You *et al. BioData Mining* (2026) 19:31
https://doi.org/10.1186/s13040-026-00531-7

BioData Mining

RESEARCH Open Access

Multi-output LSTM-based prediction of postoperative delirium: integrating baseline and perioperative data for enhanced risk stratification in older spine surgery patients

Jungmin You[1†], Jeongmin Kim[2,3†], Jeongeun Choi[1,4], Bon-Nyeo Koo[2,3*] and Hyangkyu Lee[1,5*]

†Jungmin You and Jeongmin Kim contributed equally to this work.

*Correspondence:
Bon-Nyeo Koo
koobn@yuhs.ac
Hyangkyu Lee
HKYULEE@yuhs.ac
[1]Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea
[2]Department of Anesthesiology and Pain Medicine, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea
[3]Anesthesiology and Pain Research Institute, Yonsei University College of Medicine, Seoul, Republic of Korea
[4]College of Nursing and Brain Korea 21 FOUR Project, Yonsei University, Seoul, Republic of Korea
[5]Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, Republic of Korea

Abstract

Introduction Postoperative delirium (POD) adversely affects clinical outcomes among older adults undergoing spine surgery. However, existing predictive models often neglect multidimensional nature of delirium, including its clinical subtype, duration, severity, and timing. This study developed a multi-output Long Short-Term Memory (LSTM) neural network that integrates preoperative baseline characteristics and intraoperative acute stressors to predict multiple clinical dimensions of POD in elderly patients undergoing spinal surgery.

Methods This prospective observational study included 536 patients aged 70 or older who underwent elective spine surgery between November 2019 and May 2023. Comprehensive assessments were conducted during both the preoperative and intraoperative phases. The multi-output LSTM model incorporated preoperative baseline variables (demographic, frailty scores, cognitive function, medication count, and laboratory parameters) and intraoperative data (surgical invasiveness, duration of surgery and anesthesia, intraoperative fluid management, immediate postoperative medication use). Outcomes comprised delirium occurrence, subtype, duration, severity, and onset timing. Model performance was evaluated via accuracy, precision, recall, F1-score, and ROC curve analyses. SHapley Additive exPlanations (SHAP) analysis enhanced clinical interpretability.

Results Using solely preoperative baseline data, the model demonstrated strong predictive performance with an overall AUC of 0.76, particularly for delirium occurrence (AUC = 0.68), the duration (AUC = 0.80), and severity (AUC = 0.79). Incorporating intraoperative data substantially enhanced model performance, increasing the overall AUC to 0.81, notably improving predictions for delirium subtype (AUC up to 0.84), duration (AUC = 0.81), and onset timing (AUC up to 0.87). SHAP analysis consistently identified frailty, polypharmacy, cognitive impairment, nutritional deficiencies, and acute perioperative factors—such as surgical invasiveness, pain management—as pivotal predictors across delirium dimensions.

Conclusion The proposed multi-output LSTM model predicted multiple clinical dimensions of postoperative delirium, highlighting baseline health status as a primary determinant. Strategic integration of comprehensive baseline assessments with acute perioperative data substantially enhances predictive accuracy, informing personalized delirium prevention and management strategies for improved perioperative outcomes in older spine surgery patients.

Keywords Postoperative delirium, Spine surgery, Older adults, Multi-output prediction, Long short-term memory (LSTM), Machine learning, SHAP (SHapley Additive exPlanations), Frailty, Clinical decision support, Perioperative management

Introduction

Postoperative delirium (POD) is an acute cognitive disorder primarily characterized by memory deficits and impaired consciousness, typically occurring between 2 and 5 days after surgery [1, 2]. The incidence rates of POD ranges from 5% to 50% across cohort studies and increases substantially with advancing age [3]. Given the heightened vulnerability of older surgical patients, POD has emerged as a critical clinical concern in geriatric perioperative care.

Spine surgery is among the most frequently performed surgical procedures in older adults, ranking among the top five in individuals aged 65 to 80 years [4]. According to two meta-analyses, POD incidence in older adults undergoing spine surgery is approximately 8% and 13%, respectively, although estimates vary across studies [5, 6]. The global rise in aging populations and advances in surgery have increased the number of spinal procedures among older adults [7], consequently expanding the population at risk for POD.

POD is associated with a wide range of adverse outcomes, including prolonged hospital stays, higher readmission rates, and elevated healthcare expenditure [8]. Beyond these short-term effects, POD also poses adverse long-term risks. A prospective cohort study observed that older patients with POD following orthopedic surgery experienced greater declines in activities of daily living and higher mortality rates over a 24- to 36-month follow-up period compared with those without POD [9]. Furthermore, recent meta-analyses indicate that POD is associated with subsequent cognitive decline [10] and an increased risk of dementia [11].

Approximately 30% to 40% of POD cases in older adults are preventable and reversible when timely interventions are implemented before onset [3]. Therefore, early prediction is crucial for reducing both the incidence and severity of POD. The identified risk factors for POD after spine surgery include advanced age, pain, and prolonged operative time [8].

Machine learning-based prediction models offer valuable tools for estimating POD risk and supporting preoperative screening in spine surgery patients, enabling targeted interventions for high-risk individuals [12, 13]. However, most existing models have been developed for ICU or general inpatient populations and are not specifically designed for older adults undergoing spinal surgery [14]. A recent systematic review on machine learning-based POD prediction models observed that only 1 of 23 studies focused on spine surgery patients [15], highlighting a lack of tailored approaches for this growing population. Moreover, current models for spine surgery patients primarily predict POD occurrence, neglecting critical dimensions such as severity, onset timing, or

clinical subtypes [13]. This limitation results in an incomplete approach to comprehensive risk assessment.

Identifying baseline patient conditions and vulnerabilities during the preoperative phase is crucial for early delirium risk assessment and establishing preventive strategies [16]. Additionally, evaluating acute physiological changes and medication use during the intraoperative phase facilitates more precise delirium prediction and individualized management [12]. Therefore, an integrated approach combining preoperative baseline factors and intraoperative acute stressors significantly enhances the strategic value of accurately predicting and managing not only the occurrence of delirium but also its clinical subtypes, timing, and severity.

To address these gaps, we analyzed a prospectively enrolled cohort of spine-surgery patients aged seventy years or older, a methodological focus directly informed by the predisposing and precipitating risk factor framework established in a recent meta-analysis of geriatric spine surgery patients [6]. Although international guidelines provide varying definitions for the geriatric population, our selection of the seventy-year threshold was guided by evidence showing that delirium incidence increases exponentially in this specific subgroup, reaching over 40% compared to younger geriatric cohorts [17]. Furthermore, large-scale spine-surgery data confirm that delirium is significantly more frequent in these older patients and is linked to increased length of stay and mortality, supporting the need for specialized prediction tools tailored to this highest-risk population [18, 19]. By prospectively validating variables identified as significant predictors in prior systematic reviews, such as preoperative opioid use and operative time, we aimed to refine risk stratification and optimize clinical outcomes for this vulnerable surgical group.

This study aimed to develop a prediction model specifically tailored to older adults undergoing spine surgery using a long short-term memory (LSTM) neural network. The LSTM network captures temporal dependencies in clinical data, which are essential for modeling sequential patterns influencing patient outcomes over time [12, 20]. The proposed model utilized data from both the preoperative phase—including baseline patient conditions such as demographic variables, functional status, frailty scores, medication use, and laboratory results—and intraoperative phase, which incorporated preoperative phase data in conjunction with acute surgical stressors such as surgical invasiveness, intraoperative hemodynamic instability, and immediate postoperative medication use. This multi-phase approach was designed to estimate POD risk and predict its clinical subtypes, duration, severity, and onset timing, thereby enabling more refined risk stratification and supporting individualized perioperative management.

Methods

Ethics approval

This was a secondary analysis of a previous prospective observational study that primarily focused on postoperative delirium and was approved by the local institutional review board (Severance Hospital 4-2019-0654; ClinicalTrials.gov Identifier: NCT04120272). All participants provided written informed consent.

