



Received: March 8, 2022

Accepted: March 15, 2022

Corresponding author:

Hyung-Chul Lee, M.D., Ph.D.

Department of Anesthesiology and Pain
Medicine, Seoul National University College of
Medicine, 101 Daehak-ro, Jongno-gu, Seoul
03080, Korea

Tel: +82-2-2072-0723

Fax: +82-2-747-8363

Email: vital@snu.ac.kr

ORCID: <https://orcid.org/0000-0003-0048-7958>

Artificial intelligence in perioperative medicine: a narrative review

Hyun-Kyu Yoon¹, Hyun-Lim Yang^{1,2}, Chul-Woo Jung^{1,3},
Hyung-Chul Lee^{1,3}

¹Department of Anesthesiology and Pain Medicine, Seoul National University Hospital, ²Biomedical Research Institute, Seoul National University Hospital, ³Department of Anesthesiology and Pain Medicine, Seoul National University College of Medicine, Seoul, Korea

Recent advancements in artificial intelligence (AI) techniques have enabled the development of accurate prediction models using clinical big data. AI models for perioperative risk stratification, intraoperative event prediction, biosignal analyses, and intensive care medicine have been developed in the field of perioperative medicine. Some of these models have been validated using external datasets and randomized controlled trials. Once these models are implemented in electronic health record systems or software medical devices, they could help anesthesiologists improve clinical outcomes by accurately predicting complications and suggesting optimal treatment strategies in real-time. This review provides an overview of the AI techniques used in perioperative medicine and a summary of the studies that have been published using these techniques. Understanding these techniques will aid in their appropriate application in clinical practice.

Keywords: Artificial intelligence; Deep learning; Machine learning; Outcome assessment; Perioperative care; Risk assessment.

Introduction

Recent advancements in computing power, data storage, and the accumulation of clinical data in electronic health records (EHRs), as well as picture archiving and communication systems, have played a major role in introducing artificial intelligence (AI) into various fields of medicine [1]. Numerous studies have been published that use AI techniques in radiology [2], pathology [3], cardiology [4], and surgery [5]. For perioperative medicine, AI models for perioperative risk stratification, intraoperative monitoring, and intensive care management have been studied [6,7]. In some cases, these models outperform conventional statistical models and even human experts [8–10]. Many of these models can be used in clinical practice if their performance is maintained in future prospective validation studies and their clinical utility is confirmed by randomized controlled trials.

This narrative review addresses the various AI techniques used in clinical studies. Additionally, existing evidence from clinical studies that have used AI for important perioperative outcomes is summarized.

Overview of AI techniques

Modeling algorithms

Machine learning, which can learn patterns from data, is the most widely used AI algo-



© The Korean Society of Anesthesiologists, 2022

© This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

algorithm in perioperative medicine [11]. Machine learning algorithms are typically classified into three categories: supervised, unsupervised, and reinforcement learning (Fig. 1).

Supervised learning algorithms learn patterns from pairs of input and output variables. Supervised learning algorithms are typically divided into classification and regression algorithms. Gradient boosting machine (GBM) and random forest (RF) are widely used supervised learning-based classification algorithms with excellent performance. GBM and RF use collections of decision trees whose results are summed and averaged to produce a single

result. Gradient-boosted regression trees and RF regressors are regression variants of GBM and RF. For survival analyses, they are called gradient-boosting survival trees and random survival forests. Deep learning (DL) is another technique widely used in supervised learning-based classification and regression algorithms that uses a network of mathematical models of a neuron (perceptron). A multilayer perceptron (MLP) is the most basic DL model and consists of multiple fully connected layers of perceptrons. A convolutional neural network (CNN), which is frequently used in image analyses and biosignal processing, uses perceptrons that are

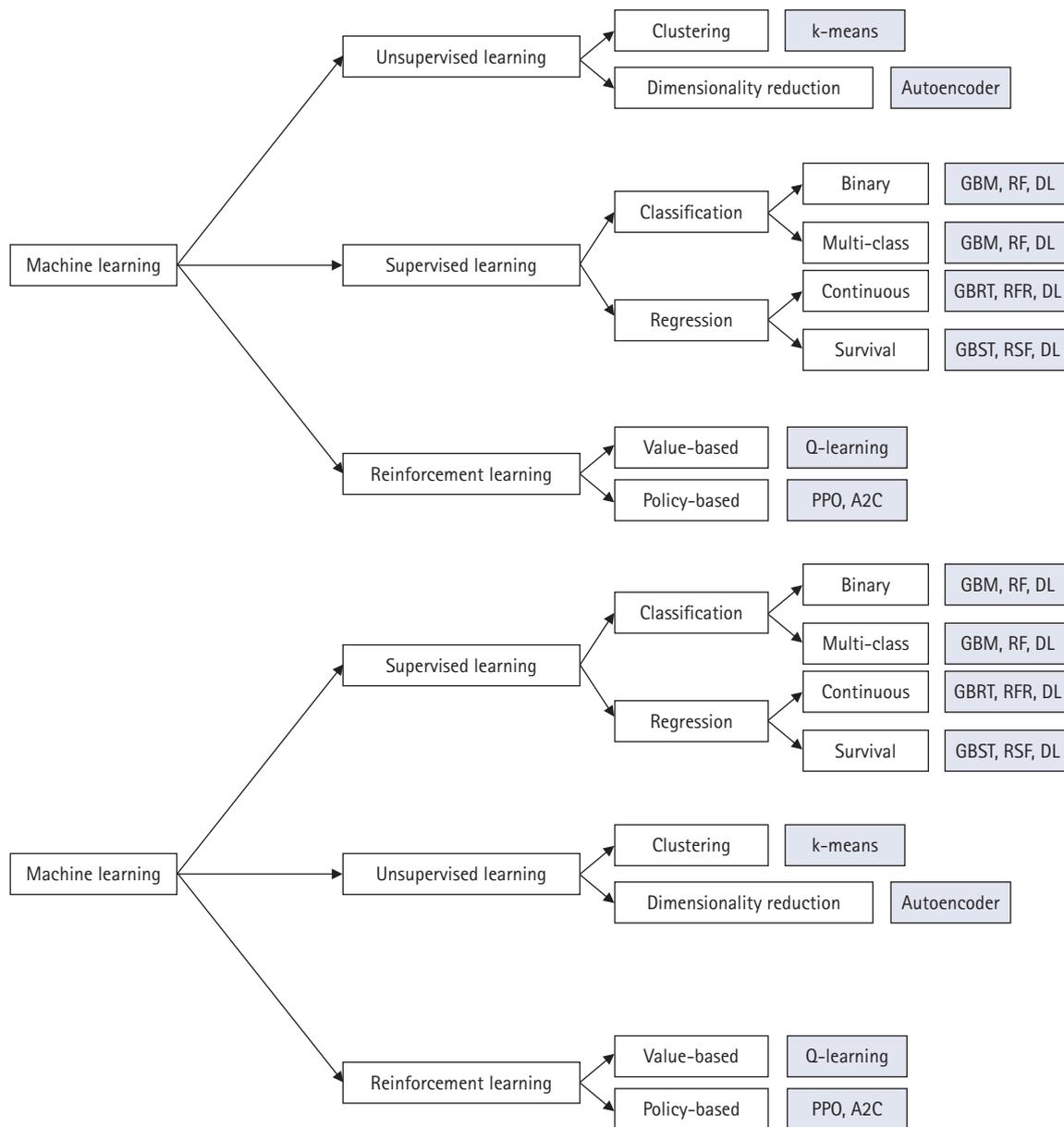


Fig. 1. Classification of machine learning algorithms. GBM: gradient boosting machine, RF: random forest, DL: deep learning, GBRT: gradient boosted regression tree, RFR: random forest regressor, GBST: gradient boosting survival tree, RSF: random survival forest, PPO: proximal policy optimization, A2C: advantage actor-critic algorithm.

activated only by specific patterns of geographically adjacent neurons in the previous layer. A recurrent neural network (RNN), which is frequently used in natural language processing and time-series data analyses, includes long short-term memory (LSTM) or gated recurrent units. RNNs include a recurrent loop for analyzing time-dependent sequences in a network. Recently, new structures of the DL model, including a self-attention layer, have emerged and have shown better performance than canonical RNN or CNN models [12–15].

