REVIEW

CrossMark

The efficacy and effectiveness of machine learning for weaning in mechanically ventilated patients at the intensive care unit: a systematic review

Man Ting Kwong[1] · Glen Wright Colopy[2] · Anika M. Weber[3] · Ari Ercole[4] · Jeroen H. M. Bergmann[1]

Abstract

Weaning from mechanical ventilation in the intensive care unit (ICU) is a complex clinical problem and relevant for future organ engineering. Prolonged mechanical ventilation (MV) leads to a range of medical complications that increases length of stay and costs as well as contributes to morbidity and even mortality and long-term quality of life. The need to reduce MV is both clinical and economical. Artificial intelligence or machine learning (ML) methods are promising opportunities to positively influence patient outcomes. ML methods have been proposed to enhance clinical decisions processes by using the large amount of digital information generated in the ICU setting. There is a particular interest in empirical methods (such as ML) to improve management of "difficult-to-wean" patients, due to the associated costs and adverse events associated with this population. A systematic literature search was performed using the OVID, IEEEXplore, PubMed, and Web of Science databases. All publications that included (1) the application of ML to weaning from MV in the ICU and (2) a clinical outcome measurement were reviewed. A checklist to assess the study quality of medical ML publications was modified to suit the critical assessment of ML in MV weaning literature. The systematic search identified nine studies that used ML for weaning management from MV in critical care. The weaning management application areas included (1) prediction of successful spontaneous breathing trials (SBTs), (2) prediction of successful extubation, (3) prediction of arterial blood gases, and (4) ventilator setting and oxygenation-adjustment advisory systems. Seven of the nine studies scored seven out of eight on the quality index. The remaining two of the nine studies scored one out of eight on the quality index. This scoring may, in part, be explained by the publications' focus on technical novelty, and therefore focusing on issues most important to a technical audience, instead of issues most important for a systematic medical review. This review showed that only a limited number of studies have started to assess the efficacy and effectiveness of ML for MV in the ICU. However, ML has the potential to be applied to the prediction of SBT failure, extubation failure, and blood gases, and also the adjustment of ventilator and oxygenation settings. The available databases for the development of ML in this clinical area may still be inadequate. None of the reviewed studies reported on the procedure, treatment, or sedation strategy undergone by patients. Such information is unlikely to be required in a technical publication but is potentially vital to the development ML techniques that are sufficiently robust to meet the needs of the "difficult-to-wean" patient population.

Keywords Artificial intelligence · Patient-centred medicine · Clinical trial · Critical care · Data science

✉ Jeroen H. M. Bergmann
  jeroen.bergmann@eng.ox.ac.uk

[1] Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

[2] snap40 Inc, 125 Princess Street, Edinburgh EH2 4AD, UK

[3] CRUK/MRC Oxford Institute for Radiation Oncology, Roosevelt Drive, Oxford OX3 7DQ, UK

[4] Division of Anaesthesia, Addenbrooke's Hospital, University of Cambridge, Hills Road, Cambridge CB2 0QQ, UK

Introduction

Organ engineering combines engineering, cell biology, and material sciences to improve or replace the functions of tissues and organs [1]. The ability to engineer whole organs

🌀 Springer

provides a great prospect to patients who are awaiting a lung transplant. Mechanical ventilation (MV) is very important for intraoperative and early post-operative management of those who are undergoing a lung transplantation [2]. The use of MV to assist patients in breathing is a life-preserving procedure in an intensive care unit (ICU) setting. MV is commonly required for patients with a wide range of life-threatening pulmonary, neurological, neuromuscular, and cardiac conditions, as well as to facilitate surgery under anaesthesia. Whilst MV is an immediate necessity to preserve life, extended MV is associated with a number of complications. MV-associated complications are common and pose a significant clinical risk as they increase morbidity and mortality amongst ICU patients [3–6]. The weaning process may account for up to 40–50% of the total duration of MV [3, 7]. Patients receiving prolonged MV account for only 6% of all ICU patients, yet they consume 37% of ICU resources when adding general hospitalisation costs to the costs of MV [8]. An epidemiological study, using data from 2009, estimated 310 per 100,000 persons in the adult population undergo invasive ventilation for nonsurgical indications each year in the USA [9]. An earlier study, using data from 2005, estimated 270 per 100,000 persons in the adult population each year. Across the USA, approximately 800,000 patients require MV each year with estimated national costs of $27 billion. This accounts for 12% of total hospital costs [10]. Furthermore, ventilation-associated pneumonia (VAP) is estimated to be developed in 9–27% of ICU patients. The additional cost of VAP, due to increased medication, staff, diagnostic tests, and hospitalisation was estimated to be approximately US$40,000 per hospital per year [9, 10].

Early recognition of patients who are capable of some level of independent respiration is necessary to begin to gradually liberate the patient from MV and ultimately achieve full independent respiratory function. This process of liberating the patient from mechanical support and the endotracheal tube is commonly referred to as "weaning". Weaning is an essential element in the care of critically ill intubated patients receiving MV, yet there is uncertainty and controversy as to the best methods for conducting this process. Weaning management is therefore an important clinical issue for both patients and clinicians alike.

