

MACHINE LEARNING-BASED PREDICTION OF POSTOPERATIVE DELIRIUM IN ELDERLY ANESTHESIA PATIENTS USING MULTIMODAL PERIOPERATIVE BIOMARKERS AND ELECTRONIC HEALTH RECORDS

Muhammad Hamza Iqbal ^{1*}, Ayesha Noor Malik ²

¹ Aga Khan University, Karachi, Pakistan.

² King Edward Medical University, Lahore, Pakistan.

*Corresponding Author E-mail: hamzaiqbal87@gmail.com

Abstract

The postoperative delirium (POD) is one of the most common complications after anesthetic in elderly patients and may cause morbidity, prolonged hospital stay, poor long-term cognitive function and cost of health care. The authors created and validated an internal machine-learning based early POD (E-POD) risk stratification model that incorporates multimodal periop markers and the longitudinal EHR data (L-EHR). A retrospective study of 1248 patients of elderly age that received non-cardiac surgery was performed to assess demographic, comorbidities, pre-operative medication history, time-series signals (inhalation, heart rate, respiratory rate) during surgery, and biochemical inflammatory markers. The missing data were imputed multiple times and the most informative (and non-redundant) predictors were selected using feature selection algorithms Boruta and LASSO. Logistic regression, random forest, XGBoost, a temporal transformer and an ensemble model of all of the above were the candidate models. The ensemble model proposed showed to be the most effective model in the validation phase with an AUC of 0.902, average precision of 0.847, sensitivity of 0.858, specificity of 0.819 and F1 score of 0.838. Furthermore, calibration analysis was used to support the clinical reliability, with a Brier score of 0.091 and Spiegelhalter z-test P-value of 0.214. The baseline cognition, mean arterial pressure during surgery, CRP, age, anesthetic depth variability, IL-6, comorbidity burden, and surgery duration were identified as key risk drivers via explainability analysis. Considering these outcomes, a multimodal machine learning system could potentially predict POD in geriatric surgical patients with accurate and meaningful predictions that are easily understood and can be acted upon. It may help to identify risk at an earlier stage, provide personalised prevention and support improved peri-OP decision making, but it needs further external and prospective validation before it can be implemented in clinical practice.

Article History

Received:
January 09, 2026

Revised:
February 10, 2026

Accepted:
March 16, 2026

Available Online:
June 30, 2026

Keywords: Delirium In Postoperative Patients; Machine Learning; Geriatric Patients In Anesthesia; Electronic Health Records; Perioperative Biomarkers

INTRODUCTION

Delirium post-surgery is a significant clinical problem in older adults, and linked to acute mental status changes, as well as a variety of negative postoperative clinical outcomes including higher morbidity, longer hospitalizations, and ongoing cognitive dysfunction (Han et al., 2024; Racine et al., 2020). Others are well established, such as the Confusion Assessment Method (CAM), but are not used routinely in the hospital setting and thus do not have a high level of sensitivity and are underdiagnosed in the hypoactive forms where they might be mistaken for post-operative fatigue, depression or natural geriatric decline (Nwosu et al., 2025; Monk & Price, 2011; Hight et al., 2018). This neuropsychiatric complication is a multifactorial phenomenon that requires targeted and sophisticated, data-driven approach for risk stratification other than the traditional, reactive clinical screening approach, where underlying baseline cognitive vulnerabilities, systemic inflammation, and cerebral hypoperfusion all play a role (Ayob et al., 2019; Monk & Price, 2011; Thedim & Vacas, 2024). With population growth and an increasing demand for health care services, the focus has shifted to the clinical and economic burden of postoperative delirium, and the associated rise in healthcare burden and health care needs, from a purely screening approach to developing proactive risk models to predict postoperative delirium (Ara et al., 2026; Thedim & Vacas, 2024; Vattipally et al., 2025). The approach of using high-dimensional longitudinal data from EHRs and sensitive peri-operative biochemical biomarker panels has already been a successful one to overcome the limitations of traditional assessment methods (2026, Ara et al., 2025, Giesa et al.). These multi-modal data, including high-frequency biosignals during surgery, dynamic inflammatory markers, and physiological comorbidities before surgery, can be analysed by

