

RESEARCH ARTICLE

# AI-delirium guard: Predictive modeling of postoperative delirium in elderly surgical patients

Sri Harsha Boppana<sup>1</sup>, Divyansh Tyagi<sup>2</sup>, Sachin Komati<sup>3</sup>, Sri Lasya Boppana<sup>4</sup>, Ritwik Raj<sup>5</sup>, C. David Mintz<sup>6\*</sup>

**1** Department of Internal Medicine, Nassau University Medical Center, East Meadow, New York, United States of America, **2** Department of Applied Physics, Delhi Technological University, Delhi, India, **3** Department of Computer Science, Florida International University, Miami, Florida, United States of America, **4** Department of Internal Medicine, Alluri Sitarama Raju Academy of Medical Sciences, Eluru, India, **5** Johns Hopkins University Zanvyl Krieger School of Arts and Sciences, Baltimore, Maryland, United States of America, **6** Department of Anesthesiology & Critical Care Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland, United States of America

☞ These authors contributed equally to this work.

\* [cmintz2@jhmi.edu](mailto:cmintz2@jhmi.edu)



**OPEN ACCESS**

**Citation:** Boppana SH, Tyagi D, Komati S, Boppana SL, Raj R, Mintz CD (2025) AI-delirium guard: Predictive modeling of postoperative delirium in elderly surgical patients. PLoS One 20(6): e0322032. <https://doi.org/10.1371/journal.pone.0322032>

**Editor:** Pedro Kallas Curiati, Hospital Sirio-Libanes, BRAZIL

**Received:** October 20, 2024

**Accepted:** March 15, 2025

**Published:** June 5, 2025

**Copyright:** © 2025 Boppana et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data availability statement:** The data used in this study were obtained from the American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP). Access to the NSQIP Participant Use Data File (PUF) is available to researchers through an application process managed by the American College of

## Abstract

### Introduction

In older patients, postoperative delirium (POD) is a major complication that can result in greater morbidity, longer hospital stays, and higher healthcare expenses. Accurate prediction models for POD can enhance patient outcomes by guiding preventative strategies. This study utilizes advanced machine learning techniques to develop a predictive model for POD using comprehensive perioperative data.

### Methods

We examined information from the National Surgical Quality Improvement Program (NSQIP), which included 17,000 patients who were over 65 and undergoing different types of surgery. The dataset included variables such as patient demographics (age, sex), comorbidities (diabetes, cardiovascular diseases, pre-existing dementia), surgical details (type, duration), anesthesia type and dosage, and postoperative outcomes. Categorical variables were encoded numerically, and data standardization was applied to ensure normal distribution. A range of machine learning approaches were assessed such as Decision Trees and Random Forests. Based on the greatest Area Under the Curve (AUC) from Receiver Operating Characteristic (ROC) analysis, the final model was chosen. Hyperparameter tuning was performed using Grid-SearchCV, optimizing parameters like max\_depth, min\_child\_weight, and gamma for XGBoost model.

Surgeons. Interested researchers may request access to the data at the following link: <https://www.facs.org/quality-programs/data-and-registries/acs-nsqip/participant-use-data-file/> Other researchers can access the NSQIP dataset in the same manner as the authors by submitting a data use request through the same portal. The authors did not receive any special access privileges that others would not have.

**Funding:** The author(s) received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

## Results

The optimized XGBoost model demonstrated superior performance, achieving an AUC of 0.85. Key hyperparameters included  $\text{min\_child\_weight} = 1$ ,  $\text{max\_depth} = 5$ ,  $\text{gamma} = 0.3$ ,  $\text{subsample} = 0.9$ ,  $\text{colsample\_bytree} = 0.7$ ,  $\text{reg\_alpha} = 0.0007$ ,  $\text{learning\_rate} = 0.14$ , and  $\text{n\_estimators} = 123$ . The model exhibited an accuracy of 0.926, recall of 0.945, precision of 0.934, and an F1-score of 0.939, depicting a higher level of predictive accuracy & balance between sensitivity and specificity.

## Conclusion

This study proposes a strong XGBoost-based model to predict POD in older surgical patients, demonstrating the potential of Machine Learning (ML) in clinical risk assessment. With the help of the model's balanced performance indicators and high accuracy, physicians may identify high-risk patients and promptly execute interventions in clinical settings. Subsequent investigations ought to concentrate on integration into clinical workflows and external validation.

## Introduction

Following surgery, POD is a common and serious complication [1]. Numerous adverse outcomes have been associated with it, such as extended hospital stays, an increased likelihood of being transferred to institutional care upon release, increased readmission rates, functional deterioration, dependency on daily tasks, and cognitive loss [2–9]. Multicomponent non-pharmacologic therapies [10], the Hospital Elder Life Program [11], or perioperative geriatric consults [12] can prevent many cases of POD. There have been reports of successful perioperative strategies that combine targeted delirium prevention practices with delirium risk stratification [13,14], though improving the accuracy of the delirium risk stratification tool that had been utilized in this strategy could improve the distribution of scarce resources to the most vulnerable. Optimizing and automating risk classification processes is crucial since preventative interventions need a significant amount of time and effort from busy professionals [15].

ML-derived risk prediction models had been created to predict delirium in patients who have been hospitalized [16], POD in certain patient categories [17], non-delirium-related intra-op complications [18,19], post-op mortality [20], and other situations range [21]. Existing POD risk prediction models may benefit from the application of the ML approach in several important areas. While age, as well as cognitive impairment, are two well-established risk factors that are frequently included in delirium prediction models [22–24], machine learning (ML) can reveal intricate patterns and interactions between variables within large datasets that are not immediately visible when using more conventional, linear approaches to data analysis [24,25].

Using patient data from the NSQIP, this study evaluates multiple machine learning algorithms with the aim of identifying the most effective POD predicting model. The first step involved preprocessing raw data. Categorical variables such as demographics and anesthesia type were systematically encoded into numerical formats

appropriate for machine learning algorithms. Standardization was applied to ensure consistent scaling across all features and minimize bias in model predictions. After comparing the multiple ML model's performance based on higher AUC from ROC analysis, two models (Random Forest and XGBoost) were identified as the most effective. Further optimization through hyperparameter tuning led to the selection of XGBoost as the algorithm for implementation. The resulting model is a reliable and scalable tool that has the ability to better inform patient care.

## Methods

### Data collection source description

This study draws upon data collected from the NSQIP predominantly U.S.-based, over the period from 2014 to 2020, offering an extensive dataset that includes essential patient demographics and clinical variables. Since this study utilized data from the NSQIP, ethical approval was not necessary. NSQIP is a de-identified, open, publicly accessible database created especially for surgical outcome research and quality improvement.

The use of NSQIP data complies with privacy regulations and ethical guidelines for research involving human participants' data, eliminating the need for additional ethics review. The NSQIP data set was accessed on 23rd August 2023. The information encompasses key demographic factors like age and sex, along with significant clinical details such as the presence of diabetes, hypertension, cardiovascular conditions, pre-existing dementia, psychiatric disorders, and preoperative delirium. Additionally, the dataset provides insights into the types of surgeries performed, their durations, pain management protocols, anesthesia types and dosages, and overall preoperative health status. A critical focus is placed on the occurrence of postoperative delirium.

