#### Check for updates

#### **OPEN ACCESS**

EDITED BY Longxiang Su, Peking Union Medical College Hospital (CAMS), China

REVIEWED BY Wei Jun Dan Ong, National University Health System, Singapore Alexander Prokazyuk, Semey State Medical University, Kazakhstan

\*CORRESPONDENCE Ming Zhang, ☑ zhangming197866@163.com

RECEIVED 16 May 2025 ACCEPTED 04 July 2025 PUBLISHED 14 July 2025

#### CITATION

Liu Z, Huang Y, Li L, Xu Y, Wu P, Zhang Z, Han T, Zhang L and Zhang M (2025) Development and clinical application of an automated machine learning-based delirium risk prediction model for emergency polytrauma patients. *Front. Physiol.* 16:1629329. doi: 10.3389/fphys.2025.1629329

#### COPYRIGHT

© 2025 Liu, Huang, Li, Xu, Wu, Zhang, Han, Zhang and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Development and clinical application of an automated machine learning-based delirium risk prediction model for emergency polytrauma patients

Zhenyi Liu<sup>1</sup>, Yihao Huang<sup>2</sup>, Long Li<sup>1</sup>, Yisha Xu<sup>3</sup>, Peng Wu<sup>4</sup>, Zhigang Zhang<sup>5</sup>, Tingyong Han<sup>6</sup>, Liangjie Zhang<sup>7</sup> and Ming Zhang<sup>1</sup>\*

<sup>1</sup>Department of Emergency and Critical Care Medicine, The 945th Hospital of the Joint Logistics Support Force of the Chinese People's Liberation Army, Ya'an, Sichuan, China, <sup>2</sup>Department of Psychosomatic Medicine, The 945th Hospital of the Joint Logistics Support Force of the Chinese People's Liberation Army, Ya'an, Sichuan, China, <sup>3</sup>Emergency Department, Ya'an People's Hospital, Ya'an, Sichuan, China, <sup>4</sup>Emergency Department, Yucheng District People's Hospital of Ya'an, Ya'an, Sichuan, China, <sup>5</sup>Emergency Department, Mingshan District People's Hospital of Ya'an, Ya'an, Sichuan, China, <sup>6</sup>Emergency Department, Affiliated Hospital of Ya'an Polytechnic College, Ya'an, Sichuan, China, <sup>7</sup>Emergency Department, Ya'an Hospital of Traditional Chinese Medicine, Ya'an, Sichuan, China,

**Objective:** To address the limitations of conventional delirium prediction models in emergency polytrauma care, this study developed an interpretable machine learning (ML) framework incorporating trauma-specific biomarkers and advanced optimization algorithms for risk stratification of delirium in emergency polytrauma patients.

Methods: This multi-center retrospective observational cohort study was conducted across six hospitals in the Ya'an region. A total of 956 polytrauma patients admitted between January 2020 and December 2024 were enrolled, complying with the American Association for the Surgery of Trauma (AAST) diagnostic criteria for polytrauma. Demographic, clinical (e.g., Glasgow Coma Scale [GCS], Injury Severity Score [ISS]), and laboratory data (e.g., fibrin degradation products [FDP], lactate) were systematically collected. To address high-dimensional clinical heterogeneity, an Improved Flood Algorithm (IFLA)-enhanced with sine mapping initialization and Cauchy mutation perturbations-was integrated into an automated machine learning (AutoML) framework for simultaneous feature selection and hyperparameter optimization. Model performance was benchmarked against conventional algorithms (logistic regression [LR], support vector machine [SVM], extreme gradient boosting [XGBoost], LightGBM) using five-fold cross-validation. The SHapley Additive exPlanations (SHAP) framework quantified predictor contributions, and a MATLAB-based clinical decision support system (CDSS) was implemented for real-time risk stratification.

**Results:** The improved algorithm significantly outperformed other algorithms on 12 standard test functions. The automated machine learning (AutoML) model achieved ROC-AUC and PR-AUC values of 0.9690 and 0.9611, respectively, on the training set, and 0.8929 and 0.8487, respectively, on

the test set, both significantly higher than those of four other prediction models. The AutoML model identified 5 important features: Glasgow Coma Scale (GCS) score, lactate level, Clinical Frailty Scale (CFS), body mass index (BMI), and fibrin degradation products (FDP). The decision support system demonstrated clinical utility with net benefit across risk thresholds.

**Conclusion:** This study provides a trauma-specific, interpretable ML tool that integrates GCS scoring and dynamic biomarker monitoring, enabling early delirium risk identification in emergency polytrauma. The framework demonstrates feasibility for integration into clinical workflows to improve trauma care quality.

KEYWORDS

delirium, polytrauma, machine learning, predictive model, explainable artificial intelligence

# 1 Introduction

Trauma-related disorders have emerged as a critical global public health challenge. According to World Health Organization statistics, the socioeconomic burden attributable to traumatic injuries has risen to become the second leading contributor to the global disease burden (Galbraith et al., 2023). As a distinct subtype of trauma, polytrauma is characterized by complex pathophysiology, multi-system complications, and prolonged hospitalization, necessitating multidisciplinary collaborative care throughout treatment (Sumann et al., 2020). While damage control resuscitation (DCR) protocols significantly improve hemodynamic stability and survival rates, we observe increased neurological complication rates in this surviving cohort-particularly in patients requiring  $\geq 6$  units of blood transfusion. This complication profile reflects emergent pathophysiological perturbations in severely injured patients who survive initial resuscitation, rather than a direct consequence of DCR strategy (Kruithof et al., 2020). Among these, delirium-a severe neuropsychiatric syndrome with substantial prognostic implications-has been reported to affect 24% of emergency polytrauma patients. This condition not only prolongs mechanical ventilation duration and increases unplanned extubation risks but also induces long-term cognitive impairment, severely compromising patients' quality of life (Von Rueden et al., 2017).