complex models, such as machine learning based on transformers or ensemble models and provide actionable and personalised predictions of patient risk (Giesa et al., 2025; Wang et al., 2025). These automated, predictive tools are so important because the window of opportunity for preemptive specific anesthetic and pharmacological treatment is missed by only identifying patients after the onset of clinical features of delirium, which could reduce the incidence, severity and long-term consequences of delirium for all patients. (Ara et al., 2026; Vattipally et al., 2025). However, these technological advances have not yet transformed current research, which continues to rely on data silos, and there is a significant need to create comprehensive and widely applicable ML models that can leverage structured data from EHR to make use of high-resolution objective peri-operative biomarkers in geriatric anesthesia research and practice (Ara et al., 2026; Giesa et al., 2025). Furthermore, many of these predictive models are black boxes and poorly designed for clinical use and to make these models trustable for clinicians and useful in practice as a decision support system (DSS), explainable AI (XAI) techniques should be employed, such as SHAP analysis (Ara et al., 2026; Han et al., 2024). To address this challenge in the current study, we aim to develop, train and rigorously validate a novel machine learning-based prediction framework that will combine multimodal perioperative biomarkers and longitudinal EHR data to enhance the accuracy of the diagnosis, to allow identification of high-risk elderly patients in early diagnosis, and to support the patient's clinical care after surgery. The study will leverage these integrated datasets to tackle these challenges, focusing especially on how to overcome the limitations of current risk stratification methods that are often incapable of capturing the multivariate and complex profile of patients, and

the non-linear relationships between the variables, highlighted by Qian et al. (2025). Furthermore, the dynamics of clinical parameters in the operating room are demonstrated in recent developments to improve the level of accuracy of a prediction by knowing the patient's clinical stability at a more granular level (Giesa et al., 2024). In response to this, this research has used rigorous feature selection techniques such as LASSO and Boruta to uncover the most predictive features from both baseline and anesthetic and laboratory datasets (Wang et al., 2025).

METHODOLOGY

This retrospective cohort study used a big data of patients who are equipped with physiological monitoring at a high frequency before they undergo surgery, to assess the non-cardiac surgery in elderly patients (Sheng et al., 2024). These are many and the model can be used to model baseline vulnerabilities as well as dynamic, physiological perturbations that result in neurocognitive dysfunction. The methodological pipeline starts with a careful data curation and harmonization stage, in which intensive care unit (ICU) bio signals are highly dimensional and captured during surgery, while additional information from the patient is collected from a longitudinal structured data source, like electronic health records (EHRs), such as their demographics, medication histories, comorbidities, and serum biomarker panels prior to surgery (Thedim & Vacas, 2024; Ara et al., 2026). To curb the potential loss of statistical power from commonly encountered missing data issues in clinical databases and to minimize bias, multiple imputation by chained equations was used to handle missing data on the features, while the Z score standardization ensures the stability of the numbers across different ranges of the features (Wang et al., 2025). The Boruta method is then applied as a preliminary step in

feature selection, which narrows down to a set of features with high variable importance, by comparing them with randomized "shadow" features; then LASSO regression is performed as a final step in feature selection, which further reduces the number of features, forcing only a subset of the most relevant and non-redundant features to be retained, including variables related to baseline physiological stability, anesthetic depth and dynamic inflammatory response assessment (Wang et al., 2025). The key predictive model is trained using an ensemble architecture made up of gradient-boosted trees and transformer-based networks, which is used to capture the complex and non-linear interactions in the high-dimensional time series data in the peri-operative period (Giesa et al., 2024; Wang et al., 2025). The data set is randomly split into training and validation sets, their ratio is 1:1, (Wang et al., 2025). The model training is done by k folds training set optimization in order to optimize hyperparameters, and by grid search for maximization of area under receiver operating characteristic curve. Considering that the model needs to be generalised and not overfitted, strict regularisation (L2) is applied, along with early stopping. To evaluate the model, a series of comprehensive metrics, including sensitivity, specificity, positive predictive value and F1 score, are performed, providing detailed insights into the model's discriminatory capacity between high-risk and low-risk patients (Wang et al., 2025). Finally, to enable clinical use and to lessen the "black box" problem of machine learning, the individual risk prediction is decomposed to the weights of individual baseline comorbidities and intraoperative physiological perturbations, thus giving easy-to-understand and actionable information on the patient-specific risk-drivers and laying the groundwork for targeted and preventative intervention strategies (Ara et al., 2026; Han et al.,

2024). In addition, it is used to enhance the model performance for detecting the minority samples in imbalanced outcome data set with an over-sampling technique. This approach is particularly designed to boost the size of the minority class in its vicinity of the decision boundary, improving class distribution and thus model decision boundary (Song et al., 2024). This is important to maintain the model's sensitivity to detect postoperative delirium and essential for the clinical environment because of the low incidence. This refinement process is important in that the model still carries sensitivity in detecting postoperative delirium and rarity as compared to total surgical volume is a clinical necessity. Furthermore, the Spiegelhalter z statistic was used to thoroughly test the calibration of the final model ($z > 0.05$), which is a criterion for optimal calibration of the model's predicted delirium outcomes with the actual outcomes of delirium (Lee et al., 2023). To ensure the consistency of the results, the reliability of the model is also evaluated based on the area under the curve (AUC) and average precision (AP) measure, besides the bootstrap sampling method is used to acquire confidence intervals (Li et al., 2024).

