

ORIGINAL RESEARCH

Machine Learning-Based Prediction Model for Infectious Complications in Trauma and Its Association With In-Hospital Mortality

Youngmin Kim¹ | Jiyeon Oh^{2,3} | Hyunjee Kim^{2,4}  | Dongjin Yeo^{2,3} | Hyeon Jin Kim^{2,4} | Soeun Kim^{2,4} | Hayeon Lee^{2,5,6}  | Seung Ha Hwang^{2,5,6} | Jaehyeong Cho² | Kisoo Kim⁶ | Jinseok Lee^{5,6}  | Eunche Hyeon⁶ | Byungchul Yu^{1,7} | Do Wan Kim⁸ | Junepill Seok^{9,10} | Kyoungwan Kim¹¹ | Dong Keon Yon^{2,3,4,12}  | Wu Seong Kang^{1,7}

¹Department of Trauma Surgery, Gachon University Gil Medical Center, Incheon, South Korea | ²Center for Digital Health, Medical Science Research Institute, Kyung Hee University College of Medicine, Seoul, South Korea | ³Department of Medicine, Kyung Hee University College of Medicine, Seoul, South Korea | ⁴Department of Precision Medicine, Kyung Hee University College of Medicine, Seoul, South Korea | ⁵Department of Electronics and Information Convergence Engineering, Kyung Hee University, Yongin, South Korea | ⁶Department of Biomedical Engineering, Kyung Hee University, Yongin, South Korea | ⁷Traumatology, Gachon University College of Medicine, Incheon, South Korea | ⁸Department of Thoracic and Cardiovascular Surgery, Chonnam National University Hospital and Chonnam National University Medical School, Gwangju, South Korea | ⁹Department of Thoracic and Cardiovascular Surgery, Chungbuk National University Hospital, Cheongju, South Korea | ¹⁰Department of Thoracic and Cardiovascular Surgery, Chungbuk National University College of Medicine, Cheongju, South Korea | ¹¹Department of Trauma Surgery, Jeju Regional Trauma Center, Cheju Halla General Hospital, Jeju, South Korea | ¹²Department of Pediatrics, Kyung Hee University Medical Center, Kyung Hee University College of Medicine, Seoul, South Korea

Correspondence: Dong Keon Yon (yonkkang@gmail.com) | Wu Seong Kang (wuseongkang@naver.com)

Received: 25 October 2025 | **Revised:** 1 April 2026 | **Accepted:** 20 April 2026

Keywords: artificial intelligence | infectious complications | machine learning | trauma

ABSTRACT

Background: Infectious complications, such as sepsis or catheter-related infections, are common and serious sequelae after trauma. Despite their clinical significance, existing risk-prediction models are limited by reliance on in-hospital data that fail to capture complex physiological interactions. Thus, this study aimed to develop and validate an interpretable ensemble machine learning (ML) model integrating both prehospital and in-hospital clinical data to predict infectious complications after trauma. **Methods:** We used data from the Korean Trauma Data Bank, comprising patients admitted to all 19 trauma centers from 2017 to 2022 in South Korea (discovery; $n = 227,567$) and from four additional centers added in 2023 for external validation ($n = 8867$). Trauma cases were defined utilizing S or T diagnostic codes based on the 7th Korean Standard Classification of Diseases, and infectious complications were defined as a composite outcome of pneumonia, urinary tract infection, catheter-related bloodstream infection, surgical site infection (deep, organ, and superficial), osteomyelitis, or severe sepsis. A total of 33 prehospital and in-hospital features were used in ML model training, and the top-performing models were ensembled to construct the final model. Model performance was evaluated through five-fold cross-validation, internal testing, and external validation. Shapley Additive Explanations (SHAP) were applied to assess predictor importance, and predicted risks were categorized into tertiles (T1–T3) to examine associations with in-hospital mortality and presented adjusted odds ratios (aORs) with 95% confidence intervals (CIs).

Dong Keon Yon and Wu Seong Kang jointly supervised this work.

Youngmin Kim, Jiyeon Oh, and Hyunjee Kim contributed equally as the first authors.

Results: Among 88,899 eligible patients with trauma in the discovery cohort, the soft-voting ensemble model integrating logistic regression, categorical boosting, and extreme gradient boosting achieved the best discrimination, with an area under the receiver operating characteristic curve of 0.796 in the discovery cohort and 0.717 in the external validation cohort. SHAP analysis identified age, accident type, Glasgow Coma Scale verbal response, and sex as the most influential variables. Higher tertiles of predicted infection risk were strongly associated with mortality, with aORs of 2.52 (95% CI, 2.12–2.99) for T1, 4.65 (3.96–5.47) for T2, and 6.19 (5.02–7.62) for T3.

Conclusion: This interpretable model, which integrates prehospital and in-hospital data available within the first 24 h of admission, presented robust predictive performance for post-traumatic infectious complications. The proportional association between predicted infection risk and mortality highlights its clinical relevance, as even modest increases in predicted risk may carry meaningful implications for patient outcomes and early intervention strategies.

1 | Introduction

Infectious complications, including sepsis or catheter-related bloodstream infection, represent a major cause of secondary morbidity and mortality among patients with trauma [1]. These infections frequently occur during the post-acute phase of trauma care, often prolonging hospitalization, increasing the need for intensive care, and substantially elevating healthcare costs [2, 3]. Early identification of patients at high risk for infectious deterioration is therefore essential for optimizing antibiotic stewardship, guiding surveillance, and prioritizing clinical sources in trauma settings [1, 4].

Several studies have recently explored the use of machine learning (ML) to predict sepsis or infection-related outcomes among patients with trauma [5–7]. However, most have been limited in scope—focusing on hospitalized or intensive care unit cohorts, or on specific subgroups such as patients with traumatic brain injury [5–7]. Because these studies depended solely on in-hospital measurements, they could not capture early physiologic signals available in the prehospital phase, when preventive action might be most effective. In addition, previous models often relied on narrow outcomes, lacked external validation, and provided limited interpretability. Although one intensive care unit (ICU)-based ML study used Shapley Additive Explanations (SHAP) analysis to identify key predictors such as FiO_2 and body mass index, its findings were not generalizable to broader trauma populations. These limitations highlight the need for a more comprehensive and clinically interpretable model capable of integrating early, real-world data from the entire trauma continuum of care.