### Population criteria variables

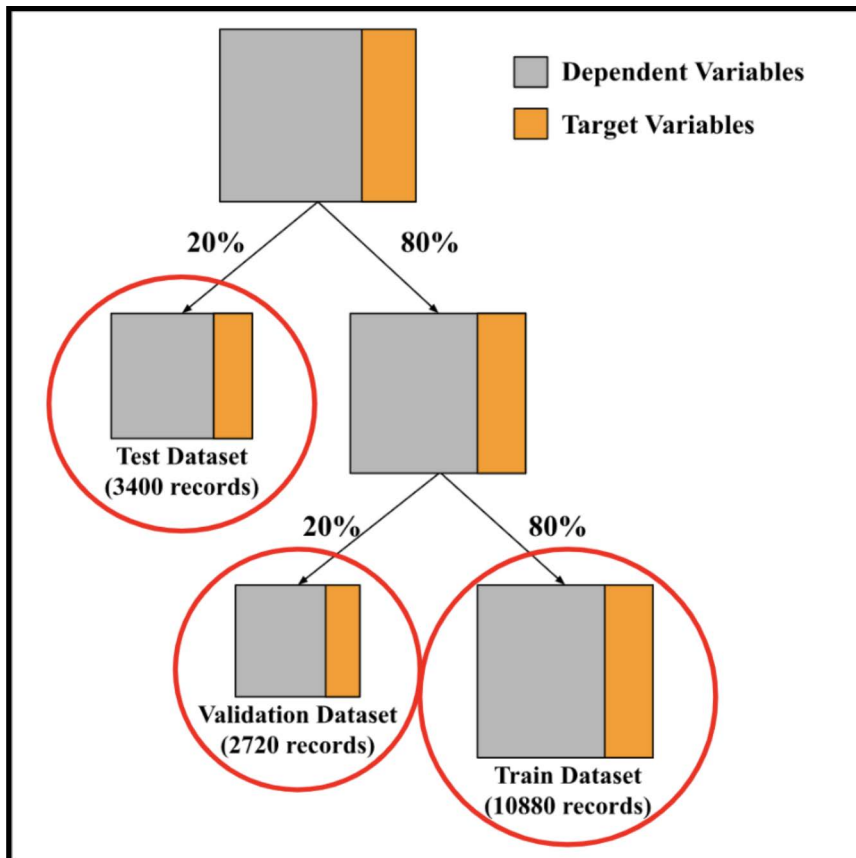
The study's population comprises individuals aged 65 years and older who underwent surgical procedures requiring anesthesia. The key variables analyzed include patient sex, age, inpatient or outpatient status, transfer status, type of anesthesia administered, discharge destination, patient height and weight, surgical specialty, whether the surgery was elective, smoking status, presence of dyspnea, cancer diagnosis, diabetes status, levels of preoperative sodium and albumin, hematocrit levels, whether the procedure was an emergency, surgery duration, presence of renal insufficiency, use of steroids, wound classification, presence of wound infection, and the occurrence of sepsis.

### Datasets

This study's main objective is to analyze the risk factors for POD in light of these characteristics. The dataset, which contains 17,000 records, was divided into two classes: 10,867 records where postoperative delirium was diagnosed, and 6,133 records where it was not. To develop and test predictive models, the data was split into a training set (80percent of total records), a test set (20 percent of the total records), and a validation set (20 percent of the training data), resulting in 10,880 records in the training set, 3,400 in the test set, and 2,720 in the validation set, as illustrated in [Fig 1](#).

In preparing the dataset for analysis, categorical variables were converted into numeric formats, which are necessary for use in most machine learning models. For instance, the variable 'transfer status' was encoded into categories such as not transferred, transferred from an acute care hospital, and other relevant categories. Similarly, 'discharge destination,' 'anesthesia type,' 'surgical specialty,' and 'diabetes status' were systematically encoded into numerical values. Binary variables, such as sex, inpatient or outpatient status, elective surgery, smoking status, cancer diagnosis, presence of wound infection, steroid use, and emergency status, were also encoded numerically to streamline the analysis.

Furthermore, normalization was applied to the dataset to ensure that every feature had a mean of 0 and an SD of 1. This is a crucial step since similar-scale input data reduces the likelihood of biased findings and enhances the performance of many ML algorithms. The process of data preparation, including encoding and standardization, is summarized in the study's workflow diagrams in [Fig 2](#).



**Fig 1. Distribution of data.**

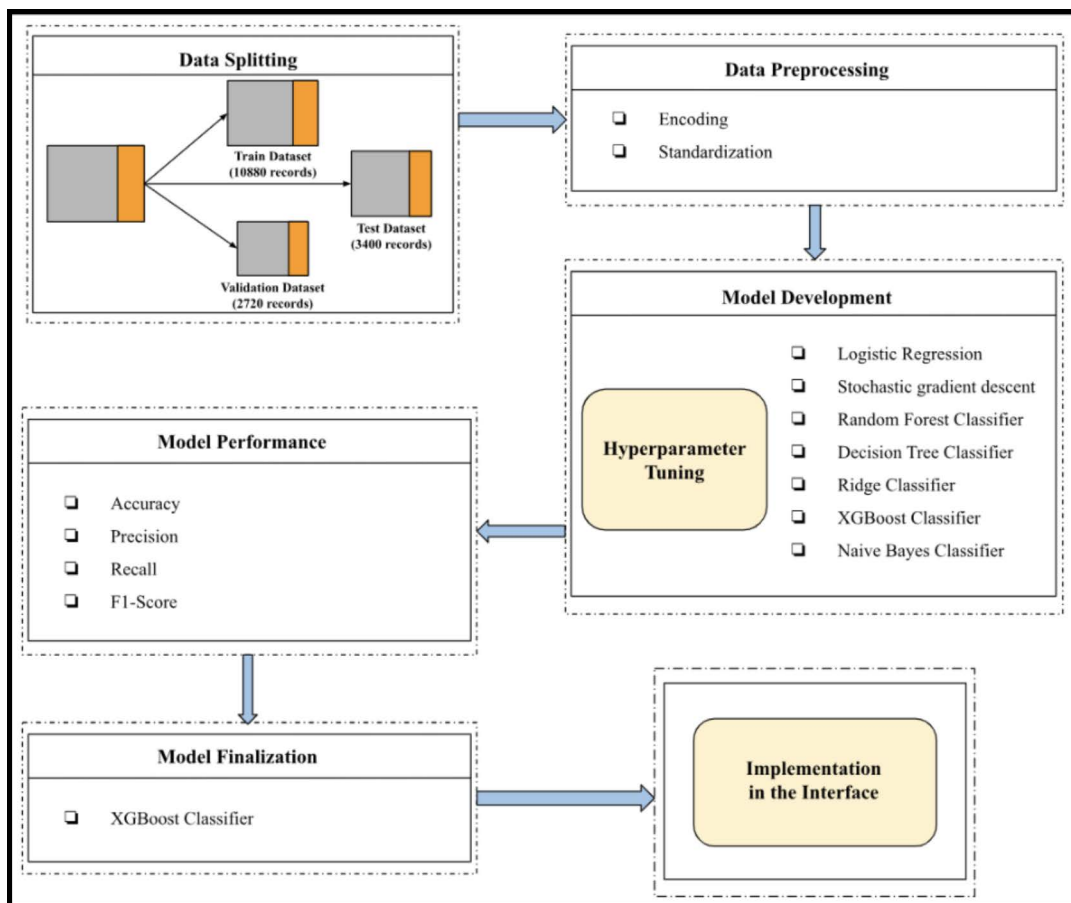
<https://doi.org/10.1371/journal.pone.0322032.g001>

### Data preprocessing

Since many machine learning methods demand numerical inputs, it was required to convert categorical data into numerical representations in order to prepare for machine learning analysis. Specific encoding methods were applied to different categorical variables. For the `transt` variable, patients admitted from home were assigned a value of 0, those transferred from an acute care hospital inpatient setting were coded as 1, from an external emergency department as 2, and other sources of transfer as 3. In the `dischdest` variable, representing discharge destinations, home was coded as 0, other destinations as 1, rehabilitation centers as 2, and skilled care facilities as 3.

The type of anesthesia used, recorded in the `anesthes` variable, was numerically encoded with general anesthesia as 0, MAC/IV sedation as 1, spinal anesthesia as 2, and other forms as 3. Surgical specialties were categorized in the `surgspec` variable, where urology surgery was assigned 0, orthopedics surgery 1, general surgery 2, neurosurgery 3, thoracic surgery 4, head and neck surgery (ENT) 5, vascular surgery 6, plastic surgery 7, cardiac surgery 8, and gynecology surgery 9.