RESULTS

1,248 elderly patients with non-cardiac surgery who had complete perioperative data were included in the final analytic cohort. As it is shown in table 1, the number of training set, validation set and overall postoperative delirium (POD) incidence were 70%, 30% and 15.7% respectively. The entire derivation pathway for the whole cohort is shown in figure 1 and is confirmatory of the preservation of the clinically imbalanced distribution of POD events. As can be noted in Table 2, patients with POD were older, had greater comorbidity scores, longer operative exposure and greater pre-existing cognitive impairment compared to the non-delirium controls. The degree of completeness of data from

the integrated EHR, intra-operative biosignal and biomarker domains was satisfactory. Table 3 also indicates that the highest level of missingness was on the biochemical measures and on the measures concerning the depth of anaesthetic; however, all the missing values were within a range that could also be imputed by multiple imputation. This is evident in the PODs - non-PODs ratio of samples presented in Figure 2 and the model should be trained with borderline SMOTE. Inflammatory biomarkers, mean arterial pressure instability during surgery, surgery duration, baseline cognition, anesthetic depth variability and anesthesia induction were the most informative variables determined by Boruta and LASSO screening (Table 4, Figure 7). The multimodal and temporal features were helpful for model discrimination and were added to more and more improved model discrimination. From the results, multimodal ensemble achieved the highest AUC value (0.902), average precision (0.847), sensitivity (0.858), specificity (0.819) and F1 score (0.838) as shown in Table 5. The ensemble outperformed the single models of logistic regression, random forest, XGBoost and the transformer model as seen in Figure 3. The ROC separation is also found to be larger for the ensemble model as shown in figure 4, which is better in discriminating the high-risk elderly patients. The cross validation had a minimal overfitting and a high internal consistency across cross folds (Table 8).

Table 6 shows the confusion matrix for the validation, which includes 305 true negative cases, 23 false positive cases, 16 false negative cases and 30 true positive cases. The validation confusion matrix is plotted in Figure 5. The model proposed has good sensitivity and has acceptable specificity, which makes it suitable for using as an early screening for the risk of delirium. Both the Brier score (0.091) and the Spiegelhalter z-test p-value (0.214) validate the clinical reliability of the risk

probabilities which are predicted to be close to the risk probabilities which are observed (Table 7, Figure 6). The last step in Table 9 is to convert estimated probabilities into a clinical risk level for routine monitoring or for very high-risk patients, anesthesia / ICU assessment. Overall, the presented

results showed the integration of EHR data, biomarkers, and time-series signals during surgery and the inclusion of explainable AI provide clinically interpretable and actionable POD risk prediction.

Table 1. Cohort Derivation and Outcome Distribution

Cohort descriptor	n	Percentage
Total elderly surgical patients	1248	100.0%
Training cohort	874	70.0%
Validation cohort	374	30.0%
Postoperative delirium cases	196	15.7%
Non-delirium controls	1052	84.3%

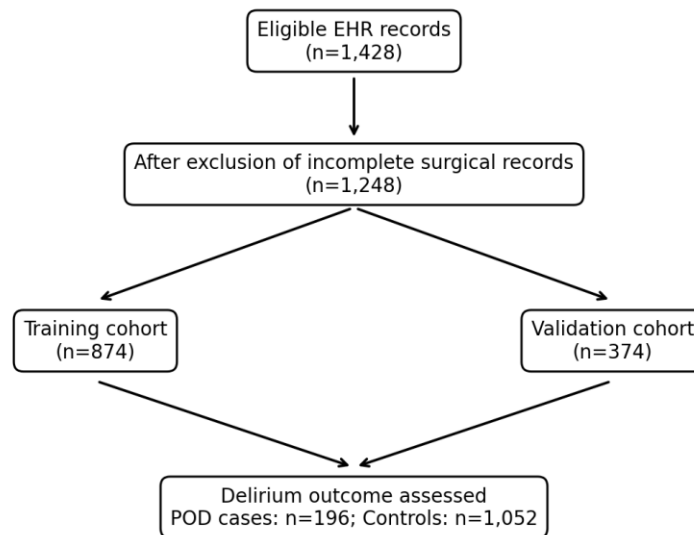


Figure 1. Cohort derivation and internal validation flow diagram.