The clinical goals of weaning from MV are twofold: first, to promptly identify those patients who are ready to begin the process of weaning, and second, to optimise the weaning regime to reduce the transition time from dependence to independence from MV. MV can generally be withdrawn in a phased approach. However, a withdrawal that is too rapid may precipitate respiratory collapse that hinders the patient's recovery. An approach that is too conservative risks failures to exploit the patient's full physiological potential and extends the duration of ventilation. Either failure carries the associated risks of VAP or other ventilator-induced lung injuries. Accurate and robust prediction of a patient's reaction to a particular weaning strategy is not currently possible. Computational intelligence presents an opportunity to assess the effect of a multitude of clinical factors on the outcomes of patient populations. Furthermore, computational intelligence may facilitate the development of patient-specific models to better-tailor clinical inference to individualised evidence.

To improve current clinical practice, a number of studies have compared the use of clinical guidelines (protocols) and/or automated weaning systems to the common clinical practice of leaving the clinician to decide when to wean. Overall, it was demonstrated that the introduction of protocols and/or the use of automated weaning systems reduced (1) the average total time spent on MV, (2) the duration of the weaning process, and (3) the overall length of time spent in the ICU [11–13]. Whilst demonstrating the importance of considering weaning potential in all patients, many studies lacked sufficient detail about usual care practices (against which the protocol-driven results were compared). Furthermore, clinical studies often exclude the "difficult-to-wean" population, since the modification to protocols may have an adverse effect on the patient outcome. This leads to a high risk of bias in the clinical research, where findings cannot be generalised for the population who consumes the most MV related resources.

Furthermore, there is significant inter-protocol variation, as well as significant variation in the magnitude of improvement achieved in different studies. No studies compared multiple rigorous protocols (comparisons were only made to protocol-free practice); therefore, no clear consensus has been reached as to which protocols will work best for particular patients [11–13]. A further confounding factor in the comparison of studies' results is the existence of dozens of different ventilator systems currently in clinical use, with hundreds of different possible modes of ventilation. Variability in studies and baseline practices makes it difficult to generalise findings across ICUs.

In practice, weaning is generally performed by reducing ventilatory support and assessing the effect. The outcome of such a trial may be quantified by a multitude of vital sign parameters such as respiratory rate, heart rate, blood oxygen saturation, or carbon dioxide tension. Subjective bedside assessments include respiratory distress. Heterogeneity in both patients and clinical practice means that these clinical parameters are variable in their predicative value. Even the definition of clinical outcome is uncertain given that weaning failure may occur over a range of timescales. The difficulty for a human expert to evaluate such multidimensional longitudinal set of predictors motivates a computational approach to learning relations between predictors and outcomes.

Computational intelligence in weaning

Many commercially available ventilators have in-built automated weaning modes [13]. Such automated systems are a simple form of automated computational intelligence or AI. These systems are model-based systems, in which a model of physiology is specified a priori, and the parameters of this physiological model are subsequently calculated using data acquired from the MV patient [14]. Model-based systems may produce suboptimal outcomes in a practical clinical environment depending on the data available and the model's assumptions. For example, when the model's assumptions are grossly violated or when (unbiased) inference of the parameters from the data is impossible. On the other end of the spectrum are the data-driven systems employing machine learning (ML) algorithms. Knowledge-based systems rely on data-driven knowledge extraction and model specification. In the case of weaning from MV, the factors governing the respiratory system in an ICU environment are multifactorial and it is not feasible to create a model that is effective across a diffuse patient population without the use of knowledge-based AI techniques.

ML methods can be categorised into supervised, unsupervised, and reinforcement learning. In supervised learning, a set of inputs (or predictor variables) is mapped to a set of outputs (or outcome variables), where the mapping is specified via one or more mathematical functions [15]. For example, respiratory tidal volumes (volume of air being inhaled and exhaled with each respiratory cycle [16]) can be used as input parameters to predict the success or failure of a spontaneous breathing trial (SBT) as an outcome [17, 18]. Unsupervised learning is where learning is carried out with only a set of inputs, and where no predefined outcome is provided, outcomes or features of interest are defined implicitly from the relation between various inputs.

A subfield of ML techniques, known as novelty detection or one-class-classification, draws from elements of both supervised and unsupervised learning. Novelty detection is popular in instances where there is insufficient data to reliably define the relationship between predictors and all outcomes of interest. This is common, for example, in the case of medical data where data sets for acutely ill patients are rarely available, compared to the relative abundance of data for healthy patients. Novelty detection learns the (unsupervised) interrelation of predictive inputs for a predefined (supervised) class of outcomes. Novel data are then defined by this deviation from these learned relations with explicit assignment to a separate class [19].

Reinforcement learning is a decision-making model that learns the optimal sequence of decisions in an environment based on rewards received to achieve a task [15]. ML techniques that are currently used in the field of weaning management are commonly supervised methods.