Unsupervised learning algorithms can learn patterns from unlabelled data. Because there are no labels to learn, unsupervised learning algorithms use the distributions or patterns of the samples in the training dataset. For example, the k-means clustering algorithm uses sample distributions to classify the data into a specific number of groups. An autoencoder is a DL model that uses the input data as the labels. However, as the autoencoder has a structural bottleneck, it can reduce the dimensions of the dataset. Autoencoders can be used to detect abnormalities in a sample and remove noise from the biosignal [16].

Reinforcement learning algorithms can learn the optimal policy from data. Because it is impossible to build a model to simulate the strategy, a model-free offline reinforcement learning algorithm is used for most medical problems. Value-based algorithms, such as Q-learning, learn the value of each action in each status [17]. Conversely, policy-based algorithms, such as proximal policy optimization (PPO) and the advantage actor-critic (A2C), learn the optimal action in each status [18].

Hyperparameters

The number of neurons in each layer of the MLP and the number of decision trees in the GBM are examples of hyperparameters, whose values are used to control or tune the learning process of AI algorithms. Although AI models can automatically learn patterns from the input data, the range of these hyperparameters should be specified by humans. A grid search, which simply searches through grids in the search space, is the traditional method for determining the best hyperparameter. Random search [19] and Bayesian optimization [20] can be used to achieve better results in a limited number of searches.

Outcome variables

Clinical outcomes are the most common output variables of AI models in the medical field. They are selected based on clinical requirements. The label should be determined by expert consensus, because the performance of the supervised learning model

depends on the quality of the labeling, and several decisions must be made even for studies with simple outcomes. For example, for a study of in-hospital mortality within 30 postoperative days, researchers must decide whether the date is based on the beginning or end of the surgery and whether the time of death is defined as the time of the declaration or certificate. Additionally, researchers must decide whether mortality should be treated as a binary or survival outcome, which includes the censoring time.

Input variables

All variables that can affect the outcome should be considered as input variables to improve the model performance. However, any variable that is affected by the outcome variable itself should be removed as an input variable since this could result in a causality problem, which decreases the external validity [21]. Examples include using the fraction of inspired oxygen to predict intraoperative hypoxia or using postoperative pain to predict postoperative nausea and vomiting.

Reducing the number of input variables by removing irrelevant variables can improve the performance, robustness, and interpretability of the model while reducing the learning time [22]. Several techniques have been suggested for optimal feature selection, such as recursive feature estimation or the Boruta algorithm [23].

In linear regression, multicollinearity between the input variables can cause algorithm instability and distortion of parameter estimates. However, most AI algorithms can converge even with correlated input variables. Nevertheless, removing collinear variables can help improve performance by reducing the number of input variables. Additionally, the multicollinearity of the input variables can have a significant impact on feature importance in explainable AI algorithms [24]. Therefore, if the effect of a specific input variable on the outcome is the opposite of what is expected, it may be the result of multicollinearity.

Study population

As with the outcome variable, the study population selected has a significant influence on the performance of the AI model. Having clear and appropriate inclusion and exclusion criteria is essential to determine the scope of the model. Because the AI model learns the pattern of the training dataset, it is critical to create a training set with as many clinical scenarios as possible. Therefore, multicenter and multinational data are preferable, particularly for AI studies.

The test dataset should be used only to evaluate the performance of the final model. It is imperative to confirm that patients

in the training dataset are excluded from the test dataset. For example, if a patient undergoes surgery twice, the test dataset randomized to surgery may contain the same patient's data in the training dataset.

Performance metrics

Accuracy is an inappropriate metric to use when evaluating the performance of a model on an imbalanced dataset, where the frequency of events is significantly lower than that of nonevents. For example, if the event frequency of the test dataset is 1%, even the all-negative predictor has an accuracy of 99%. This problem is important when dealing with rare complications, such as postoperative mortality or organ failure. The F1-score, which is the harmonic average of the precision and recall, or the balanced accuracy, which is the arithmetic mean of the recall for each class, might be better indicators in this case.

If the label is a binary variable and the output of the model is the risk of an event expressed as a continuous variable, the area under the receiver operating characteristic (AUROC) curve can be used as a performance metric. However, in severely imbalanced data, the AUROC easily becomes a high value; therefore, the area under the precision-recall curve (AUPRC) may be a better choice. Survival analysis performance is usually evaluated using the Harrell's concordance index (C-index).

In the reinforcement learning model, performance is evaluated by comparing the expected reward between the AI-suggested policy and the clinician's policy by replaying the clinically-defined trajectory.

AI models for perioperative risk stratification

Accurate perioperative risk stratification is important for facilitating shared decision-making and the allocation of medical resources. Several preoperative risk scores have been developed and used in clinical practice, including the American Society of Anesthesiologists Physical Status (ASA-PS) classification [11], American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) surgical risk calculator [25], surgical Apgar score [26], and Risk Stratification Index [27].

However, recent studies have shown that AI models for perioperative risk stratification have excellent performance (Table 1) in evaluating the risk of postoperative complications.

Prediction of mortality risk

Since its proposal in 1963, the ASA-PS classification for preop-

erative risk assessment has been used for most surgical patients. However, the limitations of the ASA-PS classification include the subjective nature of the clinicians' evaluations and high inter-rater variability [28].

Lee et al. [29] developed a DL model to predict postoperative in-hospital mortality using features extracted at the end of surgery from the data of 59,985 patients. This model used 45 intraoperative features and the ASA-PS classification to achieve an AUROC of 0.91, which is comparable to that with existing methods, such as logistic regression (LR).

Hill et al. [30] developed a fully automated score to predict postoperative in-hospital mortality using RF from the EHRs of 53,097 surgical patients. This model consisted of 58 preoperative variables that were automatically obtained from the EHR. The AUROC of the model (0.93) was larger than that of the existing risk scores (the ASA-PS, PreOperative Score to predict PostOperative Mortality [POSPOM], and Charlson Comorbidity Index scores).

Fritz et al. [31] constructed a CNN model to predict postoperative 30-day mortality from the data of 95,907 patients who underwent surgery under general anesthesia with tracheal intubation. The model consisted of 54 preoperative parameters, including patient characteristics, comorbidities, laboratory values, and 28 intraoperative variables. Its performance was compared with that of other algorithms, such as DL, RF, support vector machine (SVM), and LR. According to the results of the study, the CNN model had the best performance when using time-series data.