Although the American Guidelines for Critical Care Medicine explicitly recommend incorporating delirium screening into routine ICU care protocols, significant diagnostic gaps persist in clinical practice (Safdar et al., 2024). Studies indicate that healthcare providers actively identify delirium in 15%-20% of cases (Tonna et al., 2021). This disparity between knowledge and implementation may stem from three interrelated challenges: (1) the heterogeneous clinical manifestations driven by delirium's complex pathophysiological mechanisms; (2) the high expertise requirements for administering validated assessment tools like the Confusion Assessment Method for the ICU (CAM-ICU); and (3) the inadequacy of traditional risk factor analysis in addressing dynamically evolving clinical features of polytrauma patients. Current delirium prediction models predominantly focus on geriatric or elective postoperative populations, with scarce systematic investigations into personalized model development for emergency trauma cohorts. This knowledge gap substantially hinders evidence-based implementation of precision preventive strategies (Heinrich et al., 2022).

Emerging evidence highlights the unique value of machine learning (ML) in predicting critical illness outcomes (Li et al., 2024; Fan et al., 2023; Tobin et al., 2024). Gong et al. developed a predictive model achieving an AUC of 0.845 (95% CI: 0.831–0.859), demonstrating the clinical potential of risk stratification in delirium management (Gong et al., 2023). However, significant challenges arise when adapting such models to emergency polytrauma scenarios. Key limitations include: (1) omission of trauma-specific indicators such as Injury Severity Score (ISS); (2) insufficient capacity of linear regression methods to capture complex variable interactions; and (3) unresolved conflicts between rapid decision-making demands and model usability in emergency settings (Rostam Niakan Kalhori, 2022). These shortcomings underscore the urgent need for context-specific predictive tools.

Building upon this rationale, our study innovatively integrates three pivotal components: (1) comprehensive trauma care cycle data collection; (2) adaptive ML algorithms optimized for dynamic clinical environments; and (3) implementation of the SHapley Additive exPlanations (SHAP) framework for transparent interpretation of model decisions. By synergizing advanced information technologies with traditional clinical research paradigms, this multidisciplinary approach aims to provide an intelligent solution for delirium prevention and management in emergency polytrauma patients, ultimately advancing the quality of trauma care delivery.

# 2 Methods

#### 2.1 Study design

This multicenter retrospective observational study was conducted across six hospitals in Ya'an, China. As a retrospective analysis, the requirement for informed consent was waived, and the study protocol received ethical approval from all participating institutions. We enrolled polytrauma patients admitted to these hospitals between January 2020 and December 2024. After applying inclusion and exclusion criteria, 956 patients were



included in the final analysis (see Figure 1 for the patient selection flowchart).

Inclusion criteria: (1) Hospitalization for polytrauma meeting the diagnostic criteria of the American Association for the Surgery of Trauma (LaGrone et al., 2024); (2) ICU stay duration  $\geq$ 48 h.

Exclusion criteria: (1) Age <18 years; (2) Patients who were comatose on admission and unable to verbally communicate with healthcare providers or family members until discharge; (3) Pregnant women; (4) Individuals unable to communicate in Chinese; (5) Subjects with a history of psychiatric disorders; (6) Patients with incomplete medical records upon admission.

## 2.2 Delirium diagnosis and data collection

All patient data were collected and assessed by psychosomatic physicians for delirium through medical record analysis and Confusion Assessment Method for the ICU (CAM-ICU) scoring conducted at 48 h post-ICU admission. This scoring method evaluates the following four criteria: (1) Acute onset or fluctuating mental status; (2) Inattention; (3) Altered level of consciousness; (4) Disorganized thinking. A diagnosis of delirium was confirmed if a patient met the first two criteria and one of the latter two criteria (Chen et al., 2021). All delirium diagnoses were established through retrospective chart review by two boardcertified psychosomatic physicians following a standardized protocol. The primary physician conducted assessments using CAM-ICU criteria applied to medical records, with a secondary physician independently validating diagnoses across all identified cases to ensure inter-rater reliability.

Data collection: Patient data were retrieved via electronic medical record systems from multiple hospitals, consolidated, and uniformly processed by a single researcher. The dataset included: (1) Demographic information: Age, sex, height, weight, body mass index (BMI), Clinical Frailty Scale (CFS) (Shimura et al., 2017; Church et al., 2020), Charlson Comorbidity Index (CCI), smoking history, and alcohol use history; (2) Clinical parameters: Blood pressure, heart rate, body temperature, Glasgow

Coma Scale (GCS) score, Revised Trauma Score (RTS), Injury Severity Score (ISS), and presence of traumatic brain injury (TBI); Traumatic brain injury (TBI) diagnosis was established through admission cranial CT scans interpreted by board-certified radiologists, with severity quantified using the Abbreviated Injury Scale (HEAD-AIS) specifically targeting neuroanatomical damage. Patients were classified as TBI-positive when HEAD-AIS ≥3 (moderate-to-severe injury), consistent with AAST/WSES organ injury grading standards. (3) Laboratory data: Fibrinogen, fibrin degradation products (FDP), hemoglobin, C-reactive protein (CRP), and lactate levels. All clinical parameters (including blood pressure, heart rate, GCS score, RTS, and ISS) were documented during the initial emergency department assessment immediately following patient admission. All laboratory biomarkers (fibrinogen, FDP, hemoglobin, CRP, lactate) were measured from venous blood samples collected at triage prior to any therapeutic interventions. TBI diagnosis was based on admission CT scans.

Missing data handling: The overall data completeness rate for the 956 included polytrauma patients was 97.43%. Missing rates varied across variables, with FDP exhibiting the highest missing rate ( $\leq$ 1% for other variables). Missing values were imputed using median replacement for continuous variables and mode substitution for categorical variables.

