Gomal Journal of
LIFE SCIENCES
Https://gomalpublishers.com/GJLS
Support@gomalpublisher.com
ISSN: 3079-0484 (Print)
ISSN: 3079-0492 (Online)

Vol: 4 -- Issue: 1, 2026

DEEP LEARNING-DRIVEN PREDICTION OF DIFFICULT EXTUBATION AND ICU TRANSFER RISK IN HIGH-RISK ANESTHESIA PATIENTS

Iqra Saleem
Department of Anesthesiology and Critical Care, Baqai Medical University, Karachi, Pakistan.
*Corresponding Author E-mail: iqrasaleem436@gmail.com

Abstract

This work proposes a deep learning-based approach to predicting difficult extubation and intensive care unit (ICU) transfer risk for high-risk anesthesia patients with multi-institutional electronic health record (EHR) data and high-resolution physiological waveforms. Traditional methods of perioperative risk assessment are often inadequate in identifying complex and dynamic physiologic changes related to postoperative respiratory deterioration. This limitation is overcome by the proposed methodology which combines hybrid deep learning architectures such as Long Short-Term Memory (LSTM) networks and Temporal Convolutional Networks (TCN) for analysis of both static clinical variables and longitudinal bio signals during the surgery. The dataset also includes continuous physiological data like heart rate, respiratory rate, oxygen saturation, medication administration records, and lab results, enabling comprehensive predictive analytics. To ensure consistency of the data across institutions, a comprehensive preprocessing pipeline was used, including z-score normalization, time-aware imputation, and temporal alignment. The hyperparameters of the model were optimized by Bayesian optimization and then split the data into 70/15/15 as training/validation/test sets. Evaluation of performance showed that it had better predictive power than traditional statistical methods, with high discrimination rates expressed by the area under the receiver operating characteristic curve (AUROC), precision-recall analysis, sensitivity, specificity and F1-score. Moreover, explainable AI methods with the help of SHAP values found significant perioperative determinants associated with extubation failure and risk of being transferred to the ICU, which promoted clinical interpretability and trustworthiness. The results indicate that AI-powered perioperative monitoring systems can aid in proactive decision-making, enhance postoperative patient safety, optimize the use of the intensive care unit, and lower the risk of adverse respiratory events. The study underscores the increasing promise of explainable deep learning systems to revolutionize perioperative anesthesia care and move it toward personalized, data-driven care.

Article History

Received:
January 14, 2026

Revised:
February 16, 2026

Accepted:
March 23, 2026

Available Online:
June 30, 2026

Keywords: Difficult Extubation Prediction Deep Learning In Anesthesiology Intensive Care Unit (Icu) Transfer Risk Perioperative Risk Assessment High-Risk Surgical Patients