Additionally, the presence and type of diabetes were categorized in the `diabetes` variable, with no diabetes encoded as 0, non-insulin-dependent diabetes as 1, and insulin-dependent diabetes as 2. Dyspnea, reflected in the `dyspnea` variable, was encoded as no dyspnea (0), dyspnea on moderate exertion (1), and dyspnea at rest (2). Wound classification, indicated in the `wndclas` variable, was categorized with clean/contaminated wounds as 0, clean wounds as 1, dirty/infected wounds as 2, and contaminated wounds as 3.



**Fig 2. Workflow of the model.**

<https://doi.org/10.1371/journal.pone.0322032.g002>

Demographic and procedural variables were also encoded, with `sex` categorized as male (0) and female (1), and `inout` as inpatient (0) and outpatient (1). Binary variables, such as `electsurg`, `smoke`, `discancr`, `wndinf`, `steroid`, `prsepis`, and `emergency`, were uniformly coded as no (0) and yes (1).

### Standardization

The standardization of data was an essential step to enhance the performance of machine learning models, particularly since different scales among features can lead to biased predictions. This process involved scaling the data so that the mean of each feature was adjusted to 0 and the standard deviation to 1. This transformation ensured that the features were on a common scale, thereby improving the accuracy as well as reliability of the model's predictions.

### Model development

**Algorithm selection.** In the development of our predictive model, we evaluated multiple machine learning algorithms to determine the most effective approach for our specific dataset. The algorithms explored included LR, SVM, DT Classifier, Stochastic Gradient Descent, Ridge Classifier, Random Forest Classifier, Extreme Gradient Boosting Classifier, and Naive Bayes Classifier. Each model was implemented and assessed based on its performance metrics, particularly focusing on its discriminative ability.

The ROC curves for these models, as depicted in Fig 3, were analyzed with the AUC serving as the key criterion for model selection. Based on this analysis, the RF Classifier and Extreme Gradient Boosting Classifier were identified as the top-performing models, achieving AUC values of 0.85 and 0.84, respectively. These models were thus selected for further refinement and validation.

Fig 3 shows the ROC curve of the models implemented and the AUC value is used to choose the best model. Among the various models, we have chosen the Random Forest Classifier and Extreme Gradient Boosting Classifier having AUC 0.85 & 0.84 respectively.

**Hyperparameter tuning.** To further enhance the selected model's performance, hyperparameter tuning was conducted using GridSearchCV. With this approach, one can perform a thorough search across designated parameter grids and methodically assess various combinations to determine the ideal configurations. A further step in the GridSearchCV process is K-fold cross-validation, which splits the dataset into the k subsets. Each iteration ensures a model performance thorough evaluation across many parameter configurations by setting aside one subset for validation and using the remaining k-1 subsets for training.

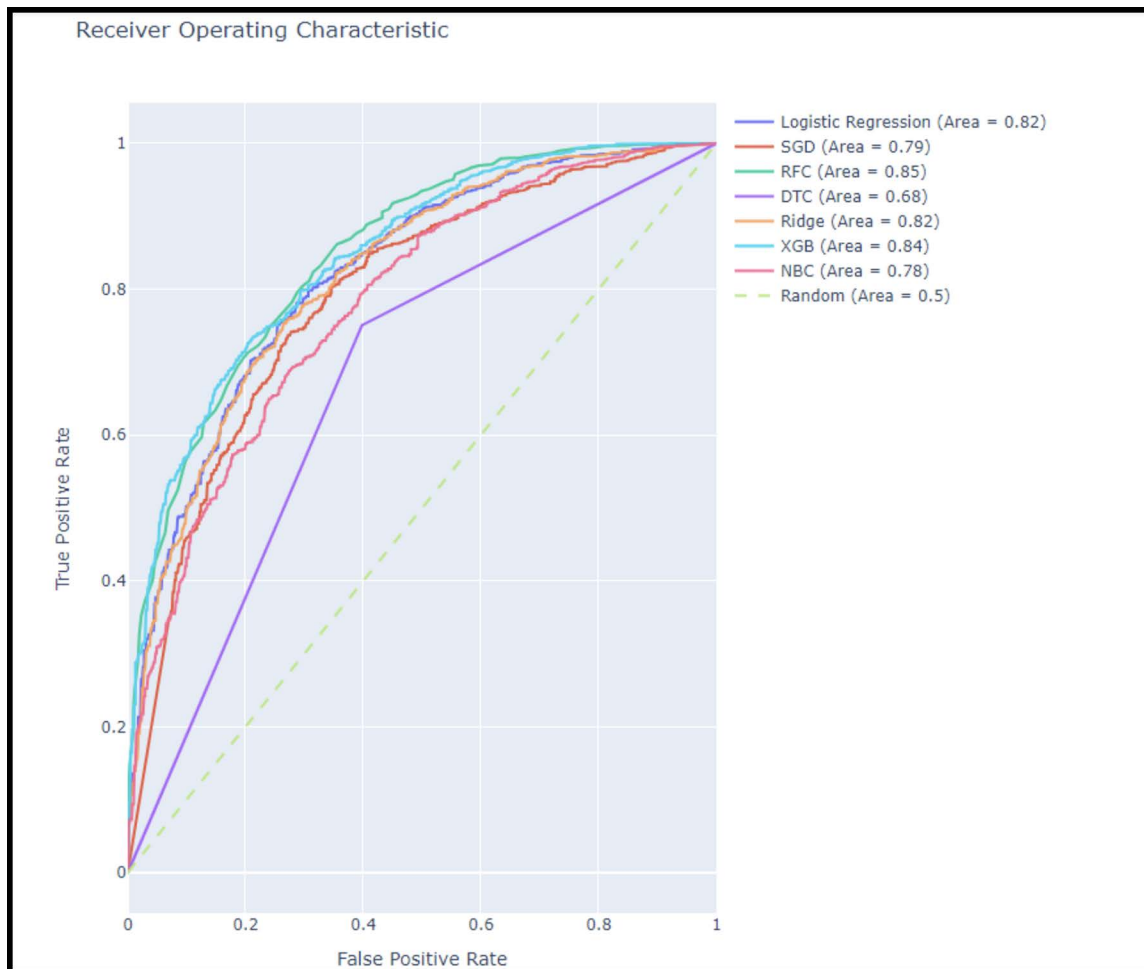


Fig 3. ROC-AUC curve of models implemented.

<https://doi.org/10.1371/journal.pone.0322032.g003>

For the Extreme Gradient Boosting (XGBoost) model, the optimal hyperparameters identified through this process were as follows: `min_child_weight=1`, `max_depth=5`, `colsample_bytree=0.7`, `gamma=0.3`, `subsample=0.9`, `learning_rate=0.14`, `reg_alpha=0.0007`, and `n_estimators=123`. Similarly, for the Random Forest model, the best hyperparameters were found to be `max_depth = 40`, `n_estimators = 174`, `max_features='sqrt'`, `max_leaf_nodes = 3500`, `min_samples_leaf=1`, and `min_samples_split=2`. These optimized parameters significantly improved the predictive accuracy and overall performance of the models, validating the effectiveness of our approach.

## Results

### Model performance

The models' efficacy has been analyzed by utilizing 4 primary classification metrics: F1-score, accuracy, recall, and precision. The cases accurately percentage classified relative to the total number of cases is known as accuracy. Recall assesses the model's capacity to find out the positive cases, whereas precision measures the positive prediction accuracy. When working with unbalanced datasets, the F1-score offers a comprehensive classification statistic by providing a harmonic mean of precision & recall.