To address these gaps, we developed and validated an explainable ML model to predict infectious complications among trauma patients using both prehospital and in-hospital clinical data from the nationwide Korean Trauma Data Bank (KTDB). We compared models trained exclusively on prehospital variables with those incorporating data obtained after hospital arrival to assess the added value of early physiological information. Furthermore, we examined the association between model-predicted infection risk and in-hospital mortality to determine its prognostic significance. Finally, we applied SHAP to identify the most influential predictors, providing transparent and clinically meaningful insights into the model's decision process.

2 | Methods

2.1 | Data Source and Study Population

This retrospective cohort study was performed in accordance with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis + Artificial Intelligence checklist [8]. We utilized data from the KTDB, a national trauma registry developed to improve the quality of trauma care through the integration of registries from 19 major trauma centers across South Korea [9–11]. In South Korea, patients with severe trauma are generally triaged in the pre-hospital setting and transported to designated trauma centers by emergency medical services. The study population included 227,567 patients with trauma recorded between January 1, 2017, and December 31, 2022 (dataset identification number K2024-03-2-02-29). Trauma cases were identified using S and T codes from the Korean Standard Classification of Diseases, 7th version (KCD-7), which is based on the 10th revision of the International Classification of Diseases [12]. Infectious complications were defined as the occurrence of any of the following during the index hospitalization following trauma admission: pneumonia, urinary tract infection, catheter-related bloodstream infection, surgical site infection (deep, organ, and superficial), osteomyelitis, or severe sepsis. After applying the exclusion criteria, the remaining dataset was used as the discovery cohort for model development, training, and internal validation.

For external validation, data were obtained from four major trauma centers in South Korea: Cheju Halla General Hospital, Gachon University Gil Medical Center, Chonnam National University Hospital, and Chungbuk National University Hospital. To provide temporal external validation, the validation cohort was constructed using data collected during a different time period than the discovery cohort. The external validation cohort comprised 8867 patients with traumatic injuries enrolled between January 1 and December 31, 2023. Trauma cases were identified using S and T codes from the KCD-7, and the same exclusion criteria and definition of infectious complications were consistently applied to the validation cohort. The research protocol was approved by the Institutional Review Boards of the Gachon University Gil Medical Center (GFIRB2024-357), Chonnam National University Hospital (CNUH-2024-016), Chungbuk National University Hospital (CBNUH 2024-10-017), and Cheju Halla General Hospital (CHH-2024-L03). Written informed consent was not required in accordance with ethical guidelines for

the secondary use of administrative data. The study was conducted in compliance with the Declaration of Helsinki.

2.2 | Data Preprocessing

We excluded individuals according to the following criteria: (1) age younger than 15 years, and (2) missing values in key variables required for model development (e.g., on-site vital sign measurements). Imputation was not performed due to the high proportion of missing data. The dataset was randomly divided into training and testing sets in an 8:2 ratio. The testing set was exclusively used to evaluate model performance and was not involved in model training or internal validation.

To develop a predictive model for infectious complications among trauma patients before hospital arrival, 33 metadata features encompassing demographic, accident-related, pre-hospital, and in-hospital variables were utilized. These included vital signs, consciousness status, emergency treatments, and Glasgow Coma Scale components. Detailed definitions and descriptions of all variables are provided in Supporting Information S1: Tables S1 and S2.

To address class imbalance and ensure feature comparability, the synthetic minority over-sampling technique and edited nearest neighbors algorithm was applied for resampling and min-max normalization was performed on all variables, with detailed procedures described in the Supporting Information S1: Supplementary Methods.

2.3 | Model Development and Training

We utilized six ML models to predict the likelihood of infectious complications: adaptive boosting (AdaBoost), extreme gradient boosting (XGBoost), LightGBM, categorical boosting (CatBoost), random forest (RF), and logistic regression (LR) [13]. Each model was optimized using a brute-force grid search to identify the optimal hyperparameters for the prediction task [14]. Fine-tuning these parameters improved both the predictive accuracy and stability of the individual models, thereby enhancing the overall performance of the ensemble. An ensemble approach was subsequently applied by evaluating all possible combinations of the top three among the six models. To enhance predictive performance and generalizability, we implemented a voting-based ensemble of XGBoost, CatBoost, and LR [15]. The finalized hyperparameters for each model are summarized in Supporting Information S1: Table S3, and additional methodological details are provided in the Supplementary Methods.

2.4 | Model Evaluation and External Validation

Model performance was evaluated using an independent test dataset that was not involved in model training or cross-validation to ensure an unbiased assessment of generalization. Evaluation metrics included sensitivity, specificity, accuracy, balanced accuracy, and the area under the receiver operating characteristic curve (AUROC), with balanced accuracy and AUROC designated as the primary metrics owing to class imbalance. Optimal likelihood thresholds, derived from AUROC analyses, were applied consistently across all models and evaluation procedures.

The generalizability of the model was further evaluated using external datasets derived from four independent regional trauma centers. Given the substantial class imbalance in the external validation cohort, balanced accuracy and AUROC were applied as the primary evaluation metrics, consistent with the internal validation [16]. Based on performance across all evaluation stages, the model representing consistently high balanced accuracy and AUROC was selected as the final proposed ML model. Calibration and explainability analyses were performed using Platt scaling, Brier scores, decision curve analysis, and SHAP values to assess model reliability and interpretability. All models were implemented using Python (version 3.11.11) with the following libraries: TensorFlow (2.19.0), Keras (3.10.0), NumPy (2.1.0), Pandas (2.2.3), Matplotlib (3.10.1), and Scikit-learn (1.6.1). Detailed descriptions of these procedures are provided in the Supporting Information S1: Supplementary Methods.