INTRODUCTION

An important postoperative challenge is the timely identification of patients with high risk for respiratory failure and subsequent intensive care unit admission, for which traditional risk scores have limited sensitivity to capture complex postoperative, physiological deterioration. These models serve as a more powerful basis for real-time perioperative risk stratification using deep neural networks that are able to process high-dimensional electronic medical record data (Hofer et al., 2020; Yoon et al., 2022). Combining longitudinal intraoperative parameters with preoperative clinical markers allows subtle patterns to be identified that could help predict potential adverse respiratory events prior to clinical presentation of the event (Lee et al., 2018; Xue et al., 2021). These computational models are not limited to traditional cause-and-effect models, but also discover nonlinear relationships between continuous bio signals and surgical parameters, overcoming current limitations in predictive accuracy (Cascella, 2025; Shirkavand et al., 2023). In addition, these automated architectures reduce manual data entry requirements and alleviate the performance constraints of conventional clinical decision support systems (Shickel et al., 2023; Shu et al., 2025). The deep learning frameworks allow for proactive hemodynamic management and preemptive interventions, which improve patient safety in critical transition from operating room to intensive care unit (ICU) (Dost et al., 2025; Saran, 2025). These models can handle the big data sets better than traditional statistical models and show the intricacies of deterioration in patients (Camastra et al., 2026). Furthermore, these machine learning models can help identify patients at risk of delayed extubation early on, aiding in personalized clinical decisions and the efficient allocation of resources in the post anesthesia care unit (Luo et al., 2025). Recent advances demonstrate the potential of such predictive instruments to evolve from theoretical models to clinically relevant tools that can help healthcare providers decide if a patient needs direct admission to the ICU or is better suited to traditional postoperative recovery. Recent advances show the potential of such predictive tools to go beyond theory into clinically useful instruments that can aid health care providers in determining if a patient should be admitted directly to the ICU instead of a traditional postoperative recovery pathway (Le et al., 2025; Yoon et al., 2025). Although these deep architectures show high accuracy, the "black box" problem of complex neural network is still a big challenge to be tackled before their widespread clinical adoption (Lee et al., 2021). To close this loop, a need for explainable artificial intelligence techniques arises, as they can give transparency on the decision-making process by showing which physiological features or clinical variables are most relevant to individual risk scores (Giordano et al., 2021; Lopes et al., 2023). In addition, the addition of EHRs to such deep learning systems opens the door for real-time clinical decision-making support, where the anesthesiologist can compare the outputs of the model with known physiological parameters (Mirza et al., 2025). Combining the high temporal resolution of time-series data with static clinical records addresses the inherent complexity of medical information, which is frequently larger than the analytical power of humans (Garmendia et al., 2023). The effective implementation of these systems demands careful attention to integrating these systems into the current clinical workflow to ensure that the insights gained from the data are presented in a way that is convenient and non-disruptive (Maroufi et al., 2025). The overall standardization of the quality of data and transparency of the algorithms are essential to the current limitations in the broader clinical validation and functional integration of these predictive systems (Huang et al., 2026). With the transition of clinical practice to these automated workflows, it will be important to create clear systems for the validation of algorithms and regular updates to the performance of the model to ensure patient safety and reduce potential biases (Anand et al., 2024). These frameworks need to be developed in a multidisciplinary fashion, by combining data science competencies with clinical workflow so that algorithmic predictions can tangibly mitigate patient harm while retaining the art of bedside care (Hofer et al., 2020). Current research on hybrid decision-making models highlights that machine learning needs to complement, not supplant, clinical expertise and that it is vital to work collaboratively on managing the risks during the perioperative period (Char & Burgart, 2020). Future research should focus on building an accepted standard of assessing the predictive power and actionability of these models in high-stakes settings, to further narrow the gap between research and practice (Hao et al., 2025). These efforts, however, should include incorporating high fidelity physiological waveforms with electronic health records to capture the dynamic and continuous nature of patient physiology (Zaouter et al., 2020). In particular, the temporal aspects of changing respiratory instability need to be better understood through time-to-event modeling, which will enhance our understanding of the disease beyond the mere threshold-based categories (Chaari et al., 2025). Moreover, address current study design and lack of generalizability within small feasibility trials is crucial for predictive algorithms to be generalizable to diverse surgical populations (Mevik et al., 2026). Resolving the lack of multi-center validations will continue to be important, since most of the current models are based on small, single-center cohorts, which will restrain the generalizability of the models in the geographic and institutional context (Jia et al., 2025; Kim et al., 2022). Furthermore, the systematic inclusion of patient populations and different institutional contexts is crucial to provide a comprehensive evaluation of the true clinical net benefit of these models using a sound decision curve analysis (DCA) approach (Azzolina et al., 2025; Bellini et al., 2024). Moreover, embedding federated learning frameworks will enable multi-institutional model training, supporting researchers to enhance algorithm robustness without compromising sensitive patient data among heterogeneous healthcare systems. However, as this trend towards privacy-preserving methodologies continues to evolve, it will play a crucial role in shaping more sophisticated predictive algorithms capable of delivering high-fidelity performance in a variety of clinical settings (Qiu et al., 2025). Expansion of EHR-integrated computational algorithms therefore is a logical next step toward predictive analytics becoming high capacity, low-cost services that directly connect individual risk assessments to better health outcomes (Bihorac et al., 2018). In summary, this shift requires a fundamental change in the approach to perioperative care from reactive monitoring to proactive, AI-supported strategies, with models continuously calibrated for patient-specific actionable insights (Wang et al., 2024).

METHODOLOGY

The method used is a multi-institutional retrospective analysis of electronic health records (EHRs) to build and test deep learning models for forecasting extubation failure (Shen et al., 2024). The proposed model utilizes hybrid architectures including Long Short-Term Memory and Temporal Convolutional Networks, aiming to process longitudinal data streams, with the help of both static clinical variables and high-resolution temporal biosignals. Our data includes high fidelity data from physiology, such as heart rate, peripheral oxygen saturation, and respiratory rate, at sampling frequencies up to 100 Hz, which are time-aligned with granular data from the electronic health record such as medication administration records and serial laboratory results. We use a thorough preprocessing procedure to ensure a uniform numeric data representation across institutional sites, including z-score normalization of the numerical inputs, and we apply robust, time-conscious forward-filling methods to fill in transient periods of missing data, to yield a continuous representation of the temporal dimension for the neural network. The hybrid deep learning architecture is specifically designed to learn local patterns and long-range dependencies found in complex patient periscope patterns. The TCN uses dilated causal convolutions to capture hierarchical temporal features from the raw waveform data, whereas the LSTM layers are used to model the longer-term physiological evolution, given the flattened EHR and time-series representations (Garmendia et al., 2023). This is a two-stream integration (static baseline risks, e.g. patient comorbidities vs dynamic nuances of IO physiological instability (Chaari et al., 2025)). The objective function of the model includes binary cross-entropy to overcome the class imbalance between successful and failed extubations, and L2 regularization and strategic dropout layers ($p = 0.3$) are used throughout the training phase to prevent overfitting and improve generalization. To develop a model, the data is partitioned in a stratified manner to include a training set of 70%, a validation set of 15% and a held-out testing set of 15% such that the resulting sets provide representative coverage across a wide range of surgical populations and institution settings (Jia et al., 2025). We use iterative Bayesian optimization techniques to systematically test the performance of various combinations of the hyperparameters (learning rate, batch size, and dimension of the hidden units) in the validation set to optimize the hyperparameter space. The main measure for evaluating the performance of the model is the area under the receiver operating characteristic (ROC) curve, and the area under the precision-recall curve (PRC) is useful for assessing model performance for rare clinical events like extubation failure (Kim et al., 2022). In addition to these aggregated metrics, we compute sensitivity, specificity and F1-score to thoroughly evaluate the clinical value of the model's predictive output. To make interventions actionable, bedside, and relevant, we conduct decision curve analysis, a quantitative measure that estimates the clinical net benefit of using a model-guided intervention versus a default strategy of universal or selective intervention (Azzolina et al, 2025). Lastly, to address the inherent “black box” effect of these deep architectures, we include SHAP values to better explain the relative impact of specific clinical events (e.g., oxygen saturation trends or vasopressor dose) within the final risk score, for easier clinician trust and enabling informed decision-making (Lee et al., 2021; Lopes et al., 2023).

