



OPEN ACCESS

EDITED BY

Rahul Kashyap,
WellSpan Health, United States

REVIEWED BY

Abhinav Hoskote,
WellSpan Health, United States
Smitesh Padte,
WellSpan Health, United States

*CORRESPONDENCE

Yong-Goo Shin
✉ ygshin92@korea.ac.kr
Seung Park
✉ spark.cbnuh@gmail.com

†These authors have contributed equally to this work

RECEIVED 31 July 2024

ACCEPTED 17 December 2024

PUBLISHED 07 January 2025

CITATION

Kim J, Kim G-H, Kim J-W, Kim KH, Maeng J-Y, Shin Y-G and Park S (2025) Transformer-based model for predicting length of stay in intensive care unit in sepsis patients.
Front. Med. 11:1473533.
doi: 10.3389/fmed.2024.1473533

COPYRIGHT

© 2025 Kim, Kim, Kim, Kim, Maeng, Shin and Park. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Transformer-based model for predicting length of stay in intensive care unit in sepsis patients

Jeesu Kim^{1,2†}, Geun-Hyeong Kim^{1,2†}, Jae-Woo Kim^{1,2}, Ka Hyun Kim^{1,2}, Jae-Young Maeng¹, Yong-Goo Shin^{3*} and Seung Park^{2,4*}

¹Medical Artificial Intelligence Center, Chungbuk National University Hospital, Cheongju, Republic of Korea, ²College of Medicine, Chungbuk National University, Cheongju, Republic of Korea,

³Department of Electronics and Information Engineering, Korea University, Sejong, Republic of Korea,

⁴Department of Biomedical Engineering, Chungbuk National University Hospital, Cheongju, Republic of Korea

Introduction: Sepsis, a life-threatening condition with a high mortality rate, requires intensive care unit (ICU) admission. The increasing hospitalization rate for patients with sepsis has escalated medical costs due to the strain on ICU resources. Efficient management of ICU resources is critical to addressing this challenge.

Methods: This study utilized the dataset collected from 521 patients with sepsis at Chungbuk National University Hospital between July 2020 and August 2023. A transformer-based deep learning model was developed to predict ICU length of stay (LOS). The model incorporated global and local input data analysis through classification and feature-wise tokens, based on sequential organ failure assessment (SOFA) criteria. Model performance was evaluated using four-fold cross-validation.

Results: The proposed model achieved a mean absolute error (MAE) of 2.05 days for predicting ICU LOS. The result demonstrates the ability of the proposed model to provide accurate and reliable predictions.

Discussion: The proposed model offers valuable insights for healthcare resource management by optimizing ICU resource allocation and potentially reducing medical expenses. These findings highlight the applicability of the proposed model to efficient healthcare cost management.

KEYWORDS

sepsis, intensive care unit, length of stay, sequential organ failure assessment, transformer, tabular data

1 Introduction

Sepsis, a life-threatening condition, arises when the body's response to infection induces widespread inflammation (1). This inflammatory response can damage multiple organ systems, leading to severe multi-organ failure (2). The rapid progression of sepsis can result in death without timely and appropriate treatment (3). Global guidelines often recommend intensive care for patients with sepsis (2). Despite advancements in medical technology, including early diagnostic methods, rapid antibiotic administration (4), and advanced supportive care such as mechanical ventilation and extracorporeal membrane oxygenation (5), sepsis continues to pose a significant healthcare challenge worldwide due to increasing mortality and morbidity rates (6–11).

Treating sepsis in the intensive care unit (ICU) incurs significantly higher costs than treating other diseases owing to the need for advanced life support measures, prolonged hospitalization, and the complexity of managing multi-organ failure (3, 12, 13). Recent statistics indicate that the annual cost of treating sepsis in the United States exceeds \$24 billion, making it the most expensive treatment option for hospitals (13). This high cost is primarily driven by the length of stay (LOS) in the ICU because patients with sepsis often require prolonged intensive care. Sepsis accounts for 4.7–42.2% of global ICU utilization because ICU admission is recommended as an aggressive treatment regimen (14). Additionally, sepsis readmission rates are alarmingly high, with approximately 19% of survivors readmitted within 30 days, further escalating healthcare expenditures (15). Accurate prediction of ICU LOS for sepsis patients is crucial, as it enables healthcare facilities to optimize resource allocation, such as bed utilization, staffing, and equipment availability. By improving care efficiency, hospitals can reduce operational costs, enhance patient turnover rates, and ultimately contribute to cost savings for both healthcare providers and the broader system (16, 17).

