



Artificial Intelligence in Perioperative Risk Prediction: A Systematic Review of Clinical, Laboratory and Anesthesiology Applications

Reem ObeidAllah Almeahmadi¹, Miaad hasan alharbi¹, Rawya Abdullah Alrishi¹, Farooq Ahmed Hussien Mojamami², Sultan Homayyan Alharbi³, Abdullah Hamdan Mobark Alahmadi⁴, Abdullah Salem Jubran Aljohani⁴, Trahib Fahad Zaben Almutairi⁵, Zain Saier Mulu Alotaibi⁶, Ibrahim Awad Saud Al-otaibi⁷, Muath Abdulrahman M Alkuwaiz⁸, Sultan Eid Abdul Rahman Al-Osaimi Al-Otaibi⁹

¹Laboratory specialist, Maternity and children hospital, Saudi Arabia.

²Anesthesia Technician, Samtah General Hospital (JAZAN) , Saudi Arabia.

³Aneasthesia Technician, Alrass General Hospital, Saudi Arabia.

⁴laboratory technician, Prince AbdulMohsen Hospital in Al-Ula, Saudi Arabia.

⁵Labrotory Technician, Afif Hospital, Saudi Arabia.

⁶Nurse technician, Afif hospital, Saudi Arabia.

⁷Nursing technician, Deaths Department, Al-Bajadiya General Hospital, Saudi Arabia.

⁸Technician-public health, Ad Diriyah Hospital, Saudi Arabia.

⁹Technician-public health ,Al-Dawadmi Hospital, Saudi Arabia.

Abstract

The use of artificial intelligence (AI) in perioperative medicine is rapidly growing, with several applications in the field of risk stratification, clinical decision making and patient management, tailored to each stage of the perioperative journey, from pre- to intra- to post-operation. This systematic review explores the latest evidence surrounding the use of AI in perioperative risk prediction, specifically focused on perioperative clinical, laboratory, and anesthesiology applications. The review summarizes the evidence from studies that examined the use of machine learning (ML), deep learning (DL), natural language processing (NLP) and other AI methods to predict postoperative complications, mortality, hemodynamic instability, difficult airway, prolonged hospitalization, intensive care admission, and other clinically relevant outcomes. Special focus is placed on the incorporation of demographic data, comorbidities, physiological parameters, laboratory markers, electronic health records, imaging data, and intraoperative monitoring data into predictive models. The literature reviewed suggests that AI models can detect patterns and complex, nonlinear relationships that could supplement traditional models used in the perioperative setting. Laboratory-based and multimodal models can also be used to help identify patients at higher risk of adverse outcomes in the early stages, and anesthesiology-related applications have the potential to predict intraoperative hypotension, postoperative respiratory complications, and other anesthesia-related complications. But there's lots of variability in datasets, outcome definitions, model architectures, validation strategies and reporting. Routine use is hindered by issues of external validity, data quality, interpretability of results, algorithmic bias, privacy, workflow integration and prospective clinical testing. In summary, AI is a promising addition to contemporary perioperative risk stratification, however, its clinical utility will ultimately rely on thorough external validation, reporting, multi-disciplinary oversight, and thoughtful application alongside the clinical acumen of the provider, rather than an autonomous decision-making tool.

Keywords: Artificial intelligence; Perioperative risk prediction; Machine learning; Anesthesiology; Laboratory biomarkers.

1. Introduction

Perioperative care is a complex continuum from preoperative assessment and optimization, to intraoperative management, to recovery after surgery. There are many complications that can arise in patients that are undergoing surgery, such as cardiovascular events, respiratory events, acute kidney injury, surgical site infection, undetected thromboembolism, extended hospitalization, unplanned admission to the intensive care unit, or death. Identification of patients at high risk, prior to or at the time of surgery, is therefore an important part of contemporary perioperative medicine (Yoon, 2022). Traditionally, perioperative risk assessment is based on medical history, physical examination, laboratory, functional and surgical evaluations and scoring systems. These methods have proven to be useful clinically, but may be limited in their ability to address the nonlinear relationship or complex interactions between demographic, clinical, physiological, laboratory, anesthetic, and procedural variables.

To circumvent these limitations, artificial intelligence (AI) has become an important analytical approach. AI in healthcare also includes machine-learning and deep-learning algorithms that can analyze vast amounts of complex

data and derive patterns or insights to make predictions or inform clinical decisions. AI-based prediction is an ideal application in peri-operative medicine, where the enormous amount of multidimensional information generated by the various records systems and monitors of the surgical care process make it particularly appropriate (Bellini, 2022). Machine-learning models have the potential to combine these data sources to provide more dynamic estimates of the individual risks than existing models that are mostly based on a limited number of pre-defined variables and linear statistical relationships.

Recent clinical applications of AI in perioperative risk prediction have significantly increased. Various models have been developed and validated before surgery to forecast other outcomes like postoperative death, major complications, prolonged hospitalization, intensive care unit use, readmission, and procedure-specific adverse events. The models could include age, comorbidities, previous diagnosis, medication exposure, functional status, surgical urgency, and characteristics of the planned procedure (Zeng, 2024). A key promise of AI is not just that it can analyze more variables, but that it can also uncover interactions and patterns that might not be obvious from the sum of its parts, such as risk scores. The predictive performance, however, may differ widely depending on the patient population, the surgical specialty, extent of data quality, outcome definition and extent of external validation.