The models were validated using a separate dataset, with the results illustrated in the confusion matrices ([Fig 4](#) for the XGBoost model and [Fig 5](#) for the Random Forest model). These matrices are essential for visualizing the model's classification outcomes. A summary of the performance metrics for both models is presented in [Fig 6](#). From this comparison ([Fig 7](#)), it is evident that the XGBoost model surpasses the Random Forest model in all evaluated metrics. Consequently, XGBoost was chosen as the final model to be integrated into the user interface designed to predict POD risk in elderly patients

### Interface description

The interface is designed to assist healthcare providers in the early detection of post-op delirium, offering quick and accurate risk predictions based on the data entered. The interface is user-friendly, featuring multiple input methods tailored for different data types. For categorical variables, users can select options from drop-down menus, while numerical inputs are entered directly into designated fields. For ease of use, height is entered in separate fields for feet and inches, and weight can be entered in either kilograms or pounds, based on the user's preference. After all data has been entered, the 'Predict' button is pressed to generate a prediction. The interface then displays a message indicating either a high likelihood ("You are likely to experience postoperative delirium after surgery") or a low likelihood ("You likely will not experience postoperative delirium after surgery") based on the model's output.

### Model implementation in the interface

User input data collected through the interface is first compiled into a data frame. Since the data includes both categorical and numerical variables, which have not been standardized, preprocessing is necessary. Categorical data is encoded in the same manner as during the initial model development, ensuring consistency. Numerical data is standardized using parameters saved during the model's training phase. The processed data is then passed by the trained XGBoost model to generate a prediction. Depending on the outcome, the interface displays a corresponding message: either "You are likely to experience POD after surgery" for a high-risk prediction or "You likely will not experience postoperative delirium after surgery" for a low-risk prediction.

**Interface Link:** <https://ai-delirium-guard.streamlit.app>

## Discussion

The accurate prediction of POD in elderly surgical patients is a significant clinical challenge, with existing methods primarily reliant on subjective clinical assessments. These traditional approaches, while valuable, often lack the precision

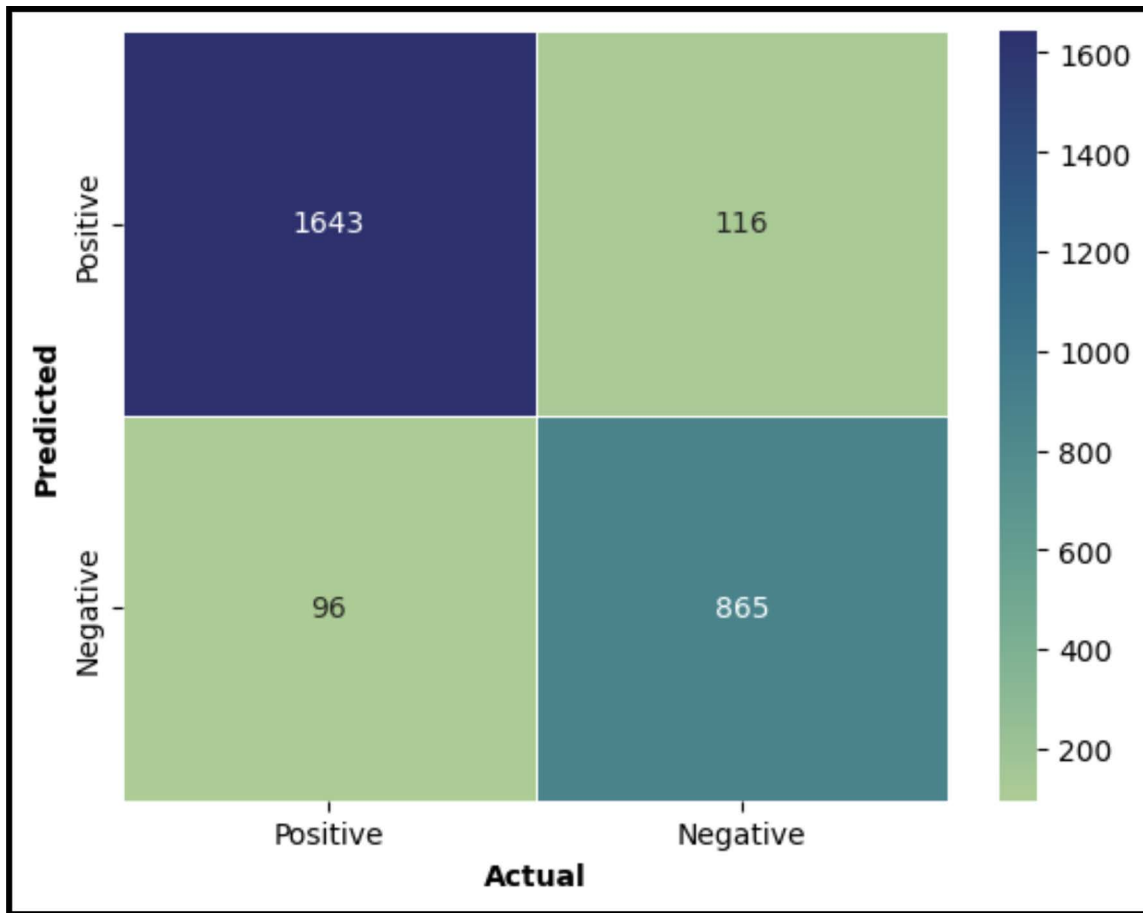


Fig 4. Confusion Matrix for XGBoost Classification Model.

<https://doi.org/10.1371/journal.pone.0322032.g004>

needed to consistently identify at-risk patients before the onset of delirium. The findings of our research highlight the potential of advanced ML techniques, particularly the XGBoost algorithm, to significantly enhance the predictive accuracy for POD, offering a more reliable tool for clinical decision-making.

Postoperative delirium has been the subject of extensive research due to its association with increased morbidity, extended hospital stays, and elevated healthcare costs. Traditional methods for predicting POD typically involve clinical scoring systems that assess known risk factors like age, baseline cognitive function, and the presence of comorbidities. Tools like the Confusion Assessment Method (CAM) have been commonly utilized to diagnose delirium postoperatively but are not designed to predict its occurrence beforehand (28).

The utilization of ML models to forecast postoperative outcomes, like POD, has gained popularity recently. For instance, Bishara et al. explored the use of ML algorithms to predict postoperative delirium, finding that models like Random Forests and Gradient Boosting Machines could outperform traditional risk assessments in accuracy. (29) Nevertheless, a lot of this research has been constrained by single-center datasets or small sample numbers, which can limit how broadly the results can be applied.

Machine learning-derived models demonstrated high AUC-ROCs; these models outperformed numerous risk prediction models, especially those designed for postoperative delirium [21,26], and they were on par with other published ML-derived risk prediction models [15,27–29]. Unlike most previously published POD risk prediction methods, which focus on