Efforts to predict ICU LOS have significantly advanced in recent years. In 2022, Wu et al. (18) demonstrated the utility of machine learning techniques by predicting ICU LOS (area under the receiver operating characteristic curve (AUROC) = 0.742) using gradient boosting decision trees (GBDTs). In the same year, Deng et al. (19) improved accuracy (AUC = 0.765) by utilizing temporal data and focusing on the changes in progression according to treatment stages using gated recurrent units (GRU) and long short-term memory (LSTM) networks. In 2023, the emphasis shifted to simpler, clinically interpretable models, such as linear regression models utilizing the sequential organ failure assessment (SOFA) score. Zangmo and Khwannimit (20) developed a model to classify sepsis patients with ICU LOS exceeding 3 days (AUC = 0.530), while Farimani et al. (21) proposed a model to predict ICU LOS in cardiac surgery patients (root mean square error (RMSE) = 5.181). Despite the advances, existing models have struggled to effectively capture the complex feature interactions inherent in structured data. GBDT emphasizes individual feature importance through splits (22), making it less effective at explicitly modeling complex interactions or high-dimensional relationships. On the other hand, GRU and LSTM models are optimized for processing sequential data, the models exhibit structural limitations in learning complex inter-variable relationships in structured datasets (19). Similarly, linear regression assumes linear relationships between variables and is, therefore, unable to capture nonlinear interactions (23). The limitations underscore the necessity for innovative methods capable of effectively learning complex and high-dimensional data structures. Hence, we present a transformer-based solution to address the limitations of structured data analysis by simultaneously capturing nonlinear feature interactions and learning global relationships through attention mechanisms.

Transformers have shown promise in structured data analysis by incorporating innovative mechanisms such as the classification (CLS) token, a functionality originally introduced in the bidirectional encoder representations from transformers (BERT) (24) in 2018. The CLS token serves as a global representation of the input sequence, summarizing overall patterns in structured data through self-attention, and allows Transformers to effectively capture global dependencies across features. In 2021, Models such as self-attention and intersample attention transformer (SAINT) (25) and feature tokenizer (FT)-Transformer (26) successfully leveraged CLS tokens, achieving performance improvements in tabular datasets. SAINT

enables feature-to-feature and sample-to-sample interactions, while the FT-Transformer captures intricate inter-feature relationships and global patterns. However, while the CLS token excels at capturing global information, achieving a comprehensive analysis of attention mechanisms using CLS tokens can be challenging, as attention mechanisms may ignore important local feature details, particularly in datasets with complex interdependencies (27–29). The limitation of the CLS token underscores a persistent challenge in Transformer-based models when applied to highly intricate structured data.

This study focuses on predicting the ICU LOS for patients with sepsis using a transformer-based DL model applied to SOFA-based tabular data. The proposed model uses an attention mechanism and a skip-connected token process, integrating global information from a CLS token and local information from feature-wise tokens during the final classification. This approach adds to the growing body of work on applying DL techniques to tabular data in predicting ICU LOS for patients with sepsis.

2 Methods

2.1 Dataset information

2.1.1 Study population

To develop the DL model, we constructed a dataset from patients treated for sepsis at Chungbuk National University Hospital (Cheong-Ju, Korea) between July 3, 2020 and August 3, 2023. The study, conducted following the principles of the Declaration of Helsinki, received approval from the Institutional Review Board of Chungbuk National University Hospital (IRB no. CBNUH 2021-02-034-001). Patient information was anonymized and de-identified prior to analysis.