# 2.3 Model algorithm optimization and validation

To address the complexity of high-dimensional clinical data, we employed an automated machine learning (AutoML) model based on an optimization algorithm, which simultaneously performed feature selection and hyperparameter tuning. Traditional machine learning models were also included for performance comparison. All analyses were conducted in MATLAB 2024b. The Flood Algorithm (FLA) (Ghasemi et al., 2024), a novel swarm intelligence algorithm inspired by the complex movements of water masses, was used to optimize the AutoML framework. To enhance optimization performance, we improved the original FLA by

10.3389/fphys.2025.1629329

integrating sine mapping initialization and Cauchy mutation perturbation strategies, resulting in the Improved Flood Algorithm (IFLA). The optimization capability of IFLA was validated using 12 standard benchmark functions from the IEEE CEC-2017 test suite (Sharma and Raju, 2024), including multimodal, hybrid, and composite functions such as Schwefel (F15), Rosenbrock (F6), and Lunacek Bi-Rastrigin (F23). Testing parameters: variable dimension = 10, population size = 30, maximum iterations = 500, with 30 independent runs for statistical robustness. Notably, these benchmark functions were used solely to evaluate the optimization performance of the swarm intelligence algorithm and did not participate in AutoML model training. The fitness function was defined as a direct mapping of the objective function value, with the optimization goal set to minimize the fitness value. Thus, a reduction in fitness value signifies improved algorithmic performance.

## 2.4 Model training and evaluation

To assess model quality in terms of performance, computational efficiency, interpretability, and robustness against underfitting/overfitting, we implemented five-fold cross-validation. The dataset was split into an 80% training set (for cross-validation) and a 20% test set. This approach effectively mitigated overfitting during training and improved prediction accuracy on the test set. We compared the performance of widely adopted and robust machine learning models, including logistic regression (LR), support vector machines (SVM), extreme gradient boosting (XGBoost), and LightGBM. These models were selected based on their proven performance and reliability in predictive analytics tasks. Quantitative evaluation metrics: Sensitivity (SEN), precision (PRE), specificity (SPE), accuracy (ACC), error rate (ER), and F1score (F1). Primary comprehensive metrics: Area under the receiver operating characteristic curve (ROC-AUC) and precision-recall curve (PR-AUC). All metrics range from 0 to 1, with higher values indicating superior classification performance.

#### 2.4.1 Interpretability analysis

SHAP (SHapley Additive exPlanations) analysis, rooted in game-theoretic Shapley values, was employed to quantify feature contributions to model predictions. This method provides both global (model-wide) and local (individual sample-level) interpretability. Two types of SHAP visualizations were generated: (1) SHAP Summary Plot: Each point represents a feature's SHAP value for a specific sample, color-mapped to reflect feature magnitude (blue: high values, white: low values), illustrating positive/negative relationships between features and predictions. (2) SHAP Importance Plot: Features are ranked by global importance based on absolute SHAP values, highlighting key predictors.

#### 2.4.2 Clinical decision system development

An interactive clinical decision support system was developed using MATLAB 2024a App Designer. This system integrates the prediction model, enabling clinicians to input clinical parameters via a structured interface and receive real-time predictions with therapeutic recommendations. The tool provides reliable and transparent decision-making assistance for clinical practice.

## 2.5 Statistical analysis

IBM SPSS v25.0 was used for conventional statistical analysis (significance: p < 0.05). Continuous variables were expressed as mean  $\pm$  SD (normally distributed, Kolmogorov-Smirnov test) or median (IQR), and categorical variables as percentages.

# **3** Results

#### 3.1 Baseline characteristics of study cohorts

The study included 956 patients, with 326 cases (34.1%) diagnosed with delirium. The dataset was randomly divided into a training set (80%, n = 764, delirium: 250 cases) and a test set (20%, n = 192, delirium: 76 cases). Baseline characteristics of both cohorts are summarized in Table 1.

# 3.2 Algorithm improvement performance evaluation

Based on 30 independent optimization runs, boxplots were generated to assess algorithm stability (Figure 2). The Improved Flood Algorithm (IFLA) demonstrated superior optimization stability compared to the original FLA and other benchmark algorithms across most test functions. Further convergence curve analysis (Figure 3) revealed that IFLA achieved faster convergence rates while maintaining the lowest risk of entrapment in local optima during iterations. These findings robustly validate IFLA's enhanced global search capability and computational efficiency.

## 3.3 Model training performance

The AutoML model exhibited optimal predictive performance on the training set: ROC-AUC: 0.9690; PR-AUC: 0.9611 (Table 2; Figure 4). Key features selected during model optimization included: Glasgow Coma Scale (GCS) score, lactate level, Clinical Frailty Scale (CFS), body mass index (BMI), and fibrin degradation products (FDP).

## 3.4 Test set validation

The AutoML model maintained strong generalizability on the independent test set: ROC-AUC: 0.8929; PR-AUC: 0.8487 (Table 3; Figure 5).

## 3.5 Interpretability analysis

SHAP analysis quantified feature importance as follows (descending order): 1-GCS score; 2-Lactate level; 3-Clinical Frailty Scale; 4-BMI; 5-FDP (Figure 6).