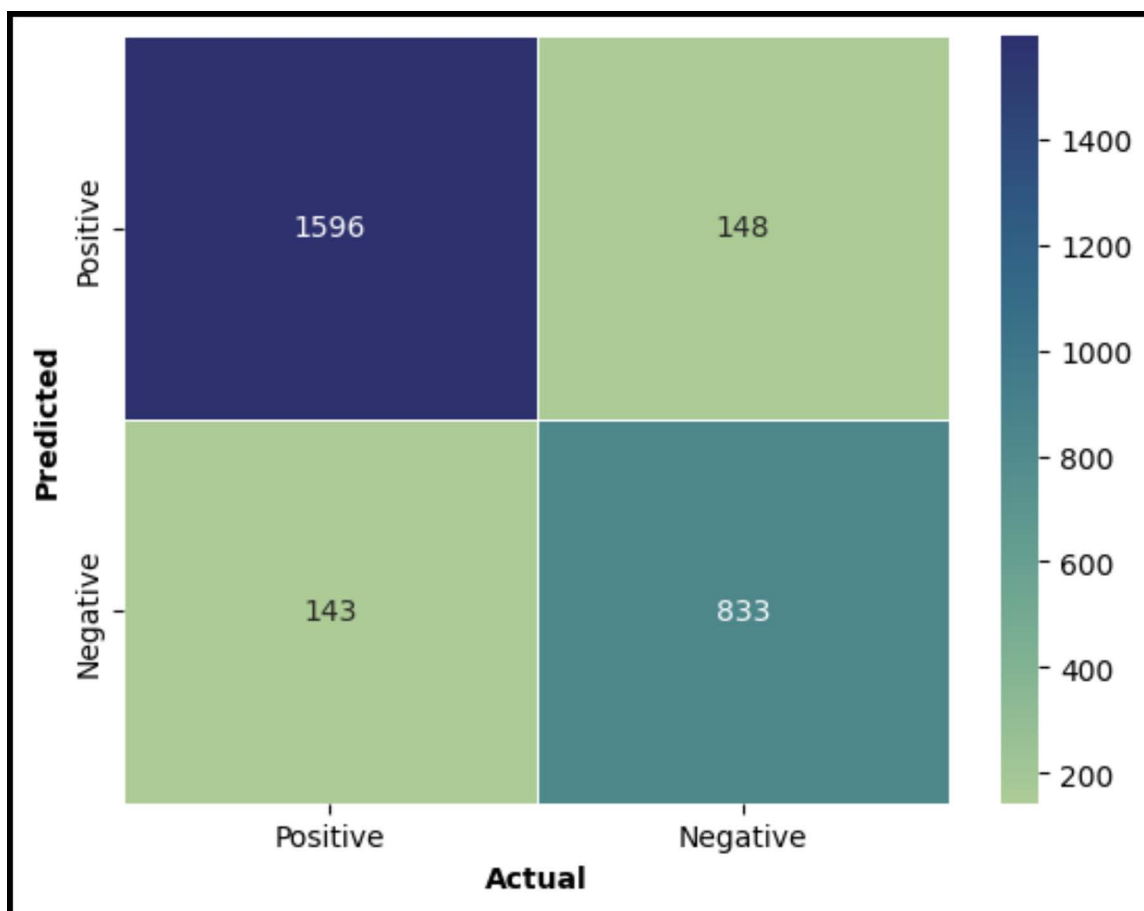


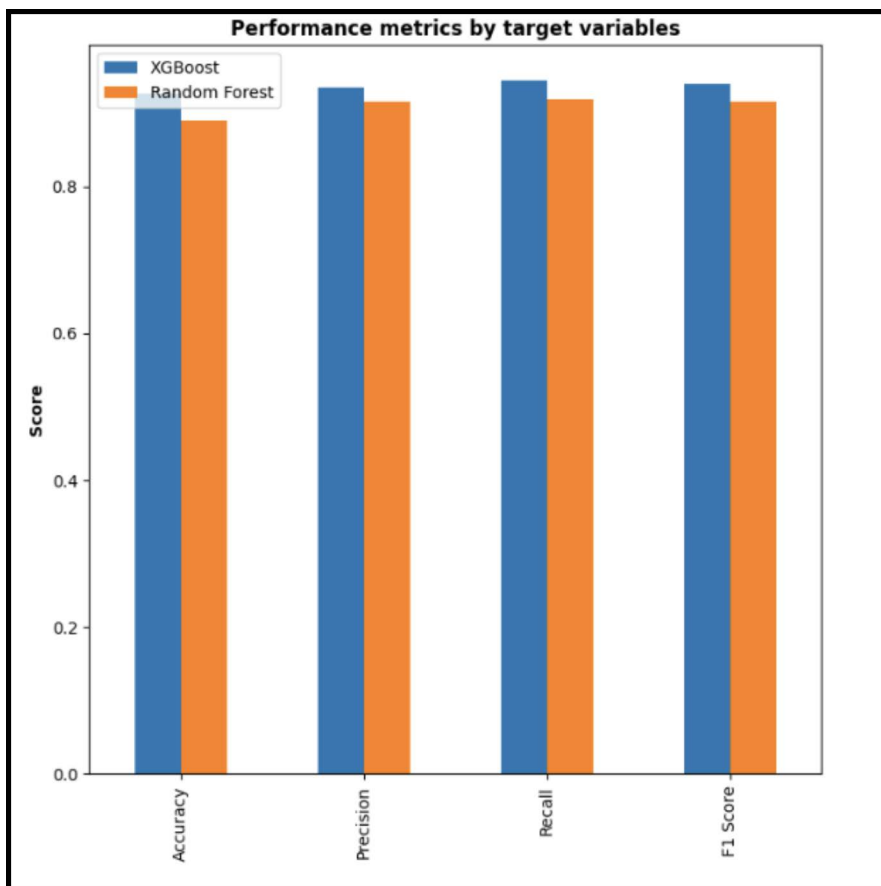
Fig 5. Confusion Matrix for Random Forest Classification Model.

<https://doi.org/10.1371/journal.pone.0322032.g005>

	XGBoost	Random Forest
<b>Accuracy</b>	0.922	0.893
<b>Precision</b>	0.934	0.915
<b>Recall</b>	0.945	0.918
<b>F1 Score</b>	0.939	0.916

Fig 6. Classification metrics for XGBoost and RF: The table below depicts the values of metrics for XGBoost and Random Forest model.

<https://doi.org/10.1371/journal.pone.0322032.g006>



**Fig 7. Comparison of both models: It can be understood that the XGBoost outperforms Random Forest in terms of classification metrics. So, we have chosen XGBoost as our final model to be used in the interface. The interface helps the user to predict the possibility of postoperative delirium.**

<https://doi.org/10.1371/journal.pone.0322032.g007>

a specific surgical group, this model aims to predict delirium over a wide range of variables [22]. In a comparable non-specific perioperative population from our institution, ML models significantly enhanced discrimination when compared to a basic regression technique [26].

Our study advances this field by utilizing a large, multicenter dataset from the NSQIP that encompasses a wide range of perioperative variables. By employing the XGBoost algorithm, which has been optimized through rigorous hyperparameter tuning, we achieved an AUC of 0.85, a substantial improvement over the AUC values typically reported for logistic regression models used in similar contexts [13]. This performance indicates that the XGBoost model can capture complex interactions between variables that traditional models may miss, offering a more nuanced and accurate risk prediction.

Giving patients who have been at risk for delirium the best care possible requires a lot of resources. Interventions like multicomponent non-pharmacologic nursing care bundles along with time-constrained consultations from physicians like pharmacists and rehabilitation specialists are part of this process. The published guidelines for the prevention of POD [1,14,22,30] served as the paradigm for the care treatments. Preoperative cognitive screenings and avoiding high-risk drugs are advised by the majority of guidelines. The postoperative elements of our approach, such as multicomponent bundles that address early mobility, medication review, and underlying reasons, are in line with postoperative care guidelines [1,14,22,30]. It is essential for the identification of patients who are most likely to have delirium to facilitate their admission into this treatment path. Enhancing the efficacy of delirium screening models could facilitate more efficient

resource allocation for high-risk patients, hence improving the overall value of healthcare. The model will next be operationalized through external and prospective validation, and then it will be integrated into the EHR for real-time use. This technique is feasible, as evidenced by earlier descriptions of comparable real-time ML model implementations [31].

Furthermore, while previous studies have focused on general surgical populations, our research specifically targets elderly patients, a demographic particularly vulnerable to delirium due to factors such as frailty, polypharmacy, and multimorbidity [17]. The inclusion of a comprehensive set of perioperative variables, such as anesthesia type, surgical duration, and comorbid conditions, further enhances the model's applicability in clinical practice.

This conclusion is shown by contrasting our results with a recent investigation by Racine, et al. [32]. The study assessed the ability of five machine learning algorithms to forecast POD in a far smaller subset of surgery patients—560 older adults drawn from an already-existing dataset. Despite employing medical document checks and in-person examinations by skilled interviewers utilizing the CAM [17] to diagnose delirium, machine learning algorithms showed AUC-ROC values of 0.53–0.71, which did not beat the logistic regression model presented in the research. This comparison emphasizes how crucial it is to have a large enough training dataset [33], among other things, in order to perform excellent machine learning studies. Even with our enormous sample size, our results are fairly rare, thus the model's performance in terms of discrimination, positive predictive value, and calibration would probably improve further with more data and a larger number of occurrences per predictor [33].