Another crucial aspect of AI-driven perioperative prediction is laboratory data. Laboratory data is also a significant factor in AI-driven perioperative prediction. Physiological reserve and underlying disease is obtained from the objective data obtained from routine measurements like the hemoglobin, platelet count, serum creatinine, electrolytes, glucose, liver function markers, coagulation markers, inflammatory markers, and other biochemical markers. AI algorithms will be able to examine patterns of abnormal laboratory values and their changes over time, not just the values that are abnormal. This method can enhance the identification of patients who are likely to develop complications after surgery, such as kidney injury, infection, cardiovascular problems, and longer recovery times (Maheshwari, 2023). However, missing data, variations in testing practices, when tests are made, reference ranges, and institutional laboratory procedures all can influence the generalizability of the model.

In the field of anesthesiology, where patient safety is paramount, AI can offer another area of support for predictive perioperative risk assessment. Intraoperative care generates continuous information on physiology (heart rate, blood pressure, oxygen saturation, respiratory parameters, electrocardiographic signals, anesthetic administration, fluid therapy, administration of vasopressors, other monitored parameters, etc.). These data can be used by an AI system to detect the progressive development of physiological instability and anticipate events like intraoperative hypotension, postoperative respiratory complications, hemodynamic deterioration, and other complications (Pardo, 2024). Dynamic pre-operative scores are not feasible, but models based on information collected throughout the operation may enable risk estimates to be adapted as the patient's physiology changes. These functions may be useful in the early detection of deterioration and more individual anesthetic and postoperative care.

Notwithstanding these advances, methodological and clinical challenges still remain. A lot of AI research is conducted in a retrospective fashion and uses datasets from a single center which might have limited external validity. Direct comparisons of models are challenging due to variations in patient populations, surgical techniques, definition of complications, prediction time horizons, data preprocessing, and performance measures. Moreover, if predictive discrimination is high, it does not imply clinical usefulness. In addition to appropriate calibration, the model needs to be robust, transparent, interpretable, and to perform well in relevant clinical and demographic subgroups (Shimada, 2024). When AI-driven predictions can shape perioperative care, the issue of algorithmic bias, data drift, data gaps, interoperability, privacy, and the introduction of AI into the clinical process are especially relevant.

With the growing availability of multimodal clinical data, there is the opportunity to integrate data from across the peri-operative period. Combining these and other baseline clinical characteristics with laboratory trends, anesthetic variables, physiological monitoring, imaging, and other electronic health-record data could enable more comprehensive and continuously updated risk assessment. But the evidence of whether multimodal AI models consistently show clinical usefulness in comparison to well-validated risk prediction models is currently mixed (Lopes, 2024).

2. Methodology

2.1 Study Design

A systematic review design was used to synthesize the existing knowledge on the application of artificial intelligence (AI) in perioperative risk prediction, specifically clinical applications, laboratory applications, and anesthesiology applications. The review was designed to detect, assess and incorporate published evidence around the application of machine learning, deep learning, natural language processing and other AI-driven methods in the prediction of perioperative complications, death, adverse events, postoperative outcomes, and anesthetic risk. The methodology followed the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework to increase the transparency, reproducibility, and methodological rigor.

2.2 Search Strategy

The literature was comprehensively surveyed with a large number of major electronic databases with a focus on medicine, clinical sciences, anesthesiology, laboratory medicine, and health informatics. The following databases were searched: PubMed/MEDLINE, Scopus, Web of Science, ScienceDirect, SpringerLink and Google Scholar.

Searches included both controlled and free-text terms for perioperative care, risk prediction, anesthesia, laboratory investigations, and surgical outcomes, all in relation to AI. Common search terms that were used include "artificial intelligence," "machine learning," "deep learning," "neural networks," "perioperative risk prediction," "preoperative risk," "intraoperative prediction," "postoperative complications," "anesthesiology," "laboratory biomarkers," "surgical outcomes," and "perioperative mortality. Boolean operators, including AND and OR, were used to synthesize ideas and expand or restrict the search, as needed.

2.3 Eligibility Criteria

Studies were eligible if they studied or reviewed an AI technique to predict, stratify or estimate perioperative risk in a patient undergoing a surgical or anesthetic procedure. Studies involving either an adult or pediatric surgical population, and applications that used routinely collected clinical information, EHR, physiologic data, laboratory data, imaging-derived data, anesthetic data, or a combination of these data sources were eligible. Studies were also included if they reported other outcomes including, but not limited to, postoperative complications, mortality, readmission, length of hospital stay, intensive care unit admission, hemodynamic instability, respiratory complications, acute kidney injury, cardiovascular events, or other clinically relevant complications that occurred in the peri-operative period. Only peer-reviewed original research or systematic reviews and other scholarly publications that contained substantive evidence pertinent to the review question were considered, excluding publications with insufficient methodological information.