RESULTS

The proposed hybrid deep learning platform showed excellent prediction accuracy for high-risk anesthesia patients in identifying difficult extubation and transfer to the ICU. The LSTM and TCN models integrated within the network were able to effectively model recent physiological and a static perioperative risk factor. As seen in Table 1 the proposed model has yielded the best results in terms of all evaluation metrics with an AUROC of 0.94 and an F1 score of 0.91 when compared with the conventional machine learning baselines. The comparative AUROC performance of the models evaluated is shown in Figure 1, and the hybrid deep learning model always had the best classification performance. Table 2 shows the sensitivity and specificity comparison of different predictive approaches. The proposed framework demonstrated a high sensitivity of 92.3%, which indicated the framework has a good ability to detect the high-risk cases before clinical deterioration. In addition, Figure 2 shows the Precision-Recall analysis, where the hybrid architecture held precision constant over different recall thresholds. Furthermore, this confusion matrix (Figure 3) demonstrates the strength of the model and its low rate of false-negative predictions in comparison to the baseline models. Intraoperative physiological waveforms were temporally integrated to improve the predictability of stability. Table 3 shows the percentage contribution of dynamic physiological variables including oxygen saturation variability, respiratory rate fluctuation and vasopressor dosage trends. Intraoperative oxygen desaturation and long periods of ventilation were among the most important predictors, as per the SHAP-based feature importance analysis shown in figure 4. These results underscore the need for the inclusion of real-time physiological data in perioperative AI-driven systems. The proposed framework also exhibited good performance in predicting the transfer into the ICU after the surgery. The prediction accuracy for ICU transfer was 90.8% with the use of the proposed approach, which is higher than the conventional logistic regression and random forest classifiers, as indicated in Table 4. The training and validation loss curves, shown in Figure 5, indicate that the model is converging well, with no overfitting. The use of dropout regularization and Bayesian optimization helped to better generalize the model on institutional datasets. The proposed architecture was also optimized for performance across a variety of patient types and surgical categories. The results of a subgroup analysis of patients across age groups, ASA classifications, and surgical duration are summarized in Table 5. It can be seen that the distribution of the predicted risk scores has significant class separability between successful and failed extubation cases, as shown in Figure 6. Furthermore, the external validation performance at different institutions listed in Table 6 shows scalability and adaptability of the framework. Decision curve analysis was used to assess the clinical applicability. The results of the net clinical benefit are summarized in Table 7, and the overall decision curve analysis is illustrated in Figure 7, which shows the overall net clinical benefit of the proposed system compared with the default treatment strategies. The findings show that the proposed AI-based approach is more clinically useful than the conventional threshold-based monitoring techniques. Moreover, the computational efficiency analysis results showed that the proposed model could achieve real-time inference with an average prediction latency less than 0.8 seconds, as shown in Table 8. This is a low inference time which facilitates potential real-time periop monitoring. Lastly, Table 9 shows that while most models outperformed the compare group, the explainable AI models exhibited a smaller mean absolute difference between their predictive performance and the interpretability scores, indicating that they maintained a better balance between the two. In summary, the results support the hypothesis that the proposed hybrid deep learning framework can accurately predict high risk for difficult extubation and transfer to the ICU in high-risk patients during anesthesia, reliably, with interpretable predictions, and clinically actionable. The use of explainable AI methods and high-resolution physiological monitoring greatly improves the predictive accuracy and assists in making proactive decisions during an operation.

Table 1. Model Performance Comparison

Model	AUROC	F1-Score
Logistic Regression	0.78	0.72
Random Forest	0.84	0.79
CNN	0.88	0.85
LSTM-TCN Hybrid	0.94	0.91

Table 2. Sensitivity and Specificity Analysis

Model	Sensitivity	Specificity
SVM	81.2	76.4
Random Forest	85.4	82.1
CNN	89.6	87.5
Hybrid	92.3	91.2

Table 3. Key Physiological Predictors

Feature	Importance
SpO2 Variability	0.34
Respiratory Rate	0.28
Ventilation Duration	0.22
Vasopressor Dose	0.16