| Factor                            | Training set (n = 764) |                          |         | Testing set (n = 192) |                          |         |  |
|-----------------------------------|------------------------|--------------------------|---------|-----------------------|--------------------------|---------|--|
|                                   | Delirium (n =<br>250)  | No delirium (n =<br>514) | P value | Delirium (n = 76)     | No delirium (n =<br>116) | P value |  |
| Age (year)                        | 77.37 ± 8.42           | 58.28 ± 9.44             | < 0.001 | 77.81 ± 10.15         | 59.68 ± 8.49             | < 0.001 |  |
| Sex, male (%)                     | 149 (59.6)             | 304 (59.1)               | 0.904   | 47 (61.8)             | 73 (62.9)                | 0.879   |  |
| BMI (kg/m <sup>2</sup> )          | 21.14 ± 4.32           | 22.6 ± 3.88              | <0.001  | 21.18 ± 4.53          | 22.71 ± 4.04             | <0.001  |  |
| CFS (score)                       | 4.58 ± 1.15            | 3.23 ± 0.68              | < 0.001 | $4.61 \pm 1.21$       | $3.28 \pm 0.84$          | < 0.001 |  |
| Charlson index (score)            | 0 (0.2)                | 0 (0.1)                  | < 0.001 | 1 (0.1)               | 0 (0.1)                  | 0.022   |  |
| Smoking (%)                       | 43 (17.2)              | 101 (19.6)               | 0.417   | 14 (18.4)             | 23 (19.8)                | 0.809   |  |
| Alcohol (%)                       | 87 (34.8)              | 185 (35.6)               | 0.747   | 28 (36.8)             | 42 (36.2)                | 0.929   |  |
| Systolic blood pressure<br>(mmHg) | 137.75 ± 23.16         | 140.85 ± 24.16           | 0.092   | 141.13 ± 25.05        | 142.31 ± 24.68           | 0.748   |  |
| GCS score (score)                 | 14 (13–15)             | 15 (14–15)               | <0.001  | 14 (13–15)            | 15 (14–15)               | <0.001  |  |
| GCS score (score)                 | 14.31 ± 2.52           | 15.85 ± 2.43             | <0.001  | 14.38 ± 2.62          | 15.79 ± 2.51             | <0.001  |  |
| Heart rate (bpm)                  | 84 (71,101)            | 82 (71.93)               | 0.168   | 82 (75,102)           | 81 (70.95)               | 0.427   |  |
| Body temperature (°C)             | 36.3 (36, 36.8)        | 36.5 (36.1.36.9)         | 0.747   | 36.4 (36, 36.8)       | 36 (36.2, 37)            | 0.814   |  |
| Hemoglobin (g/L)                  | 121.83 ± 20.67         | 137.56 ± 17.89           | <0.001  | 122.74 ± 19.52        | 138.15 ± 20.24           | <0.001  |  |
| Fibrinogen (mg/dL)                | 259.78 ± 42.17         | 257.88 ± 44.35           | 0.537   | 263.53 ± 51.05        | 259.76 ± 48.35           | 0.606   |  |
| FDP (mg/L)                        | 60.56 ± 18.79          | $16.74 \pm 8.22$         | <0.001  | 62.74 ± 17.15         | 17.18 ± 6.85             | <0.001  |  |
| Lactate (mmol/L)                  | 2.3 (1.5, 4)           | 2.1 (1.6, 2.9)           | <0.001  | 2.5 (1.7, 3.4)        | 1.9 (1.4, 2.7)           | <0.001  |  |
| CRP (mg/L)                        | 9 (3.40)               | 5 (2.14)                 | <0.001  | 8 (2.35)              | 5 (2.13)                 | <0.001  |  |
| RTS (score)                       | 7.83 (7.24–7.88)       | 7.86 (7.84–7.94)         | < 0.001 | 7.84 (7.34–7.86)      | 7.86 (7.84–7.94)         | <0.001  |  |
| ISS (score)                       | 18 (10–27)             | 12 (9–19)                | <0.001  | 18 (10–26)            | 12 (9–19)                | <0.001  |  |
| TBI(%)                            | 117 (46.8)             | 215 (41.8)               | 0.193   | 34 (44.7)             | 46 (39.7)                | 0.485   |  |

TABLE 1 Baseline demographics and clinical characteristics of training and test sets.

# 3.6 Clinical utility

#### 3.6.1 Decision Curve Analysis

The decision curve (Figure 7) demonstrated that applying the AutoML model to predict delirium risk provided greater clinical net benefit compared to alternative strategies across threshold probabilities.

#### 3.6.2 Decision support system

To address barriers in translating AI models to clinical practice (e.g., operational complexity), we developed an intuitive decision support system using MATLAB 2024a. The system allows clinicians to: Input patient features via a structured interface; Obtain real-time delirium risk predictions at the click of "Start Prediction"; Review evidence-based therapeutic recommendations. This tool significantly lowers implementation thresholds while ensuring interpretability and clinical relevance (Figure 8).

# 4 Discussion

Our study employed a multicenter retrospective design to develop an adaptive machine learning model based on an improved flood optimization algorithm (IFLA) for predicting delirium risk in emergency department (ED) patients with multiple trauma. Results demonstrated that the optimized IFLA model significantly outperformed traditional models (e.g., logistic regression and XGBoost) in key metrics including AUC and F1 scores. By integrating sine mapping initialization and Cauchy mutation perturbation strategies, the IFLA algorithm successfully overcame the local optimum trapping inherent in the conventional FLA, a





| Models   | PRE    | SEN    | SPE    | ACC    | F1     | ROC-AUC | PR-AUC |
|----------|--------|--------|--------|--------|--------|---------|--------|
| LR       | 0.5781 | 0.2960 | 0.8949 | 0.6990 | 0.3915 | 0.7129  | 0.5518 |
| SVM      | —      | 0.0000 | 1.0000 | 0.6728 | —      | 0.6177  | 0.4598 |
| XGBoost  | 0.5929 | 0.5360 | 0.8210 | 0.7277 | 0.5630 | 0.7222  | 0.5360 |
| LightGBM | 0.9722 | 0.5600 | 0.9922 | 0.8508 | 0.7107 | 0.9482  | 0.9309 |
| AutoML   | 0.8169 | 0.9280 | 0.8988 | 0.9084 | 0.8689 | 0.9690  | 0.9611 |

TABLE 2 Cross-validation performance metrics on the training set.