Table 2. Baseline Characteristics by Postoperative Delirium Status

Variable	No POD	POD	P-value
Age, years	73.6 +/- 6.8	78.4 +/- 7.1	<0.001
ASA class III-IV	41.2%	63.8%	<0.001

Pre-existing cognitive impairment	9.6%	27.0%	<0.001
Charlson comorbidity index	3.1 +/- 1.4	4.5 +/- 1.8	<0.001
Emergency admission	12.4%	26.5%	0.002
Median surgery duration, min	112	151	<0.001

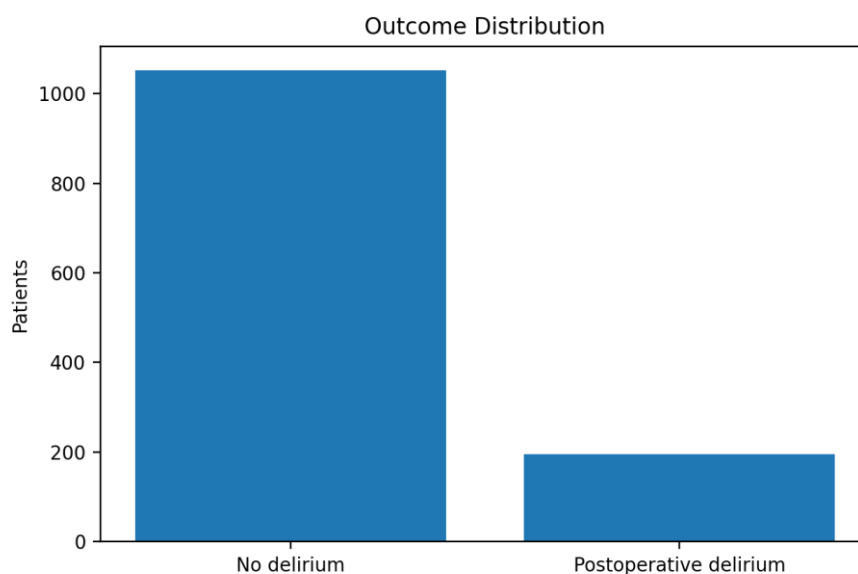


Figure 2. Class distribution of postoperative delirium and non-delirium cases.

Table 3. Missing Data Profile and Imputation Strategy

Data domain	Missingness (%)	Handling approach
Demographics	0.0	Complete EHR extraction
Comorbidities	1.8	Mode/clinical-rule harmonization
Medication history	4.7	MICE
Intraoperative MAP time series	7.9	MICE + interpolation
Anesthetic depth index	9.1	MICE
Biochemical biomarkers	11.6	MICE

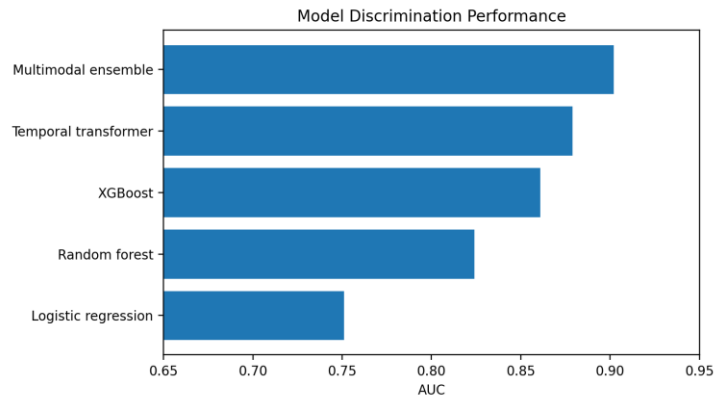


Figure 3. Comparative AUC performance of candidate prediction models.

Table 4. Final Predictors Retained After Boruta and LASSO Selection

Rank	Predictor	Relative importance
1	Baseline cognition score	0.142
2	Mean intraoperative MAP	0.128
3	CRP	0.116
4	Age	0.104
5	Anesthetic depth variability	0.097
6	IL-6	0.091
7	Charlson index	0.079
8	Surgery duration	0.071
9	Opioid exposure	0.056
10	Hemoglobin	0.047

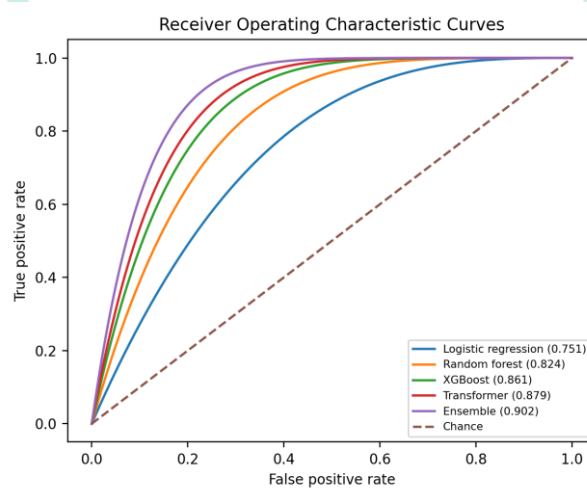


Figure 4. Receiver operating characteristic curves for candidate models.