2.5 | Risk Stratification and In-Hospital Mortality Analysis

In the discovery cohort, patients with predicted probabilities above the optimized ensemble threshold (0.47) were classified as the exposed group, and those below as the reference group. Logistic regression analysis was performed to estimate adjusted odds ratios (aORs) and 95% confidence intervals (CIs) for the association between predicted infection risk and in-hospital mortality, adjusting for all covariates [13]. Detailed analytical procedures are described in the Supporting Information S1: Supplementary Methods.

3 | Results

3.1 | Study Population

From 2017 to 2022, in the discovery cohort, among 227,567 patients who were admitted to major trauma centers in South Korea, 88,899 were eligible for analysis after excluding 10,415 patients aged < 15 years and 128,253 patients with missing information (Figure 1). In the validation cohort, 4042 were eligible after excluding those with missing vital sign measurements or not meeting age criteria. Baseline characteristics for both cohorts included in the model development are presented in Supporting Information S1: Table S2.

3.2 | Five-Fold Cross-Validation and Internal Validation

The ML model performance was assessed through five-fold cross-validation using multiple evaluation indices, including sensitivity, specificity, balanced accuracy, and AUROC (Table 1). Across the six individual ML algorithms, XGBoost, CatBoost, and LR showed superior performance as individual models, achieving the highest AUROC values.

Among all ensemble configurations, the model combining LR, CatBoost, and XGBoost achieved the most favorable and well-balanced performance. It recorded a sensitivity of 0.497 (95% CI, 0.466–0.528), a specificity of 0.865 (0.858–0.872), an accuracy of 0.852 (0.844–0.860), a balanced accuracy of 0.681 (0.663–0.699), and an AUROC of 0.777 (0.756–0.797). By integrating both linear and nonlinear learners, the ensemble

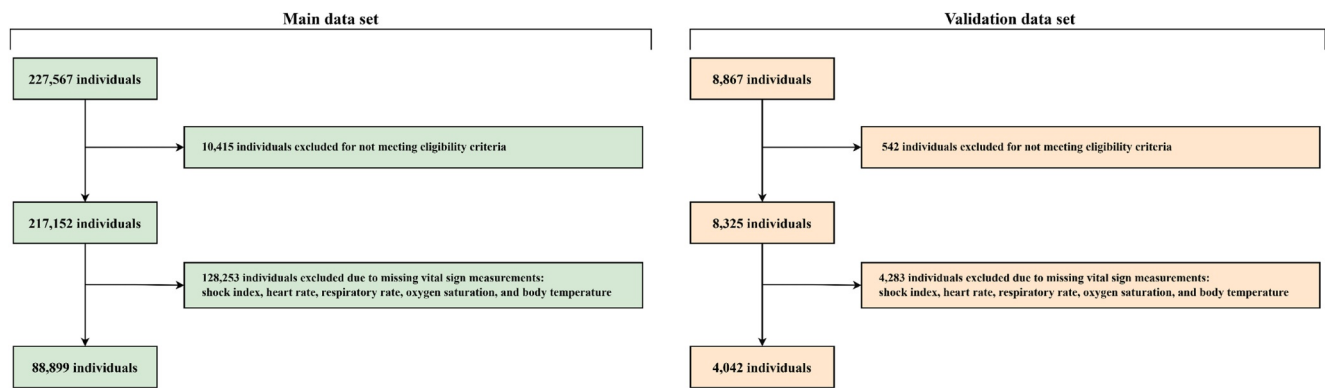


FIGURE 1 | Flow diagram of study exclusion criteria in the discovery and validation data set.

effectively captured diverse patterns in clinical data, improving overall discrimination. Considering the inherent class imbalance, balanced accuracy and AUROC were prioritized as the key evaluation metrics. The LR-CatBoost-XGBoost ensemble consistently showed the highest sensitivity and AUROC across all cross-validation folds and was selected as the final model for predicting post-traumatic infectious complications. Its overall structure and workflow are depicted in Figure 2.

During internal validation using an internal test subset from the discovery cohort, the LR-CatBoost-XGBoost ensemble maintained high and stable predictive performance, achieving a sensitivity of 0.516, specificity of 0.868, accuracy of 0.856, balanced accuracy of 0.692, and AUROC of 0.796 (Table 1). The consistent results across cross-validation and internal testing suggested strong generalizability, with no evidence of overfitting.

3.3 | External Validation

External validation was conducted using data obtained from four major trauma centers to evaluate the generalizability of the proposed model. The ensemble model combining LR, CatBoost, and XGBoost showed acceptable discriminatory ability across the external datasets, with a sensitivity of 0.289, specificity of 0.900, accuracy of 0.869, balanced accuracy of 0.595, and an AUROC of 0.717 (Table 1). Although the performance declined compared with the internal validation results, the model preserved adequate classification capability, particularly in maintaining high specificity.

3.4 | Model Interpretation

To improve clinical interpretation of the model and identify the most influential predictors of infectious complications, we performed a SHAP analysis (Figure 3). Age was the strongest contributor to the model's output, followed by the accident type, Glasgow Coma Scale (GCS) eye response score, and sex. Additional influential features included in-hospital oxygen administration and in-hospital conscious status. Overall, the model attributed higher infection risk to advanced age, impaired neurological status (lower GCS verbal, eye, and motor response), and reduced consciousness during prehospital or in-hospital assessment. In contrast, interventions such as oxygen administration and massive transfusion were associated with higher predicted risk, likely reflecting underlying severity.

Beyond discrimination metrics, we also evaluated model calibration using mean calibration, calibration slope, and plots. We assessed clinical applicability through decision curve analysis. As shown in Supporting Information S1: Figure S1, the model showed notably enhanced reliability and practical value, with predicted risks more closely following the ideal calibration line and maintaining higher net benefit across a wide range of decision thresholds. The mean calibration was 0.000, indicating negligible systematic bias, while the calibration slope of 1.004 reflected strong agreement between predicted and observed risks, with only a slight tendency toward overestimation.