Clinical efficacy and effectiveness

ML methods need to prove clinical efficacy and effectiveness in order to be applicable in clinical practice. Efficacy refers to the performance of the method under ideal and controlled conditions, whilst effectiveness is its performance under typical clinical conditions [20]. Trials that evaluate efficacy and effectiveness in isolation are not common [21]. In a clinical trial, the difference between usual or retrospective "uncontrolled clinical practice" would be compared with a controlled protocolised practice [11, 22]. Measurable clinical outcomes by which to evaluate weaning include weaning duration, rate of successful weaning trials, reintubation rates, length of stay in hospital, and mortality [23–25]. In accordance with medical publications, efficacy and effectiveness will be assessed via a systematic review, with predefined inclusion and exclusion criteria. Relevant reviews without these inclusion/exclusion criteria include [11, 26, 27]. Prior systematic reviews assessed extensively whether "protocolised versus non-protocolised" [11] and "automated versus non-automated" [13] weaning reduced the duration of mechanical ventilation. ML techniques have shown potential in various clinical applications [28–30]; the aim of this paper therefore assesses the field's current knowledge of the efficacy and effectiveness of machine learning for weaning in mechanically ventilated ICU patients.

Methods

Search strategy

A systematic search on OVID, IEEEXplore, PubMed, and Web of Science was performed according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines.

A combination of 30 keywords was devised to capture all studies that included ML applications in MV weaning management in an ICU setting with an outcome measure. In the PubMed search, a MeshTerm was used when available to maximise the coverage of the keywords. For the IEEEXplore search, a refined collection of keywords was used due to the restriction on the number of keywords. The searches were performed to include all publications up to September 2018. A detailed search strategy is available in the supplementary material.

Study selection

Weaning management procedures in ICU may include (1) sequential graded reduction in ventilator setting, (2) oxygenation management approach, (3) SBT, and until (4) extubation. Studies were included if it presented a ML model

applied in any of the above weaning-related procedures, as well as if the article explicitly mentions weaning. Studies must include ML to be included in this review; studies on common knowledge-based automated weaning methods were not included as those have previously been extensively reviewed by Rose et al. [13]. The other important inclusion criteria were that the study must provide a clinically relevant way to evaluate their ML methodology. This inclusion criterion was designed to be as generic as possible, as there is a wide range of weaning management procedures and the outcome measure of each of these areas may not be easily unified. Therefore, the focus here is to include studies that provided an evaluation of their ML framework either through comparing with retrospective ground truth based on a clinical outcome (e.g. failed extubation, blood gas measurements), or with decisions made by a clinician.

A double-blinded method was used for reviewing the publication: two reviewers (MTK and GWC) independently reviewed the title and abstract of each publication. Where there was a conflict on inclusion/exclusion, reviewers resolved by discussion of their reasoning, before reaching a final decision to include or exclude.

Inclusion criteria were:

1. Must be written in the English language.
2. Must be peer reviewed.
3. Must contain an outcome measure, methodology alone did not suffice.
4. Must have ML models applied to weaning management in an ICU.

Characteristics and data extracted from each study were (a) year of publication, (b) ML model used, (c) size of training/validation data set, (d) method of validation, (e) size of test set, and (f) type of ground truth used.

Quality assessment

The quality of each article was assessed by adapting the Joanna Briggs Institute (JBI) critical appraisal checklist for cross-sectional research [31] to ML in MV literature. This methodology was used, e.g. in Islam et al. [32] to review the application of data mining on healthcare analytics. The checklist was adapted to be applicable for ML applications for weaning in an ICU setting, and the modified checklist can be found in Table 1.

Results

A total of 417 records were identified from the search from OVID, IEEEXplore, PubMed, and Web of Science. After removing a single duplicate, 416 abstracts were screened by

two authors (MTK and GWC) in a double-blinded manner (see Fig. 1). From the initial screening phase, 20 studies were included for the full-article screening phase based on the study title and abstract. A total of nine studies met the inclusion criteria after screening the full-articles. A summary of study characteristics is presented in Table 2.

Descriptive summary of results

Application area

A range of applications was reviewed: weaning management can be further subcategorised into (1) sequential ventilation and oxygenation adjustments/decision support ($n = 3$) [14, 21, 33], (2) T-piece SBT test ($n = 4$) [17, 18, 35, 36], (3) blood gas prediction ($n = 1$) [34], and (4) extubation prediction ($n = 1$) [37].

Parameters collected and used

Since the area of application ranged from ventilator and oxygenation adjustments to SBT and extubation success prediction, a broad range of ventilator and vital signs parameters were used across the studies. The most common ventilator parameter used was tidal volume ($n = 6$). Belal et al. used the largest number of vital sign parameters, a total of eleven [33]. Mikhno et al. used most of the parameters that can be acquired from either a mechanical ventilator or a vital signs bedside monitor, these parameters included white cell count, PaO_2/FiO_2 ratio, work of breathing index and rapid shallow breathing index [37]. The parameters collected by the studies reviewed are summarised in Table 3.

ML models used

Neural network-based models, ANN ($n = 3$) and ANFIS ($n = 4$), were the most common ML models used in the studies presented. Other algorithms include SVM ($n = 1$) and LR ($n = 1$). Table 4 summarises the ML algorithm used in the reviewed studies.

Databases

In terms of data source, one study reviewed was based on the MIMIC-II database [40], the only publicly available database. Four studies were based on WEANDB database collected in ICU of Hospital de la Santa Creu i Sant Pau and the Hospital Universitario de Getafe in Spain, whilst the other studies were based on data collected in ICU units based in UK hospitals (adult general ICU of the Sheffield Royal Hallamshire Hospital and neonatal ICU of the Royal Liverpool University Hospital).