In machine learning studies, it is common to criticize reporting model discrimination as the only performance indicator [34], since other performance measures that take occurrence into account might more accurately reflect clinical application & significance. Large positive probability ratios for XGBoost allow for the delivery of resource-intensive therapy to patients who are most at risk of growing delirium. We used thresholds for optimizing sensitivity and positive probability ratio rather than PPV because the former is largely influenced by disease prevalence. Furthermore, compared to the expert clinician regression model, the ML hybrid model that utilizes predictors from the XGBoost feature importance summary performs better. This suggests that ML-based models could be able to identify higher-order interactions between variables that are hard to locate with conventional regression approaches. The majority of the XGBoost-identified predictors, such as age, type of surgery, cognitive impairment, and comorbidities, are in line with previously established POD factors [35–37]. Not often mentioned but crucial to the model are protective factors against delirium such as private insurance, alcohol consumption self-reported history, and a self-reported history of recent falls. These attributes probably point to relationships that the model does not account for. People who own private insurance, for instance, might be better educated and/or have a higher social standing. Individuals who are able to self-report falls may have superior cognitive performance over those who are not. Whilst self-reported alcohol use has been associated with better functional outcomes, such as reduced frailty in females [38] and a lower probability of mobility or arm function restrictions independent of muscle strength in older men [39], alcohol consumption is a known risk factor for delirium [40]. Some authors suggest that the protective effect of moderate alcohol use may be explained by social or lifestyle characteristics that are not included in other components of the ML model [39]. Further research is necessary because the precise mechanisms behind these relationships are unknown.

By utilizing the XGBoost model, clinicians can identify high-risk patients with greater accuracy and tailor interventions accordingly. For example, patients identified as having a higher risk of POD could benefit from enhanced postoperative monitoring, targeted pharmacological management to reduce delirium risk and the implementation of non-pharmacologic strategies such as cognitive stimulation and early mobilization. These interventions, when applied selectively to those most likely to benefit, could decrease the incidence of POD, shorten hospital stays, and lower healthcare costs [41–49].

## Limitations

This study has certain drawbacks that need to be acknowledged, despite its positives. First, while the NSQIP dataset provides a robust foundation for model development, it may not capture all relevant variables that influence POD, such

as specific intraoperative events or detailed cognitive assessments. Additionally, the study's retrospective design, while necessary for model training and validation, may introduce biases related to data quality and completeness.

Second, although the model demonstrated high predictive accuracy in the validation cohort, external validation using data from other institutions is necessary to confirm its generalizability. The model's performance may vary in different clinical settings, particularly those with different patient demographics, surgical practices, or perioperative care protocols.

Third, even while XGBoost and other machine learning models have a lot to offer in terms of prediction accuracy, their complexity can make them difficult to understand. Clinicians may be hesitant to rely on models that function as “black boxes” without a clear understanding of how predictions are generated. Therefore, efforts to enhance the transparency and interpretability of these models are crucial to their successful adoption in clinical practice.

The NSQIP dataset is limited in its geographical scope and temporal range, and this study employed it for both model training and testing. This raises important questions about the model's capability to adapt to different clinical settings and whether its predictive accuracy can be maintained across diverse patient populations and healthcare environments.

Moreover, the successful application of this approach relies on the extensive availability, of patient-specific data that are consistent with the variables recorded in NSQIP. However, not all healthcare systems have access to such detailed data, and when these variables are available, their measurement—often binary in nature (e.g., history of stroke)—may vary. These differences could influence the model's performance and limit its generalizability across different healthcare settings.

### Future directions

To expand on the results of this study, future research should concentrate on a number of important areas. First, external validation of the XGBoost model across diverse patient populations and healthcare settings is essential to ensure its broader applicability. Prospective studies that integrate the model into clinical workflows and evaluate its impact on patient outcomes will also be important for demonstrating its practical utility.

Additionally, further refinement of the model to incorporate real-time data inputs, such as intraoperative monitoring or postoperative cognitive assessments, could enhance its predictive accuracy and clinical relevance. Exploring the integration of other ML approaches, such as deep learning or ensemble models, may also provide additional insights and improve prediction capabilities.

Finally, there is a need for ongoing research into the pathophysiological mechanisms underlying POD, which could inform the development of more targeted and effective interventions. ML models, with their capability for the identification of complex patterns in large datasets, might play an important role in advancing our understanding of these mechanisms and ultimately improving patient care.

To extend the implications of our findings, we propose that an in-depth analysis of the decision-making processes within the XGBoost model could offer novel insights into the underlying pathophysiology of postoperative delirium (POD). By examining the factors and patterns that the model identifies as significant, researchers may uncover previously unrecognized mechanisms contributing to POD. This methodology has promise in filling in knowledge gaps and opening doors for the creation of more specialized and potent treatment approaches.

### Conclusion

This study provides compelling evidence that machine learning, particularly the XGBoost algorithm, can significantly enhance POD prediction in elderly surgical patients. The creation of a dependable, computer-based prediction model is a promising development in perioperative treatment that could lead to better patient outcomes by identifying POD early and preventing it specifically. To ensure this model's successful integration into clinical practice and maximize its impact on patient care, future research should work to validate and improve it.