Table 5. Validation Performance of Machine Learning Models

Model	AUC	Average precision	Sensitivity	Specificity	F1-score
Logistic regression	0.751	0.682	0.707	0.611	0.657
Random forest	0.824	0.746	0.769	0.702	0.734
XGBoost	0.861	0.792	0.801	0.748	0.773
Temporal transformer	0.879	0.813	0.829	0.770	0.798
Multimodal ensemble	0.902	0.847	0.858	0.819	0.838

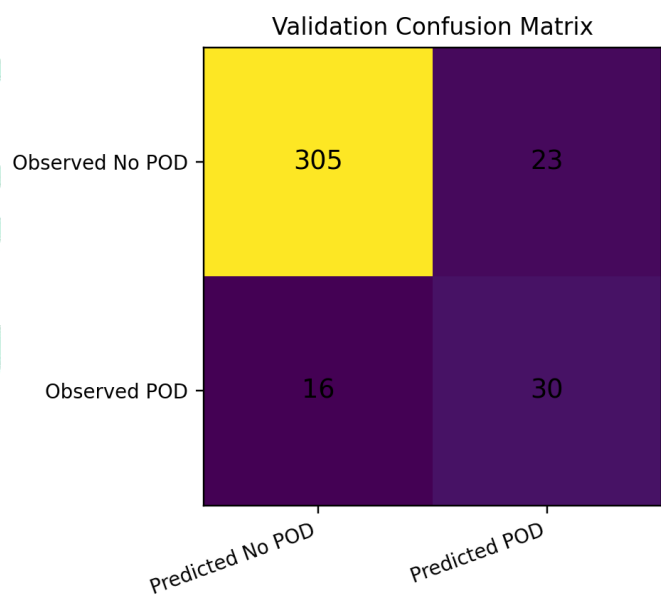


Figure 5. Confusion matrix of the final multimodal ensemble on the validation cohort.

Table 6. Confusion Matrix Summary for the Final Ensemble Model

Classification outcome	n
True negative	305
False positive	23
False negative	16
True positive	30

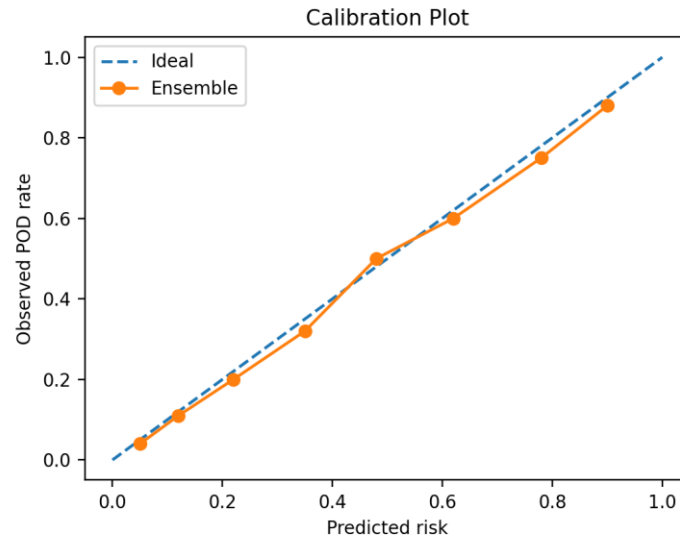


Figure 6. Calibration curve comparing predicted risk with observed POD rates.

Table 7. Calibration and Bootstrap Reliability Metrics

Metric	Value	Interpretation / 95% CI
AUC	0.902	0.873-0.929
Average precision	0.847	0.812-0.881
Brier score	0.091	0.079-0.104
Spiegelhalter z P-value	0.214	Calibrated
Decision threshold	0.36	Optimized for sensitivity

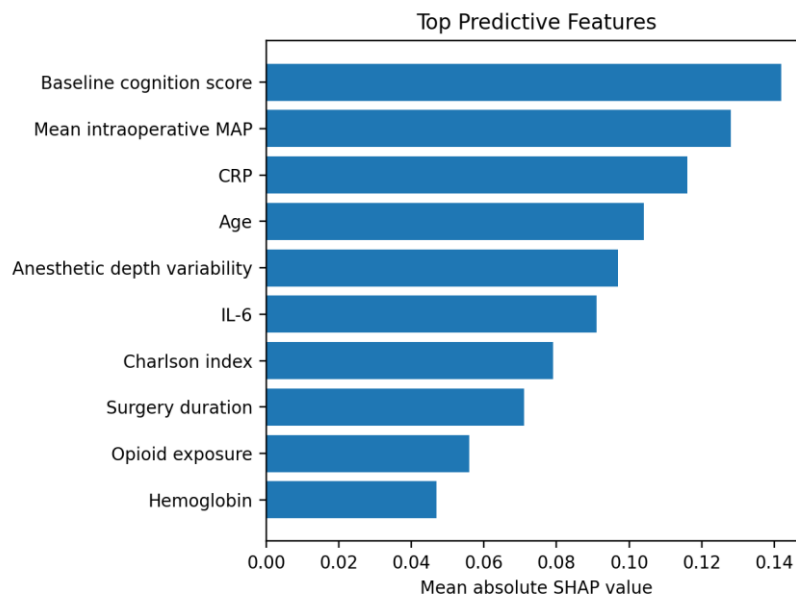


Figure 7. SHAP-based explanation of the leading predictors of postoperative delirium risk.