Table 1 Modified JBI critical appraisal checklist for analytical cross-sectional studies and summary of quality scores

Item	Checklist	Giraldo [17]	Giraldo [18]	Kwok [14]	Belal [33]	Wang [34]	Wang [21]	Arizmendi [35]	Arizmendi [36]	Mikhno [37]
1	A clear objective and description of inclusion criteria of the study population (i.e. selection of patient group, or time series data) Or statement of aim prior data to collection?	1	1	0	1	1	0	1	1	1
2	A detailed description of the study population (i.e. how are the patients recruited? Which ICU, what procedures, treatment and sedation strategy?)	0	0	0	0	0	0	0	0	0
3	A clear description of the data source and how data was collected (i.e. clearly describe method of measurement. Machine used, or manual notes. Reproducible measurements? Artefacts?)	1	1	0	1	1	0	1	1	1
4	A valid and reproducible data collection and measurement method (i.e. Was the data measured in a reliable way?)	1	1	0	1	1	0	1	1	1
5	Attainment of ethical approval. Was the ethical issue (patient confidentiality) considered?	1	1	0	1	1	0	1	1	1
6	Were findings and implications discussed in detail?	1	1	1	1	1	1	1	1	1
7	Were the outcomes (performance and result of ML tools) measured in a valid and reliable way?	1	1	0	1	1	0	1	1	1
8	Was appropriate cross-validation and evaluation method used?	1	1	0	1	1	0	1	1	1

Table modified from the JBI. The appraisal checklist for analytical cross-sectional studies was applied to the studies included in this review and quality scores allocated to each study. Available online: http://joannabriggs.org/research/critical-appraisal-tools.html. (Accessed on cited 24/09/2018)
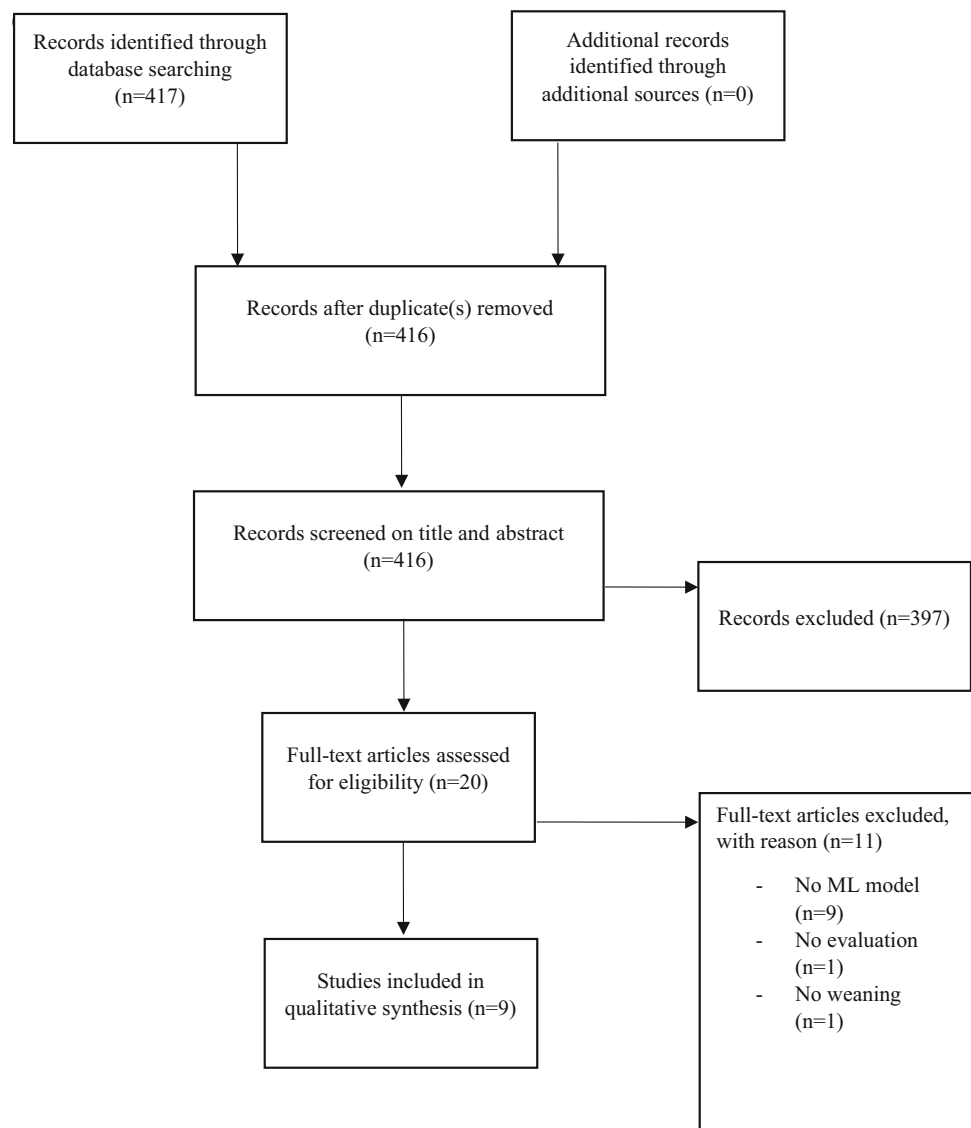
Evaluation ground truth

To evaluate ground truths, four studies, which focussed on SBT, used trial success or failure (and reintubation) as an outcome. These studies did not require explicit expert annotation. Similarly, for the study focussing on extubation prediction, the outcome measure is governed by the success or failure of extubation. The study on blood gas prediction used measured blood gases values as a ground truth. The three studies on advisory systems relied on expert agreement on real or simulated clinical scenarios as a means of validation.