## Author contributions

**Conceptualization:** Sri Harsha Boppana, C. David Mintz.

**Investigation:** Sri Harsha Boppana, Divyansh Tyagi, C. David Mintz.

**Methodology:** Divyansh Tyagi.

**Project administration:** Sri Harsha Boppana, Divyansh Tyagi, C. David Mintz.

**Resources:** Sri Harsha Boppana, Sachin Komati, Sri Lasya Boppana, Ritwik Raj.

**Software:** C. David Mintz.

**Supervision:** Sachin Komati, C. David Mintz.

**Validation:** Sri Harsha Boppana, Sachin Komati, Sri Lasya Boppana, C. David Mintz.

**Visualization:** Sri Harsha Boppana, C. David Mintz.

**Writing – original draft:** Sri Harsha Boppana, Divyansh Tyagi, C. David Mintz.

**Writing – review & editing:** Sri Harsha Boppana, Sri Lasya Boppana, Ritwik Raj, C. David Mintz.

## References

1. American Geriatrics Society Expert Panel on Postoperative Delirium in Older Adults. American Geriatrics Society abstracted clinical practice guideline for postoperative delirium in older adults. *J Am Geriatr Soc.* 2015;63(1):142–50. <https://doi.org/10.1111/jgs.13281> PMID: [25495432](https://pubmed.ncbi.nlm.nih.gov/25495432/)
2. Gleason LJ, Schmitt EM, Kosar CM, Tabloski P, Saczynski JS, Robinson T, et al. Effect of delirium and other major complications on outcomes after elective surgery in older adults. *JAMA Surg.* 2015;150(12):1134–40. <https://doi.org/10.1001/jamasurg.2015.2606> PMID: [26352694](https://pubmed.ncbi.nlm.nih.gov/26352694/)
3. Neufeld KJ, Leoutsakos J-MS, Sieber FE, Wanamaker BL, Gibson Chambers JJ, Rao V, et al. Outcomes of early delirium diagnosis after general anesthesia in the elderly. *Anesth Analg.* 2013;117(2):471–8. <https://doi.org/10.1213/ANE.0b013e3182973650> PMID: [23757476](https://pubmed.ncbi.nlm.nih.gov/23757476/)
4. Brown CH 4th, LaFlam A, Max L, Wyrobek J, Neufeld KJ, Kebaish KM, et al. Delirium after spine surgery in older adults: incidence, risk factors, and outcomes. *J Am Geriatr Soc.* 2016;64(10):2101–8. <https://doi.org/10.1111/jgs.14434> PMID: [27696373](https://pubmed.ncbi.nlm.nih.gov/27696373/)
5. Rudolph JL, Inouye SK, Jones RN, Yang FM, Fong TG, Levkoff SE, et al. Delirium: an independent predictor of functional decline after cardiac surgery. *J Am Geriatr Soc.* 2010;58:643–9. <https://doi.org/10.1111/j.1532-5415.2010.02762.x>
6. Peiffer-Smadja N, Rawson TM, Ahmad R, Buchard A, Georgiou P, Lescure F-X, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. *Clin Microbiol Infect.* 2020;26(5):584–95. <https://doi.org/10.1016/j.cmi.2019.09.009> PMID: [31539636](https://pubmed.ncbi.nlm.nih.gov/31539636/)
7. Inouye SK, Marcantonio ER, Kosar CM, Tommet D, Schmitt EM, Trivison TG, et al. The short-term and long-term relationship between delirium and cognitive trajectory in older surgical patients. *Alzheimers Dement J Alzheimers Assoc.* 2016;12:766–75. <https://doi.org/10.1016/j.jalz.2016.03.005>
8. Hshieh TT, Yue J, Oh E, Puelle M, Dowal S, Trivison T, et al. Effectiveness of multicomponent nonpharmacological delirium interventions. *JAMA Intern Med.* 2015;175(4):512. <https://doi.org/10.1001/jamainternmed.2014.7779>
9. Hshieh TT, Yang T, Gartaganis SL, Yue J, Inouye SK. Hospital elder life program: systematic review and meta-analysis of effectiveness. *Am J Geriatr Psychiatry.* 2018;26(10):1015–33. <https://doi.org/10.1016/j.jagp.2018.06.007> PMID: [30076080](https://pubmed.ncbi.nlm.nih.gov/30076080/)
10. Marcantonio ER, Flacker JM, Wright RJ, Resnick NM. Reducing delirium after hip fracture: a randomized trial. *J Am Geriatr Soc.* 2001;49(5):516–22. <https://doi.org/10.1046/j.1532-5415.2001.49108.x> PMID: [11380742](https://pubmed.ncbi.nlm.nih.gov/11380742/)
11. Moyce Z, Rodseth RN, Biccard BM. The efficacy of peri-operative interventions to decrease postoperative delirium in non-cardiac surgery: a systematic review and meta-analysis. *Anaesthesia.* 2014;69(3):259–69. <https://doi.org/10.1111/anae.12539> PMID: [24382294](https://pubmed.ncbi.nlm.nih.gov/24382294/)
12. Donovan AL, Braehler MR, Robinowitz DL, Lazar AA, Finlayson E, Rogers S, et al. An implementation-effectiveness study of a perioperative delirium prevention initiative for older adults. *Anesth Analg.* 2020;131(6):1911–22. <https://doi.org/10.1213/ANE.0000000000005223> PMID: [33105281](https://pubmed.ncbi.nlm.nih.gov/33105281/)
13. Whitlock EL, Braehler MR, Kaplan JA, Finlayson E, Rogers SE, Douglas V, et al. Derivation, validation, sustained performance, and clinical impact of an electronic medical record-based perioperative delirium risk stratification tool. *Anesth Analg.* 2020;131(6):1901–10. <https://doi.org/10.1213/ANE.0000000000005085> PMID: [33105280](https://pubmed.ncbi.nlm.nih.gov/33105280/)
14. Hughes CG, Bonczyk CS, Culley DJ, Fleisher LA, Leung JM, McDonagh DL, et al. American society for enhanced recovery and perioperative quality initiative joint consensus statement on postoperative delirium prevention. *Anesth Analg.* 2020;130(6):1572–90. <https://doi.org/10.1213/ANE.0000000000004641> PMID: [32022748](https://pubmed.ncbi.nlm.nih.gov/32022748/)
15. Wong A, Young AT, Liang AS, Gonzales R, Douglas VC, Hadley D. Development and validation of an electronic health record-based machine learning model to estimate delirium risk in newly hospitalized patients without known cognitive impairment. *JAMA Netw Open.* 2018;1(4):e181018. <https://doi.org/10.1001/jamanetworkopen.2018.1018> PMID: [30646095](https://pubmed.ncbi.nlm.nih.gov/30646095/)