3.5 | Association Between Predicted Risk of Infectious Complications and In-Hospital Mortality

Table 2 presents in-hospital mortality outcomes stratified by tertiles (T1–T3) of predicted complication probability in the discovery, cross-validation, and internal test datasets. Across all datasets, a higher predicted risk of infectious complications was strongly associated with increased odds of death. In the main dataset, aORs for in-hospital mortality were 2.52 (95% CI, 2.12–2.99) for T1, 4.65 (3.96–5.47) for T2, and 6.19 (5.02–7.62) for T3 (all $p < 0.001$) compared with patients below the threshold. Similar gradients were observed in the cross-validation dataset and the test dataset.

4 | Discussion

4.1 | Key Findings

This study developed and validated an interpretable ML model to predict the likelihood of infectious complications following trauma, using data from the KTDB, a nationwide registry encompassing patients from all major trauma centers in South Korea. Among several algorithms tested, a soft-voting ensemble combining LR, CatBoost, and XGBoost achieved the highest overall performance, with an AUROC of 0.777 in internal cross-validation and 0.717 in external validation. Model interpretability analysis using SHAP values identified that age, accident type, and GCS eye response score were the most influential predictors. This suggests that both baseline neurologic function and early physiological stability are pivotal determinants of infection risk in patients with trauma, providing transparent and clinically interpretable insight into the ensemble model's decision process. Furthermore, an increasing tertile of predicted infection probability was associated with a stepwise rise in

TABLE 1 | Comparison of the performance of the prediction models on the training and test datasets.

	Sensitivity	Specificity	Accuracy	Balanced accuracy	AUROC
Cross validation, mean (95% CI)					
AdaBoost	0.535 (0.478–0.593)	0.831 (0.822–0.841)	0.821 (0.814–0.828)	0.683 (0.658–0.708)	0.764 (0.743–0.786)
RF	0.474 (0.429–0.519)	0.864 (0.850–0.878)	0.851 (0.837–0.864)	0.669 (0.646–0.692)	0.764 (0.740–0.788)
LightGBM	0.415 (0.385–0.445)	0.890 (0.884–0.896)	0.874 (0.867–0.880)	0.652 (0.636–0.669)	0.762 (0.742–0.782)
XGBoost	0.378 (0.340–0.415)	0.906 (0.898–0.914)	0.888 (0.880–0.896)	0.642 (0.623–0.661)	0.764 (0.742–0.786)
Catboost	0.394 (0.365–0.424)	0.907 (0.900–0.915)	0.890 (0.882–0.897)	0.651 (0.634–0.668)	0.772 (0.752–0.792)
LR	0.736 (0.700–0.771)	0.682 (0.674–0.689)	0.683 (0.675–0.691)	0.709 (0.688–0.729)	0.772 (0.751–0.793)
XGBoost & Catboost ^a	0.370 (0.336–0.403)	0.913 (0.906–0.920)	0.894 (0.886–0.901)	0.641 (0.622–0.660)	0.772 (0.751–0.794)
Catboost & LR ^a	0.538 (0.507–0.568)	0.843 (0.835–0.851)	0.832 (0.824–0.841)	0.690 (0.672–0.709)	0.776 (0.755–0.796)
LR & XGBoost ^a	0.553 (0.513–0.594)	0.832 (0.827–0.837)	0.823 (0.817–0.829)	0.693 (0.670–0.716)	0.775 (0.755–0.796)
XGBoost & Catboost & LR ^a	0.497 (0.466–0.528)	0.865 (0.858–0.872)	0.852 (0.844–0.860)	0.681 (0.663–0.699)	0.777 (0.756–0.797)
Test data set					
AdaBoost	0.566	0.838	0.829	0.702	0.777
RF	0.523	0.861	0.849	0.692	0.784
LightGBM	0.426	0.897	0.880	0.661	0.784
XGBoost	0.400	0.911	0.893	0.655	0.786
Catboost	0.406	0.910	0.892	0.658	0.791
LR	0.741	0.688	0.690	0.715	0.789
XGBoost & Catboost ^a	0.400	0.911	0.893	0.655	0.790
Catboost & LR ^a	0.579	0.843	0.834	0.711	0.795
LR & XGBoost ^a	0.602	0.833	0.825	0.717	0.794
XGBoost & Catboost & LR ^a	0.516	0.868	0.856	0.692	0.796
Extra validation data set					
AdaBoost	0.397	0.871	0.847	0.634	0.728
RF	0.348	0.890	0.862	0.619	0.724
LightGBM	0.309	0.906	0.875	0.607	0.720
XGBoost	0.270	0.919	0.886	0.594	0.719
Catboost	0.250	0.926	0.892	0.588	0.716
LR	0.456	0.802	0.785	0.629	0.704
XGBoost & Catboost ^a	0.216	0.928	0.892	0.572	0.703
Catboost & LR ^a	0.324	0.894	0.863	0.584	0.696
LR & XGBoost ^a	0.299	0.881	0.852	0.590	0.694
XGBoost & Catboost & LR ^a	0.289	0.900	0.869	0.595	0.717

Note: Data in bold indicates the best performance.

Abbreviations: AdaBoost, adaptive boosting; AUROC, area under the receiver operating characteristic curve; CatBoost, categorical boosting; CI, confidence interval; Light GBM, light gradient boosting machine; LR, logistic regression; RF, random forest; XGBoost, extreme gradient boosting.

^aAll ensemble models applied a soft-voting method.