Quality index

Most studies (7 of the 9) scored seven out of eight on the quality scoring system. These studies were able to provide information on the checklist except for details on the procedures, treatment, and/or sedation strategy the patients had

Fig. 1 PRISMA flowchart of systematic review: identification, screening, eligibility, and inclusion. The number of studies identified is given for each level

```
┌─────────────────────────┐         ┌─────────────────────────┐
│ Records identified      │         │ Additional records      │
│ through database        │         │ identified through      │
│ searching (n=417)       │         │ additional sources (n=0)│
└────────────┬────────────┘         └────────────┬────────────┘
             │                                   │
             └─────────────┬─────────────────────┘
                           ▼
             ┌─────────────────────────────────┐
             │ Records after duplicate(s)      │
             │ removed (n=416)                 │
             └────────────────┬────────────────┘
                              ▼
             ┌─────────────────────────────────┐         ┌─────────────────────────┐
             │ Records screened on title and   │────────▶│ Records excluded (n=397)│
             │ abstract (n=416)                │         └─────────────────────────┘
             └────────────────┬────────────────┘
                              ▼
             ┌─────────────────────────────────┐         ┌─────────────────────────┐
             │ Full-text articles assessed     │────────▶│ Full-text articles      │
             │ for eligibility (n=20)          │         │ excluded, with reason   │
             └────────────────┬────────────────┘         │ (n=11)                  │
                              ▼                           │  - No ML model (n=9)    │
             ┌─────────────────────────────────┐         │  - No evaluation (n=1)  │
             │ Studies included in qualitative │         │  - No weaning (n=1)     │
             │ synthesis (n=9)                 │         └─────────────────────────┘
             └─────────────────────────────────┘
```

undergone. The remaining two studies scored one out of eight on the quality score, because they did not detail information of the data that was used, nor was effective validation carried out. A summary of the scores for each study can be found in Table 1.

Discussion

This is the first systematic review that focuses on the clinical effectiveness and efficacy of the application of ML to the management of weaning from MV in ICU. Publications describing novel ML methodology for medical applications are typically developed from retrospectively collected data. This makes a direct assessment of the ML method's effect on clinical outcomes impossible. Instead, such papers either focus on (1) a performance metric that acts as a near-facsimile

of effect on clinical outcome, or (2) a performance metric indirectly related to clinical outcome. We use the term "near-facsimile" (to describe attempts to measure clinical efficacy in a retrospective study) since the variability and confounding factors in clinical practice make it impossible to deduce the exact clinical effect of machine-generated information. For example, using an ML system that provides early warning that a patient is becoming fatigued during an SBT, we may (retrospectively) assess with accuracy the timeliness of such alarms. However, we may not deduce from this information whether such alarms were acted on by clinical staff or, in turn, the timeliness of the ML-induced intervention.

We delineate between (1) the assessment of protocols at a targeted clinical task (i.e. improvement of one or more clinical outcomes) and (2) the assessment of ML algorithms at ML tasks (usually inductive/generalised performance for the task at which the ML algorithm is trained). The articles

Table 2 Studies characteristics

First author, year of publication	Experts	ML models	Training set size	Validation method	Test set size	Ground truth (performance measure provided by the study)
Application area						
SBT/T-piece test						
Giraldo, 2006 [17]	N/A[a]	ANN	56	8-FCV	8	Success or failure of SBT
Giraldo, 2006 [18]	N/A	SVM	146	LOOCV	–	Success or failure of SBT
Arizmendi, 2009 [35]	N/A	ANN	149	10-FCV	–	Success or failure of SBT, or reintubated within 48 h
Arizmendi, 2018 [36]	N/A	ANN	8	5-FCV	–	Success or failure of SBT, or reintubated within 48 h
Advisory system						
Belal, 2005 [33]	Two registrars and three nurses	ANFIS	7	–	–	91% expert agreement on ventilator setting and 85% on oxygenation
Kwok, 2004 [14]	ICU consultant	ANFIS	10	–	–	≥ 75% expert agreement
Wang, 2010 [21]	–	ANFIS	–	–	–	Expert agreement on simulated clinical scenarios
Blood gas prediction						
Wang, 2010 [34]	–	ANFIS	25	–	5	Measured arterial blood gases ($PaCO_2$ and PaO_2)
Extubation prediction						
Mikhno, 2012 [37]	N/A	LR	179	Bootstrapped ROC curves	–	Success of extubation failure

ANN artificial neural networks, *ANFIS* adaptive neurofuzzy inference system, *LR* logistic regression, *LOOCV* leave-one-out cross-validation, *FCV* fold cross-validation, *N/A* not applicable