Chiew et al. [32] tested various machine learning algorithms to predict postoperative 30-day mortality and intensive care unit (ICU) stay > 24 h using data from 90,785 patients who underwent non-cardiac and non-neurological surgeries. GBM outperformed all other machine learning algorithms, with an AUPRC of 0.23.

Bertsimas et al. [33] developed a surgical risk calculator to predict postoperative 30-day mortality and 18 postoperative complications using optical classification trees from the data of 382,960 emergency surgery patients in the ACS-NSQIP database. The AUROC of the model for mortality was 0.92. The predictive performance of this calculator was tested in other populations, such as patients aged > 65 years [34] and patients undergoing emergency general surgery and laparotomy [35], and its performance in predicting mortality remained stable.

Lee et al. [36] developed an interpretable neural network to predict postoperative in-hospital mortality from the data of 59,985 surgical patients using generalized additive models (GAMs) with neural networks. The model had an AUROC of 0.92. To improve the interpretability and transparency of the pre-

Table 1. AI-based Perioperative Risk Stratification Models

Author	Year	Outcome variable	AUC	Population
Wu [60]	2016	Postoperative nausea and vomiting	0.93	Single center
Lee [29]	2018	Postoperative in-hospital mortality	0.91	Single center
Lee [52]	2018	AKI after cardiac surgery	0.78	Single center
Lee [53]	2018	AKI after liver transplantation	0.86	Single center
Bertsimas [33]	2018	Postoperative 30-day mortality & morbidity (POTTER)	0.84–0.92	Multi-center
Chen [59]	2018	Postoperative bleeding	0.82	Single center
Fritz [31]	2019	Postoperative 30-day mortality	0.87	Single center
Bihorac [42]	2019	Mortality; AKI; sepsis; VTE; ICU > 48 h; MV > 48 h; & wound, neurologic, cardiovascular complication (MySurgeryRisk)	0.77–0.94	Single center
Lei [55]	2019	AKI after major non-cardiac surgery	0.82	Single center
Hill [30]	2019	Postoperative in-hospital mortality	0.93	Single center
Adhikari [54]	2019	Postoperative AKI	0.86	Single center
Bolourani [48]	2020	Postoperative respiratory failure	NA	Multi-center
Tseng [56]	2020	AKI after cardiac surgery	0.78–0.84	Single center
Rank [10]	2020	AKI after cardiothoracic surgery	0.89	Single center
Hofer [57]	2020	Postoperative mortality, AKI, and reintubation	0.79–0.91	Single center
Mathis [43]	2020	Postoperative heart failure	0.87	Single center
Chiew [32]	2020	Postoperative 30-day mortality & ICU admission	0.96	Single center
Chen [49]	2021	Pneumonia after liver transplantation	0.73	Single center
Xue [58]	2021	Postoperative pneumonia, AKI, DVT, PE, delirium	0.76–0.91	Single center
Lee [36]	2021	Postoperative in-hospital mortality	0.92	Single center

All of these studies are retrospective. AUC: area under curve, AKI: acute kidney injury, POTTER: machine-learning-based Predictive Optimal Trees in Emergency Surgery Risk, VTE: venous thromboembolism, ICU: intensive care unit, MV: mechanical ventilation, DVT: deep vein thrombosis, PE: pulmonary embolism, NA: not applicable.

diction model, feature contributions were visualized, enabling clinicians to better understand the model's prediction process.

Prediction of cardiac risk

The most widely used classical model in this field is the Revised Cardiac Risk Index (RCRI), which has been incorporated into the guidelines of the American College of Cardiology/American Heart Association and European Society of Cardiology/European Society of Anesthesiology [37–39]. The RCRI, which was reported in 1999 by Lee et al. [37], consists of six variables: high-risk surgery, history of congestive heart failure, history of ischemic heart disease, history of cerebrovascular disease, preoperative serum creatinine > 2.0 mg/dl, and preoperative insulin treatment. However, a recent large-scale retrospective validation study using the Danish National Patient Registry revealed that the estimated odds ratio for each variable in the RCRI varies between 1.45 for serum creatinine and 10.02 for a history of cerebrovascular disease [40]. Additionally, in a systematic review of 24 studies (792,740 patients), the RCRI showed only modest performance (AUROC = 0.75) [41].

Bihorac et al. [42] developed a machine learning model called

MySurgeryRisk for predicting eight major postoperative complications: mortality; acute kidney injury (AKI); sepsis; venous thromboembolism; ICU stay > 48 h; mechanical ventilation > 48 h; and wound, neurologic, and cardiovascular complications using the GAM. The model had an AUROC of 0.85 for predicting cardiovascular complications.

Mathis et al. [43] developed a GBM model to predict heart failure after non-cardiac surgery. Using 499 preoperative and 263 intraoperative data points from 67,697 patients, the AUROC of the model was 0.87.

Prediction of pulmonary risk

Postoperative pulmonary complications frequently develop after major surgery, and even a mild form of these complications is associated with a prolonged hospital stay and an increased mortality rate [44,45]. The most commonly used classical model in this field is the Assess Respiratory Risk in Surgical Patients in Catalonia (ARISCAT) score. Canet et al. [46] developed the ARISCAT score to predict postoperative pulmonary complications in surgical patients using LR, which is the only risk score that maintains discriminatory power for external validation [47]. In a pro-

spective validation study involving 5,859 patients, the ARISCAT score achieved an AUROC of 0.80 [47].

Additionally, Bolourani et al. [48] developed a machine learning model to predict postoperative respiratory failure in 4,062 patients who underwent pulmonary lobectomy. Although the sensitivity and specificity were 83.3% and 94.5%, respectively, the AUROC was not provided.

Chen et al. [49] investigated various machine-learning algorithms, including LR, SVM, RF, adaptive boosting, and GBM, to predict pneumonia after orthotopic liver transplantation in 786 patients. Fourteen features, which included laboratory and clinical variables, were associated with postoperative pneumonia in this study.

Prediction of AKI

AKI is associated with increased morbidity, length of hospital stay, and mortality [50]. Although there are various criteria and time frames for diagnosing AKI, a recent consensus statement recommended that postoperative AKI be defined according to the Kidney Disease: Improving Global Outcomes (KDIGO) criteria, which defines it as an AKI that occurs within the first seven days after surgery [51].

Lee et al. [52] compared the performance of various machine-learning algorithms (e.g., decision tree, RF, GBM, SVM, and DL) for predicting postoperative AKI using LR in 2,010 patients undergoing cardiac surgery. GBM showed the highest AUROC of 0.78 and the lowest error rate of 26%. This group also conducted another study using preoperative, intraoperative, and surgery-related variables to test the performance of various algorithms on AKI prediction using data from 2,911 patients undergoing liver transplantation. GBM also showed the best performance, with an AUROC of 0.90 [53].

Adhikari et al. [54] developed a machine learning model containing intraoperative time-series variables to predict postoperative AKI in 2,911 surgical patients. Compared to the model using only preoperative variables, the model that included both preoperative variables and intraoperative time-series data showed better predictive performance, with an AUROC of 0.86.