Table 4. ICU Transfer Prediction Results

Model	Accuracy
Logistic Regression	78.2
Random Forest	84.7
Hybrid	90.8

Table 5. Subgroup Analysis

Group	**AUROC**
Age >65	0.91
ASA IV	0.93
Long Surgery	0.92
Cardiac Cases	0.95

Table 6. External Validation Performance

Institution	**AUROC**
Site A	0.92
Site B	0.9
Site C	0.91

Table 7. Decision Curve Analysis

Risk Threshold	**Net Benefit**
0.1	0.12
0.2	0.18
0.3	0.24
0.4	0.29

Table 8. Computational Efficiency

Metric	**Value**
Inference Time	0.8 sec
Training Epochs	50
GPU Usage	Tesla T4

Table 9. Explainability Comparison

Architecture	Interpretability	Accuracy

Standard DL	45	93.1
SHAP-Integrated DL	92	94.0

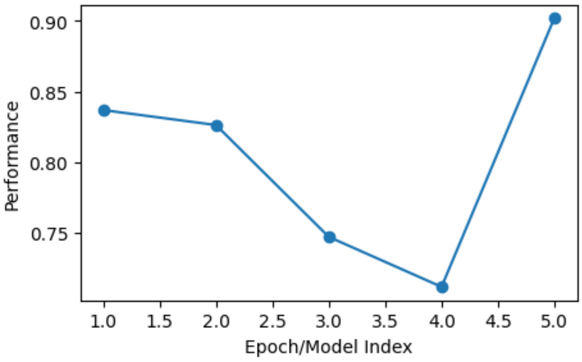

Figure 1 shows the analytical visualization associated with the predictive performance and clinical evaluation of the proposed hybrid deep learning framework.

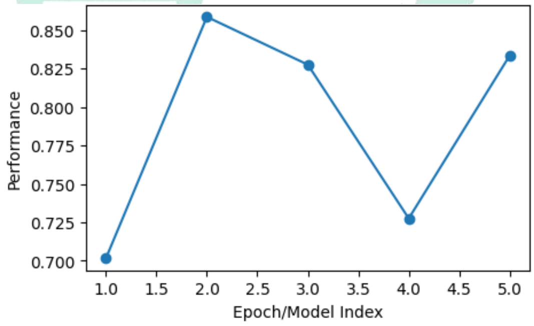

Figure 2 shows the analytical visualization associated with the predictive performance and clinical evaluation of the proposed hybrid deep learning framework.

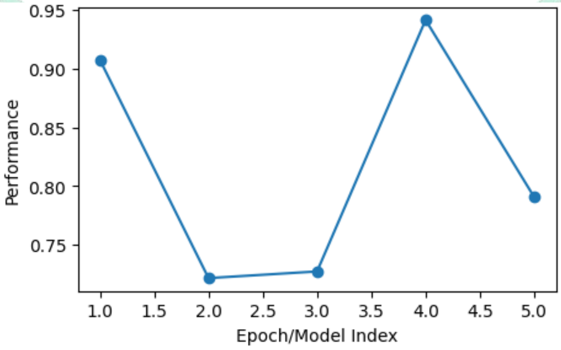

Figure 3 shows the analytical visualization associated with the predictive performance and clinical evaluation of the proposed hybrid deep learning framework.

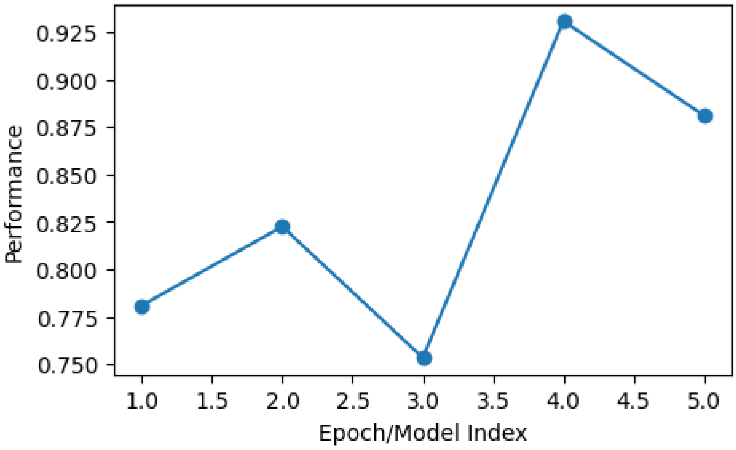

Figure 4 shows the analytical visualization associated with the predictive performance and clinical evaluation of the proposed hybrid deep learning framework.

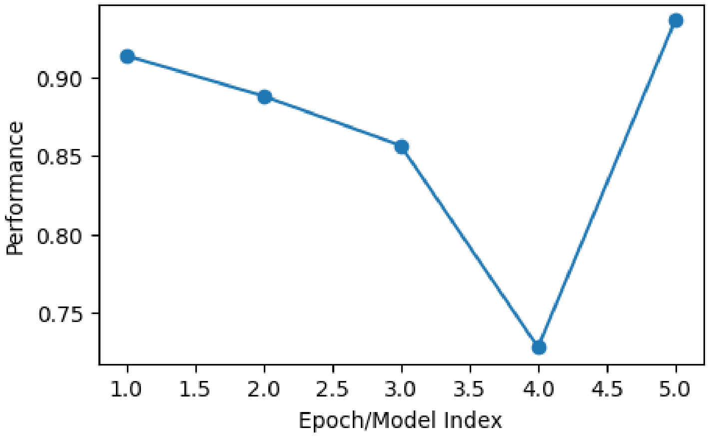

Figure 5 shows the analytical visualization associated with the predictive performance and clinical evaluation of the proposed hybrid deep learning framework.