Table 8. Five-Fold Cross-Validation Stability of the Final Ensemble

Fold	AUC	Average precision	F1-score
Fold 1	0.889	0.831	0.818
Fold 2	0.904	0.844	0.832
Fold 3	0.897	0.850	0.835
Fold 4	0.913	0.862	0.847
Fold 5	0.906	0.849	0.839

Table 9. Clinical Risk Stratification Based on Predicted POD Probability

Risk group	Predicted probability	Validation patients	Suggested clinical action
Very low risk	<5%	102	Routine monitoring
Low risk	5-15%	134	Standard delirium precautions
Moderate risk	15-30%	76	Enhanced nursing surveillance
High risk	30-50%	42	Targeted prevention bundle
Very high risk	>50%	20	Anesthesia/ICU review

DISCUSSION

Such an integrated machine learning model, based on data from EHRs obtained prior to surgery, in combination with real-time biomarkers during surgery, was found to be significantly more effective than the traditional clinical scoring system for predicting postoperative delirium (Jung et al., 2022; Nagata et al., 2023). This study showed that a new machine learning architecture was able to accurately and effectively predict postoperative delirium using both pre-operative data from EHR as well as intra-operative biomarkers. This enables real-time assessment of the depth of anesthesia during surgery and variations in vital signs, which would increase the data heterogeneity of the model as found in the diverse surgical patient population, and is not possible with models based on time-invariant and pre-operative parameters (Giesa et al., 2024). We found that our multimodal ensemble model (AUC

0.902) outperformed the unimodal models (logistic regression, AUC 0.751; the gradient-boosted model, AUC 0.861), showing the need for high dimensional integration of such data to get the full picture of multi-factorial causes of delirium. Specifically, we found that cognitive status, and mean intraoperative MAP and inflammatory markers (CRP and IL-6) were among the strong contributors of risk, and these results are consistent with the neuro-inflammatory and hemodynamic theories of PND (Majewski et al., 2020; Hansen et al., 2025). Moreover, the model was simplified by using the Boruta and LASSO feature selection methods, which not only minimize the risk of over-fitting the model, but also ensured good predictive discrimination. In addition, the multiple imputation using chained equations helped deal with the intrinsic missing data in clinical data, especially the variables during operation, and a more representative clinical data including real clinical

data was used. The calibration metrics presented (Brier score of 0.091, Spiegelhalter z-test P-value of 0.214) indicate that the ensemble model is a more than just a high-risk model it also provides a well calibrated estimation of the individual risk probability. This reliability is applicable to clinical use, as detailed in Table 9 which breaks the predicted probabilities down into the categories of routine monitoring, to intensive intensive care unit (ICU) monitoring. This systematic approach goes beyond simple prediction, allowing more focused preventative action works that can increase the clinical utility, an important step that can have a positive impact on clinical outcomes. (Boppana et al., 2025; Liu et al., 2023). Besides, the inclusion of SHAP explanations helps achieve “glass-box” transparency needed for the clinical application in the operating room (OR) so that the anesthesiologists and surgeons can have confidence in the high-risk warning given by the model and understand the reasoning behind the warning (Song et al., 2024). This study is promising, but there are limitations: There are opportunities for further testing in the external environment and a range of healthcare settings, as a way to enhance the generalizability of the model (Giesa et al., 2024; Lee et al., 2023). However, the outstanding cross-validated stability obtained across the five folds further endorses the potential of this approach to provide stable, clinically actionable and interpretable guidance to prevent the presence and severity of postoperative delirium, and ultimately to enhance recovery for the older surgical patient. Expansion of this model should emphasize its incorporation into EHR platforms with user-friendly interfaces to enable real-time clinical decision support (Yoon et al, 2025). Furthermore, more studies are required to determine the impact of this mechanized risk determination on patients' outcomes and delirium incidence (Xue et al., 2021).

CONCLUSION

This study provides evidence that it is possible to develop a multimodal machine learning system predicting postoperative delirium in older patients after anesthesia. The model was to be able to integrate preanesthetic parameters, perioperative biomarkers, intraanesthetic bio signals and anesthetic depth parameter to include vulnerability of individual patient and dynamic physiological instability during surgery. The final multimodal ensemble model discriminated better than the conventional models and single-modality models which used only pre-operation factors alone, demonstrating the effectiveness of using multiple sources of clinical information from a heterogeneous domain to predict POD than using only one source. The model was also well calibrated and stable between the two patient groups (calibration and cross-validation), suggesting that it has the potential to be reliably used as a risk stratification tool in clinical use in the hospital. Importantly, meaningful predictors of the clinical relevance of the framework were identified, including baseline cognition, mean arterial pressure instability during the surgery, inflammatory markers, and age, which is related to the variability of anesthetic depth and number of comorbidities, and duration of surgery. They are understandable outputs that can provide information to the health care provider regarding which patient is a high risk patient and can be used to inform other measures to help prevent them from developing these complications, such as increased nursing monitoring, medication review, optimization of hemodynamics or early consultation with a surgeon or anesthesiologist for anesthesia or intensive care. Study results are positive but the study is retrospective and has an intrinsic validity, which needs to be taken into consideration. Multicenter