As shown in Figure 1, we initially identified patients meeting the Sepsis-3 guidelines for suspicion or diagnosis of sepsis, defined as a quick SOFA (qSOFA) score of ≥ 2 . We sequentially excluded patients who met the following criteria: ICU admission post-surgery, readmission due to sepsis during treatment, ICU stays of less than 24 h, withdrawal of life-sustaining therapy, ICU discharge, admission with cardiogenic shock, hypovolemic shock, or acute stroke, procalcitonin level of ≤ 0.05 , missing data, death, and ICU LOS outliers. This process resulted in a dataset comprising 521 patients.

We collected various clinical and SOFA-related features to construct a sepsis-specific ICU LOS prediction model. The features included: (1) Clinical features: age, sex, body mass index (BMI), lactate, atrial fibrillation (AF), systolic blood pressure (SBP), diastolic blood pressure (DBP), mean blood pressure (MBP), partial pressure of arterial oxygen and fraction of inspired oxygen ratio (PaO₂/FiO₂, P/F ratio), Glasgow Coma Scale (GCS)]. (2) SOFA-related features: vasopressor (VASO), mechanical ventilator (MV), 24-h urinary excretion (UR), platelets (Plt), serum total bilirubin (Bil), serum creatinine (Cr)]. (3) Target feature: ICU LOS.

Table 1 presents detailed statistical information on the features used in this study. Numerical features are described using means, standard deviations, and min-max ranges, while categorical features are reported as frequencies and percentages. The Pearson correlation coefficient for numerical features and point-biserial correlation for categorical features were calculated to determine their correlation with the target feature. *p*-values in Table 1 test the null hypothesis that the correlation coefficient is zero.