16. Mufti HN, Hirsch GM, Abidi SR, Abidi SSR. Exploiting machine learning algorithms and methods for the prediction of agitated delirium after cardiac surgery: models development and validation study. *JMIR Med Inform.* 2019;7(4):e14993. <https://doi.org/10.2196/14993> PMID: [31558433](https://pubmed.ncbi.nlm.nih.gov/31558433/)
17. Wijnberge M, Geerts BF, Hol L, Lemmers N, Mulder MP, Berge P, et al. Effect of a machine learning-derived early warning system for intraoperative hypotension vs standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery: the HYPE randomized clinical trial. *JAMA.* 2020;323(11):1052–60. <https://doi.org/10.1001/jama.2020.0592> PMID: [32065827](https://pubmed.ncbi.nlm.nih.gov/32065827/)
18. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng.* 2018;2(10):749–60. <https://doi.org/10.1038/s41551-018-0304-0> PMID: [31001455](https://pubmed.ncbi.nlm.nih.gov/31001455/)
19. Hill BL, Brown R, Gabel E, Rakocz N, Lee C, Cannesson M, et al. An automated machine learning-based model predicts postoperative mortality using readily-extractable preoperative electronic health record data. *Br J Anaesth.* 2019;123(6):877–86. <https://doi.org/10.1016/j.bja.2019.07.030> PMID: [31627890](https://pubmed.ncbi.nlm.nih.gov/31627890/)
20. Hashimoto DA, Witkowski E, Gao L, Meireles O, Rosman G. Artificial intelligence in anesthesiology: current techniques, clinical applications, and limitations. *Anesthesiology.* 2020;132(2):379–94. <https://doi.org/10.1097/ALN.0000000000002960> PMID: [31939856](https://pubmed.ncbi.nlm.nih.gov/31939856/)
21. Jansen CJ, Absalom AR, de Bock GH, van Leeuwen BL, Izaks GJ. Performance and agreement of risk stratification instruments for postoperative delirium in persons aged 50 years or older. *PLoS One.* 2014;9(12):e113946. <https://doi.org/10.1371/journal.pone.0113946> PMID: [25464335](https://pubmed.ncbi.nlm.nih.gov/25464335/)
22. van Meenen LCC, van Meenen DMP, de Rooij SE, ter Riet G. Risk prediction models for postoperative delirium: a systematic review and meta-analysis. *J Am Geriatr Soc.* 2014;62(12):2383–90. <https://doi.org/10.1111/jgs.13138> PMID: [25516034](https://pubmed.ncbi.nlm.nih.gov/25516034/)
23. Pendlebury ST, Lovett N, Smith SC, Cornish E, Mehta Z, Rothwell PM. Delirium risk stratification in consecutive unselected admissions to acute medicine: validation of externally derived risk scores. *Age Ageing.* 2016;45(1):60–5. <https://doi.org/10.1093/ageing/afv177> PMID: [26764396](https://pubmed.ncbi.nlm.nih.gov/26764396/)
24. Pinsky MR, Dubrawski A. Gleaning knowledge from data in the intensive care unit. *Am J Respir Crit Care Med.* 2014;190(6):606–10. <https://doi.org/10.1164/rccm.201404-0716CP> PMID: [25068389](https://pubmed.ncbi.nlm.nih.gov/25068389/)
25. Moore A, Bell M. XGBoost, a novel explainable AI technique, in the prediction of myocardial infarction: a UK biobank cohort study. *Clin Med Insights Cardiol.* 2022;16:11795468221133611. <https://doi.org/10.1177/11795468221133611> PMID: [36386405](https://pubmed.ncbi.nlm.nih.gov/36386405/)
26. Kassie GM, Nguyen TA, Kalisch Ellett LM, Pratt NL, Roughead EE. Do risk prediction models for postoperative delirium consider patients' preoperative medication use? *Drugs Aging.* 2018;35(3):213–22. <https://doi.org/10.1007/s40266-018-0526-6> PMID: [29423780](https://pubmed.ncbi.nlm.nih.gov/29423780/)
27. Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. *Ann Intern Med.* 1990;113(12):941–8. <https://doi.org/10.7326/0003-4819-113-12-941> PMID: [2240918](https://pubmed.ncbi.nlm.nih.gov/2240918/)
28. Bishara A, Chiu C, Whitlock EL, Douglas VC, Lee S, Butte AJ, et al. Postoperative delirium prediction using machine learning models and preoperative electronic health record data. *BMC Anesthesiol.* 2022;22(1):8. <https://doi.org/10.1186/s12871-021-01543-y> PMID: [34979919](https://pubmed.ncbi.nlm.nih.gov/34979919/)
29. Corey KM, Kashyap S, Lorenzi E, Lagoo-Deenadayalan SA, Heller K, Whalen K, et al. Development and validation of machine learning models to identify high-risk surgical patients using automatically curated electronic health record data (Pythia): a retrospective, single-site study. *PLoS Med.* 2018;15(11):e1002701. <https://doi.org/10.1371/journal.pmed.1002701> PMID: [30481172](https://pubmed.ncbi.nlm.nih.gov/30481172/)
30. Aldecoa C, Bettelli G, Bilotta F, Sanders RD, Audisio R, Borozdina A, et al. European Society of Anaesthesiology evidence-based and consensus-based guideline on postoperative delirium. *Eur J Anaesthesiol.* 2017;34(4):192–214. <https://doi.org/10.1097/EJA.0000000000000594> PMID: [28187050](https://pubmed.ncbi.nlm.nih.gov/28187050/)
31. Mohanty S, Rosenthal RA, Russell MM, Neuman MD, Ko CY, Esnaola NF. Optimal perioperative management of the geriatric patient: a best practices guideline from the American College of Surgeons NSQIP and the American Geriatrics Society. *J Am Coll Surg.* 2016;222(5):930–47. <https://doi.org/10.1016/j.jamcollsurg.2015.12.026> PMID: [27049783](https://pubmed.ncbi.nlm.nih.gov/27049783/)
32. Mohanty S, Liu Y, Paruch JL, Kmiecik TE, Cohen ME, Ko CY, et al. Risk of discharge to postacute care: a patient-centered outcome for the American College of Surgeons National Surgical Quality Improvement Program surgical risk calculator. *JAMA Surg.* 2015;150(5):480–4. <https://doi.org/10.1001/jamasurg.2014.3176> PMID: [25806660](https://pubmed.ncbi.nlm.nih.gov/25806660/)
33. Racine AM, Tommet D, D'Aquila ML, Fong TG, Gou Y, Tabloski PA, et al. 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. *J Gen Intern Med.* 2021;36(2):265–73. <https://doi.org/10.1007/s11606-020-06238-7> PMID: [33078300](https://pubmed.ncbi.nlm.nih.gov/33078300/)
34. Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the confusion assessment method. In: A new method for detection of delirium.
35. Figueroa RL, Zeng-Treitler Q, Kandula S, Ngo LH. Predicting sample size required for classification performance. *BMC Med Inform Decis Mak.* 2012;12:8. <https://doi.org/10.1186/1472-6947-12-8> PMID: [22336388](https://pubmed.ncbi.nlm.nih.gov/22336388/)
36. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous end-points. *BMC Med Res Methodol.* 2014;14:137. <https://doi.org/10.1186/1471-2288-14-137> PMID: [25532820](https://pubmed.ncbi.nlm.nih.gov/25532820/)
37. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc.* 2017;24(1):198–208. <https://doi.org/10.1093/jamia/ocw042> PMID: [27189013](https://pubmed.ncbi.nlm.nih.gov/27189013/)
38. Dasgupta M, Dumbrell AC. Preoperative risk assessment for delirium after noncardiac surgery: a systematic review. *J Am Geriatr Soc.* 2006;54(10):1578–89. <https://doi.org/10.1111/j.1532-5415.2006.00893.x> PMID: [17038078](https://pubmed.ncbi.nlm.nih.gov/17038078/)
39. Robinson TN, Raeburn CD, Tran ZV, Angles EM, Brenner LA, Moss M. Postoperative delirium in the elderly: risk factors and outcomes. *Ann Surg.* 2009;249(1):173–8. <https://doi.org/10.1097/SLA.0b013e31818e4776> PMID: [19106695](https://pubmed.ncbi.nlm.nih.gov/19106695/)

40. Berian JR, Zhou L, Russell MM, Hornor MA, Cohen ME, Finlayson E, et al. Postoperative delirium as a target for surgical quality improvement. *Ann Surg*. 2018;268(1):93–9. <https://doi.org/10.1097/SLA.0000000000002436> PMID: [28742701](https://pubmed.ncbi.nlm.nih.gov/28742701/)
41. Inouye SK, Westendorp RGJ, Saczynski JS. Delirium in elderly people. *Lancet*. 2014;383(9920):911–22. [https://doi.org/10.1016/S0140-6736\(13\)60688-1](https://doi.org/10.1016/S0140-6736(13)60688-1) PMID: [23992774](https://pubmed.ncbi.nlm.nih.gov/23992774/)
42. DeClercq V, Duhamel TA, Theou O, Kehler S. Association between lifestyle behaviors and frailty in Atlantic Canadian males and females. *Arch Gerontol Geriatr*. 2020;91:104207. <https://doi.org/10.1016/j.archger.2020.104207> PMID: [32781378](https://pubmed.ncbi.nlm.nih.gov/32781378/)
43. Wang T, Sun S, Li S, Sun Y, Sun Y, Zhang D, et al. Alcohol consumption and functional limitations in older men: Does muscle strength mediate them? *J Am Geriatr Soc*. 2019;67(11):2331–7. <https://doi.org/10.1111/jgs.16082> PMID: [31373385](https://pubmed.ncbi.nlm.nih.gov/31373385/)
44. Fong TG, Tulebaev SR, Inouye SK. Delirium in elderly adults: diagnosis, prevention and treatment. *Nat Rev Neurol*. 2009;5(4):210–20. <https://doi.org/10.1038/nrneurol.2009.24> PMID: [19347026](https://pubmed.ncbi.nlm.nih.gov/19347026/)
45. Jameson JL, Longo DL. Precision medicine—personalized, problematic, and promising. *N Engl J Med*. 2015;372(23):2229–34. <https://doi.org/10.1056/NEJMs1503104> PMID: [26014593](https://pubmed.ncbi.nlm.nih.gov/26014593/)
46. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925–31. <https://doi.org/10.1093/eurheartj/ehu207> PMID: [24898551](https://pubmed.ncbi.nlm.nih.gov/24898551/)
47. Abelha FJ, Luis C, Veiga D, Parente D, Fernandes V, Santos P, et al. Outcome and quality of life in patients with postoperative delirium during an ICU stay following major surgery. *Crit Care Lond Engl*. 2013;17:R257. <https://doi.org/10.1186/cc13084> PMID: [24168808](https://pubmed.ncbi.nlm.nih.gov/24168808/)
48. Saczynski JS, Marcantonio ER, Quach L, Fong TG, Gross A, Inouye SK, et al. Cognitive trajectories after postoperative delirium. *N Engl J Med*. 2012;367(1):30–9. <https://doi.org/10.1056/NEJMoa1112923> PMID: [22762316](https://pubmed.ncbi.nlm.nih.gov/22762316/)
49. Brown CH 4th, Probert J, Healy R, Parish M, Nomura Y, Yamaguchi A, et al. Cognitive decline after delirium in patients undergoing cardiac surgery. *Anesthesiology*. 2018;129(3):406–16. <https://doi.org/10.1097/ALN.0000000000002253> PMID: [29771710](https://pubmed.ncbi.nlm.nih.gov/29771710/)