Lei et al. [55] investigated whether combining preoperative and intraoperative data could improve the prediction of postoperative AKI in 42,915 patients undergoing major noncardiac surgery. The GBM algorithm outperformed LR with elastic net selection and RF. The authors found that adding intraoperative data slightly improved the predictive performance.

Tseng et al. [56] used a machine learning model to predict postoperative AKI in cardiac surgical patients using preoperative and

intraoperative time-series hemodynamic variables. The combination of RF and GBM showed the highest AUROC (0.84). They used the SHapley Additive exPlanation method to explain how the model's predictions were made, thus improving the interpretability of the model.

Rank et al. [10] used an RNN and 96 routinely collected variables to develop a DL model for the real-time prediction of postoperative AKI (stage 2 or 3 based on the KDIGO criteria) in patients undergoing cardiothoracic surgery until discharge from the ICU or post-anesthesia care unit. They compared the predictive performance of the DL model with that of experienced clinicians and found that the DL model outperformed experienced clinicians.

Prediction of other complications

Hofer et al. [57] developed a DL model to predict multiple postoperative complications, including AKI, reintubation, and mortality, using a single-input feature set available at the end of surgery. Its performance was compared with that of the ASA-PS classification. The AUROCs of the models were 0.79, 0.88, 0.91, and 0.87 for AKI, reintubation, mortality, and the composite outcome, respectively.

Xue et al. [58] tested five machine learning algorithms (LR, SVM, RF, GBM, and DL) to predict five postoperative complications (pneumonia, AKI, deep vein thrombosis, pulmonary embolism, and delirium) using preoperative and intraoperative data. The best-performing model for each complication showed the following AUROCs: pneumonia (GBM), 0.91; AKI (GBM), 0.85; deep vein thrombosis (GBM), 0.88; pulmonary embolism (MLP), 0.83; and delirium (GBM), 0.76. A model for predicting other postoperative complications was also developed. Chen et al. [59] developed an AI model using 299 perioperative variables to predict bleeding after colorectal surgery. The GBM model had an AUROC of 0.82, which was higher than that of the LR model (AUROC = 0.74).

Wu et al. [60] developed an SVM model to predict postoperative nausea in orthopedic surgery patients receiving patient-controlled epidural analgesia. Their model showed an AUROC of 0.93, which was higher than that of the LR model (AUROC = 0.73).

AI models for intraoperative event prediction and biosignal analysis

Anesthesiologists play a pivotal role in monitoring and maintaining hemodynamic stability during surgery, which can affect

postoperative clinical outcomes [61]. Several AI algorithms that can assist anesthesiologists in intraoperative management by calculating secondary indices using real-time data have been published (Table 2). These algorithms can help anesthesiologists improve intraoperative management by predicting future events, such as intraoperative hypotension [62–65] or desaturation [8], and by processing biosignals [10,66].

Prediction of intraoperative hypotension

Kendale et al. [63] developed a machine learning model to predict post-induction hypotension, which was defined as a mean arterial pressure (MAP) < 55 mmHg within 10 min of anesthesia induction, in 13,323 surgical patients. GBM outperformed the other algorithms (e.g., LR, SVM, naive Bayes, k-nearest neighbor, linear discriminant analysis, RF, neural nets, and GBM).

Kang et al. [62] investigated four machine learning techniques (LR, naive Bayes, RF, and artificial neural network) to predict post-induction hypotension, which was defined as a systolic blood pressure < 90 mmHg or MAP < 65 mmHg occurring between tracheal intubation and surgical incision. RF showed the highest AUROC. The patients' lowest systolic blood pressure, lowest MAP, and mean SBP before tracheal intubation were the most important features in terms of prediction accuracy.

Hatib et al. [67] developed a machine learning model to predict upcoming hypotensive events (MAP < 65 mmHg) using features from arterial pulse waveforms in 1,334 patients, and the Hypotension Prediction Index (HPI) was externally validated in 204 patients from a prospectively collected cohort. The AUROC of the prediction 5–15 min before a hypotensive event was 0.95–0.97. The HPI was tested in 255 patients undergoing major surgery, and it predicted hypotension 5–15 min before a hypotensive event with an AUROC of 0.879–0.926 [68].

Wijnberge et al. [69] conducted a single-center randomized controlled study of 68 patients undergoing elective non-cardiac

surgery under general anesthesia to evaluate the effect of early prediction of hypotension using the HPI on the number of hypotensive events. The primary outcome of the study was time-weighted average hypotension (MAP < 65 mmHg). Patients were randomly assigned to two groups: those who received standard care and those who received an early warning when their HPI value exceeded 85. In this study, the HPI-guided early warning system significantly reduced intraoperative hypotension. However, Maheshwari et al. [70] failed to report the benefits of HPI-guided management during moderate- to high-risk non-cardiac surgery. In this study, the patients were randomly assigned to either the HPI-guided or HPI-unguided group. In the HPI-guided group, if the HPI exceeded 85, clinicians received electronic alerts and treatment algorithms, such as fluid administration, inotrope or vasopressor administration, or observation, which they could choose to follow or not. However, approximately half of the clinicians who received electronic alerts did not follow the recommended treatment algorithm, indicating the need for a simpler treatment algorithm and lower alert threshold.

Lee et al. [65] developed DL algorithms for real-time predictions 5–15 min before the occurrence of a hypotensive event based on biosignals collected using routine invasive and noninvasive intraoperative monitoring of 3,301 patients from the VitalDB database [71]. Using an arterial pressure waveform, electrocardiography, photoplethysmography, and capnography, the multi-channel DL model predicted hypotensive events 15 min before the occurrence of an actual hypotensive event, with an AUROC of 0.90.

Prediction of intraoperative hypoxemia

Lundberg et al. [8] developed an AI model to predict intraoperative hypoxemia, which is defined as an oxygen saturation \leq 92% within 5 min. They extracted more than 20 preoperative and 45 intraoperative features at 1-min intervals from the data of 53,126

Table 2. AI-based Intraoperative Event Prediction Models

Author	Year	Outcome variable	AUC	Population	Design
Lundberg [8]	2018	Intraoperative hypoxemia (Prescience)	0.83	Single center	Retrospective
Kendale [63]	2018	Postinduction hypotension	0.74	Single center	Retrospective
Hatib [67]	2018	Intraoperative hypotension (HPI)	0.95–0.97	Multi-center	Retrospective
Solomon [64]	2020	Intraoperative bradycardia associated with hypotension	0.89	Single center	Retrospective
Kang [62]	2020	Postinduction hypotension	0.84	Single center	Retrospective
Wijnberge [69]	2020	HPI vs. conventional	NA	Single center	RCT
Maheshwari [70]	2020	HPI vs. conventional	NA	Single center	RCT
Lee [65]	2021	Intraoperative hypotension	0.90	Single center	Retrospective

AUC: area under curve, HPI: hypotension prediction index, N/A: not applicable, RCT: randomized controlled trial.

surgical patients on EHRs and trained the GBM model. The AI model had a significantly higher AUROC than the anesthesiologists (0.81 vs. 0.66).

Model for the electroencephalography analysis

Saadeh et al. [66] developed a machine learning model based on several features from electroencephalography (EEG) signals to estimate the depth of anesthesia, irrespective of age and type of anesthetic drugs. Their model showed an average accuracy of 92% for all stages of anesthesia.