2.4 Study Selection

The retrieved records were analysed in a systematic way in various steps. Firstly, duplicate records were identified and removed. The titles and abstracts were then evaluated for inclusion in the study based on the predetermined criteria. Any publications that were found to have relevance were reviewed in full text for the inclusion criteria. Studies were rejected at the full-text stage if they failed to focus on perioperative risk prediction, did not use an AI-based analytical method, did not have clinical outcomes that were relevant, or did not include enough information to make meaningful syntheses. The overall selection process was managed using the PRISMA framework, which consists of four stages: identification, screening, eligibility, and final inclusion.

2.5 Data Extraction

A structured data-extraction approach was used to extract relevant information from each eligible publication. Characteristics extracted were publication year, study location, study population, sample size, surgical specialty, procedure type, data source, predictor variables, laboratory or physiological parameters, AI technique, AI comparator model, prediction target, validation approach, and reported performance measures. Special care was taken to identify the development of models based on pre-operative, intra-operative, and/or post-operative data. When reported, additional measures included area under the receiver operating characteristic (ROC) curve, sensitivity, specificity, accuracy, precision, calibration, positive and negative predictive values and other discrimination and clinical utility measures. Consideration of information about external validation, interpretability, missing data treatment and limitation was also made.

2.6 Evidence Synthesis and Analysis

The included studies were anticipated to have significant heterogeneity in populations, surgical procedures, predictor variables, AI algorithms, outcome definitions, and performance measures, which led to a primary synthesis using a structured narrative approach. Findings were sorted by the perioperative phase, clinical application, type of data, AI methodology, and predicted outcome. Where comparisons were possible, comparisons were made between AI-based approaches and a conventional statistical or clinical risk-prediction model. Specific focus was placed on whether AI models showed improvement in predictive discrimination, early detection of high-risk patients, individual risk stratification, and/or support for perioperative decision-making.

2.7 Ethical Considerations

This research was not a research involving the collection of identifiable human-subject data, but rather an analysis and synthesis of previous research materials and literature, so there was no need to be involved in direct recruitment of participants or primary data collection. However, ethical concerns pertinent to AI-related perioperative prediction were taken into account when interpreting the evidence. These encompassed data privacy, security of data, algorithmic bias, transparency, explainability, reproducibility, and the implications of false predictions. Also included in the review was the significance of external validation and clinical oversight prior to the implementation of AI-based risk-prediction systems into daily perioperative practice.

3. Findings and Discussion

3.1 Characteristics and Scope of the Reviewed Evidence

Evidence reviewed showed the swift growth of the use of AI and ML method for perioperative risk prediction. In both published literature, there are several applications of AI in the preoperative assessment, the intraoperative monitoring, prediction of postoperative complications, estimation of mortality and morbidity, assessment of the airway, and surveillance of recovery. A systematic review of 103 prediction studies conducted in the peri-operative setting identified that the majority of studies were single centres (around 75% of studies based on single centre validation, and only 13% were externally validated in multiple centres). Furthermore, approximately 90% of the studies were

deemed to have high or unclear overall risk of bias, which differentiates between prospective predictive performance and clinical maturity. (Arina, 2023)

3.1.1 Study Designs, Clinical Settings and Patient Populations

Most of the evidence comes from retrospective cohort studies analysing routinely collected clinical data, though prospective observational studies, studies evaluating the clinical validity of AI interventions, studies investigating the diagnostic accuracy of AI interventions and randomized studies of AI-supported interventions are starting to become more common. Large electronic health-record databases and institutional perioperative registries have offered important sources of data, as they facilitate the creation of models of thousands or in some instances hundreds of thousands of surgical patients. In a recent systematic review of machine-learning predictions for postoperative delirium, for instance, 17 studies were selected in which 205,202 surgical patients participated, and a more recent review of machine-learning models for non-cardiac surgery identified 22 eligible studies that were developed between 2014 and 2024 (Hashemi, 2024).

The populations represented in the evidence are heterogeneous. Studies have included patients undergoing general, orthopedic, cardiovascular, abdominal, thoracic, neurosurgical, and other procedures, with both elective and emergency surgical populations represented. The anesthesiology literature has also expanded rapidly, particularly in intraoperative hypotension prediction. A 2025 scoping review identified 26 studies involving 48,707 adult surgical patients, with approximately 92.3% involving non-cardiac surgery. (Huotari, 2023) This distribution suggests that AI research is increasingly moving beyond narrowly defined experimental populations toward clinically diverse surgical cohorts, although important geographical and institutional imbalances remain. Evidence from highly resourced hospitals and health systems is substantially more visible than evidence from low- and middle-income settings, creating uncertainty about how well models developed in technologically advanced institutions will perform where laboratory testing, continuous monitoring, electronic records, and perioperative staffing differ considerably.