Research design and participants

This prospective observational study was performed at an academic tertiary hospital in South Korea between November 2019 and May 2023. Participants were patients aged 70 years or older scheduled for elective spinal surgery. Prior to enrollment, cognitive screening using the Mini-Mental State Examination for Dementia Screening (MMSE-DS [21]) resulted in the initial recruitment of 600 cognitively intact individuals. Exclusion criteria were rigorously applied, resulting in the exclusion of 64 participants. These criteria included cognitive impairment as defined by the MMSE-DS, diagnosis of malignancy within the previous five years, illiteracy or significant language impairment, a documented history of neurological conditions (e.g., seizures, stroke, dementia), confirmed alcoholism or substance abuse disorders, and planned surgical procedures with an expected duration of less than two hours. This operative-time criterion was applied to focus on major elective spine procedures with more comparable anesthesia exposure and perioperative physiological stress, given that longer operative duration is a well-established risk factor for postoperative delirium in older surgical patients and has also been identified as a relevant factor in spine-surgery populations [1, 18, 22]. This selection process resulted in an initial analytic cohort of 536 patients, as summarized in Fig. 1.

Data collection

POD evaluations were systematically performed twice daily from postoperative days 1 to 3 and subsequently once daily from postoperative days 4 to 7. Trained nursing staff administered the confusion assessment method (CAM) [23]. Positive delirium assessments identified by nurses were confirmed via subsequent evaluations by experienced physicians, who finalized patient assignments to the delirium cohort. Additionally, delirium severity and symptomatic presentations were comprehensively evaluated using the Korean adaptation of the Delirium Rating Scale–Revised-98 (K-DRS-R-98 [24, 25]).

The study aimed to accurately predict five distinct POD results: the occurrence of delirium (absence or presence), clinical subtype (no delirium, hyperactive, hypoactive, or mixed-type), duration (categorized as no delirium, 1 day, 2 days, 3 days, or 4 or more days), severity based on K-DRS-R-98 scores (normal < 15, mild-to-moderate delirium

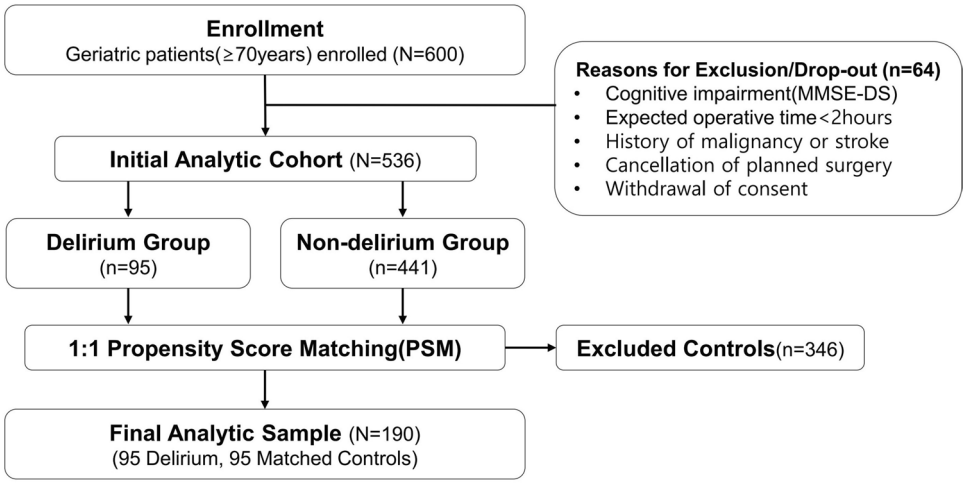


Fig. 1 Flowchart of participant selection and propensity score matching

15–25, severe delirium ≥ 26), and onset timing (no delirium, onset on the day of surgery, or onset occurring after postoperative day 1).

To provide a multi-dimensional representation of patient risk, we compiled a comprehensive dataset encompassing demographic characteristics, preoperative functional status, intraoperative physiological stressors, and an extensive panel of laboratory results. These variables were strategically selected based on their established clinical relevance to geriatric perioperative outcomes and recognized risk factors for postoperative delirium [3, 26]. Specifically, we incorporated geriatric-specific domains—such as frailty, cognitive function, and nutritional status—to establish a robust baseline of preoperative vulnerability [27, 28]. The complete list of variables, categorized by their respective clinical domains and data sources, is detailed in Supplementary Table 1.

Incremental predictive value of perioperative clinical data

To accurately assess the incremental predictive value of perioperative clinical variables, data collection was explicitly categorized into two clearly defined phases (Table 1). The preoperative phase comprised baseline patient characteristics, including sociodemographic variables, functional assessments such as frailty levels (FRAIL scale), activities of daily living (K-ADL), instrumental activities of daily living (K-IADL), detailed medication profiles, and comprehensive laboratory findings. In this study, the intraoperative phase extended the preoperative dataset by adding day-of-surgery variables and intraoperative measurements, including surgical invasiveness, intraoperative hemodynamic variability, quantified blood loss, meticulous fluid balance tracking, and immediate postoperative analgesic interventions.

To quantify the incremental predictive value across clinical periods, we implemented the phase-specific multi-output architecture using structured feature sets. While intraoperative physiologic data were available as high-frequency records, they were represented as fixed-length aggregated feature vectors for model input. Specifically, hemodynamic variables were summarized using extreme values and body temperature as the maximum value within the surgical window. Additional intraoperative stressors,

Table 1 Variables included in the prediction model development

Category	Number of variables	Examples	Data collection phase
General characteristics	20	Age, Sex, BMI, Education, Marital Status	Preoperative
Clinical characteristics	6	ASA Score, CCI, Operative Level	Preoperative & Intraoperative (Day of Surgery)
Preoperative functional status	6	K-FRAIL, GDSSF-K, MNA-S, K-ADL, K-IADL	Preoperative
Vital signs	19	Highest SBP, Lowest DBP, Body Temperature	Preoperative & Intraoperative
Pain assessment	6	NRS pain scores (rest/movement)	Intraoperative
Postoperative analgesic use	9	Total Opioid via IV PCA, Additional Analgesics	Intraoperative
Intraoperative intake/output	4	Intake (mL), Output (mL), Blood Loss (mL)	Intraoperative
Postoperative complications	1	Other complications	Intraoperative
Laboratory data	41	CBC, ESR, CRP, Glucose, BUN, Creatinine, Electrolytes	Preoperative

Note: POD = Postoperative day; ASA = American Society of Anesthesiologists; BMI = Body Mass Index; CCI = Charlson Comorbidity Index; K-FRAIL = Korean FRAIL scale; GDSSF-K = Geriatric Depression Scale Short Form (Korean); MNA-S = Mini Nutritional Assessment Short Form; K-ADL = Korean Activities of Daily Living; K-IADL = Korean Instrumental Activities of Daily Living; SBP = Systolic Blood Pressure; DBP = Diastolic Blood Pressure; NRS = Numeric Rating Scale; IV PCA = Intravenous Patient-Controlled Analgesia; CBC = Complete Blood Count; ESR = Erythrocyte Sedimentation Rate; CRP = C-Reactive Protein; BUN = Blood Urea Nitrogen

including anesthesia duration, estimated blood loss, fluid balance, and cumulative analgesic administration, were incorporated as patient-level scalars. Because the models were trained on these aggregated representations rather than multi-time-step waveforms, sequence-related preprocessing such as padding, truncation, or duration-based standardization was not required. This structured input format enabled a stable comparison of feature attributions across the preoperative and intraoperative phases.

Variables reflecting anesthetic exposure and its physiological correlations (e.g., duration of anesthesia, intraoperative hemodynamics, and postoperative opioid administration) were extracted and included as candidate predictors (Supplementary Table 1).

Data processing

Data preprocessing included categorizing variables as either continuous or categorical. Outliers in continuous variables were adjusted using the interquartile range (IQR) method [29]. Regarding missing data, the majority of clinical and demographic variables exhibited minimal missingness (less than 2%). For these and other continuous laboratory parameters, missing values were imputed using the median, while categorical variables were imputed using the mode. This approach is a validated strategy in clinical machine learning to maintain the total sample size ($N = 536$ before PSM matching) while preserving the underlying data distribution. Continuous variables were standardized with StandardScaler to enhance the stability and convergence of the multi-task LSTM model [30].