Park et al. [10] developed a DL model to perform real-time estimation of the depth of anesthesia. This model, which combined an EEG-based depth of anesthesia monitoring system with DL, had a stronger correlation with the minimum alveolar concentration than the bispectral index.

Model for anesthetic titration

Underdosing anesthetic agents can cause intraoperative awareness during general anesthesia, whereas overdosing can cause complications, such as hypotension, delayed recovery, and delirium after surgery. The target-controlled infusion (TCI) algorithm, developed by Shafer and Greg [72] in 1992, has been used to titrate fast-acting anesthetics and opioids. However, because the TCI algorithm relies on population pharmacokinetic and pharmacodynamic models that can have high inter-individual variability, control methods using EEG-based anesthesia depth monitors, such as the bispectral index have been proposed [73].

However, AI models for achieving and maintaining an appropriate depth of anesthesia without complications can be devel-

oped. Reinforcement learning algorithms for propofol titration based on Q-learning and PPO-based algorithms have been reported [74,75]. In a recent study, a deep reinforcement learning model using the A2C algorithm outperformed a proportional-integral derivative controller [76].

AI models for patients in the ICU

Several studies using AI techniques for ICU patient management have been published (Table 3). These algorithms are aimed at predicting ward or ICU complications and helping clinicians respond early. The Medical Information Mart for Intensive Care (MIMIC) database, a single-center open dataset from the Beth Israel Deaconess Medical Center (Boston, MA, USA) [77], and the eICU, a multicenter open dataset in the USA [78], are frequently used in studies in the field for model development and validation.

Prediction of mortality in the ICU

To stratify acutely ill patients and evaluate the effects of therapy, the Acute Physiology, Age, Chronic Health Evaluation (APACHE) II was developed in 1985, using 12 variables [79]. The APACHE III, a revised version of the APACHE II system using different variables, was developed from 17,440 patients in the ICU of 40 US hospitals [80]. The authors of the original study reported that the APACHE III predicted in-hospital mortality with an AUROC of 0.90. However, in a multicenter prospective study involving 1,144 patients, the AUROCs of the APACHE II and APACHE III were 0.806 and 0.847, respectively [81].

Delahanty et al. [82] developed a risk-adjustment algorithm for in-hospital mortality in 237,173 patients in 131 ICUs across 53

Table 3. AI-based Prediction Models for Intensive Care Unit Patients

Author	Year	Outcome variable	AUC	Population
Delahanty [82]	2018	ICU mortality (RIPD)	0.94	Multi-center
Rojas [84]	2018	ICU readmission	0.73	Multi-center
Mao [88]	2018	Sepsis in ICU (InSight)	0.92	Multi-center
Nemati [87]	2018	Sepsis in ICU (AISE)	0.83–0.85	Multi-center
Giannini [89]	2019	Sepsis in ICU	0.88	Multi-center
Scherpf [90]	2019	Sepsis in ICU	0.81	Single center
Kong [93]	2020	Mortality in patient with sepsis	0.83–0.85	Single center
Burdick [94]	2020	Severe sepsis and septic shock	0.83–0.93	Multi-center
He [91]	2020	Sepsis in ICU	NA	Multi-center
Hur [85]	2021	Delirium in ICU (PRIDE)	0.92	Multi-center
Goh [92]	2021	Sepsis in ICU (SERA)	0.94	Single center

All of these studies are retrospective. AUC: area under curve, ICU: intensive care unit, RIPD: risk of inpatient death, AISE: Artificial Intelligence Sepsis Expert, NA: not applicable, PRIDE: Prediction of Intensive Care Unit Delirium, SERA: sepsis early risk assessment.

hospitals. This model, which used the GBM algorithm and had 17 features, including clinical and administrative data, showed excellent discrimination, with an AUROC of 0.94.

Baker et al. [83] developed a continuous mortality risk prediction model for ICU mortality using a hybrid neural network approach that combined a CNN and bidirectional LSTM. Using the MIMIC III database, the authors predicted in-hospital mortality within 3, 7, and 14 days using vital signs over a 24 h period. This model achieved the highest AUROC of 0.88.

Prediction of ICU readmission

Rojas et al. [84] developed a prediction model for ICU readmission using data from 24,885 patients, and validated the model's performance using MIMIC data. Their GBM model showed an AUROC of 0.71, which was significantly better than the Stability and Workload Index for Transfer score (SWIFT; AUROC = 0.58) and the Modified Early Warning Score (MEWS; AUROC = 0.57).

Prediction of delirium in the ICU

Hur et al. [85] developed a model called the Prediction of Intensive Care Unit Delirium to predict the risk of delirium in patients in the ICU. This RF model used 59 variables extracted from 37,543 patients and had an AUROC of 0.72.

Jauk et al. [86] prospectively verified the performance of an RF-based delirium prediction model over seven months in internal medicine patients. The retrospective performance of this model had an AUROC of 0.91, whereas the prospective validation performance had an AUROC of 0.86.

Prediction and management of sepsis in the ICU

Several GBM-, SVN-, and DL-based prediction models for sepsis in ICU patients have been developed [87–92]. These models have been found to predict sepsis 3–12 h before onset, with an AUROC of 0.81–0.92. They showed better performance than conventional scoring systems, such as the systemic inflammatory response syndrome (SIRS), Sequential Organ Failure Assessment (SOFA), quick SOFA, and MEWS. Implementing these algorithms into EHR systems could help clinicians respond to high-risk patients early.

Kong et al. [93] evaluated various machine-learning algorithms for predicting in-hospital mortality in 16,688 ICU patients with sepsis from the MIMIC III database. Among the tested algorithms (least absolute shrinkage and selection operator, RF, GBM, and traditional LR), GBM showed the highest AUROC of 0.845.

Burdick et al. [94] created a machine learning prediction model using GBM for severe sepsis and septic shock 48 h before the onset of events from the data of 270,438 patients. The predictive performances had AUROCs of 0.83 and 0.75 for the internal test and external validation datasets, respectively. The model showed superior performance to previous prediction models, such as the SIRS, SOFA, and MEWS.

Raghu et al. [95] developed a reinforcement learning model that suggests a policy for sepsis treatment using a Q-learning algorithm. The model proposes the optimal volume of intravenous fluids and dosage of vasopressors that should be administered to improve the SOFA score and lactate concentrations. A retrospective validation of the model showed that their algorithm was expected to reduce mortality by up to 3.6%.

Models for ventilator control

Mechanical ventilation is one of the most common treatments in the ICU [96]. Although there is evidence that a lung-protective ventilatory strategy improves survival in patients with acute respiratory distress syndrome (ARDS), the optimal ventilatory strategy for patients without ARDS remains unknown [97,98]. In a recent study of AI-based ventilator control algorithms, the Q-learning-based reinforcement learning model outperformed the clinician's policy in terms of rewards defined as in-hospital or 90-day mortality [9].

Conclusion

Although the above-described AI-based predictive models showed high predictive performance in various perioperative settings, most of the results were obtained from single-center retrospective studies. Before applying AI models in clinical practice, additional external and prospective validation and randomized clinical trials are required. Reinforcement learning models may suggest optimal strategies to overcome inter-individual variability; however, their clinical utility must be verified. If the high performance of AI algorithms is well maintained in future studies, they can be widely used in clinical practice as a powerful tool to help clinicians improve patient safety and outcomes.

