective analysis. Crit Care. 2014;18:R152. https://doi.org/10.1186/cc13990.
[19] 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. The Cochrane Collaboration. BMJ. 2017;356. https://doi.org/10.1136/bmj.i6460.
[20] 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.
[21] Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PLoS One. 2021;7. https://doi.org/10.1371/journal.pone.0032844.
[22] 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.
[23] Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. BMJ. 2015;350. https://doi.org/10.1136/bmj.
[24] 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.pmed.1001744.
[25] Fleuren LM, Thoral P, Shillan D, Ercole A, PWG Elbers, 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-y.
[26] Hockstein NG, Thaler ER, Torigian D, Miller Jr WT, Deffenderfer O, Hanson CW. Diagnosis of pneumonia with an electronic nose: correlation of vapor signature with chest computed tomography scan findings. Laryngoscope. 2004;114:1701–5. https://doi.org/10.1097/00005537-200410000-00005.
[27] Hanson CW, Thaler ER. Electronic nose prediction of a clinical pneumonia score: biosensors and microbes. Anesthesiology. 2005;102:63–8. https://doi.org/10.1097/00000542-200501000-00013.
[28] Hockstein NG, Thaler ER, Lin Y, Lee D, Hanson CW. Correlation of pneumonia score with electronic nose signature: a prospective study. Ann Otol Rhinol Laryngol. 2005;114:504–8. https://doi.org/10.1177/000348940511400702.
[29] Humphreys L, RML'E Orme, Moore P, Charaklias N, Sahgal N, Pont NP, et al. Electronic nose analysis of bronchoalveolar lavage fluid. Eur J Clin Invest. 2011;41:52–8. https://doi.org/10.1111/j.1365-2362.2010.02376.x.
[30] Bos LD, Martin-Loeches I, Kastelijn JB, Gili G, Espasa M, Povoa P, et al. The volatile metabolic fingerprint of ventilator-associated pneumonia. Intensive Care Med. 2014;40:761–2. https://doi.org/10.1007/s00134-014-3260-5.
[31] Chiu SW, Wang JH, Chang KH, Chang TH, Wang CM, Chang CL, et al. A fully integrated nose-on-a-chip for rapid diagnosis of ventilator-associated pneumonia. IEEE Trans Biomed Circuits Syst. 2014;8:765–8. https://doi.org/10.1109/TBCAS.2014.2377754.
[32] Tang K, Chiu S, Shih C, Chan C-L, Yang C-M, Yao D-J, et al. A 0.5V 1.27mW nose-on-a-chip for rapid diagnosis of ventilator-associated pneumonia. Digest of Technical Papers - IEEE International Solid-State Circuits Conference, 57. ; 2014. p. 420–1.
[33] Schnabel RM, Boumans ML, Smolinska A, Stobberingh EE, Kaufmann R, Roekaerts PM, et al. Electronic nose analysis of exhaled breath to diagnose ventilator-associated pneumonia. Respir Med. 2015;109:1454–9. https://doi.org/10.1016/j.rmed.2015.09.014.
[34] Liao YH, Wang ZC, Zhang FG, Abbod MF, Shih CH, Shieh JS. Machine learning methods applied to predict ventilator-associated pneumonia with *pseudomonas aeruginosa* infection via sensor array of electronic nose in intensive care unit. Sensors (Basel, Switzerland). 2019:19. https://doi.org/10.3390/s19081866.
[35] 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. Respir Res. 2020;21:45. https://doi.org/10.1186/s12931-020-1285-6.
[36] Liao YH, Shih CH, Abbod MF, Shieh J-S, Hsiao Y-J. Development of an E-nose system using machine learning methods to predict ventilator-associated pneumonia. Microsyst Technol. 2020. https://doi.org/10.1007/s00542-020-04782-0.
[37] Shih CH, Linb Y-J, Leeb K-F, Chienb P-Y, Drake P. Real-time electronic nose-based pathogen detection for respiratory intensive care patients. Sens Actuators B. 2010;148:153–7.
[38] Gao J, Zou Y, Wang Y, Wang F, Lang L, Wang P, et al. Breath analysis for noninvasively differentiating *Acinetobacter baumannii* ventilator-associated pneumonia from its respiratory tract colonization of ventilated patients. J Breath Res. 2016;10. https://doi.org/10.1088/1752-7155/10/2/027102.