Propensity score matching and analytic dataset

To mitigate potential confounding bias and ensure baseline comparability between delirium and non-delirium groups, propensity score matching (PSM) was performed. This statistical method pairs participants from the delirium group with those from the non-delirium group based on similarity in baseline demographic and clinical characteristics, facilitating more precise evaluation of delirium results. Patients diagnosed with delirium ($n = 95$) were matched in a 1:1 ratio with 95 non-delirium controls drawn from the initial cohort of 536 older adults undergoing spine surgery. Matching used propensity scores based on key baseline covariates: age, gender, frailty index, functional assessments (ADL, IADL), cognitive function (MMSE), nutritional status (MNA), medication count, and laboratory parameters. Propensity scores were estimated using baseline covariates only, and all predictors were defined using measurements obtained prior to delirium outcome ascertainment. This approach effectively balanced covariates across groups, thereby enhancing the internal validity of the subsequent analyses [31, 32]. Importantly, the PSM-matched cohort ($n = 190$) was used as the analytic dataset for both model development (training) and evaluation (testing/validation). Model training and internal validation were conducted using an 80:20 training–validation split within the matched cohort. Following PSM, separate datasets were created for the preoperative and intraoperative phases to evaluate incremental predictive performance associated with the availability of additional perioperative data. To align with the sequential prediction design, covariate balance was reported separately by phase. Supplementary Table 2A summarizes preoperative baseline variables, including laboratory values obtained before surgery. Supplementary Table 2B summarizes variables available on the day of surgery and during surgery, including acute perioperative stressors and laboratory values measured during the intraoperative phase, which may differ from preoperative baselines. Detailed

results of the matching process and covariate balance assessments are provided in the supplementary material (Supplementary Table 2A-2B), demonstrating the effectiveness and appropriateness of the employed PSM methodology [33].

Multi-output LSTM model implementation

A multi-output LSTM neural network was developed to predict five postoperative delirium outcomes: occurrence, motor subtype, duration, severity, and onset timing in older adults undergoing spine surgery. To ensure balanced comparison and mitigate confounding, we established a cohort using propensity score matching (PSM), resulting in 95 delirium cases and 95 matched non-delirium controls ($N = 190$). As illustrated in Fig. 2, we implemented a phase-specific multi-output LSTM framework that utilizes a feature-to-sequence transformation to integrate heterogeneous clinical data.

To support multi-dimensional delirium prediction under modest sample sizes, we implemented a multi-task LSTM framework. While newer architectures such as Transformers offer high modeling capacity, recurrent baselines like LSTM are often preferred for clinical cohorts of modest size to mitigate the risk of overfitting while maintaining stable joint optimization across concurrent tasks [34]. Specifically, for the preoperative model, the 127 structured variables were concatenated into a feature vector $V_{pre} \in \mathbb{R}^{127}$. This vector was then reshaped into a three-dimensional tensor of (1,127), treating the clinical snapshot as a single-step sequence. Similarly, for the intraoperative-phase model, 135 variables were transformed into a (1, 135) input tensor, treating each clinical snapshot as a single-step sequence. We retained an LSTM backbone to provide a shared representation across tasks and to support stable joint optimization in the multi-output setting, consistent with multitask learning principles and established clinical time-series benchmarks [35]. This representation allows the shared LSTM layer with 64 units and hyperbolic tangent activation to function as a high-dimensional feature extractor, capturing non-linear interactions across the entire feature set within a single temporal gate [35–37]. This shared layer connected to task-specific dense layers tailored to each delirium-related outcome. The optimization of the multi-task LSTM model was conducted using task-specific loss functions and probability thresholds. The primary delirium occurrence output utilized a Sigmoid activation function paired with Binary Cross-Entropy loss, with a standard classification threshold of 0.5. In contrast, the secondary outcomes were modeled using Softmax activation and Categorical Cross-Entropy loss, with predictions determined by the maximum class probability.

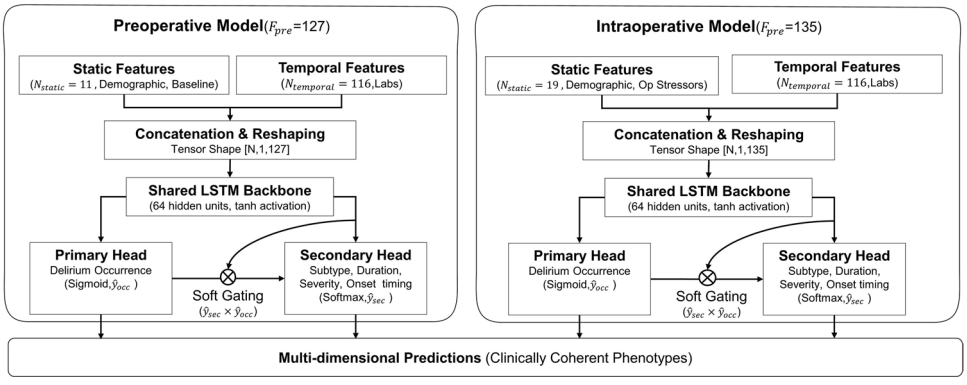


Fig. 2 Phase-specific multi-task LSTM framework with soft conditional gating

To minimize bias from class imbalance, particularly for motor subtypes, a class-weighted loss strategy was implemented. The final optimization utilized a joint loss function, defined as the sum of individual task-specific losses, allowing for a stable and shared feature representation across all delirium-related dimensions. Model optimization involved a systematic grid search of hyperparameters, including L2 regularization (10^{-7} to 10^{-1}), dropout rates, and learning rates. The model was trained with a stratified 80:20 training–validation split based on the primary outcome, ensuring the 1:1 matched ratio was maintained across sets [14, 38]. The final hyperparameter settings and the specific class distributions for all five outputs are detailed in Supplementary Table 3. Early stopping was implemented after twenty epochs without improvement to ensure robustness and prevent overfitting. A key feature was a soft conditional gating mechanism, whereby secondary results were dynamically scaled by the primary delirium probability, maintaining clinical coherence and enabling effective model optimization [39, 40]. The integrated multi-task design enhanced predictive accuracy and provided clinically actionable insights beyond traditional binary assessments [41, 42].

Performance metrics and interpretation

The multi-task LSTM model was evaluated using several key metrics, including classification accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). ROC curve analysis enabled visual assessment of predictive performance consistency across perioperative phases. Additionally, SHapley Additive exPlanations (SHAP) analysis quantified the relative importance of clinical predictors, improving interpretability and clinical relevance.

Results

Demographic and clinical characteristics

A total of 536 older adults who underwent spine surgery were included in the final analysis. The median participant age was 75 years (interquartile range [IQR]: 72–78 years), with 342 (63.8%) being female. Baseline demographic and clinical characteristics, including the frailty index, activities of daily living (ADL), instrumental activities of daily living (IADL), mini nutritional assessment (MNA) scores, cognitive function assessed by the mini-mental state examination (MMSE), baseline medication count, and key surgical characteristics are summarized in Table 2. Statistically significant differences between the delirium and non-delirium groups were observed in age ($p = 0.027$) and MNA scores ($p = 0.023$). No significant differences were found in other baseline demographic or clinical characteristics.

Propensity score matching

Propensity score matching (PSM) yielded a balanced analytic cohort comprising 95 delirium cases and 95 matched non-delirium controls. Preoperative baseline balance was assessed using absolute standardized mean differences (|SMD|) and group comparisons (Supplementary Table 2A). Overall balance was acceptable, with 52 of 69 variables achieving |SMD| <0.10 and 66 of 69 achieving |SMD| <0.20. A small number of covariates showed residual imbalance, most notably vision impairment (|SMD|=0.235), cigarettes per day (|SMD|=0.233), and serum sodium (|SMD|=0.228).