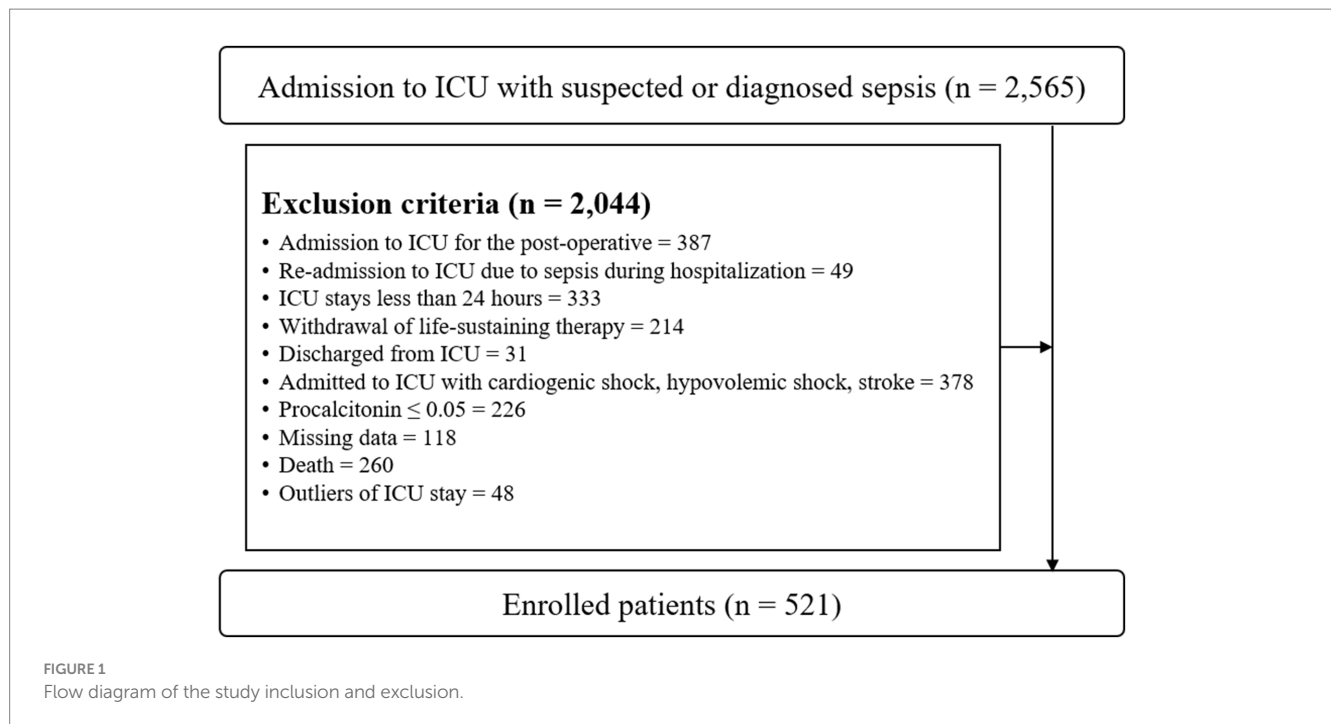
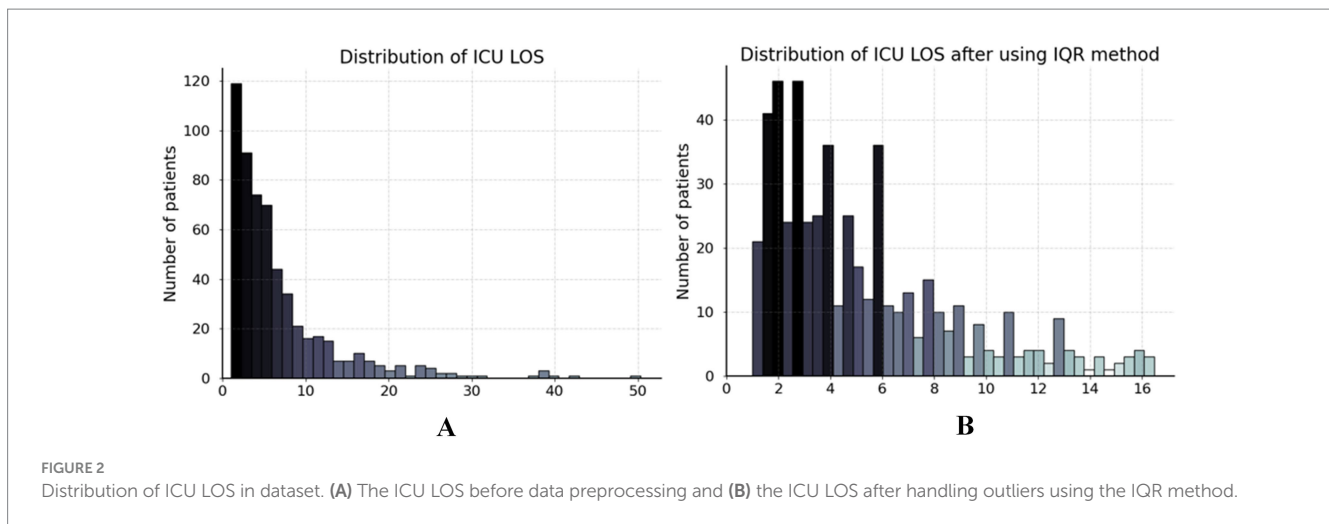


FIGURE 1
Flow diagram of the study inclusion and exclusion.

TABLE 1 Statistical information of features in our dataset.