TABLE 3 Predictive performance metrics on the testing set.

| Models   | PRE    | SEN    | SPE    | ACC    | F1     | ROC-AUC | PR-AUC |
|----------|--------|--------|--------|--------|--------|---------|--------|
| LR       | 0.5294 | 0.2368 | 0.8621 | 0.6146 | 0.3273 | 0.6484  | 0.5234 |
| SVM      | —      | 0.0000 | 1.0000 | 0.6042 | _      | 0.6130  | 0.4972 |
| XGBoost  | 0.5333 | 0.4211 | 0.7586 | 0.6250 | 0.4706 | 0.6098  | 0.4421 |
| LightGBM | 0.7333 | 0.5789 | 0.8621 | 0.7500 | 0.6471 | 0.8394  | 0.8166 |
| AutoML   | 0.9286 | 0.6421 | 0.9828 | 0.7292 | 0.7351 | 0.8929  | 0.8487 |

finding corroborated by standard benchmark function tests. This innovative approach aligns with the algorithmic enhancements proposed by Gao et al. in COVID-19 prediction models (Gao et al., 2022). Through SHAP interpretability analysis, five critical predictors were identified: Glasgow Coma Scale (GCS) score, fibrin degradation products (FDP), lactate levels, body mass index (BMI), and Clinical Frailty Scale (CFS). Notably, GCS score exhibited the highest SHAP value contribution (26.8%). The real-time decision support system embedded in our model demonstrated favorable clinical acceptance during ED validation, indicating substantial translational potential.

Current delirium prediction research primarily focuses on medical or postoperative cohorts (Tobin et al., 2024; Gong et al., 2023; Rostam Niakan Kalhori, 2022; LaGrone et al., 2024; Chen et al., 2021; Shimura et al., 2017; Church et al., 2020; Ghasemi et al., 2024; Sharma and Raju, 2024; Gao et al., 2022; Liu et al., 2025; Shpakov et al., 2023; Saviano et al., 2023), with limited models specifically designed for trauma







Decision Curve Analysis for Predictive Models. Note: (A) Training set; (B) Testing set. Net benefit (Y-axis) calculated against two extreme scenarios: "treat all" (red dashed) and "treat none" (black dashed).



populations. Conventional linear regression approaches frequently exhibit inadequate predictive performance (AUC typically <0.85 (Matsuoka et al., 2021)) due to their limited capacity for modeling nonlinear relationships. Compared to rapid decision tree models developed in previous studies (Xie et al., 2022), our model demonstrated superior adaptability for polytrauma patients through the inclusion of trauma-specific indicators such as Injury Severity Score (ISS). While ISS significantly discriminated between groups, its utility as a reliable predictor was constrained by the limited critical trauma representation in our cohort, necessitating exclusion during model optimization. Future large-scale studies should validate its reintegration to enhance trauma-specific applicability. While Dana's emergency informatics framework emphasizes data acquisition efficiency (Im et al., 2023), our study achieved concurrent feature selection and parameter optimization using AutoML technology, significantly enhancing computational efficiency. Importantly, previous research has largely overlooked the predictive value of coagulation markers (Stollings et al., 2021), whereas our findings highlight the critical role of FDP dynamics in delirium risk stratification, potentially mediated by neuroinflammatory cascades secondary to microcirculatory dysfunction in polytrauma (Bramley et al., 2021). Although Kang et al.'s sleep quality intervention reduced delirium incidence (Kang et al., 2023), its reliance on subjective clinician assessments contrasts with our objective predictive model that enables early targeted interventions.

Model refinement and SHAP analysis identified five core predictors, with their pathophysiological implications analyzed as follows: (1) GCS score: As a standardized consciousness assessment tool, GCS showed an inverse correlation with delirium risk. Severe brain injury (GCS  $\leq$ 8) may trigger thalamocortical feedback loop dysregulation (attentional deficits), locus coeruleus norepinephrine system hyperactivation (neurotransmitter imbalance), and bloodbrain barrier disruption (neuroinflammation via IL-6/TNF-a infiltration) (Raquer et al., 2024). For ED physicians, dynamic GCS monitoring (particularly in TBI patients) facilitates early identification of high-risk individuals (GCS ≤12), enabling timely preventive measures. (2) Lactate levels: This biomarker of tissue hypoperfusion quantifies oxygen metabolism dysregulation. Levels >2 mmol/L promote delirium via three pathways: 1) astrocytic glutamate uptake inhibition (excitotoxicity); 2) microglial TLR4/NF-κB pathway activation (neuroinflammation); 3) cerebral acidosis impairing neurotransmitter dynamics (Qian et al., 2024). The sharp SHAP value increase at >4 mmol/L suggests a threshold effect. Integrating central venous oxygen saturation monitoring for fluid resuscitation optimization (as shown by Taylor et al. (Taylor et al., 2022)) could reduce delirium incidence by 19%. (3) Clinical Frailty Scale (CFS): Scores ≥5 indicate depleted physiological reserves, amplifying trauma effects through immunosenescence (sustained inflammation), autonomic dysregulation (circadian disruption), and altered pharmacokinetics (sedative accumulation) (Zhang et al., 2021; Mazzola et al., 2021). Our model ranks CFS third in SHAP importance, warranting "precision trauma care" strategies including nutritional support (protein  $\geq 1.2$  g/kg/day), early mobilization (bedside sitting within 24h), and benzodiazepine restriction. (4) BMI: The U-shaped delirium risk (optimal range 18.5-24.9 kg/m<sup>2</sup> (Feinkohl et al., 2023)) reflects dual mechanisms: low BMI exacerbates catabolism (neurotransmitter precursor deficiency), while obesity induces leptin resistance (insulin resistance/BBB disruption). Obese patients require vigilance for occult hypoperfusion from intraabdominal hypertension, whereas underweight patients may benefit from enteral nutrition with branched-chain amino acids (Fu et al., 2024). (5) FDP: Elevated FDP (>20 µg/mL) signals coagulopathy via complement C5a activation (microvascular NETosis) and competitive fibrinogen inhibition (hemorrhagic risk) (Payne et al., 2024). Six-hourly FDP monitoring combined with tranexamic acid administration may mitigate microcirculatory dysfunction-related delirium (Lu et al., 2024).