3.1.2 Types of Artificial Intelligence and Machine-Learning Approaches

A large degree of methodological variation was observed in the reviewed studies. It is important to contrast the conventional statistical modelling, especially logistic regression, as it is familiar to clinicians and gives relatively 'transparently' estimated association between predictors and outcomes. But many studies have used random forests, support vector machines (SVMs), gradient-boosted algorithms, extreme gradient boosting, neural networks, deep learning architectures, and ensemble methods (Rellum, 2021). Random forests and gradient-boosting methods are especially popular in the context of structured clinical data sets since they can model nonlinear relationships and interactions between variables without the investigator having to determine every interaction first. In the field of research on postoperative delirium, for instance, random forests have shown good predictive performance, with a recent meta-analysis reporting the pooled AUROC value for random-forest models of 0.89. (Elfanagely, 2021)

When the prediction of the desired variable is carried out in the perioperative setting, it is particularly important to use deep learning, if the variable is a high dimensional physiological signal, image or other complex data. In recent years, for instance, convolutional neural nets have been widely employed in predicting the airway in airway image analysis. We performed a systematic review and meta-analysis of 13 studies involving difficult airways, which yielded an AUROC of 0.84 for the VGG deep-learning model and also showed good discrimination for other traditional methods like SVM and naïve Bayes. Overall, therefore, the evidence does not suggest that more complex algorithms necessarily lead to a better clinical prediction (Singh, 2022). Model choice, on the other hand, seems to be determined by the type, amount, time series and modality of the available data. For many common clinical uses, there might be a sweet spot for a relatively interpretable model that has good performance, is easy to implement, and doesn't require a huge amount of computing.

3.1.3 Data Sources and Predictor Variables

AI models have increasingly leveraged on a wide variety of perioperative data. These are typical risk factors: age, sex, body characteristics, comorbidities, previous surgeries, drug exposure, ASA physical status, vital signs, lab values, renal function, hemoglobin, glucose, coagulation markers, inflammatory markers, and prior healthcare utilization (Menon, 2024). More sophisticated models include continuously recorded blood pressure, heart rate, oxygen saturation, respiratory waveforms, administration of anesthetic to the patient, physiological monitoring, imaging, and longitudinal electronic health-record data.

The main benefit of these databases is that they have the ability to model the nature of perioperative risk as a complex rather than a handful of discrete clinical factors. However, there is a significant variation in institutions with regards to the availability of predictors. Some models utilize high frequency arterial-waveform data or specialized monitoring, which may not be available at most hospitals. This has led to concerns in the perioperative literature regarding applicability, if models rely on predictors not typically gathered in healthcare environments. Model portability may also be further limited due to missing observations, lack of uniformity of laboratory measurement intervals, coding differences, changes in clinical protocols, and differences in outcome definitions. Therefore, the amount of data shouldn't automatically be considered synonymous with data quality.

3.2 AI Applications in Clinical Perioperative Risk Prediction

3.2.1 Prediction of Postoperative Complications and Adverse Outcomes

One key insight from the summarized evidence is the widespread use of AI in postoperative complications. Several models have been created for acute kidney injury (AKI), infection, cardiovascular complications, respiratory deterioration, thromboembolic events, delirium, prolonged hospitalization and other adverse outcomes. The application of AKI is especially important as there are interactions between the preoperative, intraoperative and postoperative factors. For instance, in the case of cardiac surgery, the prediction of AKI is a specific clinical problem that has been targeted in a 2026 systematic review and meta-analysis, showing the evolving specialisation of the AI models towards clinically relevant organ-specific outcomes. (Li, 2022)

The case of postoperative delirium is another good example. When performing a meta-analysis over 17 studies with over 205,000 surgical patients, the pooled AUROC was 0.83, the pooled sensitivity was 0.73 and the pooled specificity was 0.79. Advanced age, cognitive impairment before surgery, burden of comorbidities, anemia and hypoalbuminemia were common predictors. According to the findings, AI systems could be able to synthesize the traditional risk factors in a meaningful personalised estimation. The performance, however, varies with the population, surgical specialty, predictor availability, definition of outcome and validation strategy. Therefore, a high level of discrimination in one setting is not necessarily reflective of the same level of discrimination for all postoperative complications or surgical populations.

3.2.2 Prediction of Mortality, Morbidity and Hospital Readmission

AI models have been explored as well for perioperative mortality, significant morbidity, intensive-care admission, long hospital stay, and readmission. The value that can be derived is significant, as these outcomes are related to several interacting clinical factors which may not be fully captured by traditional additive scores. However, there is less consistency in the data surrounding mortality than for some other more common outcomes. The systematic review of perioperative prediction models showed that many of the models of mortality were poorly discriminating, and the overall evidence was also severely hampered by methodological flaws and poor external validation. (Paiste, 2024)

The case study of readmission prediction shows how specialist-specific validation is important. In recent years, a large number of ML models have been developed for prediction of the readmission in THA and TKA, but systematic evaluation is still ongoing in order to uncover problems with model stability and methodological robustness. Likewise, there are benefits that continue to exist for conventional mortality tools like the established peri-operative risk scores that are familiar, have been externally validated, and are clinically usable. (Zhu, 2024) Thus, AI must be viewed as a chance to complement existing Perioperative Risk Models, rather than a replacement for them.

3.2.3 Patient-Specific Risk Stratification and Clinical Decision Support

The main clinical value of AI is that it can turn many of these diverse patient factors into a risk score for each patient. The different parameters included in the checklist, such as age, comorbidities, laboratory parameters, physiological parameters, and clinical events, are not necessarily independent; indeed, they can interact, and a ML model can detect the interactions and relationships between these different parameters that are not linear (Cascella, 2023). This provides opportunities for more individualised pre-operative counselling, optimization of modifiable risk factors, and postoperative monitoring intensity, and the allocation of critical-care resources.