[a]Each study based on SBT does not require expert annotation, since the outcome is defined based on success, failure, or reintubation. Therefore, these studies are marked as N/A

reviewed demonstrated proof-of-concept rather than assessing efficacy. The results show that only a few papers have started to assess the degree of beneficial effect of ML under "real world" clinical conditions for MV. Results indicate that ML has the potential to be applied to important clinical issues in weaning, such as SBT and extubation failure predictions, blood gases predictions, and ventilator settings (and oxygenation) adjustments. It is worth mentioning that ML development focussing on advisory systems tended to have a less-robust validation method. This is due to (1) the difficulty in testing the system in a clinical environment and (2) the subjectivity and impracticality of ground truth annotations from human experts on the long list of decisions the algorithm makes. More research may be required to devise an effective way to validate the performance of such systems. A previous systematic review on automated weaning discussed the need to further develop technology in the neurosurgical population

and also for studies to examine sedation strategy [11]. In this review, no study reported the procedure, treatment, or sedation strategy undergone by patients. Furthermore, the sample sizes of most of these studies are relatively small. Rigorous evaluation of system performance may be challenging in the presence of "difficult-to-wean" patients since these patients (1) may require a more specialised weaning strategy and (2) have less associated data from which to train an effective ML algorithm. This highlights the fact that the data collected around the development of ML technology for this application is incomplete for important portions of the MV patient population. This is an issue that must be overcome for ML technology to progress for this clinical area.

The publications included in this review have several items of commonality that warrant discussion. First, within the nine papers included in the systematic review, there are two groups with significant overlap in authorship:

Table 3 Ventilator and vital sign parameters

Parameters	
Ventilator	*Vital signs bedside monitor*
Inspiratory time (T_I) [17, 18, 33, 35, 36]	Cardiac interbeat duration (RR-interval) [35, 36]
Expiratory time (T_E) [17, 18, 33, 35, 36]	Plethysmogram waveform [33]
Breath duration (T_{Tot}) [17, 18, 35, 36]	Respiration waveform [33]
	Oxygen saturation (SaO_2) [33]
Tidal volume (V_T) [17, 18, 21, 34–36],	Heart rate (HR) [33, 37]
Fractional inspiratory time (T_I/T_{Tot}) [17, 18, 35, 36]	Pulse rate (PR) [33]
Mean inspiratory flow (V_T/T_T) [17, 18, 35, 36]	Respiratory rate [33]
Frequency–tidal volume ratio (f/V_T) [17, 18, 35, 36]	Transcutaneous O_2 ($tcpO_2$) [33]
	Transcutaneous CO_2 ($tcpCO_2$) [33]
Ventilatory rate (V_{rate}) [14, 21, 33, 34]	Invasive blood pressure (INBP) [33]
Peak inspiratory pressure (PIP) [14, 33]	Non-invasive blood pressure (BP) [33]
Positive end expiratory pressure (PEEP) [14, 21, 33, 34]	Temperature (TEMP) [33]
Mean airway pressure (MEAN) [14, 33]	
Fraction of inspired oxygen (FIO_2) [14, 21, 33, 34, 37]	*Other*
Inspiration-to-expiration ratio (V_I:V_E) [21, 34]	Arterial blood gases (PaO_2, $PaCO_2$) [21, 34, 37]
	white blood cell count [37]
Relative dead space (Kd) [21, 34]	PaO_2/FiO_2 ratio [37]
Total minute volume [21, 34]	Work of breathing index [37]
	Rapid shallow breathing index [37]
	Age, gender, weight and height [21, 34]

Table 4 Summary of ML algorithms

Algorithm	Description
Artificial neural networks (ANN)	Algorithm inspired by the neural networks of the brain. A network consists of layers of nodes in which each node is connected to each other by a weighted link. With many parameters that can be tuned and optimised, it can handle variability and noisy data. However, without proper regularisation or too little data to exemplify variability, ANN can be prone to overfitting [15], leading to poor generalisation
Support vector machine (SVM)	This is a powerful classification method that partitions the predictive variable space by determining a hyperplane that optimally separates the outcome classes according to a loss function. SVMs are effective on high dimensional data but are most effective on a small data set with minimal class-overlap or noise-corruption [38]
Adaptive network-based fuzzy inference systems (ANFISs)	This method provides fuzzy membership functions between input and output parameters within a neural network. This method requires a domain expert to predefine the system's constituent membership functions and rule structure [39]
Logistic regression (LR)	This is a predictive statistical method for dealing with binary outcomes. LR defines a linear relation between the independent variable and the odds ratio of the binary [15]. LR is a simple yet powerful method; however, it can only be used to classify binary outcomes. The multi-class equivalent of logistic regression is called "multinomial logistic regression"

The first group of significant overlap in authorship comprises of Giraldo [17, 18] and Arizmendi [35, 36]. Each publication in this first group was a conference paper and used an identical data source (WEANDB). These publications contained strong overlap in the derivation and tuning of predictive features and each paper aimed to predict the clinical outcome of SBT via cross-validation methods.