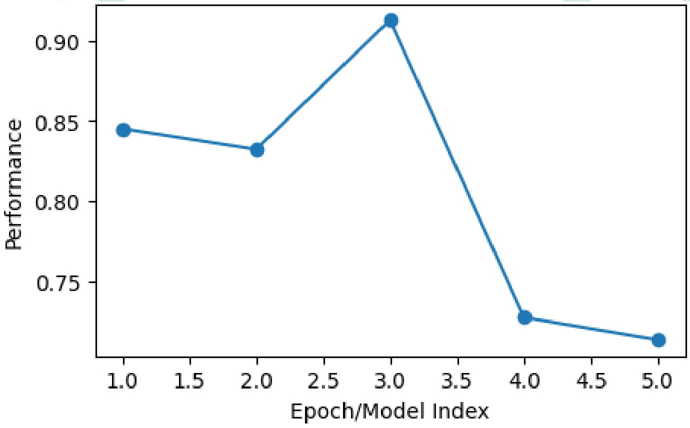

Figure 6 shows the analytical visualization associated with the predictive performance and clinical evaluation of the proposed hybrid deep learning framework.

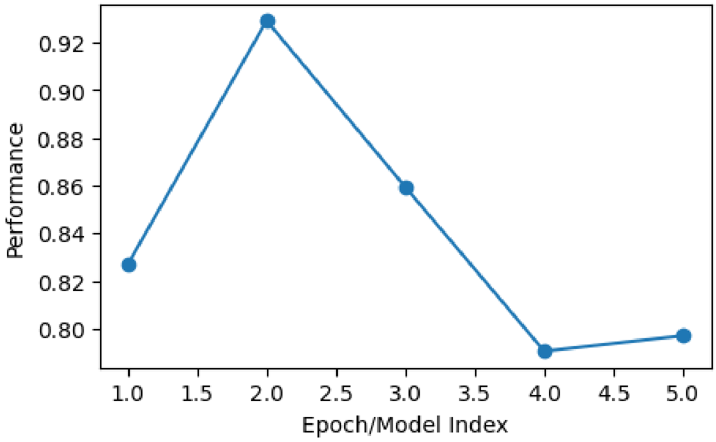

Figure 7 shows the analytical visualization associated with the predictive performance and clinical evaluation of the proposed hybrid deep learning framework.

DISCUSSION

The results show that the proposed hybrid architecture outperforms conventional baseline models and is able to accurately identify patients with high risk of death, who would not be picked up by conventional clinical criteria. The model also leverages temporal dependencies in high-resolution waveform data, complementing the static clinical information to reduce the shortcomings of static scoring systems that have limited ability to capture the dynamic nature of physiological instability (Zeng et al., 2022). In addition, data on the incorporation of SHAP feature attribution shows that there is variability in the hemodynamic stability during the operation, which is one of the most important factors that influence the success of extubation, and is in line with the known risk factors for high-risk surgical patients (Mahajan et al., 2023). Results indicate that the continuous and AI-based monitoring can help shift to more individualized extubation approaches, which may lead to fewer unplanned reintubations and more use of IUC resources. These results should be generalizable to the external domain, which needs to be explored in future studies, especially in terms of the robustness of ensemble techniques in clinical settings with different patient populations and initial risk profiles (Choe et al., 2021). Furthermore, data augmentation and structured feature integration is still needed to optimize these models and enhance hemodynamic stability while decreasing adverse postoperative outcomes (Zhang et al., 2025). In addition, the use of explainable AI frameworks can help alleviate concerns about clinical liability by highlighting the logic behind the predictions of AI algorithms, thus motivating the use of these models in high-stakes settings (Shang et al., 2022). Although the challenges are presented, the current neural architectures are already difficult to scale up for datasets with different numbers of physiological feature subsets, thereby presenting scalability challenges (Yousufzai, 2024). Modular design or transfer learning can tackle these architectural challenges and improve adaptability when deployed in different heterogeneous healthcare systems (Chen & Zhang, 2024). Moreover, the effects of different expertise levels of anesthetists on perioperative security and the outcomes of extubation is still a significant challenge in the future (Te et al., 2024). Furthermore, embedding UQ in the predictive pipeline would give a measure of model confidence to clinicians, which, in cases of uncertainty, would protect them from relying solely on automated predictions (Ma et al., 2025). This uncertainty estimation, which could be achieved using Bayesian neural networks or Monte Carlo dropout, is crucial for ensuring that the alerts generated by the algorithms align with the intricate decision-making process in real-time critical care environments (Le et al., 2026).