This potential is also reflected in the emerging literature on how AI can be used to support perioperative assessment. In 2026, 11 studies with 227,059 patients were included in a systematic review of AI-powered chatbots for perioperative risk stratification, which found good agreement between the AI system and the clinician's assessment of routine ASA classification and risk stratification. As complexity of patients grew, however, performance fell and there was very low certainty of evidence due to geographic bias, inconsistent results, sample size imbalance, methodological flaws, and other limitations. These findings underscore the necessity of a supervised approach to clinical decision support with AI, as opposed to using AI as an alternative to personalized clinical judgment.

3.3 AI Applications Using Laboratory and Physiological Data

3.3.1 Laboratory-Based Prediction of Perioperative Complications

Lab values are also crucial for an AI model used for perioperative predictions, as they offer objective data about organ function and physiology (Pham, 2023). Hemoglobin provides information on anemia and oxygen carrying capacity; creatinine can provide information on renal function; glucose reflects metabolic status; electrolytes can inform about disturbances of cardiovascular and neuromuscular function; coagulation variables can provide information on bleeding risk; and markers of inflammation or biochemical markers can provide information on underlying physiological stress. AI techniques are especially useful when multiple abnormalities are found in the laboratory. For instance, each of the following markers of mild renal dysfunction, anemia, hypoalbuminemia, and electrolyte abnormalities might have small associations with postoperative complications, but the combination of all three might be a much different risk profile. This is exemplified by the postoperative delirium prediction, in which both anemia and hypoalbuminemia have been consistently shown to be important predictors in addition to age, cognitive status and comorbidities. However, laboratory-based models are still reliant on the timeliness and completeness of the measurements taken. Laboratory testing done just prior to surgery may not be identical to tests obtained several days earlier or after clinical deterioration.

3.3.2 Integration of Laboratory, Vital-Sign and Electronic Health Record Data

One of the most significant advances of perioperative AI is multimodal prediction. Preoperative variables are not the only factors used in models; laboratory results, vital signs, medication exposure, diagnoses, prior hospitalizations,

procedural details, and longitudinal electronic health-record data can all be incorporated. Models can include laboratory results, vital signs, medication exposures, diagnosis, prior admissions, procedure information, and longitudinal electronic health-record data, in addition to relying on static preoperative variables (Alexander, 2020). This method allows for both recognition of the patient's inherent risk and changes that may happen during the perioperative phase.

In the prediction during surgery, the temporal aspect of the information is especially important. Algorithms which predict hypotension can analyse a series of arterial pressure and other physiological signals, rather than a single blood-pressure reading. The pooled AUROC of Hypotension Prediction Index algorithms in the systematic review of AI-based intraoperative hypotension prediction was around 0.89 and 0.79 for other AI models. These results show that physiological data may be used to get predictive signals which are not available from the normal preoperative assessment.

3.3.3 Early Detection of Physiological Deterioration

Continuous and repeated measures provide a chance to identify deterioration before a clinical event. They can generate subtle trends in blood pressure, oxygenation, heart rate, breathing, renal function, temperature or other parameters (Lonsdale, 2020). These systems may be useful for monitoring changing respiratory instability, hemodynamic status, infection and organ dysfunction.

One of the most clear cut examples is the intraoperative hypotension literature. Machine-learning systems can determine characteristics of waveforms before a hypotensive episode, giving clinicians a chance to intervene before blood pressure becomes clinically significant. However, accurate predictions do not always lead to better results (Cai, 2024). While there is strong evidence of the predictive capabilities of AI systems for hypotension, the evidence supporting the prediction to improve actual outcome after surgery is less clear-cut. Future research should, therefore, investigate if deterioration can be predicted, and if not, whether prediction leads to timely and beneficial clinical intervention (Dumitru, 2022).

3.4 AI Applications in Anesthesiology and Perioperative Monitoring

3.4.1 Prediction of Intraoperative Hemodynamic and Respiratory Events

Anesthesiology is one of the most technologically advanced fields of Research on Perioperative AI, due to the high frequency of physiological data continuously generated during the period of anesthesia. The development of models to predict hypotension, hypertension, hypoxemia, respiratory deterioration, arrhythmia, etc. during the operation (Han, 2024). Of these, intraoperative hypotension has been studied in particular, as it has been shown to be associated with poor outcomes following surgery.

The 2024 systematic review and meta-analysis of 43 studies showed that Hypotension Prediction Index (HPI) showed high AUROC of 0.89 while other AI models showed pooled AUROC of 0.79. Another scoping review with 48,707 patients confirmed that models based on HPI were often investigated, and that another class of models, neural networks, was also important (Ocampo Osorio, 2024). (Duran, 2023) These results show that it is technically possible to predict, but the clinical utility will be determined by whether and how clinicians respond to predictions without causing too many false alarms or unnecessary interventions.