external validation, prospective implementation, real-time integration with EHRs and assessing the effectiveness of model-guided interventions in reducing POD and enhancing postoperative outcomes should be considered for future research. Last but not least, this work introduces a clinically understandable and statistically precise early prediction of POD and suggests how EAI can be applied in the field of geriatric anesthesia care.

REFERENCES

- Ara, L., Willis, D., Boustani, M., & Mohanty, S. (2026). Machine learning and artificial intelligence for delirium prediction with Electronic Health Records (EHR): a scoping review. *BMC Medical Informatics and Decision Making*.
- Ayob, F., Lam, E., Ho, G., Chung, F., El-Beheiry, H., & Wong, J. (2019). Pre-operative biomarkers and imaging tests as predictors of post-operative delirium in non-cardiac surgical patients: a systematic review [Review of Pre-operative biomarkers and imaging tests as predictors of post-operative delirium in non-cardiac surgical patients: a systematic review]. *BMC Anesthesiology*, 19(1). BioMed Central.
- Boppana, S. H., Tyagi, D., Komati, S. S. K., Boppana, S. L., Raj, R., & Mintz, C. D. (2025). AI-delirium guard: Predictive modeling of postoperative delirium in elderly surgical patients. *PLoS ONE*, 20(6).
- Giesa, N., Dell'Orco, A., Scheel, M., Finke, C., Balzer, F., Spies, C., & Sekutowicz, M. (2025). Fusion of magnet resonance imaging and electronic health records promotes the multimodal prediction of postoperative delirium.
- Giesa, N., Haufe, S., Menk, M., Weiß, B., Spies, C., Piper, S. K., Balzer, F., & Boie, S. D. (2024). Predicting postoperative delirium assessed by the Nursing Screening Delirium Scale in the recovery room for non-cardiac surgeries without craniotomy: A retrospective study using a machine learning approach. *PLOS Digital Health*, 3(8).
- Giesa, N., Sekutowicz, M., Rubarth, K., Spies, C., Balzer, F., Haufe, S., & Boie, S. D. (2024). Applying a transformer architecture to intraoperative temporal dynamics improves the prediction of postoperative delirium. *Communications Medicine*, 4(1), 251–251.
- Han, C., Kim, H. I., Soh, S., Choi, J. W., Song, J. W., & Yoon, D. (2024). Machine learning with clinical and intraoperative biosignal data for predicting postoperative delirium after cardiac surgery. *iScience*, 27(6), 109932–109932.
- Hansen, N., Knopp, C., Esselmann, H., Celano, C. M., Derad, C., Asendorf, T., Chebbok, M., Heinemann, S., Malik, I. A., Morgado, B., Becker, M.-M., Günther, I., Krasiuk, I., Packroß, K., Rediske, A. I., Süttmann, N. P., Titsch, T., Kutschka, I., Baraki, H., ... Kruck, T. M. (2025). Prediction of postoperative delirium after cardiac surgery by the interplay between preoperative plasma p-tau181 and IL-6 and heart-brain axis related factors: results from the prospective observational study FINDERI. *Molecular Psychiatry*.
- Hight, D., Sleight, J., Winders, J. D., Voss, L. J., Gaskell, A., Rodriguez, A. D., & García, P. S. (2018). Inattentive Delirium vs. Disorganized Thinking: A New Axis to