Despite constructing this automated prediction model, limitations persist in data quality and clinical implementation: Data source bias: Though standardized across six regional hospitals, geographical disparities in trauma protocols and monitoring standardization may introduce bias. While multicenter recruitment enhances external validity, the moderate cohort size limited subgroup analyses for rare trauma phenotypes. We also recognize inherent inter-hospital variability in scoring systems despite standardized training. Future large-scale validation should prioritize algorithmic adaptation to institution-specific documentation patterns using federated learning frameworks. While median imputation mitigates leakage risk, FDP remains susceptible to bias due to its higher missing rate (2.6%). Future studies should employ advanced methods like multiple imputation chained equations (MICE) for variables exceeding 2% missingness. Retrospective acquisition of FDP values, though demonstrating critical prognostic value, necessitates future validation of our dynamic prediction updating protocol in prospective studies employing point-ofcare viscoelastic testing to eliminate turnaround delays. Model constraints: Missing core variables and incomplete capture of nonlinear interactions may reduce sensitivity in complex trauma scenarios. Temporal resolution: Static-input prediction systems face intrinsic latency in dynamic ED environments requiring real-time biomarker feedback (e.g., rapidly changing lactate/FDP). Study design limitations: While retrospective validation provides preliminary evidence, prospective cohorts remain essential for examining delirium's temporal progression and intervention dynamics.

The Intensive Care Big Data Steward Consensus publishes future industry standards in this area (Su et al., 2024), this consensus makes 29 recommendations on the following five parts: Concept of intensive care big data, Important scientific issues, Standards and principles of database, Methodology in solving big data problems, Clinical application and safety consideration of intensive care big data. Aligned with the Intensive Care Big

Data Consensus, our future research framework will embed its 29 evidence-based recommendations across five core dimensions: establishing harmonized multimodal trauma databases adhering to standardized ICU data protocols, implementing federated learning architectures for privacy-preserving multicenter integration, applying advanced AutoML optimization for feature engineering, developing clinical translation pathways within evidence-based safety parameters, and creating real-time SHAP interpretability dashboards for predictive governance. This structured methodology will operationalize the consensus guidelines-particularly regarding scientific question formulation, database standardization, and ethical computational methods-as applied to dynamic delirium prediction in trauma ecosystems. It includes the following aspects: (1) Data integration: Establish multimodal trauma databases incorporating real-time vital signs, continuous EEG, and cytokine profiles to transcend retrospective "time-slice" limitations. Our realtime data pipeline implements sliding-window RNNs for hourly risk-score updates coupled with automatic quarterly calibration audits against AAST/WSES standards, ensuring temporal relevance through federated learning with patient-level partitioning. (2) Algorithm enhancement: Develop spatiotemporal architectures (e.g., temporal convolutional networks for biomarker trends, graph neural networks for multi-organ injury topology) to transition from "point prediction" to "process warning." We implement federated learning and ensemble transition strategies where legacy models are progressively weighted with PAN-GAN-synthesized newer cohorts, enabling continuous adaptation to clinical practice shifts during model development. (3) Clinical translation: Implement edge computing-embedded decision systems integrated with bedside monitors/laboratory streams during the "golden hour" of trauma care. Edge-computing-embedded decision systems integrated with bedside monitors/laboratory streams during the "golden hour" now incorporate SHAP-based performance dashboards triggering alerts for critical predictor drift (e.g., >1.50 change in GCS or FDP contributions). Future iterations should integrate multimodal neurological assessments such as the Full Outline of UnResponsiveness (FOUR) scale to enhance sensitivity in patients with communication barriers (e.g., intubation, aphasia). To advance translational implementation, our research road now explicitly prioritizes EHR interoperability through three parallel initiatives: Development of HL7 FHIR-compliant APIs enabling automated data exchange with hospital information systems at participating centers; Design of clinician-centered mobile interfaces with offline functionality to support bedside risk stratification during resuscitation, featuring real-time SHAP visualizations when FDP trends exceed >1.5 $\sigma$  baseline deviations; Prospective workflow integration trials launching Q4-2026 to quantify adoption metrics and time-motion efficiency gains using the System Usability Scale across three trauma networks. This aligns with our prioritization of spatiotemporal feature engineering and edge-computing integration, potentially improving real-time risk stratification during the "golden hour" of trauma care. Synergizing evidencebased medicine with AI could enable personalized interventions (e.g., circadian modulation for high-CFS patients, anticoagulant optimization for coagulopathic cases), ultimately creating a closedloop "prediction-intervention-verification" ecosystem through SHAP-guided precision pathways.

# Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

# **Ethics statement**

The studies involving humans were approved by As a retrospective analysis, the requirement for informed consent was waived, and the study protocol received ethical approval from all participating institutions. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because As a retrospective analysis, the requirement for informed consent was waived.

# Author contributions

ZL: Conceptualization, Data curation, Investigation, Writing – original draft. YH: Data curation, Formal Analysis, Investigation, Methodology, Writing – review and editing. LL: Formal Analysis, Funding acquisition, Methodology, Software, Supervision, Writing – review and editing. YX: Data curation, Investigation, Writing – review and editing. PW: Data curation, Investigation, Writing – review and editing. ZZ: Data curation, Investigation, Writing – review and editing. TH: Data curation, Investigation, Writing – review and editing. LZ: Data curation, Investigation, Writing – review and editing. LZ: Data curation, Investigation, Writing – review and editing. MZ: Formal Analysis, Methodology, Project administration, Resources, Writing – review and editing.