Funding

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare (HI21C1074) and the National Research Foundation of

Korea (NRF) grant, funded by the Ministry of Science and ICT, Republic of Korea (NRF-2020R1C1C1014905).

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

Author Contributions

Hyun-Kyu Yoon (Writing – original draft)

Hyun-Lim Yang (Writing – review & editing)

Chul-Woo Jung (Writing – review & editing)

Hyung-Chul Lee (Writing – original draft)

ORCID

Hyun-Kyu Yoon, <https://orcid.org/0000-0001-5424-3559>

Hyun-Lim Yang, <https://orcid.org/0000-0003-2221-3042>

Chul-Woo Jung, <https://orcid.org/0000-0001-7876-8659>

Hyung-Chul Lee, <https://orcid.org/0000-0003-0048-7958>

References

1. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019; 25: 44-56.
2. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial intelligence in radiology. *Nat Rev Cancer* 2018; 18: 500-10.
3. Salto-Tellez M, Maxwell P, Hamilton P. Artificial intelligence-the third revolution in pathology. *Histopathology* 2019; 74: 372-6.
4. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, et al. Artificial intelligence in cardiology. *J Am Coll Cardiol* 2018; 71: 2668-79.
5. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg* 2018; 268: 70-6.
6. Hashimoto DA, Witkowski E, Gao L, Meireles O, Rosman G. Artificial intelligence in anesthesiology: current techniques, clinical applications, and limitations. *Anesthesiology* 2020; 132: 379-94.
7. Chae D. Data science and machine learning in anesthesiology. *Korean J Anesthesiol* 2020; 73: 285-95.
8. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng* 2018; 2: 749-60.
9. Peine A, Hallawa A, Bickenbach J, Dartmann G, Fazlic LB, Schmeink A, et al. Development and validation of a reinforcement learning algorithm to dynamically optimize mechanical ventilation in critical care. *NPJ Digit Med* 2021; 4: 32.
10. Rank N, Pfahringer B, Kempfert J, Stamm C, Kühne T, Schoenrath F, et al. Deep-learning-based real-time prediction of acute kidney injury outperforms human predictive performance. *NPJ Digit Med* 2020; 3: 139.
11. Yoon D, Jang JH, Choi BJ, Kim TY, Han CH. Discovering hidden information in biosignals from patients using artificial intelligence. *Korean J Anesthesiol* 2020; 73: 275-84.
12. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv e-prints 2014 [Internet]. New York: Cornell University; 2014 Sep 1 [updated 2016 May 19; cited 2022 Mar 8]. Available from <https://arxiv.org/abs/1409.0473>.
13. Luong MT, Pham H, Manning CD. Effective approaches to attention-based neural machine translation. arXiv e-prints 2015 [Internet]. New York: Cornell University; 2015 Aug 17 [updated 2015 Sep 20; cited 2022 Mar 8]. Available from <https://arxiv.org/abs/1508.04025>.
14. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. arXiv e-prints 2017 [Internet]. New York: Cornell University; 2017 Jun 12 [updated 2017 Dec 6; cited 2022 Mar 8]. Available from <https://arxiv.org/abs/1706.03762>.
15. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. arXiv e-prints 2020 [Internet]. New York: Cornell University; 2020 Oct 22 [updated 2021 Jun 3; cited 2022 Mar 8]. Available from <https://arxiv.org/abs/2010.11929>.
16. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015; 61: 85-117.
17. Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, et al. Playing Atari with Deep Reinforcement Learning. arXiv e-prints 2013 [Internet]. New York: Cornell University; 2013 Dec 19 [cited 2022 Mar 8]. Available from <https://arxiv.org/abs/1312.5602>.
18. Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. arXiv e-prints 2017 [Internet]. New York: Cornell University; 2017 Jul 20 [updated 2017 Aug 28; cited 2022 Mar 8]. Available from <https://arxiv.org/abs/1707.06347>.
19. James B, Yoshua B. Random search for hyper-parameter optimization. *J Mach Learn Res* 2012; 13: 281-305.
20. Jasper S, Hugo L, Ryan AP. Practical Bayesian optimization of

- machine learning algorithms. arXiv e-prints 2012. Available from <https://arxiv.org/abs/1206.2944>.
21. Griffith GJ, Morris TT, Tudball MJ, Herbert A, Mancano G, Pike L, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat Commun* 2020; 11: 5749.
 22. Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: A new perspective. *Neurocomputing* 2018; 300: 70-9.
 23. Kursu MB, Rudnicki WR. Feature selection with the Boruta package. *J Stat Softw* 2010; 36: 1-13.
 24. Basu I, Maji S. Multicollinearity correction and combined feature effect in Shapley values. arXiv e-prints 2020 [Internet]. New York: Cornell University; 2020 Nov 3 [cited 2022 Mar 8]. Available from <https://arxiv.org/abs/2011.01661>.
 25. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg* 2013; 217: 833-42.e1-3.
 26. Gawande AA, Kwaan MR, Regenbogen SE, Lipsitz SA, Zinner MJ. An Apgar score for surgery. *J Am Coll Surg* 2007; 204: 201-8.
 27. Sessler DI, Sigl JC, Manberg PJ, Kelley SD, Schubert A, Chamoun NG. Broadly applicable risk stratification system for predicting duration of hospitalization and mortality. *Anesthesiology* 2010; 113: 1026-37.
 28. De Cassai A, Boscolo A, Tonetti T, Ban I, Ori C. Assignment of ASA-physical status relates to anesthesiologists' experience: a survey-based national-study. *Korean J Anesthesiol* 2019; 72: 53-9.
 29. Lee CK, Hofer I, Gabel E, Baldi P, Cannesson M. Development and validation of a deep neural network model for prediction of postoperative in-hospital mortality. *Anesthesiology* 2018; 129: 649-62.
 30. Hill BL, Brown R, Gabel E, Rakocz N, Lee C, Cannesson M, et al. An automated machine learning-based model predicts postoperative mortality using readily-extractable preoperative electronic health record data. *Br J Anaesth* 2019; 123: 877-86.
 31. Fritz BA, Cui Z, Zhang M, He Y, Chen Y, Kronzer A, et al. Deep-learning model for predicting 30-day postoperative mortality. *Br J Anaesth* 2019; 123: 688-95.
 32. Chiew CJ, Liu N, Wong TH, Sim YE, Abdullah HR. Utilizing machine learning methods for preoperative prediction of post-surgical mortality and intensive care unit admission. *Ann Surg* 2020; 272: 1133-9.
 33. Bertsimas D, Dunn J, Velmahos GC, Kaafarani HM. Surgical risk is not linear: derivation and validation of a novel, user-friendly, and machine-learning-based predictive optimal trees in emergency surgery risk (POTTER) calculator. *Ann Surg* 2018; 268: 574-83.
 34. Maurer LR, Chetlur P, Zhuo D, El Hechi M, Velmahos GC, Dunn J, et al. Validation of the ai-based predictive optimal trees in emergency surgery risk (POTTER) calculator in patients 65 years and older. *Ann Surg* 2020. Advance Access published on Dec 23, 2020. doi: 10.1097/SLA.0000000000004714.
 35. El Hechi MW, Maurer LR, Levine J, Zhuo D, El Moheb M, Velmahos GC, et al. Validation of the artificial intelligence-based predictive optimal trees in emergency surgery risk (POTTER) calculator in emergency general surgery and emergency laparotomy patients. *J Am Coll Surg* 2021; 232: 912-9.e1.
 36. Lee CK, Samad M, Hofer I, Cannesson M, Baldi P. Development and validation of an interpretable neural network for prediction of postoperative in-hospital mortality. *NPJ Digit Med* 2021; 4: 8.
 37. Lee TH, Marcantonio ER, Mangione CM, Thomas EJ, Polanczyk CA, Cook EF, et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation* 1999; 100: 1043-9.
 38. Fleisher LA, Fleischmann KE, Auerbach AD, Barnason SA, Beckman JA, Bozkurt B, et al. 2014 ACC/AHA guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation* 2014; 130: 2215-45.
 39. Kristensen SD, Knuuti J, Saraste A, Anker S, Botker HE, Hert SD, et al. 2014 ESC/ESA Guidelines on non-cardiac surgery: cardiovascular assessment and management: The Joint Task Force on non-cardiac surgery: cardiovascular assessment and management of the European Society of Cardiology (ESC) and the European Society of Anaesthesiology (ESA). *Eur Heart J* 2014; 35: 2383-431.
 40. Andersson C, Wissenberg M, Jørgensen ME, Hlatky MA, Mérie C, Jensen PF, et al. Age-specific performance of the revised cardiac risk index for predicting cardiovascular risk in elective non-cardiac surgery. *Circ Cardiovasc Qual Outcomes* 2015; 8: 103-8.
 41. Ford MK, Beattie WS, Wijeyesundera DN. Systematic review: prediction of perioperative cardiac complications and mortality by the revised cardiac risk index. *Ann Intern Med* 2010; 152: 26-35.
 42. Bihorac A, Ozrazgat-Baslanti T, Ebadi A, Motaei A, Madkour M, Pardalos PM, et al. MySurgeryRisk: Development and validation of a machine-learning risk algorithm for major complications and death after surgery. *Ann Surg* 2019; 269: 652-62.
 43. Mathis MR, Engoren MC, Joo H, Maile MD, Aaronson KD,