Table 2 Baseline demographic and clinical characteristics of study participants

	Delirium (*N* = 95)	Non-delirium (*N* = 441)	*p*-value
Age(year)	76 (73, 79)	75(72, 78)	0.027*
Sex(M/F)	30/65 (31.6%)	164/277 (37.2%)	0.637
Height(cm)	154.0 (150.4, 162.5)	154.5 (150.0, 161.5)	0.721
BMI	24.09 (21.66, 26.56)	24.21 (22.90, 26.58)	0.305
Charlson Comorbidity Index	4.0 (4.0, 5.0)	4.0 (3.0, 4.5)	0.052
Frailty index	4.0 (2.0, 8.0)	3.0 (1.0, 5.0)	0.089
Geriatric Depression Scale	2.0 (1.0, 3.0)	2.0 (1.0, 3.0)	0.159
ADL	7.0 (7.0, 8.0)	7.0 (7.0, 7.0)	0.081
IADL	11.0 (10.0, 16.0)	10.0 (10.0, 13.0)	0.054
Mini Nutritional Assessment	12.0 (10.0, 13.5)	13.0 (12.0, 14.0)	0.023*
Education			0.640
0–3 years	12 (12.6%)	42 (9.5%)	
4–6 years	32 (33.7%)	145 (32.9%)	
7–12 years	40 (42.1%)	187 (42.4%)	
Over 13 years	11 (11.6%)	67 (15.2%)	
Baseline Medication Count	7.0 (5.0, 11.0)	6.0 (3.0, 10.0)	0.090
MMSE	27.0 (25.5, 29.0)	27.0 (26.0, 29.0)	0.203
Surgical Complexity Level	2.0 (1.0, 3.0)	1.0 (1.0, 2.0)	0.064
Duration of Surgery (min)	176.0 (128.5, 221.0)	168.0 (125.5, 228.5)	0.966

Note: Values are presented as the median (Q1, Q3) or number of patients (%)

* *p*-value for the Mann-Whitney U analysis for continuous or ordinal variables and the chi-square test or Fisher's exact test for categorical variables, as appropriate BMI, Body Mass Index; ADL, Activities of Daily Living score; IADL, Instrumental Activities of Daily Living; MMSE, Mini-Mental State Examination

Intraoperative-phase variables, including day-of-surgery/intraoperative measures and perioperative laboratory values up to the immediate postoperative period, were evaluated separately (Supplementary Table 2B). Overall balance remained acceptable, with most variables achieving |SMD| < 0.20 and no variables exceeding |SMD| ≥ 0.20; the largest residual imbalance was observed for [variable] (|SMD|=[value]).

Overall model performance

The predictive performance of the multi-output LSTM model was systematically assessed across two distinct perioperative phases, reflecting incremental availability of clinical data: (1) the preoperative phase, using baseline clinical variables exclusively, and (2) the intraoperative phase, integrating baseline, intraoperative, and immediate postoperative clinical data. Comprehensive performance metrics, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC) with 95% confidence intervals (CIs) are presented in Table 3. The corresponding receiver operating characteristic (ROC) analyses and AUC curves for each perioperative phase are illustrated in Table 4, while detailed class-specific performance metrics are provided in Supplementary Table 4.

During the preoperative phase, the multi-output LSTM model demonstrated meaningful predictive capability with an overall AUC of 0.76 (95% CI, 0.73–0.79). Notably, delirium occurrence achieved an AUC of 0.68, underscoring the relevance of baseline demographic, functional, and clinical factors. Furthermore, clinical subtype predictions exhibited robust performance (AUC 0.82), highlighting the substantial predictive value of baseline data in distinguishing delirium subtypes. Duration and severity predictions were similarly effective, with AUC values of 0.80 and 0.79, respectively, while onset timing predictions provided good discriminative capability with an AUC of 0.78.

Table 3 Performance metrics of multi-output LSTM model by clinical period

Phase	Output	Accuracy	Recall	Precision	F1-score	AUC(95% CI)
Preoperative Phase	Overall	0.61	0.61	0.59	0.57	0.76 (0.73–0.79)
	Delirium Occurrence	0.58	0.58	0.58	0.58	0.68 (0.60–0.75)
	Clinical Subtype	0.59	0.59	0.63	0.57	0.82 (0.77–0.86)
	Duration	0.53	0.53	0.33	0.40	0.80 (0.75–0.84)
	Severity	0.58	0.58	0.62	0.59	0.79 (0.73–0.85)
	Onset Timing	0.76	0.76	0.83	0.71	0.78 (0.73–0.83)
Intraoperative Phase	Overall	0.64	0.65	0.63	0.59	0.81(0.78–0.84)
	Delirium Occurrence	0.68	0.68	0.69	0.68	0.74 (0.67–0.81)
	Clinical Subtype	0.61	0.61	0.74	0.61	0.84 (0.80–0.88)
	Duration	0.61	0.61	0.49	0.56	0.81 (0.76–0.85)
	Severity	0.55	0.55	0.59	0.59	0.87 (0.82–0.91)
	Onset Timing	0.74	0.74	0.71	0.67	0.87 (0.83–0.91)

Note: AUC, area under the receiver operating characteristic curve; CI, confidence interval. Multi-output/class targets where metrics are reported as weighted averages to account for class support

The intraoperative phase, incorporating intraoperative and immediate postoperative clinical data, substantially enhanced the predictive performance of the model, elevating the overall AUC to 0.81 (95% CI, 0.78–0.84). The delirium occurrence prediction improved, achieving an AUC of 0.74 with increased accuracy (0.68) and precision (0.69) Clinical subtype predictions showed further refinement with an AUC of 0.84, highlighting the significant contribution of acute perioperative data. Notably, severity and onset timing predictions showed significant enhancement, both achieving an AUC of 0.87, further validating the exceptional predictive capacity of the model through integrated perioperative data.

SHAP importance analysis

The SHAP method enhanced the clinical interpretability of the multi-output LSTM model by identifying key predictors for each outcome. The top 20 clinical features for each outcome were analyzed, comparing results from the preoperative and intraoperative phases (Table 5).

Delirium occurrence In the preoperative phase, key predictors for delirium occurrence included the frailty index (FRAIL), medication count (med_count), alkaline phosphatase (alk_phosphatase), IADL, and large unstained cells count (luc_num). FRAIL and med_count showed the most substantial predictive impact, indicating that baseline frailty and polypharmacy significantly contribute to delirium risk, reflecting underlying patient vulnerabilities [43]. In the intraoperative phase, FRAIL and IADL remained essential predictors. Concurrently, preoperative chloride level (Cl), additional narcotic analgesic administration on the day of surgery (extra_narc0_count), and lowest preoperative systolic blood pressure (lowest sbp) also emerged as significant predictors. Low chloride levels may reflect electrolyte imbalances leading to neurological instability [3], and narcotic analgesics can exacerbate delirium risk by affecting central nervous system functions [43]. These findings suggest that baseline health status remains crucial, with acute perioperative factors incrementally enhancing delirium prediction.

Clinical subtype In the preoperative phase, key features for predicting clinical subtypes included medication count, FRAIL, Mini Nutritional Assessment (MNA), renal function