# References

Bramley, P., McArthur, K., Blayney, A., and McCullagh, I. (2021). Risk factors for postoperative delirium: an umbrella review of systematic reviews. *Int. J. Surg.* 202 (93), 106063. doi:10.1016/j.ijsu.2021.106063

Chen, T. J., Chung, Y. W., Chang, H. R., Chen, P. Y., Wu, C. R., Hsieh, S. H., et al. (2021). Diagnostic accuracy of the CAM-ICU and ICDSC in detecting intensive care unit delirium: a bivariate meta-analysis. *Int. J. Nurs. Stud.* 113, 103782. doi:10.1016/j.ijnurstu.2020.103782

Church, S., Rogers, E., Rockwood, K., and Theou, O. (2020). A scoping review of the clinical frailty scale. *BMC Geriatr.* 20 (1), 393. doi:10.1186/s12877-020-01801-7

Fan, Z., Jiang, J., Xiao, C., Chen, Y., Xia, Q., Wang, J., et al. (2023). Construction and validation of prognostic models in critically ill patients with sepsis-associated acute kidney injury: interpretable machine learning approach. *J. Transl. Med.* 21 (1), 406. doi:10.1186/s12967-023-04205-4

Feinkohl, I., Janke, J., Slooter, A. J. C., Winterer, G., Spies, C., Pischon, T., et al. (2023). Metabolic syndrome and the risk of postoperative delirium and postoperative cognitive dysfunction: a multi-centre cohort study. *Br. J. Anaesth.* 131 (2), 338–347. doi:10.1016/j.bja.2023.04.031

Fu, J., Zhang, X., Zhang, G., Wei, C., Fu, Q., Gui, X., et al. (2024). Association between body mass index and delirium incidence in critically ill patients: a retrospective cohort study based on the MIMIC-IV database. *BMJ Open* 14 (3), e079140. doi:10.1136/bmjopen-2023-079140

Galbraith, C. M., Wagener, B. M., Chalkias, A., Siddiqui, S., and Douin, D. J. (2023). Massive trauma and resuscitation strategies. *Anesthesiol. Clin.* 41 (1), 283–301. doi:10.1016/j.anclin.2022.10.008

Gao, C., Zhang, R., Chen, X., Yao, T., Song, Q., Ye, W., et al. (2022). Integrating internet multisource big data to predict the occurrence and development of COVID-19 cryptic transmission. *NPJ Digit. Med.* 5 (1), 161. doi:10.1038/s41746-022-00704-8

# Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the Hospital Management Fund of the 945 Hospital of the Joint Logistics Support Force of the Chinese People's Liberation Army (No. 2023945YG-09).

# **Conflict of interest**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

# Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

# Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Ghasemi, M., Golalipour, K., Zare, M., Mirjalili, S., Trojovský, P., Abualigah, L., et al. (2024). Flood algorithm (FLA): an efficient inspired meta-heuristic for engineering optimization. *J. Supercomput* 80, 22913–23017. doi:10.1007/s11227-024-06291-7

Gong, K. D., Lu, R., Bergamaschi, T. S., Sanyal, A., Guo, J., Kim, H. B., et al. (2023). Predicting intensive care delirium with machine learning: model development and external validation. *Anesthesiology* 138 (3), 299-311. doi:10.1097/ALN.00000000004478

Heinrich, M., Woike, J. K., Spies, C. D., and Wegwarth, O. (2022). Forecasting postoperative delirium in older adult patients with fast-and-frugal decision trees. *J. Clin. Med.* 11 (19), 5629. doi:10.3390/jcm11195629

Im, D. D., Scott, K. W., Venkatesh, A. K., Lobon, L. F., Kroll, D. S., Samuels, E. A., et al. (2023). A quality measurement framework for emergency department care of psychiatric emergencies. *Ann. Emerg. Med.* 81 (5), 592–605. doi:10.1016/j.annemergmed.2022.09.007

Kang, J., Cho, Y. S., Lee, M., Yun, S., Jeong, Y. J., Won, Y. H., et al. (2023). Effects of nonpharmacological interventions on sleep improvement and delirium prevention in critically ill patients: a systematic review and meta-analysis. *Aust. Crit. Care* 36 (4), 640–649. doi:10.1016/j.aucc.2022.04.006

Kruithof, N., Polinder, S., de Munter, L., van de Ree, C. L. P., Lansink, K. W. W., de Jongh, M. A. C., et al. (2020). Health status and psychological outcomes after trauma: a prospective multicenter cohort study. *PLoS One* 15 (4), e0231649. doi:10.1371/journal.pone.0231649

LaGrone, L. N., Stein, D., Cribari, C., Kaups, K., Harris, C., Miller, A. N., et al. (2024). American association for the surgery of trauma/american college of surgeons committee on trauma: clinical protocol for damage-control resuscitation for the adult trauma patient. *J. Trauma Acute Care Surg.* 96 (3), 510–520. doi:10.1097/TA.000000000004088 Li, L., Han, X., Zhang, Z., Han, T., Wu, P., Xu, Y., et al. (2024). Construction of prognosis prediction model and visualization system of acute paraquat poisoning based on improved machine learning model. *Digit. Health* 10, 20552076241287891. doi:10.1177/20552076241287891

Liu, X., Huangfu, Z., Zhang, X., and Ma, T. (2025). Global research trends in postoperative delirium and its risk factors: a bibliometric and visual analysis. *J. Perianesth Nurs.* 40 (2), 400–414. doi:10.1016/j.jopan.2024.04.002

Lu, Z., Wang, B., Liu, M., Yu, D., and Li, J. (2024). Correlation analysis between plasma biomarkers albumin, fibrinogen, and their ratio with postoperative delirium in patients undergoing non-cardiac surgery: a systematic review and meta-analysis. *Am. J. Transl. Res.* 16 (2), 363–373. doi:10.62347/AEHR2759

Matsuoka, A., Miike, T., Miyazaki, M., Goto, T., Sasaki, A., Yamazaki, H., et al. (2021). Development of a delirium predictive model for adult trauma patients in an emergency and critical care center: a retrospective study. *Trauma Surg. Acute Care Open* 6 (1), e000827. doi:10.1136/tsaco-2021-000827