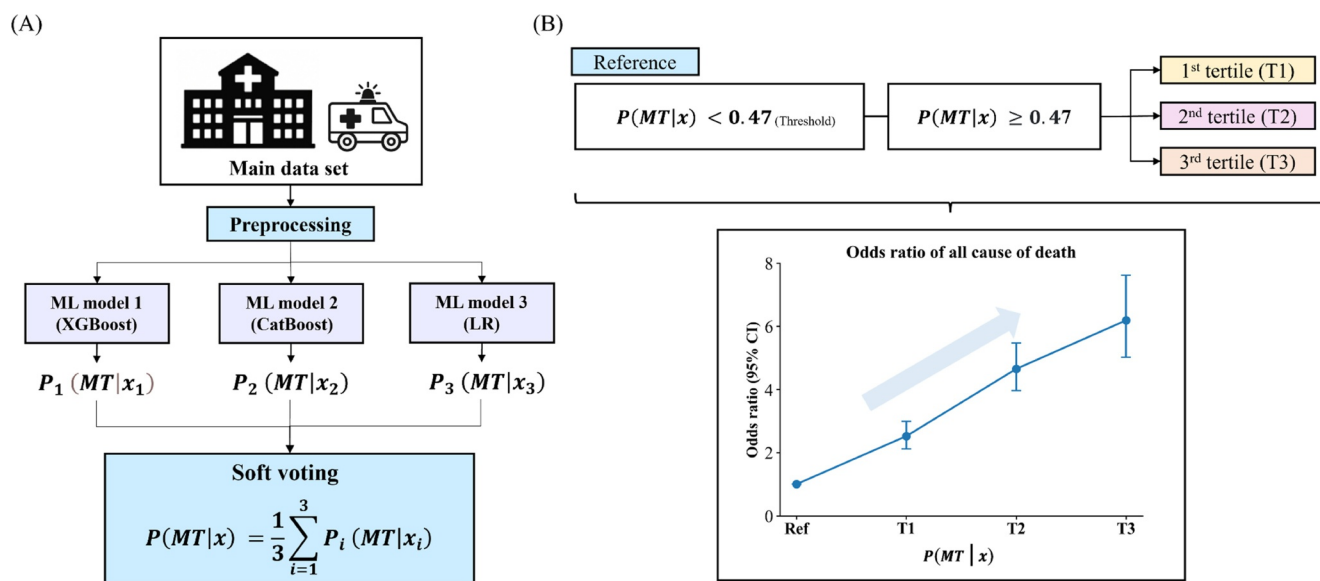


FIGURE 2 | Model structure of (A) the proposed ensemble learning model to predict infectious complications and (B) mortality risk based on the model probability.

in-hospital mortality events, showing a clear association between model-predicted risk and adverse outcomes. Together, these findings highlight that the proposed LR–CatBoost–XGBoost ensemble provides a robust, interpretable, and externally generalizable framework for early identification of patients with trauma at high risk of infection and subsequent mortality.

4.2 | Comparison With Previous Studies

Several studies have developed ML models to predict the risk of sepsis among trauma patients [5–7]; however, the study populations have generally focused on hospitalized or ICU cohorts, or on specific injury types, such as traumatic brain injury, thereby limiting their applicability to early, prehospital prediction. Because previous models relied almost exclusively on in-hospital measurements [5–7], they were unable to incorporate real-time physiological data collected during the prehospital phase, which is a critical window for early recognition and risk stratification.

Our study extends the field by developing a model that integrates prehospital and in-hospital clinical variables, representing one of the few attempts to predict post-traumatic infectious complications using information available before hospital arrival. This approach allows for the potential implementation of real-time infection risk alerts during early trauma management, when timely preventive action may be most effective [17].

While a few previous sepsis-related ML studies have applied SHAP analysis for model interpretation [18, 19], these were limited to ICU-admitted patients and identified predictors such as FiO2 and body mass index as the most influential factors [20]. In contrast, our SHAP findings revealed that neurological and physiological variables—particularly age, accident type, GCS verbal response score, and sex—were the strongest contributors to infection risk [21]. This difference underscores the distinctive nature of trauma-specific infection dynamics and the importance of early physiologic instability as a predictor of downstream

complications [22]. Furthermore, unlike many earlier models that were not externally validated, our ensemble model was tested across multiple trauma centers, showing its reproducibility and external generalizability. Collectively, these features distinguish our study as a step toward clinically interpretable, prehospital-ready infection prediction within the trauma continuum of care.

4.3 | Clinical Implications

Infectious complications are among the most serious sequelae following trauma, frequently leading to prolonged hospitalization and preventable death [23]. Early detection and timely intervention remain critical yet challenging, as infection risk is often underestimated during the initial triage or resuscitation phases [24]. The present study provides clinically actionable insight by introducing an ML model capable of predicting infectious complications based on routinely collected prehospital and in-hospital variables. All input features used in our model are obtainable within 24 h of patient admission [25].

Unlike traditional scoring systems, such as the Injury Severity Score or other rule-based indices [26], that rely on linear assumptions and limited variable interactions, the proposed LR–CatBoost–XGBoost ensemble integrates nonlinear patterns within multidimensional data. This approach presented improved predictive performance and interpretability through SHAP-based explanations, allowing clinicians to understand which features most strongly influence risk estimation. With further validation and real-time integration into trauma registries or electronic health records, this model may serve as an early-warning decision aid to guide infection surveillance, optimize antibiotic timing, and prioritize high-risk patients for targeted intervention [27].

4.4 | Limitations

Despite the inherent strengths of the study, being the first to develop and validate an ML model of the likelihood of infectious