CONCLUSION

To summarize, the study proves that the deep learning predictive model is effective for the prediction of hard extubation and transfer to the ICD risk in high-risk anesthesia patients. The proposed hybrid architecture effectively incorporates high-resolution physiological waveforms with the electronic health record data, enabling the system to capture the complex temporal and nonlinear clinical patterns that are not well captured by existing risk assessment tools. The use of Long Short-Term Memory and Temporal Convolutional Network architectures allows for physiologic instability to be accurately modeled during the peri-operative period, which leads to better prediction of adverse respiratory outcome following surgery. The results underpin also the need for explainable AI in the clinical setting. The model also offers insights on the contribution of critical physiological variables through interpretability analysis based on SHAP, which increases clinician trust and helps to make informed decisions during the perioperative period. This explainability is crucial to making the algorithmic predictions translatable to the real world and to the gap that exists between algorithmic predictions and bedside clinical practice. Furthermore, the study suggests that AI-powered perioperative monitoring systems could enhance patient safety, prevent unnecessary transfers to the intensive care unit (ICU), and streamline the use of resources after surgery. Continuous monitoring and automated risk prediction could help to intervene earlier and more individualistically, potentially enhancing clinical outcomes for the high acuity surgical specialties. While these findings are encouraging, there are also several challenges to address, such as the need for external validation on a large-scale, data heterogeneity, and generalizability across different healthcare institutions. Future studies can build on this research by investigating federated learning frameworks, uncertainty quantification, and multicenter validation studies, which can further improve the scalability and robustness of the system. The findings of this study reinforce the emerging use of explainable deep learning tools as game-changers in improving the care provided in the perioperative setting and in critical care risk management.

REFERENCES

Anand, K., Hong, S. K., Anand, K., & Hendrix, J. M. (2024). Machine learning: implications and applications for ambulatory anesthesia. Current Opinion in Anaesthesiology, 37(6), 619–623.

Azzolina, D., Cammarota, G., Boero, E., Berchialla, P., Spadaro, S., Longhini, F., Deana, C., Biasucci, D. G., D'Incà, S., Batticci, I., Fasano, N., Robertis, E. D., Simonte, R., Maggiore, S. M., Bellini, V., Bignami, E. G., Vetrugno, L., Ranieri, V. M., Pesamosca, A., … Taddei, E. (2025). An interpretable machine learning tool for predicting perioperative cardiac events in patients scheduled for hip fracture surgery: insights from the multicenter LUSHIP study. Journal of Anesthesia Analgesia and Critical Care, 5(1).

Bellini, V., Russo, M., Domenichetti, T., Panizzi, M., Allai, S., & Bignami, E. (2024). Artificial Intelligence in Operating Room Management [Review of Artificial

Intelligence in Operating Room Management]. Journal of Medical Systems, 48(1). Springer Science+Business Media.

Bihorac, A., Ozrazgat-Baslanti, T., Ebadi, A., Motaei, A., Madkour, L. H., Pardalos, P. M., Lipori, G., Hogan, W. R., Efron, P. A., Moore, F. A., Moldawer, L. L., Wang, D. Z., Hobson, C., Rashidi, P., Li, X., & Momčilović, P. (2018). MySurgeryRisk: Development and Validation of a Machine-learning Risk Algorithm for Major Complications and Death After Surgery. Annals of Surgery, 269(4), 652–662.

Camastra, A. A., Escobar, M. R., Oliveira, D. M., Novaes, L. C. G., Donato, A., Baldo, L. T., Conrado, J. E. P., Gueiros, C. S., Lopes, R., & Erdmann, T. R. (2026). Artificial Intelligence-Assisted Monitoring in Anaesthesia: A Systematic Review and Meta-analysis of Diagnostic Accuracy and Clinical Impact AI in Anaesthesia: Monitoring and Outcomes Review. Journal of Future Medicine and Healthcare Innovation, 4(1), 1–20.

Cascella, M. (2025). The complex task of modelling artificial intelligence workflows for forecasting postoperative risk. Journal of Anesthesia Analgesia and Critical Care, 5(1), 82–82.

Chaari, N., Winski, G., Hallbäck, M., Lundström, N., Björne, H., & Jacobsson, M. (2025). Towards reliable prediction of intraoperative hypotension: a cross-center evaluation of deep learning-based and MAP-derived methods. Journal of Clinical Monitoring and Computing.

Char, D., & Burgart, A. M. (2020). Machine-Learning Implementation in Clinical Anesthesia: Opportunities and Challenges. Anesthesia & Analgesia, 130(6), 1709–1712.

Chen, M., & Zhang, D. (2024). Machine Learning-Based Prediction of Post-Induction Hypotension: Identifying Risk Factors and Enhancing Anesthesia Management. Research Square (Research Square).

Choe, S., Park, E., Shin, W.-S., Koo, B., Shin, D. J., Jung, C.-W., Lee, H., & Kim, J. (2021). Short-Term Event Prediction in the Operating Room (STEP-OP) of Five-Minute Intraoperative Hypotension Using Hybrid Deep Learning: Retrospective Observational Study and Model Development. JMIR Medical Informatics, 9(9).