Mazzola, P., Tassistro, E., Di Santo, S., Rossi, E., Andreano, A., Valsecchi, M. G., et al. (2021). The relationship between frailty and delirium: insights from the 2017 delirium day study. *Age Ageing* 50 (5), 1593–1599. doi:10.1093/ageing/afab042

Payne, T., Taylor, J., Kunkel, D., Konieczka, K., Ingram, F., Blennow, K., et al. (2024). Association of preoperative to postoperative change in cerebrospinal fluid fibrinogen with postoperative delirium. *BJA Open* 12, 100349. doi:10.1016/j.bjao.2024.100349

Qian, X., Sheng, Y., Jiang, Y., and Xu, Y. (2024). Associations of serum lactate and lactate clearance with delirium in the early stage of ICU: a retrospective cohort study of the MIMIC-IV database. *Front. Neurol.* 15, 1371827. doi:10.3389/fneur.2024.1371827

Raquer, A. P., Fong, C. T., Walters, A. M., Souter, M. J., and Lele, A. V. (2024). Delirium and its associations with critical care utilizations and outcomes at the time of hospital discharge in patients with acute brain injury. *Med. Kaunas.* 60 (2), 304. doi:10.3390/medicina60020304

Rostam Niakan Kalhori, S. (2022). Towards the application of machine learning in emergency informatics. *Stud. Health Technol. Inf.* 291, 3–16. doi:10.3233/SHTI220003

Safdar, M., Colosimo, C., Khurshid, M. H., Spencer, A. L., Hejazi, O., Castanon, L., et al. (2024). Drugs, delirium, and trauma: substance use and incidence of delirium after traumatic brain injury. *J. Surg. Res.* 301, 45–53. doi:10.1016/j.jss.2024.05.042

Saviano, A., Zanza, C., Longhitano, Y., Ojetti, V., Franceschi, F., Bellou, A., et al. (2023). Current trends for delirium screening within the emergency department. *Med. Kaunas*. 59 (9), 1634. doi:10.3390/medicina59091634

Sharma, P., and Raju, S. (2024). Metaheuristic optimization algorithms: a comprehensive overview and classification of benchmark test functions. *Soft Comput.* 28 (4), 3123–3186. doi:10.1007/s00500-023-09276-5

Shimura, T., Yamamoto, M., Kano, S., Kagase, A., Kodama, A., Koyama, Y., et al. (2017). Impact of frailty markers on outcomes after transcatheter aortic valve replacement: insights from a Japanese multicenter registry. *Ann. Cardiothorac. Surg.* 6 (5), 532–537. doi:10.21037/acs.2017.09.06

Shpakov, A. O., Zorina, I. I., and Derkach, K. V. (2023). Hot spots for the use of intranasal insulin: cerebral ischemia, brain injury, diabetes mellitus, endocrine disorders and postoperative delirium. *Int. J. Mol. Sci.* 24 (4), 3278. doi:10.3390/ijms24043278

Stollings, J. L., Kotfis, K., Chanques, G., Pun, B. T., Pandharipande, P. P., and Ely, E. W. (2021). Delirium in critical illness: clinical manifestations, outcomes, and management. *Intensive Care Med.* 47 (10), 1089–1103. doi:10.1007/s00134-021-06503-1

Su, L., Liu, S., Long, Y., Chen, C., Chen, K., Chen, M., et al. (2024). Chinese experts' consensus on the application of intensive care big data. *Front. Med. (Lausanne)* 10, 1174429. doi:10.3389/fmed.2023.1174429

Sumann, G., Moens, D., Brink, B., Brodmann Maeder, M., Greene, M., Jacob, M., et al. (2020). Multiple trauma management in Mountain environments - a scoping review: evidence based guidelines of the international commission for Mountain emergency medicine (ICAR medCom). Intended for physicians and other advanced life support personnel. *Scand. J. Trauma Resusc. Emerg. Med.* 28 (1), 117. doi:10.1186/s13049-020-00790-1

Taylor, J., Parker, M., Casey, C. P., Tanabe, S., Kunkel, D., Rivera, C., et al. (2022). Postoperative delirium and changes in the blood-brain barrier, neuroinflammation, and cerebrospinal fluid lactate: a prospective cohort study. *Br. J. Anaesth.* 129 (2), 219–230. doi:10.1016/j.bja.2022.01.005

Tobin, J. M., Lusczek, E., and Bakker, J. (2024). Artificial intelligence and machine learning in critical care research. J. Crit. Care 82, 154791. doi:10.1016/j.jcrc.2024.154791

Tonna, J. E., Dalton, A., Presson, A. P., Zhang, C., Colantuoni, E., Lander, K., et al. (2021). The effect of a quality improvement intervention on sleep and delirium in critically ill patients in a surgical ICU. *Chest* 160 (3), 899–908. doi:10.1016/j.chest.2021.03.030

Von Rueden, K. T., Wallizer, B., Thurman, P., McQuillan, K., Andrews, T., Merenda, J., et al. (2017). Delirium in trauma patients: prevalence and predictors. *Crit. Care Nurse* 37 (1), 40–48. doi:10.4037/ccn2017373

Xie, Q., Wang, X., Pei, J., Wu, Y., Guo, Q., Su, Y., et al. (2022). Machine learning-based prediction models for delirium: a systematic review and meta-analysis. J. Am. Med. Dir. Assoc. 23 (10), 1655–1668.e6. doi:10.1016/j.jamda.2022.06.020

Zhang, X. M., Jiao, J., Xie, X. H., and Wu, X. J. (2021). The association between frailty and delirium among hospitalized patients: an updated meta-analysis. *J. Am. Med. Dir. Assoc.* 22 (3), 527–534. doi:10.1016/j.jamda.2021.01.065