- Burns ML, et al. Early detection of heart failure with reduced ejection fraction using perioperative data among noncardiac surgical patients: a machine-learning approach. *Anesth Analg* 2020; 130: 1188-200.
44. LAS VEGAS Investigators. Epidemiology, practice of ventilation and outcome for patients at increased risk of postoperative pulmonary complications: LAS VEGAS - an observational study in 29 countries. *Eur J Anaesthesiol* 2017; 34: 492-507.
 45. Serpa Neto A, Hemmes SN, Barbas CS, Beiderlinden M, Fernandez-Bustamante A, Futier E, et al. Incidence of mortality and morbidity related to postoperative lung injury in patients who have undergone abdominal or thoracic surgery: a systematic review and meta-analysis. *Lancet Respir Med* 2014; 2: 1007-15.
 46. Canet J, Gallart L, Gomar C, Paluzie G, Vallès J, Castillo J, et al. Prediction of Postoperative Pulmonary Complications in a Population-based Surgical Cohort. *Anesthesiology* 2010; 113: 1338-50.
 47. Mazo V, Sabaté S, Canet J, Gallart L, de Abreu MG, Belda J, et al. Prospective external validation of a predictive score for postoperative pulmonary complications. *Anesthesiology* 2014; 121: 219-31.
 48. Bolourani S, Wang P, Patel VM, Manetta F, Lee PC. Predicting respiratory failure after pulmonary lobectomy using machine learning techniques. *Surgery* 2020; 168: 743-52.
 49. Chen C, Yang D, Gao S, Zhang Y, Chen L, Wang B, et al. Development and performance assessment of novel machine learning models to predict pneumonia after liver transplantation. *Respir Res* 2021; 22: 94.
 50. Chertow GM, Burdick E, Honour M, Bonventre JV, Bates DW. Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. *J Am Soc Nephrol* 2005; 16: 3365-70.
 51. Prowle JR, Forni LG, Bell M, Chew MS, Edwards M, Grams ME, et al. Postoperative acute kidney injury in adult non-cardiac surgery: joint consensus report of the Acute Disease Quality Initiative and PeriOperative Quality Initiative. *Nat Rev Nephrol* 2021; 17: 605-18.
 52. Lee HC, Yoon HK, Nam K, Cho YJ, Kim TK, Kim WH, et al. Derivation and validation of machine learning approaches to predict acute kidney injury after cardiac surgery. *J Clin Med* 2018; 7: 322.
 53. Lee HC, Yoon SB, Yang SM, Kim WH, Ryu HG, Jung CW, et al. Prediction of acute kidney injury after liver transplantation: machine learning approaches vs. logistic regression model. *J Clin Med* 2018; 7: 428.
 54. Adhikari L, Ozrazgat-Baslanti T, Ruppert M, Madushani RW, Paliwal S, Hashemighouchani H, et al. Improved predictive models for acute kidney injury with IDEA: intraoperative data embedded analytics. *PLoS One* 2019; 14: e0214904.
 55. Lei VJ, Luong T, Shan E, Chen X, Neuman MD, Eneanya ND, et al. Risk stratification for postoperative acute kidney injury in major noncardiac surgery using preoperative and intraoperative data. *JAMA Netw Open* 2019; 2: e1916921.
 56. Tseng PY, Chen YT, Wang CH, Chiu KM, Peng YS, Hsu SP, et al. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care* 2020; 24: 478.
 57. Hofer IS, Lee C, Gabel E, Baldi P, Cannesson M. Development and validation of a deep neural network model to predict postoperative mortality, acute kidney injury, and reintubation using a single feature set. *NPJ Digit Med* 2020; 3: 58.
 58. Xue B, Li D, Lu C, King CR, Wildes T, Avidan MS, et al. Use of Machine Learning to Develop and Evaluate Models Using Preoperative and Intraoperative Data to Identify Risks of Postoperative Complications. *JAMA Netw Open* 2021; 4: e212240.
 59. Chen D, Afzal N, Sohn S, Habermann EB, Naessens JM, Larson DW, et al. Postoperative bleeding risk prediction for patients undergoing colorectal surgery. *Surgery* 2018; 164: 1209-16.
 60. Wu HY, Gong CA, Lin SP, Chang KY, Tsou MY, Ting CK. Predicting postoperative vomiting among orthopedic patients receiving patient-controlled epidural analgesia using SVM and LR. *Sci Rep* 2016; 6: 27041.
 61. Sun LY, Wijeyesundera DN, Tait GA, Beattie WS. Association of intraoperative hypotension with acute kidney injury after elective noncardiac surgery. *Anesthesiology* 2015; 123: 515-23.
 62. Kang AR, Lee J, Jung W, Lee M, Park SY, Woo J, et al. Development of a prediction model for hypotension after induction of anesthesia using machine learning. *PLoS One* 2020; 15: e0231172.
 63. Kendale S, Kulkarni P, Rosenberg AD, Wang J. Supervised machine-learning predictive analytics for prediction of postinduction hypotension. *Anesthesiology* 2018; 129: 675-88.
 64. Solomon SC, Saxena RC, Neradilek MB, Hau V, Fong CT, Lang JD, et al. Forecasting a crisis: machine-learning models predict occurrence of intraoperative bradycardia associated with hypotension. *Anesth Analg* 2020; 130: 1201-10.
 65. Lee S, Lee HC, Chu YS, Song SW, Ahn GJ, Lee H, et al. Deep learning models for the prediction of intraoperative hypotension. *Br J Anaesth* 2021; 126: 808-17.
 66. Saadeh W, Khan FH, Altaf MA. Design and implementation of a machine learning based EEG processor for accurate estimation of depth of anesthesia. *IEEE Trans Biomed Circuits Syst* 2019; 13: 658-69.
 67. Hatib F, Jian Z, Buddi S, Lee C, Settels J, Sibert K, et al. Machine-learning algorithm to predict hypotension based on high-fidelity arterial pressure waveform analysis. *Anesthesiology* 2018; 129: 663-74.