The second group of significant overlap in authorship comprises of Kwok [14] and Wang [21, 34]. Each publication in this group was a journal article, with Wang [21, 34] being paired publications within the same issue. Whilst Kwok [14] and Wang [21] evaluated ML-performance based on an expert's (subjective) agreement with algorithmic output, Wang [21] compared predictive performance to (objective) arterial blood gas values. A further distinguishing feature of the papers by Wang and Kwok is that they each included physiological modelling (via differential equations) *in addition* to ML models, whereas the other papers included only ML models. Kwok [14] and Wang [21] were the two papers that scored one out of eight on the modified JBI checklist. Given that the partner paper of Wang 21] (i.e. Wang [34]), scored seven out of eight, it is possible that the authors may have expected readers to use the data description in Wang [34] instead of evaluating Wang [21] as a stand-alone paper. However, for the purposes of a systematic review, it seems most appropriate to score each paper independently. For Kwok [14], several of the scoring criteria did not apply, since the described work required neither patient data nor the associated ethic approval to collect such data. These items alone would account for five points on the checklist.

An important commonality of all papers included is that each was intended for a technical audience: four papers were published as IEEE conference proceedings, one as AI conference proceedings. The four journal papers were each for a technical or computational journal. It is not typical for a technical or ML publication to focus on the issues important to a medical systematic review (particularly when considering external factors such as page limits). Accordingly, the score assigned for the criteria of the systemic review should not be interpreted as an assessment of technical strength of the paper.

This study has shown that more work needs to be done to bring truly patient-centred MV technologies into the healthcare system. Patient-specific MV technology can reduce ventilator-induced lung injuries caused by generic weaning protocol; this can be beneficial specifically for the "difficult-to-wean" population where the generic weaning protocol is not effective. The ability to optimise ventilation for a specific patient will be desirable for most post-operative patients, particularly those who have undergone a lung transplantation. The ventilation needs to be tuned to the requirements of the organ and this is likely to be even more important for tissue-engineered organs. Providing organ-specific biophys-ical stimuli could greatly support the future developments and integration of engineered organs, such as lungs.

References

1. Elitok MS, Gunduz E, Gurses HE, Gunduz M (2018) Tissue engineering: towards development of regenerative and transplant medicine. In: Barh D, Azevedo V (eds) Omics technologies and bio-engineering. Academic Press, pp 471–495. https://doi.org/10.1016/B978-0-12-804659-3.00020-8

2. Barnes L et al (2015) Mechanical ventilation for the lung transplant recipient. Curr Pulmonol Rep 4(2):88–96

3. Esteban A et al (2002) Characteristics and outcomes in adult patients receiving mechanical ventilation. Am Med Assoc 287(3):345–355

4. Tobin MJ (1994) Mechanical ventilation. N Engl J Med 330(15):1056–1061

5. Melsen WG, Rovers MM, Koeman M, Bonten MJM (2011) Estimating the attributable mortality of ventilator-associated pneumonia from randomized prevention studies. Crit Care Med 39(12):2736–2742

6. Bekaert M et al (2011) Attributable mortality of ventilator-associated pneumonia: a reappraisal using causal analysis. Am J Respir Crit Care Med 184(10):1133–1139

7. McConville JF, Kress JP (2012) Weaning patients from the ventilator. N Engl J Med 367(23):2233–2239

8. Jackson M, Strang T, Rajalingam Y (2012) A practical approach to the difficult-to-wean patient. J Intensive Care Soc 13(4):327–331

9. Mehta AB, Syeda SN, Wiener RS, Walkey AJ (2015) Epidemiological trends in invasive mechanical ventilation in the United States: a population-based study. J Crit Care 30(6):1217–1221

10. Wunsch H, Linde-Zwirble WT, Angus DC, Hartman ME, Milbrandt EB, Kahn JM (2010) The epidemiology of mechanical ventilation use in the United States. Crit Care Med 38(10):1947–1953

11. Blackwood B, Burns KEA, Cardwell CR, O'Halloran P (2014) Protocolized versus non-protocolized weaning for reducing the duration of mechanical ventilation in critically ill adult patients. Cochrane Database Syst Rev. https://doi.org/10.1002/14651858.CD006904.pub3

12. Burns KEA, Lellouche F, Nisenbaum R, Lessard MR, Friedrich JO (2014) Automated weaning and SBT systems versus non-automated weaning strategies for weaning time in invasively ventilated critically ill adults. Cochrane Database Syst Rev. https://doi.org/10.1002/14651858.CD008638.pub2

13. Rose L, Schultz MJ, Cardwell CR, Jouvet P, McAuley DF, Blackwood B (2015) Automated versus non-automated weaning for reducing the duration of mechanical ventilation for critically ill adults and children: a cochrane systematic review and meta-analysis. Crit Care 19(1):1–12

14. Kwok HF, Linkens DA, Mahfouf M, Mills GH (2004) SIVA: a hybrid knowledge-and-model-based advisory system for intensive care ventilators. IEEE Trans Inf Technol Biomed 8(2):161–172

15. Alpaydin E (2010) Introduction to machine learning. The MIT Press, Cambridge

16. Hallett S, Ashurst JV (2018) Physiology, tidal volume. StatPearls. https://www.ncbi.nlm.nih.gov/books/NBK482502/. Accessed 02 Oct 2018