3.4.2 Prediction of Difficult Airway and Anesthesia-Related Risks

In addition, AI has been used for the evaluation of airway pre-op. Old-fashioned methods of evaluating airway difficulty rely on physical examination, airway dimensions, past experience and clinician's opinion. In addition to these variables, AI can add facial photographs, airway images, demographic data and clinical measurements. In 2025, 13 studies were included in a systematic review and meta-analysis that revealed the potential for deep learning algorithms, which employ image-based methods, to demonstrate useful discrimination, but with significant heterogeneity across studies. (Srinivasareddy, 2024)

This evidence indicates that AI can prove even more useful in certain situations where visual or anatomical data is relevant. The results, however, do not conclude that AI should be used in place of the trained airway assessment. Performance of the model is subject to variability with regard to image quality, patient population, equipment, airway definitions, and difficulty with intubation (Yoon, 2022). Furthermore, the impact of a misdiagnosis of the need for airway management may be severe therefore external validation and clinical supervision are of special significance.

3.4.3 Postoperative Risk Prediction and Recovery Monitoring

AI applications for surgery continue to be used for several days, from the moment of surgery. Various models have been created for postoperative respiratory complications, delirium, long-term recovery, intensive-care unit transfer and other negative outcomes. Delirium prediction is especially illustrative, as age, cognition, comorbidity, physiological disturbance, medications, and surgical factors all have an impact on onset of delirium. Recent research was able to show moderate to high discrimination across numerous ML models while revealing a dependency of performance and applicability on validation strategy and patient population. (Bellini, 2022)

The integration of postoperative monitoring systems could be used to provide risk estimates that are updated as observations are made. Patients who were classified as moderate risk can be re-classified as high risk when there are changes in their oxygen sats, blood pressure, renal function or neurological status. Risk estimates preoperatively (before surgery) may not be as clinically informative as this dynamic approach (Zeng, 2024). However,

implementation will be dependent on control of the frequency of alarms, latency of data, missing measurements and clinician workload.

3.5 Comparative Performance, Clinical Utility and Implementation Challenges

3.5.1 Predictive Performance and Comparison with Conventional Risk Models

Across the reviewed literature, AI models frequently demonstrate moderate-to-high discrimination, but performance is highly dependent on the clinical problem and validation strategy. For example, postoperative delirium models have produced pooled AUROCs around 0.82–0.83, while AI models for intraoperative hypotension have produced higher discrimination in some datasets. (Maheshwari, 2023) These differences should not be interpreted as evidence that one algorithm is universally superior because the outcomes, populations, predictors, prediction horizons, and definitions differ between studies.

The central issue is therefore incremental clinical value rather than AUROC alone. Calibration, sensitivity, specificity, positive and negative predictive values, decision-curve performance, false-alarm burden, and clinical consequences should all be considered. Established statistical scores may have lower apparent complexity but stronger evidence concerning interpretability and external validation. The literature on conventional perioperative mortality models similarly emphasizes the importance of calibration and clinical usability, not discrimination alone. (Pardo, 2024) AI models that achieve a small statistical improvement but require expensive infrastructure, extensive data preprocessing, or substantial clinician workload may have limited practical advantage.

3.5.2 Explainability, Bias, Generalizability and Data Quality

The major limitation of the evidence is the gap between model development and generalizable clinical performance. The systematic review of 103 perioperative AI studies found that 75% relied on single-centre validation and only 13% achieved external multicentre validation. Approximately 90% of studies were assessed as having high or unclear risk of bias. (Shimada, 2024) These findings indicate that impressive internal performance may partly reflect the characteristics of the development population rather than universally transferable predictive ability.

Overfitting, class imbalance, missing data, dataset shift, differences in clinical practice, inconsistent outcome definitions, and algorithmic bias can all reduce generalizability. Explainability is another important concern. Complex ensemble and deep-learning models may identify highly predictive patterns without providing clinicians with an intuitive explanation of why a particular patient has been classified as high risk (Lopes, 2024). This becomes particularly important when predictions influence decisions concerning surgery, intensive monitoring, ICU admission, or postoperative intervention.

Recent evidence concerning postoperative delirium illustrates the importance of validation. Models subjected only to internal validation showed different performance from those receiving both internal and external validation, while newer studies with more complex architectures did not necessarily demonstrate major performance gains over simpler approaches. (Kambale, 2024) Thus, increasing model complexity should not be treated as an objective in itself; robustness, calibration, interpretability, and transferability are more important indicators of clinical readiness.

3.5.3 Clinical Integration, Ethical Considerations and Future Research Priorities

The evidence indicates that successful implementation of perioperative AI requires considerably more than developing a high-performing algorithm. Models must be integrated into existing electronic health-record and anesthesia information systems, operate with reliable data feeds, provide understandable outputs, and fit naturally within clinical workflows. Clinicians should remain responsible for interpreting predictions in the context of the patient's circumstances (Arina, 2023). Cybersecurity, privacy protection, appropriate governance of patient data, transparency concerning model limitations, and monitoring for performance degradation are also essential.