Table 4 ROC curves for multi-output LSTM model by clinical period

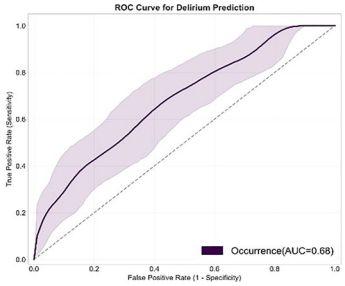
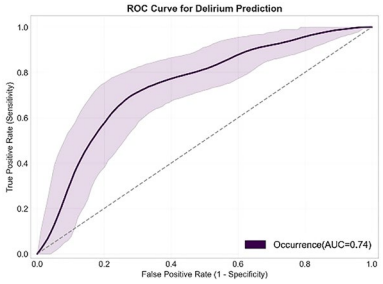
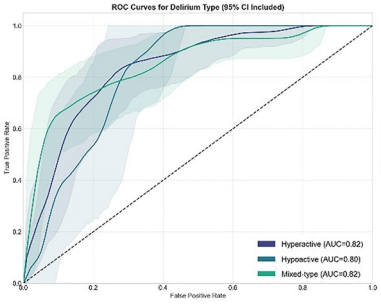
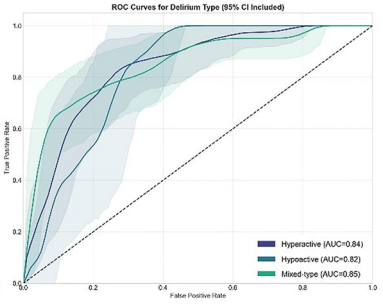
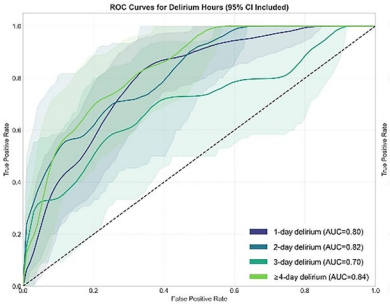
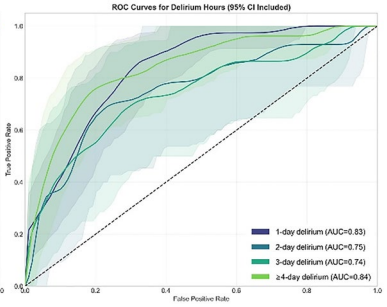
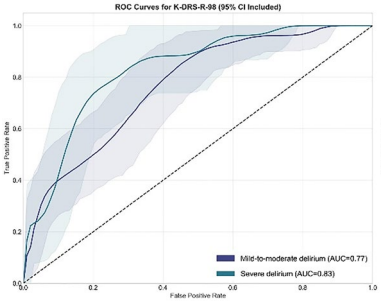
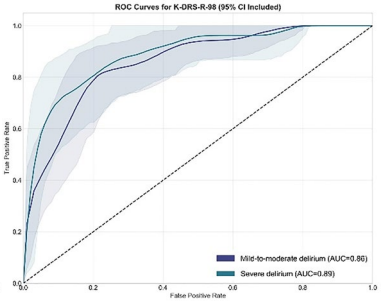
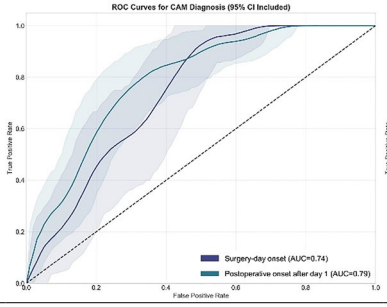
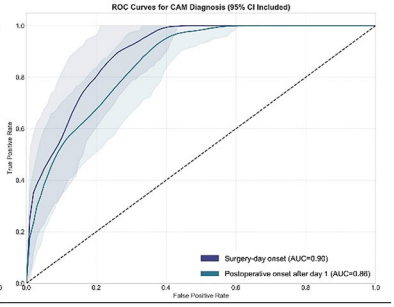
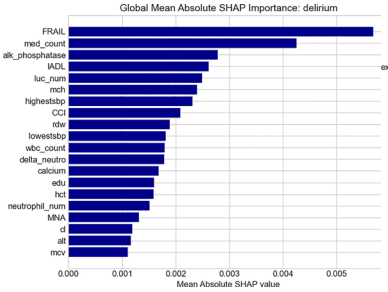
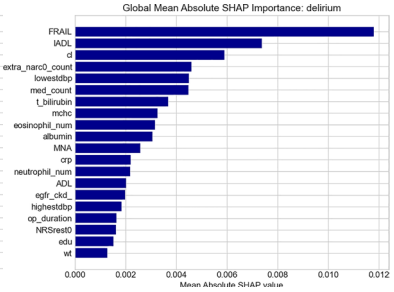
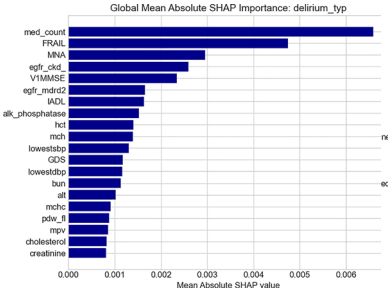
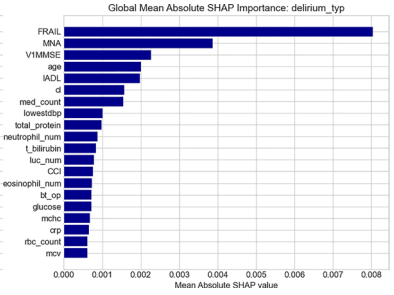
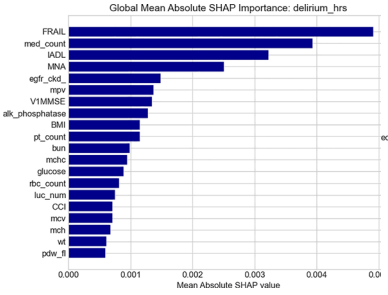
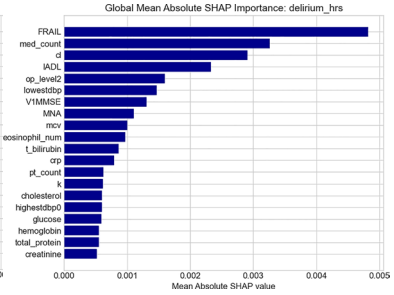
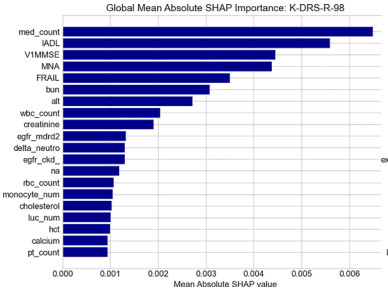
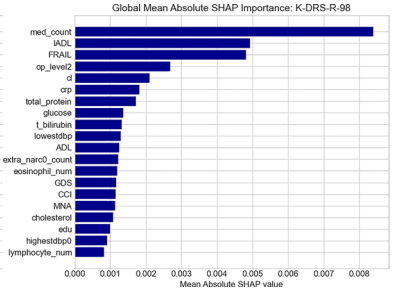
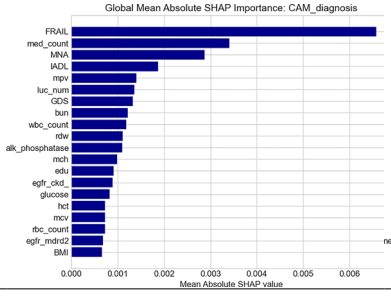
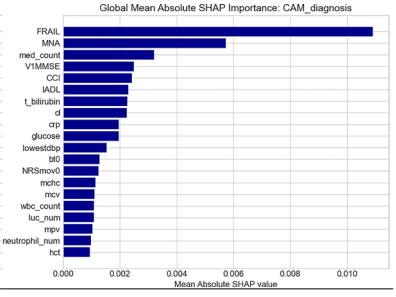
	Preoperative phase	Intraoperative phase
Delirium Occurrence		
Clinical Subtype		
Duration		
Severity		
Onset timing		

Table 5 SHAP importance for multi-output LSTM model by clinical period

	Preoperative phase	Intraoperative phase
Delirium Occurrence		
Clinical Subtype		
Duration		
Severity		
Onset Timing		

(eGFR_ckd, eGFR_mdrd2), cognitive function (MMSE), and IADL. The high importance of polypharmacy, frailty, and impaired nutritional and cognitive states emphasizes the role of baseline patient conditions in shaping delirium clinical presentation [44]. During the intraoperative phase, FRAIL, MNA, MMSE, age, and IADL continued to show strong importance, emphasizing further the critical impact of baseline vulnerability, while age emerged as additionally substantial, highlighting the amplified vulnerability of older patients under acute surgical stress.

Duration In the preoperative phase, FRAIL, medication count, IADL, MNA, renal function (eGFR_ckd), mean platelet volume (mpv), and cognitive function (MMSE) were predictive of delirium duration, reinforcing the critical role of baseline frailty, polypharmacy, and functional status in predicting prolonged delirium episodes. In the intraoperative phase, FRAIL, medication count, surgical invasiveness (op_level2), lowest preoperative systolic blood pressure, MMSE, and MNA remained important. Surgical invasiveness and low systolic blood pressure, indicating acute stress and possible cerebral hypoperfusion, emerged as additional factors influencing delirium duration [45]. Nevertheless, baseline health remained a primary determinant of delirium duration.

Severity Preoperatively, medication count, IADL, MMSE, MNA, FRAIL, and blood urea nitrogen (bun) were predominant predictors, indicating the strong correlation between baseline cognitive, functional, nutritional, and frailty status with delirium severity. In the intraoperative phase, medication count, IADL, FRAIL, surgical invasiveness, preoperative chloride, and the inflammation marker CRP emerged as significant. Elevated CRP levels likely indicate acute neuroinflammation that exacerbates delirium severity [46]. Therefore, baseline patient vulnerability and acute inflammatory responses collectively influence delirium severity.