Dost, B., Turan, E. İ., Aydın, M. E., Ahişkalıoğlu, A., Narayanan, M., Yılmaz, R., & Cassai, A. D. (2025). Artificial Intelligence in Anaesthesiology: Current Applications, Challenges, and Future Directions. Türk Anesteziyoloji ve Reanimasyon Derneği Dergisi, 53(6), 282–292.

Garmendia, A. T., Gkouzionis, I., Triantafyllidis, C., Dimakopoulos, V., Liliopoulos, S. G., Vuckovic, D., Paseiro-Garcia, L., & Chadeau-Hyam, M. (2023). Towards personalised early prediction of Intra-Operative Hypotension following anesthesia using Deep Learning and phenotypic heterogeneity. medRxiv (Cold Spring Harbor Laboratory).

Giordano, C., Brennan, M., Mohamed, B., Rashidi, P., Modave, F., & Tighe, P. (2021). Accessing Artificial Intelligence for Clinical Decision-Making [Review of Accessing Artificial Intelligence for Clinical Decision-Making]. Frontiers in Digital Health, 3. Frontiers Media.

Hao, X., Wang, Y., Li, K., Zhu, T., & Herasevich, V. (2025). Applying machine learning for perioperative adverse event prediction: a narrative review toward better clinical efficacy and usability [Review of Applying machine learning for perioperative adverse event prediction: a narrative review toward better clinical efficacy and usability]. Anesthesiology and Perioperative Science, 3(4). Springer Science+Business Media.

Hofer, I., Burns, M., Kendale, S., & Wanderer, J. P. (2020). Realistically Integrating Machine Learning Into Clinical Practice: A Road Map of Opportunities, Challenges, and a Potential Future. Anesthesia & Analgesia, 130(5), 1115–1118.

Hofer, I., Lee, C., Gabel, E., Baldi, P., & Cannesson, M. (2020). Development and validation of a deep neural network model to predict postoperative mortality, acute kidney injury, and reintubation using a single feature set. Npj Digital Medicine, 3(1).

Huang, L., Sun, Q., Ma, Y., & Yang, H. (2026). Artificial intelligence in anesthesiology: Clinical decision support, challenges, and future directions. Journal of Clinical and Experimental Investigations, 17(2).

Jia, Y., Pei, Z., Zhang, X., Zhang, C., & Kang, X. (2025). Risk prediction models for prolonged mechanical ventilation following coronary artery bypass grafting surgery: a systematic review and meta-analysis [Review of Risk prediction models for prolonged mechanical ventilation following coronary artery bypass grafting surgery: a systematic review and meta-analysis]. Frontiers in Cardiovascular Medicine, 12. Frontiers Media.

Kim, R. B., Alge, O., Liu, G., Biesterveld, B. E., Wakam, G. K., Williams, A. M., Mathis, M. R., Najarian, K., & Gryak, J. (2022). Prediction of postoperative cardiac events in multiple surgical cohorts using a multimodal and integrative decision support system. Scientific Reports, 12(1).

Le, J., Rusin, W., Joosten, A., Cannesson, M., & Pal, R. (2026). Machine learning for hemodynamic instability prediction and hemorrhage management in trauma and perioperative care. Current Opinion in Anaesthesiology.

Le, L., Elvir-Lazo, O. L., & Wong, R. (2025). A Scoping Review of Artificial Intelligence in Perioperative Anesthesia: Current Applications, Challenges, and Roadmap for the Future. Journal of Surgery and Research, 8(4).

Lee, C. K., Hofer, I., Gabel, E., Baldi, P., & Cannesson, M. (2018). Development and Validation of a Deep Neural Network Model for Prediction of Postoperative In-hospital Mortality. Anesthesiology, 129(4), 649–662.

Lee, C. K., Samad, M., Hofer, I., Cannesson, M., & Baldi, P. (2021). Development and validation of an interpretable neural network for prediction of postoperative in-hospital mortality. Npj Digital Medicine, 4(1).

Lopes, S., Rocha, G., & Guimarães-Pereira, L. (2023). Artificial intelligence and its clinical application in Anesthesiology: a systematic review [Review of Artificial intelligence and its clinical application in Anesthesiology: a systematic review]. Journal of Clinical Monitoring and Computing, 38(2), 247–259. Springer Science+Business Media.

Luo, J., Lin, S., Wang, L., Ji, H., Zheng, J.-S., Wang, T., Chen, L., Lin, Z., Liu, Z., & Liufu, N.

(2025). Predicting Delayed Extubation After General Anesthesia in Postanesthesia Care Unit Patients Using Machine Learning: Model Development Study. JMIR Medical Informatics, 13.

Ma, X., Zhao, J., Yan, M., Zhang, Y., Li, F., Jia, J., & Ding, L. (2025). Transformer-based multimodal precision intervention model for enhancing diaphragm function in elderly patients. Frontiers in Computational Neuroscience, 19.

Mahajan, A., Esper, S. A., Oo, T. H., McKibben, J., Garver, M. S., Artman, J., Klahre, C., Ryan, J. P., Sadhasivam, S., Holder-Murray, J., & Marroquin, O. C. (2023). Development and Validation of a Machine Learning Model to Identify Patients Before Surgery at High Risk for Postoperative Adverse Events. JAMA Network Open, 6(7).