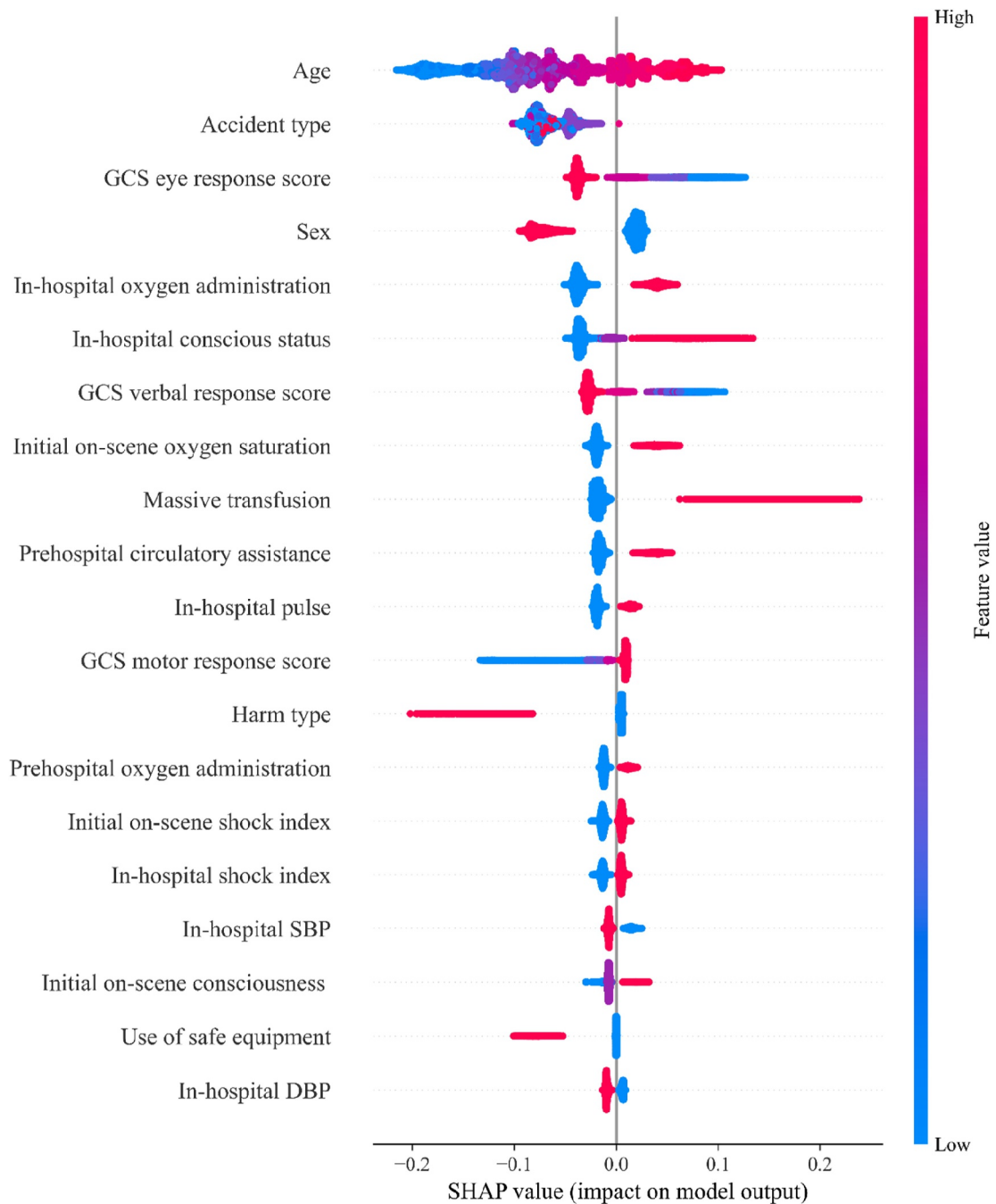


FIGURE 3 | SHAP values on our proposed ensemble learning model using the top 20 features. AED, automated external defibrillator; DBP, diastolic blood pressure; GCS, Glasgow Coma Scale; SBP, systolic blood pressure; SHAP, Shapley Additive Explanations.

TABLE 2 | ORs (95% CI) of in-hospital mortality by predicted likelihood tertiles.

Likelihood	Main data set		Cross validation data set		Test data set	
	aOR (95% CI) ^a	p-value	aOR (95% CI) ^a	p-value	aOR (95% CI) ^a	p-value
< 0.5	1.00 (reference)		1.00 (reference)		1.00 (reference)	
T1	2.52 (2.12 to 2.99)	< 0.001	3.10 (2.60 to 3.69)	< 0.001	2.27 (1.59 to 3.24)	< 0.001
T2	4.65 (3.96 to 5.47)	< 0.001	4.69 (3.89 to 5.66)	< 0.001	4.41 (3.05 to 6.37)	< 0.001
T3	6.19 (5.02 to 7.62)	< 0.001	6.12 (4.83 to 7.76)	< 0.001	4.99 (3.11 to 8.00)	< 0.001

Note: Data in bold indicate significant differences ($p < 0.05$).
Abbreviations: aOR, adjusted odds ratio; CI, confidence interval.
^aAdjusted for age, sex, prehospital cardiac arrest patient, accident type, harm type, safe equipment, harm time, initial on-scene consciousness status, initial on-scene vital signs, initial on-scene emergency treatment, In-hospital consciousness, In-hospital massive transfusion, In-hospital treatment, In-hospital vital signs, and GCS score.

complications following trauma using pre-hospital and in-hospital parameters, several limitations should be addressed. First, the analysis was conducted using data from South Korea, a relatively homogeneous population, which may constrain generalizability to trauma populations in other regions with different case-mix and care systems [28]. For instance, as penetrating injuries represent a minimal proportion of trauma cases in South Korea, we did not distinguish between blunt and penetrating trauma. Also, procedure-specific outcomes and surgical site infections (superficial, deep, and organ/space) could not be precisely captured due to data limitations. Nonetheless, the use of the KTDB, encompassing all major trauma centers in South Korea, enhances internal consistency and data quality and allows the use of pre-hospital data as well as in-hospital vital signs. Second, patients with incomplete or missing data were excluded rather than imputed, as most missingness stemmed from clinical rather than technical factors. This exclusion may have introduced selection bias; however, overfitting was mitigated through cross-validation and external testing across multiple centers. Third, because the study design was retrospective, unmeasured confounders and selection effects cannot be entirely ruled out. Although rigorous model development and prespecified analytical procedures were followed, prospective multicenter validation is warranted to confirm clinical applicability and assess real-world impact. Fourth, although four trauma centers were used for external validation, it was not possible to determine the proportion of patients contributed by these centers to the derivation cohort because participating institutions in the KTDB are anonymized. Finally, the model was trained exclusively on structured variables (e.g., demographics, vital signs, injury severity) and did not include unstructured data sources such as imaging findings, laboratory time-series, or clinical narratives. Future research incorporating multimodal data and continuous physiological monitoring could further enhance prediction accuracy and expand clinical utility [13].