Onset timing In the preoperative phase, FRAIL, medication count, MNA, IADL, mean platelet volume (mpv), large unstained cells count (luc_num), and depression score (GDS) were key predictors [47]. Elevated platelet activation and white blood cell count indicates chronic inflammation and underlying vascular dysfunction, potentially accelerating delirium onset [48]. In the intraoperative phase, FRAIL, MNA, medication count, MMSE, surgical invasiveness, preoperative chloride (Cl), CRP, and total bilirubin (t_bilirubin) were prominent predictors. Elevated CRP and bilirubin levels during acute surgical stress may trigger rapid metabolic and neurological disturbances, hastening delirium onset [46]. Nevertheless, baseline conditions remained strong predictors for onset timing.

Discussion

This study provides robust evidence demonstrating the efficacy of a multi-output LSTM neural network in accurately predicting various clinical dimensions of POD, including its occurrence, clinical subtype, duration, severity, and onset timing. By integrating data across distinct perioperative phases, notably preoperative baseline and intraoperative periods, the model achieved significant predictive precision.

Analysis limited to preoperative baseline clinical data alone demonstrated substantial predictive performance, reflected by an overall AUC of 0.76 (95% CI, 0.73–0.79). These outcomes highlight the foundational significance of baseline patient health metrics, such

as frailty, polypharmacy, cognitive function, nutritional status, and functional impairments, in delineating risk profiles for delirium. Consistently, these baseline parameters were strong predictors of delirium susceptibility, clinical course, and outcomes, aligning closely with existing literature [43, 44].

Including intraoperative and immediate postoperative variables markedly augmented the predictive capability of the model, enhancing the overall AUC to 0.81 (95% CI, 0.78–0.84). This increment highlights the significant contribution of acute perioperative factors—including surgical invasiveness, duration of anesthesia and surgery, intraoperative pain management, and immediate postoperative pharmacological interventions—to improving prediction accuracy and refining risk stratification [43, 45, 46]. Nevertheless, despite these enhancements, baseline conditions retained strong predictive value, reinforcing their fundamental role in determining delirium outcomes. Notably, baseline frailty, cognitive impairments, polypharmacy, and nutritional deficiencies remained pivotal predictors across both perioperative phases.

These findings support and extend existing research underscoring the incremental predictive value of integrating multiple perioperative clinical phases, confirming that the addition of intraoperative data significantly enhances the precision and reliability of delirium predictions [48, 49]. Importantly, this study emphasizes the key role of baseline vulnerabilities in shaping delirium outcomes. Therefore, a strategic combination of comprehensive preoperative assessments and meticulous intraoperative monitoring emerges as critical for tailored delirium prevention and management protocols, ultimately optimizing clinical outcomes for elderly patients undergoing spine surgery.

Given that frailty, polypharmacy, cognitive impairment, nutritional vulnerability, and functional limitations were consistently important predictors, these geriatric domains may be used for pragmatic preoperative risk stratification to inform targeted delirium-prevention strategies. This is consistent with prior work highlighting the importance of preoperative comprehensive geriatric assessment (CGA) in delirium prevention among older surgical patients [50]. Evidence evaluating CGA as a preoperative intervention has reported lower postoperative delirium incidence across several surgical populations, including orthopedic surgery (5.6% vs. 18.5%) [51], vascular surgery (11% vs. 24%) [52], and elective colorectal surgery (11% vs. 29%) [53]. Notably, we found that frailty was consistently among the most influential features in the SHAP-derived feature importance results, suggesting that frailty-focused screening may be particularly informative for preoperative risk stratification and targeted delirium prevention. Consistent with this, multiple meta-analyses show that preoperative frailty is associated with a higher risk of postoperative delirium across surgical populations [54–56]. Collectively, these findings support incorporating frailty-focused screening within CGA-based preoperative workflows to more effectively identify high-risk patients and guide delirium-prevention strategies.

In practice, patients identified as high risk may benefit from an intensified delirium-prevention pathway and closer perioperative surveillance. Multicomponent nonpharmacologic prevention programs and perioperative reviews support structured screening and bundled prevention approaches in older surgical patients [1, 57]. Importantly, our SHAP analysis highlighted intraoperative predictors such as additional narcotic administration and hemodynamic vulnerability (e.g., lower systolic blood pressure), which can be translated into actionable targets: adopting opioid-sparing, multimodal analgesia with

avoidance of excessive sedation, and proactive hemodynamic optimization to minimize hypotension. Spine-surgery–specific evidence has linked intraoperative hypotension and intraoperative opioid use to postoperative delirium risk, and broader perioperative data also suggest that higher intraoperative opioid exposure may increase delirium risk in susceptible older patients [58, 59]. Accordingly, targeted measures may include heightened delirium screening during the first 48–72 h, careful medication review to minimize deliriogenic agents, opioid-sparing pain management, early mobilization, sleep promotion, and orientation support [1].

Nonetheless, several limitations should be considered regarding the findings of this research.

First, the model showed robust internal performance, external validation in independent, multi-center cohorts is necessary to establish generalizability. This study analyzed a single-center prospectively enrolled cohort restricted to elective spine-surgery patients aged seventy years or older. While this targeted approach may appear to limit applicability, it represents a deliberate methodological strategy to enhance predictive precision by focusing on the highest-risk population where delirium signals are most concentrated. Rather than utilizing expansive big data which often introduces substantial confounding noise and clinical heterogeneity, we prioritized the model's discriminative power through rigorous variable control within a homogeneous high-risk group. This design aligns with established evidence indicating that advanced age and prolonged operative time are among the most critical risk factors for delirium in geriatric spine surgery [6]. By excluding procedures with an expected duration of less than two hours, we aimed to maximize internal validity for major surgical stressors and provide more reliable risk stratification for the most vulnerable patients who stand to benefit most from early intervention.

Second, despite propensity score matching, residual confounding from unmeasured or imperfectly captured perioperative factors may remain. In addition, the number of POD events—particularly for multi-output targets such as subtype, severity, duration, and timing—was relatively limited, and the matching-related reduction in sample size may have increased uncertainty and reduced model stability. Because the model was developed and internally validated within the PSM-matched cohort, the reported performance reflects internal validity within this matched analytic sample and should be interpreted cautiously in terms of broader generalizability. Subtype prediction should be interpreted cautiously because subtype classes were imbalanced. Although class-weighted loss was applied to mitigate imbalance, estimates and performance metrics may be less stable for the Hypoactive subtype, increasing the risk of both type I and type II errors. Larger cohorts with sufficient subtype counts are needed for external validation and potential recalibration of subtype prediction.

Third, the primary limitation of this study remains the omission of high-resolution intraoperative physiological monitoring data within the feature engineering framework. Specifically, dynamic fluctuations in mean arterial pressure and end-tidal carbon dioxide were not integrated despite their critical roles in maintaining cerebral perfusion and oxygen saturation [60]. At our institution, these parameters were strictly managed according to standardized anesthetic protocols to ensure hemodynamic stability and patient safety. This high level of standardization resulted in minimal inter-patient variance and likely reduced discriminative power for machine learning prediction. While

such standardized care is essential for postoperative cognitive protection [61], our analysis identified supplementary narcotic administration as a more decisive pharmacological trigger for delirium occurrence and severity [62]. This underscores the necessity for future research utilizing high-frequency waveform data to more effectively capture these complex and heterogeneous clinical interactions.

In addition, retrospective external validation was not feasible within the scope of this study because the model relied on a prospective, protocolized dataset with geriatric domains that are not routinely captured in many perioperative datasets, which limits external harmonization. Future large-scale, multi-center prospective studies with standardized variable definitions and measurement timing are warranted to enable rigorous external validation and calibration assessment, as well as model updating and evaluation of longer-term outcomes and cost-effectiveness.

Our findings support integrating baseline geriatric vulnerabilities with perioperative acute stressors for more tailored delirium prevention in older spine-surgery patients, while multi-center validation remains a key next step for broader implementation.

Conclusions

In conclusion, this study demonstrates that a multi-output LSTM model can simultaneously forecast key clinical dimensions of postoperative delirium by integrating perioperative data. Baseline health status remains a critical determinant, underscoring the value of thorough preoperative evaluation. Therefore, combining baseline assessments with acute perioperative data significantly enhances the precision and clinical relevance of delirium prediction, paving the way for more personalized, targeted, and effective perioperative care strategies in older surgical populations.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13040-026-00531-7.