The current evidence supports a cautious but constructive interpretation of AI in perioperative medicine. AI has demonstrated substantial potential for individualized risk prediction, particularly when large longitudinal datasets and continuous physiological measurements are available. However, the evidence remains stronger for predictive discrimination than for demonstrated improvements in patient outcomes. This distinction is particularly clear in intraoperative hypotension research, where strong predictive performance has not consistently translated into proven improvements in objective clinical outcomes. (Huotari, 2023)

Future research should therefore prioritize prospective multicentre validation, standardized outcome definitions, transparent reporting according to contemporary prediction-model guidelines, calibration assessment, external validation across diverse populations, and evaluation of health-equity implications. Randomized or pragmatic clinical trials should determine whether acting on AI predictions actually reduces complications, mortality, ICU utilization, readmission, or length of stay. Research should also compare AI models directly with established risk scores using identical datasets and clinically relevant endpoints (Hashemi, 2024). Finally, implementation studies should examine clinician acceptance, workflow integration, alert burden, interpretability, and patient perspectives. Overall, the evidence suggests that AI is evolving from an experimental predictive technology toward a potentially important component of perioperative decision support, but its clinical value will ultimately depend not simply on how accurately it predicts risk, but on whether those predictions are reliable across populations and can be translated into timely, equitable, and beneficial patient care. (Rellum, 2021)

4. Conclusion

Artificial intelligence is emerging as a valuable approach to perioperative risk prediction, with applications spanning clinical assessment, laboratory-based risk stratification, anesthesiology, and postoperative outcome prediction. The evidence reviewed indicates that AI models can integrate large and heterogeneous datasets, including demographic characteristics, comorbidities, vital signs, laboratory parameters, medication profiles, imaging findings, and intraoperative physiological data, to identify patients at increased risk of complications and support individualized perioperative decision-making. In anesthesiology, AI-assisted prediction may improve recognition of difficult airway, hemodynamic instability, postoperative respiratory complications, and other adverse events, while laboratory and clinical data-driven models provide opportunities for earlier identification of infection, organ dysfunction, bleeding, and prolonged recovery. However, the clinical value of these technologies depends on external validation, data quality, interpretability, fairness, interoperability, and integration into established clinical workflows. Differences in datasets, algorithms, outcome definitions, and validation strategies also limit direct comparison across studies and highlight the need for prospective multicentre research. Overall, AI should be regarded as a decision-support technology that complements rather than replaces clinical expertise, with responsible implementation requiring rigorous validation, continuous monitoring, transparent reporting, and attention to patient safety, privacy, and equity.

References

1. Abraham, J., Bartek, B., Meng, A., King, C. R., Xue, B., Lu, C., & Avidan, M. S. (2023). Integrating machine learning predictions for perioperative risk management: towards an empirical design of a flexible-standardized risk assessment tool. *Journal of biomedical informatics*, 137, 104270.
2. Alexander, J. C., Romito, B. T., & Çobanoğlu, M. C. (2020). The present and future role of artificial intelligence and machine learning in anesthesiology. *International anesthesiology clinics*, 58(4), 7-16.
3. Arina, P., Kaczorek, M. R., Hofmaenner, D. A., Pisciotto, W., Refinetti, P., Singer, M., ... & Whittle, J. (2023). Prediction of complications and prognostication in perioperative medicine: a systematic review and PROBAST assessment of machine learning tools. *Anesthesiology*, 140(1), 85.
4. Bellini, V., Valente, M., Bertorelli, G., Pifferi, B., Craca, M., Mordonini, M., ... & Bignami, E. (2022). Machine learning in perioperative medicine: a systematic review. *Journal of anesthesia, analgesia and critical care*, 2(1), 2.
5. Brydges, G., Uppal, A., & Gottumukkala, V. (2024). Application of machine learning in predicting perioperative outcomes in patients with cancer: A narrative review for clinicians. *Current Oncology*, 31(5), 2727-2747.
6. Cai, J., Li, P., Li, W., & Zhu, T. (2024). Outcomes of clinical decision support systems in real-world perioperative care: a systematic review and meta-analysis. *International journal of surgery (London, England)*, 110(12), 8057.
7. Cascella, M., Tracey, M. C., Petrucci, E., & Bignami, E. G. (2023). Exploring artificial intelligence in anesthesia: a primer on ethics, and clinical applications. *Surgeries*, 4(2), 264-274.
8. Davoud, S. C., & Kovacheva, V. P. (2023). On the horizon: specific applications of automation and artificial intelligence in anesthesiology. *Current anesthesiology reports*, 13(2), 31-40.
9. Dumitru, M., Berghi, O. N., Taciuc, I. A., Vranceanu, D., Manole, F., & Costache, A. (2022). Could artificial intelligence prevent intraoperative anaphylaxis? reference review and proof of concept. *Medicina*, 58(11), 1530.
10. Duran, H. T., Kingeter, M., Reale, C., Weinger, M. B., & Salwei, M. E. (2023). Decision-making in anesthesiology: will artificial intelligence make intraoperative care safer?. *Current opinion in anaesthesiology*, 36(6), 691.
11. Elfanagely, M. O., Toyoda, M. Y., Othman, B. S., Mellia, B. J. A., Basta, M. M., Liu, B. T., ... & Fischer, M. M. J. P. (2021). Machine learning and surgical outcomes prediction: a systematic review. *Journal of Surgical Research*, 264, 346-361.
12. Guni, A., Varma, P., Zhang, J., Fehervari, M., & Ashrafian, H. (2024). Artificial intelligence in surgery: the future is now. *European Surgical Research*, 65(1), 22-39.
13. Han, L., Char, D. S., Aghaeepour, N., & Stanford Anesthesia AI Working Group. (2024). Artificial intelligence in perioperative care: opportunities and challenges. *Anesthesiology*, 141(2), 379.
14. Hashemi, S., Yousefzadeh, Z., Abin, A. A., Ejmalian, A., Nabavi, S., & Dabbagh, A. (2024). Machine learning-guided anesthesiology: a review of recent advances and clinical applications. *J Cell Mol Anesth*, 9(1), e145369.
15. Huotari, A. (2023). Utilizing artificial intelligence in perioperative patient flow: systematic literature review.
16. Kambale, M., & Jadhav, S. (2024). Applications of artificial intelligence in anesthesia: a systematic review. *Saudi journal of anaesthesia*, 18(2), 249.
17. Li, Y. Y., Wang, J. J., Huang, S. H., Kuo, C. L., Chen, J. Y., Liu, C. F., & Chu, C. C. (2022). Implementation of a machine learning application in preoperative risk assessment for hip repair surgery. *BMC anesthesiology*, 22(1), 116.