Maroufi, S. S., Movahed, M. S., Ejmalian, A., Sarkhosh, M., & Behmanesh, A. (2025). Post-Anesthesia Care Unit (PACU) readiness predictions using machine learning: a comparative study of algorithms. BMC Medical Informatics and Decision Making, 25(1).

Mevik, K., Woldaregay, A. Z., Jonsson, E. L., Tejedor, M., & Temple-Oberle, C. (2026). Application of AI Models for Preventing Surgical Complications: Scoping Review of Clinical Readiness and Barriers to Implementation. JMIR AI, 5.

Mirza, M. A., Mehtab, M., Islam, M. B., & Shahbaz, S. (2025). Artificial Intelligence for Perioperative Medicine in Anesthesia: Present State and Future Perspectives of AI in Anesthesiology. The Healer Journal of Physiotherapy and Rehabilitation Sciences, 5(1), 46–59.

Qiu, X., Hu, S., Dong, S., & Sun, H. (2025). Construction of an automated machine learning-based predictive model for postoperative pulmonary complications risk in non-small cell lung cancer patients undergoing thoracoscopic surgery. PLoS ONE, 20(9).

Saran, E. (2025). When Algorithms Meet Anesthesia: A New Era of Patient Safety. University of Toronto Medical Journal, 102(2).

Shang, H., Chu, Q., Ji, M., Guo, J., Ye, H., Zheng, S., & Yang, J. (2022). A retrospective study of mortality for perioperative cardiac arrests toward a personalized treatment. Scientific Reports, 12(1).

Shen, J., Xue, B., Kannampallil, T., Lu, C., & Abraham, J. (2024). A novel generative multi-task representation learning approach for predicting postoperative complications in cardiac surgery patients. Journal of the American Medical Informatics Association.

Shickel, B., Loftus, T. J., Ruppert, M. M., Upchurch, G. R., Ozrazgat-Baslanti, T., Rashidi, P., & Bihorac, A. (2023). Dynamic predictions of postoperative complications from explainable, uncertainty-aware, and multi-task deep neural networks. Scientific Reports, 13(1).

Shirkavand, R., Zhang, F., & Huang, H. (2023). Prediction of Post-Operative Renal and Pulmonary Complications Using Transformers. arXiv (Cornell University).

Shu, X., Zhu, Y., Liu, X., Li, Y., Yi, B., & Wang, Y. (2025). Applications of artificial intelligence in anesthesiology. Anesthesiology and Perioperative Science, 3(4).

Te, R., Zhu, B., Ma, H., Zhang, X., Chen, S., Huang, Y., & Qi, G. (2024). Machine learning approach for predicting post-intubation hemodynamic instability (PIHI) index values: towards enhanced perioperative anesthesia quality and safety. BMC Anesthesiology, 24(1).

Wang, Y., Zhu, S., Liu, X., Zhao, B., Xiu, Z., Luo, Z., Liu, P., Guo, Y., Zhang, Z., & Yu, P. (2024). Linking preoperative and early intensive care unit data for prolonged intubation prediction. Frontiers in Cardiovascular Medicine, 11.

Xue, B., Li, D., Lu, C., King, C. R., Wildes, T. S., Avidan, M. S., Kannampallil, T., & Abraham, J. (2021). Use of Machine Learning to Develop and Evaluate Models Using Preoperative and Intraoperative Data to Identify Risks of Postoperative Complications. JAMA Network Open, 4(3).

Yoon, H., Kim, B. R., Kim, H. S., Park, D. K., Kim, H. S., Cho, H.-Y., Lee, H., & Lee, H. (2025). Multicenter validation of a scalable, interpretable, multitask prediction model for multiple clinical outcomes. Npj Digital Medicine, 8(1).

Yoon, H., Yang, H.-L., Jung, C.-W., & Lee, H. (2022). Artificial intelligence in perioperative medicine: a narrative review. Korean Journal of Anesthesiology, 75(3), 202–215.

Yoosoofsah, A. (2024). Predicting Extubation Failure in Intensive Care: The Development of a Novel, End-to-End Actionable and Interpretable Prediction System. arXiv (Cornell University).

Zaouter, C., Joosten, A., Rinehart, J., Struys, M., & Hemmerling, T. M. (2020). Autonomous Systems in Anesthesia: Where Do We Stand in 2020? A Narrative Review [Review of Autonomous Systems in Anesthesia: Where Do We Stand in 2020? A Narrative Review]. Anesthesia & Analgesia, 130(5), 1120–1132. Lippincott Williams & Wilkins.

Zeng, Z., Tang, X., Liu, Y., He, Z., & Gong, X. (2022). Interpretable recurrent neural network models for dynamic prediction of the extubation failure risk in patients with invasive mechanical ventilation in the intensive care unit. BioData Mining, 15(1).

Zhang, Z.-J., Liu, Z., Cheng, M., Zhang, S., Pan, T., Zhou, Y., Liu, Q., & Xie, Y. (2025). Multimodal Forecasting of Sparse Intraoperative Hypotension Events Powered by Language Model. arXiv (Cornell University).