68. Davies SJ, Vistisen ST, Jian Z, Hatib F, Scheeren TW. Ability of an arterial waveform analysis-derived hypotension prediction index to predict future hypotensive events in surgical patients. *Anesth Analg* 2020; 130: 352-9.
69. Wijnberge M, Geerts BF, Hol L, Lemmers N, Mulder MP, Berge P, et al. Effect of a machine learning-derived early warning system for intraoperative hypotension vs standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery: the HYPE randomized clinical trial. *JAMA* 2020; 323: 1052-60.
70. Maheshwari K, Shimada T, Yang D, Khanna S, Cywinski JB, Irefin SA, et al. Hypotension prediction index for prevention of hypotension during moderate- to high-risk noncardiac surgery. *Anesthesiology* 2020; 133: 1214-22.
71. Lee HC, Jung CW. Vital Recorder-a free research tool for automatic recording of high-resolution time-synchronised physiological data from multiple anaesthesia devices. *Sci Rep* 2018; 8: 1527.
72. Shafer SL, Gregg KM. Algorithms to rapidly achieve and maintain stable drug concentrations at the site of drug effect with a computer-controlled infusion pump. *J Pharmacokinet Biopharm* 1992; 20: 147-69.
73. American Society of Anesthesiologists Task Force on Intraoperative Awareness. Practice advisory for intraoperative awareness and brain function monitoring: a report by the American Society of Anesthesiologists Task Force on Intraoperative Awareness. *Anesthesiology* 2006; 104: 847-64.
74. Moore BL, Pyeatt LD, Kulkarni V, Panousis P, Padrez K, Doufas AG. Reinforcement learning for closed-loop propofol anesthesia: a study in human volunteers. *J Mach Learn Res* 2014; 15: 655-96.
75. Padmanabhan R, Meskin N, Haddad WM. Closed-loop control of anesthesia and mean arterial pressure using reinforcement learning. *Biomed Signal Process Control* 2015; 22: 54-64.
76. Schamberg G, Badgeley M, Meschede-Krasa B, Kwon O, Brown EN. Continuous action deep reinforcement learning for propofol dosing during general anesthesia. *Artif Intell Med* 2022; 123: 102227.
77. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016; 3: 160035.
78. Pollard TJ, Johnson AE, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci Data* 2018; 5: 180178.
79. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13: 818-29.
80. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991; 100: 1619-36.
81. Beck DH, Taylor BL, Millar B, Smith GB. Prediction of outcome from intensive care: a prospective cohort study comparing Acute Physiology and Chronic Health Evaluation II and III prognostic systems in a United Kingdom intensive care unit. *Crit Care Med* 1997; 25: 9-15.
82. Delahanty RJ, Kaufman D, Jones SS. Development and evaluation of an automated machine learning algorithm for in-hospital mortality risk adjustment among critical care patients. *Crit Care Med* 2018; 46: e481-8.
83. Baker S, Xiang W, Atkinson I. Continuous and automatic mortality risk prediction using vital signs in the intensive care unit: a hybrid neural network approach. *Sci Rep* 2020; 10: 21282.
84. Rojas JC, Carey KA, Edelson DP, Venable LR, Howell MD, Churpek MM. Predicting intensive care unit readmission with machine learning using electronic health record data. *Ann Am Thorac Soc* 2018; 15: 846-53.
85. Hur S, Ko RE, Yoo J, Ha J, Cha WC, Chung CR. A machine learning-based algorithm for the prediction of intensive care unit delirium (PRIDE): retrospective study. *JMIR Med Inform* 2021; 9: e23401.
86. Jauk S, Kramer D, Großauer B, Rienmüller S, Avian A, Berghold A, et al. Risk prediction of delirium in hospitalized patients using machine learning: An implementation and prospective evaluation study. *J Am Med Inform Assoc* 2020; 27: 1383-92.
87. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit Care Med* 2018; 46: 547-53.
88. Mao Q, Jay M, Hoffman JL, Calvert J, Barton C, Shimabukuro D, et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. *BMJ Open* 2018; 8: e017833.
89. Giannini HM, Ginestra JC, Chivers C, Draugelis M, Hanish A, Schweickert WD, et al. A machine learning algorithm to predict severe sepsis and septic shock: development, implementation, and impact on clinical practice. *Crit Care Med* 2019; 47: 1485-92.
90. Scherpf M, Gräßer F, Malberg H, Zaunseder S. Predicting sepsis with a recurrent neural network using the MIMIC III database. *Comput Biol Med* 2019; 113: 103395.
91. He Z, Du L, Zhang P, Zhao R, Chen X, Fang Z. Early sepsis prediction using ensemble learning with deep features and artificial features extracted from clinical electronic health records. *Crit*

- Care Med 2020; 48: e1337-42.
92. Goh KH, Wang L, Yeow AY, Poh H, Li K, Yeow JJ, et al. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nat Commun* 2021; 12: 711.
 93. Kong G, Lin K, Hu Y. Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU. *BMC Med Inform Decis Mak* 2020; 20: 251.
 94. Burdick H, Pino E, Gabel-Comeau D, Gu C, Roberts J, Le S, et al. Validation of a machine learning algorithm for early severe sepsis prediction: a retrospective study predicting severe sepsis up to 48 h in advance using a diverse dataset from 461 US hospitals. *BMC Med Inform Decis Mak* 2020; 20: 276.
 95. Raghu A, Komorowski M, Celi LA, Szolovits P, Ghassemi M. Continuous state-space models for optimal sepsis Treatment - a deep reinforcement learning approach. *arXiv e-prints* 2017 [Internet]. New York: Cornell University; 2017 May 23 [cited 2022 Mar 8]. Available from <https://arxiv.org/abs/1705.08422>.
 96. Slutsky AS, Ranieri VM. Ventilator-induced lung injury. *N Engl J Med* 2013; 369: 2126-36.
 97. Acute Respiratory Distress Syndrome Network, Brower RG, Matthay MA, Morris A, Schoenfeld D, Thompson BT, et al. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *N Engl J Med* 2000; 342: 1301-8.
 98. Writing Group for the PRoVENT Investigators, Simonis FD, Serpa Neto A, Binnekade JM, Braber A, Bruin KC, et al. Effect of a low vs intermediate tidal volume strategy on ventilator-free days in intensive care unit patients without ARDS: a randomized clinical trial. *JAMA* 2018; 320: 1872-80.