[39] van Oort PM, de Bruin S, Weda H, Knobel HH, Schultz MJ, Bos LD, et al. Exhaled breath metabolomics for the diagnosis of pneumonia in intubated and mechanically-ventilated Intensive Care Unit (ICU)-patients. Int J Mol Sci. 2017;18:449. https://doi.org/10.3390/ijms18020449.
[40] Daza C, Maria JS, Gomez I, Barbe M, Trincado J, Capurro D. Phenotyping intensive care unit patients using temporal abstractions and temporal pattern matching. Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, Seattle, WA, USA; 2016. p. 508–9.
[41] Ding D, Stachel A, Iturrate E. Phillips M.1184. Making pneumonia surveillance easy: automation of pneumonia case detection. Open Forum Infect Dis. 2019;6:S424–5. https://doi.org/10.1093/ofid/ofz360.1047.
[42] 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. 2012;12:79. https://doi.org/10.1186/1471-2288-12-79.
[43] 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.
[44] Schurink CAM, Visscher S, Lucas PJF, van Leeuwen HJ, Buskens E, Hoff RG, et al. A Bayesian decision-support system for diagnosing ventilator-associated pneumonia. Intensive Care Med. 2007;33:1379–86. https://doi.org/10.1007/s00134-007-0728-6.
[45] Visscher S, Kruisheer EM, Schurink CAM, Lucas PJF, Bonten MJM. Predicting pathogens causing ventilator-associated pneumonia using a Bayesian network model. J Antimicrob Chemother. 2008;62:184–8. https://doi.org/10.1093/jac/dkn141.
[46] Ramírez-Estrada S, Borgatta B, Rello J. Pseudomonas aeruginosa ventilator-associated pneumonia management. Infect Drug Resist. 2016;20:7–18. https://doi.org/10.2147/IDR.S50669.
[47] Micek ST, Wunderink RG, Kollef MH, Chen C, Rello J, Chastre J, et al. An international multicenter retrospective study of *Pseudomonas aeruginosa* nosocomial pneumonia: impact of multidrug resistance. Crit Care. 2015;19:219. https://doi.org/10.1186/s13054-015-0926-5.
[48] European Centre for Disease Prevention and Control. Surveillance of antimicrobial resistance in Europe 2018. Stockholm: ECDC; 2019.
[49] Selvaraj R, Vasa NJ, Nagendra SMS, Mizaikoff B. Advances in mid-infrared spectroscopy-based sensing techniques for exhaled breath diagnostics. Molecules. 2020;25:2227. https://doi.org/10.3390/molecules25092227.
[50] Bos LDJ, Sterk PJ, Schultz MJ. Volatile metabolites of pathogens: a systematic review. PLoS Pathog. 2013;9. https://doi.org/10.1371/journal.ppat.1003311.
[51] Time to adapt. https://newsroom.intel.com/wp-content/uploads/sites/11/2018/07/healthcare-iot-infographic.pdf.
[52] Mamdani M, Slutsky AS. Artificial intelligence in intensive care medicine. Intensive Care Med. 2021;47:147–9. https://doi.org/10.1007/s00134-020-06203-2.

[53] Shaw JA, Sethi N, Block BL. Five things every clinician should know about AI ethics in intensive care. Intensive Care Med. 2021;47:157–9. https://doi.org/10.1007/s00134-020-06277-y.
[54] Lovejoy CA, Buch V, Maruthappu M. Artificial intelligence in the intensive care unit. Crit Care. 2019;23:7. https://doi.org/10.1186/s13054-018-2301-9.
[55] Ramírez-Estrada S, Pena-Lopez Y, Rello J. The effects of sedatives, neuromuscular blocking agents and opioids on ventilator-associated events. Eur J Anesthesiol. 2020;37:67–9. https://doi.org/10.1097/EJA.0000000000001132.
[56] Schurink CA, Van Niuwenhoven CA, Jacobs JA, Rozenberg-Arska M, Joore HCA, Buskens E, et al. Clinical pulmonary infection score for ventilator-associated pneumonia: accuracy and inter-observer variability. Intensive Care Med. 2004;30:217–24. https://doi.org/10.1007/s00134-003-2018-2.