5 | Conclusion

Our proposed ML model showed high accuracy, calibration, and external validity for predicting infectious complications in trauma patients. By incorporating prehospital variables, the model enables early and interpretable risk assessment, offering a practical tool to guide infection prevention and clinical prioritization. The proportional association between predicted infection risk and mortality underscores its prognostic significance, highlighting potential utility in optimizing resource allocation and improving outcomes in trauma care.

Author Contributions

Youngmin Kim: conceptualization, data curation, formal analysis, methodology, writing – original draft preparation, writing – review and editing. **Jiyeon Oh:** data curation, formal analysis, methodology, writing – original draft preparation, writing – review and editing. **Hyunjee Kim:** data curation, formal analysis, methodology, writing – original draft preparation, writing – review and editing. **Dongjin Yeo:** writing – review and editing. **Hyeon Jin Kim:** writing – review and editing. **Soeun Kim:** writing – review and editing. **Hayeon Lee:**

writing – review and editing. **Seung Ha Hwang:** writing – review and editing. **Jaehyeon Cho:** writing – review and editing. **Kisoo Kim:** writing – review and editing. **Jinseok Lee:** writing – review and editing. **Eunche Hyeon:** writing – review and editing. **Byungchul Yu:** writing – review and editing. **Do Wan Kim:** writing – review and editing. **Junepill Seok:** writing – review and editing. **Kyounghwan Kim:** writing – review and editing. **Dong Keon Yon:** conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing – original draft preparation, writing – review and editing. **Wu Seong Kang:** conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing – original draft preparation, writing – review and editing.

Acknowledgments

The data access agreement for the database (KTDB) was valid through March 2026.

Funding

This research was supported by the Ministry of Science and ICT (Grants RS-2024-00509257 and IITP-2026-RS-2024-00438239) and the Ministry of Health & Welfare (Grant RS-2025-02220492), Republic of Korea. The funders had no role in the study design, data collection, analysis, interpretation, or manuscript preparation.

Ethics Statement

The research protocol was approved by the Institutional Review Boards of Gachon University Gil Medical Center (GFIRB2024-357), Chonnam National University Hospital (CNUH-2024-016), Chungbuk National University Hospital (CBNUH 2024-10-017), and Cheju Halla General Hospital (CHH-2024-L03). The requirement for written informed consent was waived in accordance with relevant ethical guidelines for secondary use of administrative data.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data are available upon request. Study protocol and statistical code: available from WSK (wuseongkang@naver.com). Dataset: The dataset is available from the KTDB (discovery dataset) through data use agreements. However, the validation dataset cannot be shared due to privacy concerns, as it originates from internal hospital data. The underlying code for this study [and training/validation data set] is not publicly available but may be made available to qualified researchers upon reasonable request from the corresponding author.

References

1. K. Stern, Q. Qiu, M. Weykamp, G. O'Keefe, and S. C. Brakenridge, "Defining Posttraumatic Sepsis for Population-Level Research," *JAMA Network Open* 6, no. 1 (2023): e2251445, <https://doi.org/10.1001/jamanetworkopen.2022.51445>.
2. E. Eguia, C. Bunn, S. Kulshrestha, et al., "Trends, Cost, and Mortality From Sepsis After Trauma in the United States: An Evaluation of the National Inpatient Sample of Hospitalizations, 2012–2016," *Critical Care Medicine* 48, no. 9 (2020): 1296–1303, <https://doi.org/10.1097/ccm.0000000000004451>.
3. T. Abe, A. Komori, A. Shiraishi, et al., "Trauma Complications and in-Hospital Mortality: Failure-to-Rescue," *Critical Care* 24, no. 1 (2020): 223, <https://doi.org/10.1186/s13054-020-02951-1>.