Supplementary Material 1

Acknowledgements

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education [No. 2020R1A6A1A03041989], and by the National Research Foundation of Korea (NRF) funded by the Korea government (MSIT) [No. 2022R1C1C1009622, 2023R1A2C1006054 and RS-2025-00520935].

Author contributions

J.Y. conceived and designed the study, curated and analyzed the data, conducted the investigation, developed the methodology, managed the project, validated the findings, visualized the results, and drafted and revised the manuscript. J.K. contributed to the study conception and design, data curation, investigation, methodology development, validation, visualization, and manuscript review and editing. J.C. contributed to data curation, investigation, and manuscript review and editing. B.-N.K. contributed to the study conception and design, funding acquisition, data curation, investigation, methodology development, project administration, and supervision. H.L. contributed to the study conception and design, funding acquisition, data curation, investigation, methodology development, project administration, validation, visualization, manuscript review and editing, and overall supervision of the study. All authors reviewed and approved the final manuscript.

Data availability

The datasets generated and/or analysed during the current study are not publicly available due to still being used for further analysis. However, upon approval of the final manuscript related to these data, they may be provided by the corresponding author upon reasonable request.

Declarations

Consent for publication

Not applicable.

Competing interests
The authors declare no competing interests.

Received: 27 October 2025 / Accepted: 13 February 2026
Published online: 24 March 2026

References

1. Jin Z, Hu J, Ma D. Postoperative delirium: perioperative assessment, risk reduction, and management. Br J Anaesth. 2020;125:492–504. https://doi.org/10.1016/j.bja.2020.06.063.
2. Xiao MZ, Liu CX, Zhou LG, Yang Y, Wang Y. Postoperative delirium, neuroinflammation, and influencing factors of postoperative delirium: a review. Medicine. 2023;102:e32991. https://doi.org/10.1097/MD.0000000000032991.
3. Inouye SK, Westendorp RGJ, Saczynski JS. Delirium in elderly people. Lancet. 2014;383:911–22. https://doi.org/10.1016/S0140-6736(13)60688-1.
4. Deiner S, Westlake B, Dutton RP. Patterns of surgical care and complications in elderly adults. J Am Geriatr Soc. 2014;62:829–35. https://doi.org/10.1111/jgs.12794.
5. Wu X, Sun W, Tan M. Incidence and risk factors for postoperative delirium in patients undergoing spine surgery: a systematic review and meta-analysis. BioMed Res Int. 2019;2139834. https://doi.org/10.1155/2019/2139834.
6. Baek W, Kim YM, Lee H. Risk factors of postoperative delirium in older adult spine surgery patients: A meta-analysis. AORN J. 2020;112:650–61. https://doi.org/10.1002/aorn.13252.
7. Beschloss A, Dicindio C, Lombardi J, Varthi A, Ozturk A, Lehman R, et al. Marked increase in spinal deformity surgery throughout the United States. Spine. 2021;46:1402–8. https://doi.org/10.1097/BRS.0000000000004041.
8. Zhang HJ, Ma XH, Ye JB, Liu CZ, Zhou ZY. Systematic review and meta-analysis of risk factor for postoperative delirium following spinal surgery. J Orthop Surg Res. 2020;15:509. https://doi.org/10.1186/s13018-020-02035-4.
9. Shi Z, Mei X, Li C, Chen Y, Zheng H, Wu Y, et al. Postoperative delirium is associated with long-term decline in activities of daily living. Anesthesiology. 2019;131:492–500. https://doi.org/10.1097/ALN.0000000000002849.
10. Goldberg TE, Chen C, Wang Y, Jung E, Swanson A, Ing C, et al. Association of delirium with long-term cognitive decline: a meta-analysis. JAMA Neurol. 2020;77:1373–81. https://doi.org/10.1001/jamaneurol.2020.2273.
11. Mohanty S, Gillio A, Lindroth H, Ortiz D, Holler E, Azar J, et al. Major surgery and long term cognitive outcomes: the effect of postoperative delirium on dementia in the year following discharge. J Surg Res. 2022;270:327–34. https://doi.org/10.1016/j.jss.2021.08.043.
12. Giesa N, Sekutowicz M, Rubarth K, Spies CD, Balzer F, Haufe S, et al. Applying a transformer architecture to intraoperative temporal dynamics improves the prediction of postoperative delirium. Commun Med (Lond). 2024;4:251. https://doi.org/10.1038/s43856-024-00681-x.
13. Zhang Y, Wan DH, Chen M, Li YL, Ying H, Yao GL, et al. Automated machine learning-based model for the prediction of delirium in patients after surgery for degenerative spinal disease. CNS Neurosci Ther. 2023;29:282–95. https://doi.org/10.1111/cns.14002.
14. Xie Q, Wang X, Pei J, Wu Y, Guo Q, Su Y, et al. Machine learning–based prediction models for delirium: a systematic review and meta-analysis. J Am Med Dir Assoc. 2022;23:1655–e16686. https://doi.org/10.1016/j.jamda.2022.06.020.
15. Chen H, Yu D, Zhang J, Li J. Machine learning for prediction of postoperative delirium in adult patients: A systematic review and meta-analysis. Clin Ther. 2024;46:1069–81. https://doi.org/10.1016/j.clinthera.2024.09.013.
16. Kang T, Park SY, Lee JH, Lee SH, Park JH, Kim SK, et al. Incidence and risk factors of postoperative delirium after spinal surgery in older patients. Sci Rep. 2020;10:9232. https://doi.org/10.1038/s41598-020-66276-3.
17. Brown CHIV, 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.
18. Fineberg SJ, Nandyala SV, Marquez-Lara A, Oglesby M, Patel AA, Singh K. Incidence and risk factors for postoperative delirium after lumbar spine surgery. Spine (Phila Pa 1976). 2013;38(20):1790–6. https://doi.org/10.1097/BRS.0b013e3182a0d507.
19. Susano MJ, Scheetz SD, Grasfield RH, Cheung D, Xu X, Kang JD, et al. Retrospective analysis of perioperative variables associated with postoperative delirium and other adverse outcomes in older patients after spine surgery. J Neurosurg Anesthesiol. 2019;31(4):385–91. https://doi.org/10.1097/ANA.0000000000000566.
20. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9:1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.
21. Kim TH, Jhoo JH, Park JH, Kim JL, Ryu SH, Moon SW, et al. Korean version of mini mental status examination for dementia screening and its short form. Psychiatry Invest. 2010;7:102.
22. Ravi B, Pincus D, Choi S, Jenkinson R, Wasserstein DN, Redelmeier DA. Association of duration of surgery with postoperative delirium among patients receiving hip fracture repair. JAMA Netw Open. 2019;2(2):e190111. https://doi.org/10.1001/jamanetworkopen.2019.0111.
23. 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;113:941–8. https://doi.org/10.7326/0003-4819-113-12-941.
24. Trzepacz PT, Mittal D, Torres R, Kanary K, Norton J, Jimerson N. Validation of the Delirium Rating Scale-revised-98: comparison with the delirium rating scale and the cognitive test for delirium. J Neuropsychiatry Clin Neurosci. 2001;13:229–42. https://doi.org/10.1176/jnp.13.2.229.
25. Lim KO, Kim SY, Lee YH, Lee SW, Kim JL. A validation study for the Korean version of Delirium Rating Scale-Revised-98 (K-DRS-98). J Korean Neuropsychiatr Assoc. 2006;45:518–26.
26. Aldecoa C, Bettelli G, Bilotta F, Sanders RD, Audisio R, Borozdina A, et al. European Society of Anaesthesiology evidence-based and consensus-based guidelines on postoperative delirium. Eur J Anaesthesiol. 2017;34(4):192–214. https://doi.org/10.1097/EJA.0000000000000594.
27. Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. Lancet. 2013;381(9868):752–62. https://doi.org/10.1016/S0140-6736(12)62167-9.