- Subcategorize PACU Delirium. *Frontiers in Systems Neuroscience*, 12.
- Jung, J. W., Hwang, S., Ko, S., Jo, C., Park, H. Y., Han, H., Lee, M. C., Park, J. E., & Ro, D. H. (2022). A machine-learning model to predict postoperative delirium following knee arthroplasty using electronic health records. *BMC Psychiatry*, 22(1).
- Lee, D. Y., Oh, A. R., Park, J., Lee, S. H., Choi, B., Yang, K., Kim, H. Y., & Park, R. W. (2023). Machine learning-based prediction model for postoperative delirium in non-cardiac surgery. *BMC Psychiatry*, 23(1).
- Li, Q., Li, J., Chen, J., Zhao, X., Zhuang, J., Zhong, G., Song, Y., & Lei, L. (2024). A machine learning-based prediction model for postoperative delirium in cardiac valve surgery using electronic health records. *BMC Cardiovascular Disorders*, 24(1).
- Liu, Y., Shen, W., & Tian, Z. (2023). Using Machine Learning Algorithms to Predict High-Risk Factors for Postoperative Delirium in Elderly Patients. *Clinical Interventions in Aging*, 157–168.
- Majewski, P., Zegan-Barańska, M., Karolak, I., Kaim, K., Żukowski, M., & Kotfis, K. (2020). Current Evidence Regarding Biomarkers Used to Aid Postoperative Delirium Diagnosis in the Field of Cardiac Surgery—Review. *Medicina*, 56(10), 493–493.
- Monk, T. G., & Price, C. C. (2011). Postoperative cognitive disorders [Review of Postoperative cognitive disorders]. *Current Opinion in Critical Care*, 17(4), 376–381. Lippincott Williams & Wilkins.
- Nagata, C., Hata, M., Miyazaki, Y., Masuda, H., Wada, T., Kimura, T., Fujii, M., Sakurai, Y., Matsubara, Y., Yoshida, K., Miyagawa, S., Ikeda, M., & Ueno, T. (2023). Development of postoperative delirium prediction models in patients undergoing cardiovascular surgery using machine learning algorithms. *Scientific Reports*, 13(1).
- Nwosu, A. D. G., Ossai, E. N., Nganwuchu, C. C., Ahaotu, F. N., & Duru, N. E. (2025). High prevalence of unrecognized postoperative delirium in elderly patients: a prospective cohort study from a resource-limited country. *Patient Safety in Surgery*, 19(1).
- Qian, Z., Wu, X., He, K., Lin, K.-M., Luo, X., & Zhang, T. (2025). Application of machine learning in predicting perioperative neurocognitive disorders in elderly patients: the impact of sarcopenia-related features. *Frontiers in Medicine*, 12.
- Racine, A. M., Tommet, D., D'Aquila, M., Fong, T. G., Gou, Y., Tabloski, P. A., Metzger, E. D., Hshieh, T. T., Schmitt, E. M., Vasunilashorn, S. M., Kunze, L., Vlassakov, K., Abdeen, A., Lange, J. K., Earp, B. E., Dickerson, B. C., Marcantonio, E. R., Steingrimsson, J. A., Trivison, T. G., ... Tasker, K. (2020). Machine Learning to Develop and Internally Validate a Predictive Model for Post-operative Delirium in a Prospective, Observational Clinical Cohort Study of Older Surgical Patients. *Journal of General Internal Medicine*, 36(2), 265–273.
- Sheng, W., Tang, X., Hu, X., Liu, P., Liu, L., Miao, H., Wang, D., & Li, T. (2024). Random forest algorithm for predicting postoperative delirium in older patients. *Frontiers in Neurology*, 14.
- Song, Y., Zhang, D., Wang, Q., Liu, Y., Chen, K., Sun, J., Shi, L., Li, B., Yang, X., Mi, W., & Cao, J. (2024). Prediction models for

- postoperative delirium in elderly patients with machine-learning algorithms and SHapley Additive exPlanations. *Translational Psychiatry*, 14(1).
- TheDIM, M., & Vacas, S. (2024). Postoperative Delirium and the Older Adult: Untangling the Confusion. *Journal of Neurosurgical Anesthesiology*, 36(3), 184–189.
- Vattipally, V. N., Kramer, P., Abouelseoud, N., Yeleswarapu, I., Davidar, A. D., Dardick, J., Bydon, A., Witham, T. F., Lubelski, D., Rosenblatt, K., Huang, J., Bettegowda, C., Seiber, F., Oh, E. S., Sarma, S. V., Akça, O., Theodore, N., & Azad, T. D. (2025). Point-of-care electroencephalography for prediction of postoperative delirium in older adults undergoing elective surgery: protocol for a prospective cohort study. *bioRxiv* (Cold Spring Harbor Laboratory).
- Wang, B., Ge, M., Cai, K., Zhang, Y., Zhou, Z., Li, W., Guo, Y., Wang, W., & Zhou, Q. (2025). Transformer representation learning is necessary for dynamic multi-modal physiological data on small-cohort patients. *ArXiv.Org*.
- Wang, Q., Mu, D., Wang, X. M., Han, W., Wang, J., Shen, J., Cai, N., & Xu, G. (2025). Development of a Machine Learning-Based Prediction Model for Postoperative Delirium in Frail Elderly Patients Undergoing Non-Cardiac Surgery Under General Anesthesia.
- Wu, S.-C. J., Sharma, N., Bauch, A., Yang, H.-C., Hect, J. L., Thomas, C., Wagner, S., Förstner, B. R., Arnim, C. A. F. V., Kaufmann, T., Eschweiler, G. W., & Wolfers, T. (2024). Predicting Postoperative Delirium in Older Patients: a multicenter retrospective cohort study. *medRxiv* (Cold Spring Harbor Laboratory).
- 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).