4. A. Wafaisade, R. Lefering, B. Bouillon, et al., "Epidemiology and Risk Factors of Sepsis After Multiple Trauma: An Analysis of 29,829 Patients From the Trauma Registry of the German Society for Trauma Surgery," *Critical Care Medicine* 39, no. 4 (2011): 621–628, <https://doi.org/10.1097/ccm.0b013e318206d3df>.
5. J. Li, F. Xi, W. Yu, C. Sun, and X. Wang, "Real-Time Prediction of Sepsis in Critical Trauma Patients: Machine Learning-Based Modeling Study," *JMIR Formative Research* 7 (2023): e42452, <https://doi.org/10.2196/42452>.
6. B. Sun, M. Lei, L. Wang, et al., "Prediction of Sepsis Among Patients With Major Trauma Using Artificial Intelligence: A Multicenter Validated Cohort Study," *International Journal of Surgery* 111, no. 1 (2025): 467–480, <https://doi.org/10.1097/js9.0000000000001866>.
7. L. Tang, Y. Li, J. Zhang, et al., "Machine Learning Model to Predict Sepsis in ICU Patients With Intracerebral Hemorrhage," *Scientific Reports* 15, no. 1 (2025): 16326, <https://doi.org/10.1038/s41598-025-99431-9>.
8. G. S. Collins, K. G. M. Moons, P. Dhiman, et al., "TRIPOD+AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods," *BMJ* 385 (2024): e078378, <https://doi.org/10.1136/bmj-2023-078378>.
9. M. H. Kim, H. M. Cho, S. H. Kim, Y. Kim, Y. K. Shin, and K. H. Kim, "Prevalence of and Factors Associated With Trauma Surgeons' Referral and Patients' Willingness to Acupuncture Treatment After Traumatic Rib Fractures: A Single-Center Cross-Sectional Study," *Integrative Medicine Research* 13, no. 4 (2024): 101096, <https://doi.org/10.1016/j.imr.2024.101096>.
10. B. Yu, J. Cho, H. Kim, et al., "On-Scene Machine Learning Prediction Model for Massive Transfusion in Trauma and Its Association With in-Hospital Mortality," *BJS Open* 10, no. 1 (2025): zraf167, <https://doi.org/10.1093/bjsopen/zraf167>.
11. J. Cho, J. Oh, S. Kim, D. K. Yon, and W. S. Kang, "Deep Learning-Based Survival Outcomes of REBOA vs Resuscitative Thoracotomy in Trauma: A Nationwide Cohort Study in South Korea," *International Journal of Surgery* (2026), <https://doi.org/10.1097/JS9.00000000000004925>.
12. J. Im, E. W. Seo, K. Jung, and J. Kwon, "Understanding Regional Trauma Centers and Managing a Trauma Care System in South Korea: A Systematic Review," *Annals of Surgical Treatment and Research* 104, no. 2 (2023): 61–70, <https://doi.org/10.4174/ast.2023.104.2.61>.
13. H. Lee, S. H. Hwang, S. Park, et al., "Prediction Model for Type 2 Diabetes Mellitus and Its Association With Mortality Using Machine Learning in Three Independent Cohorts From South Korea, Japan, and the UK: A Model Development and Validation Study," *EclinicalMedicine* 80 (2025): 103069, <https://doi.org/10.1016/j.eclinm.2025.103069>.
14. A. B. Ahanger, S. W. Aalam, T. A. Masoodi, et al., "Radiogenomics and Machine Learning Predict Oncogenic Signaling Pathways in Glioblastoma," *Journal of Translational Medicine* 23, no. 1 (2025): 121, <https://doi.org/10.1186/s12967-025-06101-5>.
15. A. Mohammed and R. Kora, "A Comprehensive Review on Ensemble Deep Learning: Opportunities and Challenges," *Journal of King Saud University. Computer and Information Sciences* 35, no. 2 (2023): 757–774, <https://doi.org/10.1016/j.jksuci.2023.01.014>.
16. M. G. Poirot, H. G. Ruhe, H. M. M. Mutsaerts, et al., "Treatment Response Prediction in Major Depressive Disorder Using Multimodal MRI and Clinical Data: Secondary Analysis of a Randomized Clinical Trial," *American Journal of Psychiatry* 181, no. 3 (2024): 223–233, <https://doi.org/10.1176/appi.ajp.20230206>.
17. R. Chunara, C. C. Freifeld, and J. S. Brownstein, "New Technologies for Reporting Real-Time Emergent Infections," *Parasitology* 139, no. 14 (2012): 1843–1851, <https://doi.org/10.1017/s0031182012000923>.
18. J. Gao, Y. Lu, N. Ashrafi, I. Domingo, K. Alaei, and M. Pishgar, "Prediction of Sepsis Mortality in ICU Patients Using Machine Learning Methods," *BMC Medical Informatics and Decision Making* 24, no. 1 (2024): 228, <https://doi.org/10.1186/s12911-024-02630-z>.
19. S. Diwan, V. Gandhi, K. E. Baidya, P. Khanna, and A. Mehndiratta, "Explainable Machine Learning Models for Mortality Prediction in Patients With Sepsis in Tertiary Care Hospital ICU in Low- to Middle-Income Countries," *Intensive Care Medicine Experimental* 13, no. 1 (2025): 56, <https://doi.org/10.1186/s40635-025-00765-5>.
20. L. Douda, H. Hasnat, J. Schwank, et al., "Predictors of Intensive Care Unit Admissions in Patients Presenting With Coronavirus Disease 2019," *Avicenna Journal of Medicine* 14, no. 1 (2024): 45–53, <https://doi.org/10.1055/s-0043-1778068>.
21. Y. Han, X. Xie, J. Qiu, et al., "Early Prediction of Sepsis Associated Encephalopathy in Elderly ICU Patients Using Machine Learning Models: A Retrospective Study Based on the MIMIC-IV Database," *Frontiers in Cellular and Infection Microbiology* 15 (2025): 1545979, <https://doi.org/10.3389/fcimb.2025.1545979>.
22. B. Liu, J. Zhang, C. Song, et al., "Association of Glycemic Variability With Short- and Long-Term Mortality in Critically Ill Patients With Trauma," *Scientific Reports* 16, no. 1 (2025): 2588, <https://doi.org/10.1038/s41598-025-32464-2>.
23. F. Pandolfi, C. Brun-Buisson, D. Guillemot, and L. Watier, "Care Pathways of Sepsis Survivors: Sequelae, Mortality and Use of Healthcare Services in France, 2015–2018," *Critical Care* 27, no. 1 (2023): 438, <https://doi.org/10.1186/s13054-023-04726-w>.
24. A. Komori, H. Iriyama, T. Kainoh, M. Aoki, T. Naito, and T. Abe, "The Impact of Infection Complications After Trauma Differs According to Trauma Severity," *Scientific Reports* 11, no. 1 (2021): 13803, <https://doi.org/10.1038/s41598-021-93314-5>.
25. J. Wei, J. Zhou, Z. Zhang, et al., "Predicting Individual Patient and Hospital-Level Discharge Using Machine Learning," *Communication and Medicine* 4, no. 1 (2024): 236, <https://doi.org/10.1038/s43856-024-00673-x>.
26. N. Dehouche, "The Injury Severity Score: An Operations Perspective," *BMC Medical Research Methodology* 22, no. 1 (2022): 48, <https://doi.org/10.1186/s12874-022-01528-6>.
27. A. Pinto, F. Pennisi, S. Odelli, et al., "Artificial Intelligence in the Management of Infectious Diseases in Older Adults: Diagnostic, Prognostic, and Therapeutic Applications," *Biomedicine* 13, no. 10 (2025): 2525, <https://doi.org/10.3390/biomedicine13102525>.
28. C. N. Kim and S. J. Yoon, "Impact of COVID-19 on Severe Trauma Patients in Korea: A Nationwide Regional Subgroup Analysis," *Disaster Medicine and Public Health Preparedness* 19 (2025): e111, <https://doi.org/10.1017/dmp.2025.115>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting Information S1: wjs70389-sup-0001-suppl-data.docx.