28. Ellis G, Gardner M, Tsiachristas A, Langhorne P, Burke O, Harwood RH, et al. Comprehensive geriatric assessment for older adults admitted to hospital. Cochrane Database Syst Rev. 2017;9(9):CD006211. https://doi.org/10.1002/14651858.CD006211.pub3.
29. Boukerche A, Zheng L, Alfandi O. Outlier detection: methods, models, and classification. ACM Comput Surv. 2021;53:1–37. https://doi.org/10.1145/3381028.
30. Brownlee J. Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python. Machine Learning Mastery; 2020.
31. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivar Behav Res. 2011;46:399–424. https://doi.org/10.1080/00273171.2011.568786.
32. Benedetto U, Head SJ, Angelini GD, Blackstone EH. Statistical primer: propensity score matching and its alternatives. Eur J Cardiothorac Surg. 2018;53:1112–7. https://doi.org/10.1093/ejcts/ezy167.
33. Haukoos JS, Lewis RJ. The propensity score. JAMA. 2015;314:1637–8. https://doi.org/10.1001/jama.2015.13480.
34. Liu S, Schlesinger JJ, McCoy AB, Reese TJ, Steitz B, Russo E, et al. New onset delirium prediction using machine learning and long short-term memory (LSTM) in electronic health record. J Am Med Inf Assoc. 2022;30:120–31. https://doi.org/10.1093/jamia/ocac210.
35. Harutyunyan H, Khachatrian H, Kale DC, Ver Steeg G, Galstyan A. Multitask learning and benchmarking with clinical time series data. Sci Data. 2019;6:96. https://doi.org/10.1038/s41597-019-0103-9.
36. Han C, Kim HI, Soh S, Choi JW, Song JW, Yoon D. Machine learning with clinical and intraoperative biosignal data for predicting postoperative delirium after cardiac surgery. iScience. 2024;27:109932. https://doi.org/10.1016/j.isci.2024.109932.
37. Deng Y, Liu S, Wang Z, Wang Y, Jiang Y, Liu B. Explainable time-series deep learning models for the prediction of mortality, prolonged length of stay and 30-day readmission in intensive care patients. Front Med (Lausanne). 2022;9:933037. https://doi.org/10.3389/fmed.2022.933037.
38. 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:e181018. https://doi.org/10.1001/jamanetworkopen.2018.1018.
39. Lee M, Park T, Shin J-Y, Park M. A comprehensive multi-task deep learning approach for predicting metabolic syndrome with genetic, nutritional, and clinical data. Sci Rep. 2024;14:17851. https://doi.org/10.1038/s41598-024-68541-1.
40. Kim G, Lim H, Kim Y, Kwon O, Choi JH. Intra-person multi-task learning method for chronic-disease prediction. Sci Rep. 2023;13:1069. https://doi.org/10.1038/s41598-023-28383-9.
41. Nguyen TNQ, García-Rudolph A, Saurí J, Kelleher JD. Multi-task learning for predicting quality-of-life and independence in activities of daily living after stroke: A proof-of-concept study. Front Neurol. 2024;15:1449234. https://doi.org/10.3389/fneur.2024.1449234.
42. Guo A, Beheshti R, Khan YM, Langabeer JR, Foraker RE. Predicting cardiovascular health trajectories in time-series electronic health records with LSTM models. BMC Med Inf Decis Mak. 2021;21:5. https://doi.org/10.1186/s12911-020-01345-1.
43. Hein C, Forgues A, Piau A, Sommet A, Vellas B, Nourhashémi F. Impact of polypharmacy on occurrence of delirium in elderly emergency patients. J Am Med Dir Assoc. 2014;15:e85011–5. https://doi.org/10.1016/j.jamda.2014.08.012.
44. Persico I, Cesari M, Morandi A, Haas J, Mazzola P, Zambon A, et al. Frailty and delirium in older adults: A systematic review and meta-analysis of observational studies. J Gerontol S A. 2018;73:1259–66. https://doi.org/10.1111/jgs.15503.
45. Marcantonio ER. Delirium in hospitalized older adults. N Engl J Med. 2017;377:1456–66. https://doi.org/10.1056/NEJMcp1605501.
46. Wilson JE, Mart MF, Cunningham C, Shehabi Y, Girard TD, MacLullich AMJ, et al. Delirium. Nat Rev Dis Primers. 2020;6:90. https://doi.org/10.1038/s41572-020-00223-4.
47. Gracie TJ, Caufield-Noll C, Wang NY, Sieber FE. The Association of Preoperative Frailty and Postoperative Delirium: A Meta-analysis. Anesth Analg. 2021;133(2):314–23. https://doi.org/10.1213/ANE.0000000000005609.
48. Song Y, Luo Y, Zhang F, et al. Systemic immune-inflammation index predicts postoperative delirium in elderly patients after surgery: a retrospective cohort study. BMC Geriatr. 2022;22:730. https://doi.org/10.1186/s12877-022-03418-.
49. Wang Y, Zhao Y, Wang Z, Zhang H, Wang L, Wu L. Factors influencing delirium after spinal surgery in elderly patients: A systematic review and meta-analysis. Aging Clin Exp Res. 2021;33:331–40. https://doi.org/10.1007/s40520-020-01715-y.
50. Lee A, Mu JL, Joynt GM, Chiu CH, Lai VKW, Gin T, et al. Risk prediction models for delirium in the intensive care unit after cardiac surgery: a systematic review and independent external validation. Br J Anaesth. 2017;118:391–9. https://doi.org/10.1093/bja/aew476.
51. Swarbrick CJ, Partridge JSL. Evidence-based strategies to reduce the incidence of postoperative delirium: a narrative review. Anaesthesia. 2022;77(Suppl 1):92–101.
52. Harari D, Hopper A, Dhesi J, Babic-Illman G, Lockwood L, Martin F. Proactive care of older people undergoing surgery ('POPS'): designing, embedding, evaluating and funding a comprehensive geriatric assessment service for older elective surgical patients. Age Ageing. 2007;36(2):190–6.
53. Partridge JSL, Harari D, Martin FC, Peacock JL, Bell R, Mohammed A, et al. Randomized clinical trial of comprehensive geriatric assessment and optimization in vascular surgery. Br J Surg. 2017;104(6):679–87.
54. Tarazona-Santabalbina FJ, Llabata-Broseta J, Belenguer-Varea A, Alvarez-Martinez D, Cuesta-Peredo D, Avellana-Zaragoza JA. A daily multidisciplinary assessment of older adults undergoing elective colorectal cancer surgery is associated with reduced delirium and geriatric syndromes. J Geriatr Oncol. 2019;10(2):298–303.
55. Liu CY, Gong N, Liu W. The association between preoperative frailty and postoperative delirium: a systematic review and meta-analysis. J Perianesth Nurs. 2022;37(1):53–e621.
56. Fu D, Tan X, Zhang M, Chen L, Yang J. Association between frailty and postoperative delirium: a meta-analysis of cohort study. Aging Clin Exp Res. 2022;34(1):25–37.
57. Hshieh TT, Yue J, Oh E, Puelle M, Dowal S, Travison T, Inouye SK. Effectiveness of multicomponent nonpharmacological delirium interventions: a meta-analysis. JAMA Intern Med. 2015;175(4):512–20. https://doi.org/10.1001/jamainternmed.2014.7779.
58. Jiang X, Chen D, Lou Y, Li Z. Risk factors for postoperative delirium after spine surgery in middle- and old-aged patients. Aging Clin Exp Res. 2017;29(5):1039–44. https://doi.org/10.1007/s40520-016-0640-4.
59. Davani AB, Lee HB, Marcantonio ER, et al. Kidney function modifies the effect of intraoperative opioid dosage on postoperative delirium. J Am Geriatr Soc. 2021;69(2):443–51. https://doi.org/10.1111/jgs.16870.

60. Wang F, Zhu Y, Zhu H, Zhu T. Association between intraoperative end-tidal CO2 and cerebral oxygen saturation during general anesthesia. Asian J Surg. 2025;48(12). https://doi.org/10.1016/j.asjsur.2025.06.171.
61. Peng X, Liu C, Zhu Y, Peng L, Zhang X, Wei W, Zhu T. Hemodynamic Influences of Remimazolam Versus Propofol During the Induction Period of General Anesthesia: A Systematic Review and Meta-analysis of Randomized Controlled Trials. Pain Physician. 2023;26(7):E761–73.
62. Wang F, Hao X, Zhu Y. Effects of perioperative intravenous glucocorticoids on perioperative neurocognitive disorders in adults after surgery: A PRISMA-compliant meta-analysis of randomized controlled trials. Med (Baltim). 2023;102(34):e34708. https://doi.org/10.1097/MD.0000000000034708.

Publisher's note