17. Giraldo B et al (2006) Patients on weaning trials from mechanical ventilation classified with neural networks and feature selection. In: Annual international conference of the IEEE engineering in medicine and biology society—proceedings, pp 2195–2198

18. Giraldo B, Garde A, Arizmendi C, Jane R, Benito S, Diaz I, Ballesteros D (2006) Support vector machine classification applied on weaning trials patients. In: 2006 international conference of the IEEE engineering in medicine and biology society. pp 5587–5590. https://doi.org/10.1109/IEMBS.2006.259440

19. Pimentel MAF et al (2017) Toward a robust estimation of respiratory rate from pulse oximeters. IEEE Trans Biomed Eng 64(8):1914–1923

20. Revicki DA, Frank L (1999) Pharmacoeconomic evaluation in the real world: effectiveness versus efficacy studies. Pharmacoeconomics 15(5):423–434

21. Wang A et al (2010) Intelligent model-based advisory system for the management of ventilated intensive care patients. Part II: advisory system design and evaluation. Comput Methods Programs Biomed 99(2):208–217

22. Vitacca M et al (2001) Comparison of two methods for weaning patients with chronic obstructive pulmonary disease requiring mechanical ventilation for more than 15 days. Am J Respir Crit Care Med 164:225–230

23. Fagoni N et al (2018) Comparison between a nurse-led weaning protocol and weaning based on physician's clinical judgment in tracheostomized critically ill patients: a pilot randomized controlled clinical trial. Ann Intensive Care 8(1):1–10

24. Girard TD et al (2008) Efficacy and safety of a paired sedation and ventilator weaning protocol for mechanically ventilated patients in intensive care (awakening and breathing controlled trial): a randomised controlled trial. Lancet 371(9607):126–134

25. Nava S, Ambrosino N, Clini E, Prato M, Orlando G, Vitacca M, Brigada P, Fracchia C, Rubini F (1998) Noninvasive mechanical ventilation in the weaning of patients with respiratory failure due to chronic obstructive pulmonary disease: a randomized, controlled trial. Ann Intern Med 128(9):721–728. https://doi.org/10.7326/0003-4819-128-9-199805010-00004

26. Hodgson C et al (2016) Recruitment manoeuvres for adults with acute respiratory distress syndrome receiving mechanical ventilation. Cochrane Database Syst Rev (11) (review)

27. Klingenberg C, Wheeler KI, Mccallion N, Morley CJ, Davis PG (2017) Volume-targeted versus pressure-limited ventilation in neonates. Cochrane Database Syst Rev. https://doi.org/10.1002/14651858.CD003666.pub4

28. Sen A, Rubinfeld I, Azuh O (2009) Point-of-care use of Visensia (BioSign) index score predicts poor outcomes and may benefit pre-hospital trauma triage. In: The American Association for the surgery of trauma annual meeting, p 66

29. Hravnak M, Devita MA, Clontz A, Edwards L, Valenta C, Pinsky MR (2011) Cardiorespiratory instability before and after implementing an integrated monitoring system. Crit Care Med 39(1):65–72

30. Choukalas S, Takemoto CG, Stotts J, Wallace A (2015) Identifying ICU patients at high risk for cardiac arrest: a retrospective analysis of the Visensia algorithm. In: International anesthesia research society annual meeting and international science symposium

31. Joanna Briggs Institute critical appraisal tool. http://joannabriggs.org/research/critical-appraisal-tools.html. Accessed 20 Sept 2018

32. Islam M, Hasan M, Wang X, Germack H, Noor-E-Alam M (2018) A systematic review on healthcare analytics: application and theoretical perspective of data mining. Healthcare 6(2):54

33. Belal SY, Taktak AFG, Nevill A, Spencer A (2005) An intelligent ventilation and oxygenation management system in neonatal intensive care using fuzzy trend template fitting. Physiol Meas 26(4):555–570

34. Wang A et al (2010) Intelligent model-based advisory system for the management of ventilated intensive care patients: hybrid blood gas patient model. Comput Methods Programs Biomed 99(2):195–207

35. Arizmendi C et al (2009) Data mining of patients on weaning trials from mechanical ventilation using cluster analysis and neural networks. In Annual international conference of the IEEE engineering in medicine and biology, pp 4343–4346

36. Arizmendi CJ, Solano EH, Gonzalez H, Acuna HG, Giraldo BF (2018) Analysis of cardiorespiratory interaction in patients submitted to the T-tube test in the weaning process implementing symbolic dynamics and neural networks. In: 2018 international conference on artificial intelligence and big data (ICAIBD), pp 101–105

37. Mikhno A, Ennett CM (2012) Prediction of extubation failure for neonates with respiratory distress syndrome using the MIMIC—II clinical database, pp 5094–5097

38. Singh A, Thakur N, Sharma A (2016) A review of supervised machine learning algorithms. In: Proceedings of 10th INDIACom; 2016 3rd international conference on computing for sustainable global development. INDIACom 2016, pp 1310–1315

39. Jang JSR (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23(3):665–685

40. Saeed M et al (2011) Multiparameter intelligent monitoring in intensive care II (MIMIC-II): a public-access intensive care unit database. Crit Care Med 39(February 2010):952–960