18. Lonsdale, H., Jalali, A., Ahumada, L., & Matava, C. (2020). Machine learning and artificial intelligence in pediatric research: current state, future prospects, and examples in perioperative and critical care. *The Journal of Pediatrics*, 221, S3-S10.
19. Lopes, S., Rocha, G., & Guimarães-Pereira, L. (2024). Artificial intelligence and its clinical application in anesthesiology: a systematic review. *Journal of clinical monitoring and computing*, 38(2), 247-259.
20. Maheshwari, K., Cywinski, J. B., Papay, F., Khanna, A. K., & Mathur, P. (2023). Artificial intelligence for perioperative medicine: perioperative intelligence. *Anesthesia & Analgesia*, 136(4), 637-645.
21. Mahsood, H. Y., & Farooq, M. I. (2024). Artificial intelligence in predicting anesthesia-related complications: A cross-specialty approach integrating pulmonary, renal, and cardiovascular risk factors. *Journal of Biological and Medical Innovations*, 2(2), 27-40.
22. Menon, S., Patel, R., & Musuku, S. (2024). Applications of artificial intelligence and machine learning in cardiac Anesthesia across the continuum of perioperative care. *Academic Medicine & Surgery*.
23. Ocampo Osorio, F., Alzate-Ricaurte, S., Mejia Vallecilla, T. E., & Cruz-Suarez, G. A. (2024). The anesthesiologist's guide to critically assessing machine learning research: a narrative review. *BMC anesthesiology*, 24(1), 452.
24. Paiste, H. J., Godwin, R. C., Smith, A. D., Berkowitz, D. E., & Melvin, R. L. (2024). Strengths-weaknesses-opportunities-threats analysis of artificial intelligence in anesthesiology and perioperative medicine. *Frontiers in Digital Health*, 6, 1316931.
25. Pardo, E., Le Cam, E., & Verdonk, F. (2024). Artificial intelligence and nonoperating room anesthesia. *Current Opinion in Anesthesiology*, 37(4), 413-420.
26. Pham, F. M. (2023). Artificial intelligence-supported systems in anesthesiology and its standpoint to date—a review. *Open Journal of Anesthesiology*, 13(7), 140-168.
27. Rellum, S. R., Schuurmans, J., van der Ven, W. H., Eberl, S., Driessen, A. H., Vlaar, A. P., & Veelo, D. P. (2021). Machine learning methods for perioperative anesthetic management in cardiac surgery patients: a scoping review. *Journal of thoracic disease*, 13(12), 6976.
28. Shimada, K., Inokuchi, R., Ohigashi, T., Iwagami, M., Tanaka, M., Gosho, M., & Tamiya, N. (2024). Artificial intelligence-assisted interventions for perioperative anesthetic management: a systematic review and meta-analysis. *BMC anesthesiology*, 24(1), 306.
29. Singh, M., & Nath, G. (2022). Artificial intelligence and anesthesia: A narrative review. *Saudi journal of anaesthesia*, 16(1), 86.
30. Srinivasareddy, S. (2024). Artificial intelligence in anesthesia: What might the future hold. *International Journal of Clinical Anesthesiology*, 12(1), 1131.
31. Yoon, H. K., Yang, H. L., Jung, C. W., & Lee, H. C. (2022). Artificial intelligence in perioperative medicine: a narrative review. *Korean journal of anesthesiology*, 75(3), 202-215.
32. Zeng, S., Qing, Q., Xu, W., Yu, S., Zheng, M., Tan, H., ... & Huang, J. (2024). Personalized anesthesia and precision medicine: a comprehensive review of genetic factors, artificial intelligence, and patient-specific factors. *Frontiers in medicine*, 11, 1365524.
33. Zhu, Y., Liu, X., Li, Y., & Yi, B. (2024). The applications and prospects of big data in perioperative anesthetic management. *Anesthesiology and Perioperative Science*, 2(3), 30.