| Characteristics | Categories | | Dataset (n = 521) | Correlation coefficient | p-value |
|-----------------------|------------|--------------------------|-----------------------------|-------------------------|---------|
| Clinical features | Age | Years | 69.19 ± 14.65 (19–95) | −0.09 | 0.05 |
| | Sex | Female | 239 (45.87%) | 0.01 | 0.90 |
| | | Male | 282 (54.13%) | | |
| | BMI | kg/cm ² | 22.41 ± 4.18 (11.55–43.94) | −0.07 | 0.11 |
| | LACTATE | mmol/L | 2.58 ± 2.56 (0–29) | 0.16 | < 0.01 |
| | AF | yes | 86 (16.51%) | 0.06 | 0.19 |
| | | no | 435 (83.49%) | | |
| | SBP | mmHg | 89.67 ± 19.59 (33–176) | −0.08 | 0.08 |
| | DBP | mmHg | 49.31 ± 11.46 (17–90) | −0.09 | 0.04 |
| | MBP | mmHg | 60.28 ± 11.83 (24–104) | −0.10 | 0.03 |
| PF | - | 290.69 ± 152.05 (16–943) | −0.31 | < 0.01 | |
| GCS | - | 11.08 ± 3.55 (3–15) | −0.42 | < 0.01 | |
| SOFA-related features | VASO | Yes | 298 (57.2%) | 0.11 | 0.02 |
| | | No | 223 (42.8%) | | |
| | MV | Yes | 171 (32.82%) | 0.43 | < 0.01 |
| | | No | 350 (67.18%) | | |
| | UR | cm ³ | 1803.52 ± 1332.96 (0–1,250) | −0.19 | < 0.01 |
| | Plt | ×10 ³ /μl | 154.16 ± 96.02 (4–657) | −0.06 | 0.19 |
| | Bil | mg/dl | 1.31 ± 2.09 (0.09–22.58) | −0.02 | 0.72 |
| Cr | mg/dl | 2.14 ± 2.20 (0.11–16.60) | 0.07 | 0.13 | |
| Target feature | ICU LOS | days | 5.24 ± 3.57 (1.01–16.47) | | |

SOFA, sequential organ failure assessment; BMI, body mass index; AF, atrial fibrillation; SBP, systolic blood pressure; DBP, diastolic blood pressure; MBP, mean blood pressure; P/F ratio, partial pressure of arterial oxygen and a fraction of inspired oxygen ratio (PaO₂/FiO₂); GCS, Glasgow Coma Scale; ICU LOS, length of stay in intensive care unit; VASO, vasopressor; MV, mechanical ventilator; UR, 24 h urinary excretion; Plt, platelets; Bil, serum total bilirubin; Cr, serum creatinine.



2.1.2 Data preprocessing

As depicted in [Figure 2A](#), the ICU LOS distribution in our dataset exhibited a pronounced positive skew, with a concentration of values at the lower end and a long tail extending towards higher values. This necessitated the removal of outliers prior to analysis. The interquartile range (IQR) method was employed to handle outliers in ICU LOS ([30](#)). The IQR method effectively retains most data points within a reasonable range, excluding outliers that could potentially distort the analysis ([31](#)). Specifically, the IQR is the range between the first quartile (Q1) and third quartile (Q3) of the data, with outliers defined as points below $Q1 - 1.5IQR$ or above $Q3 + 1.5IQR$ ([32](#)). Our study identified patients with an ICU LOS > 16.52 days as outliers, excluding 48 patients as shown in [Figure 2B](#). Furthermore, we standardized the dataset to ensure that all features contributed equally to the analysis and to prevent any single feature from disproportionately influencing the results due to scale differences. This procedure was applied exclusively to numerical features. The standardization formula is defined in [Equation 1](#) as follows:

$$Z = \frac{(X - \bar{X})}{s}, \quad (1)$$

where \bar{X} represents the mean and s denotes the standard deviation ([33](#)).

2.2 Model architecture

We developed a transformer-based DL model using a CLS token to predict the ICU LOS. The architecture of the proposed model is depicted in [Figure 3](#); it consists of three modules as shown in [Figure 4](#).

2.2.1 Module of concatenating CLS tokens

The input data $x \in \mathbb{R}^{col}$, where col represents the number of input features, is batch normalized before entering the “fzCLSBlock.” As represented in [Figure 4A](#), x in the fzCLSBlock is concatenated with a trainable CLS token $CLS_0 \in \mathbb{R}^1$, which is zero-initialized to ensure stable training ([24](#)). The CLS token is the first special token of every sequence and is widely used as an aggregate sequence representation for classification tasks ([26](#)). The concatenated vectors are embedded

through a dense layer to achieve a representative embedding of the input data and capture the complex relationships. This process is expressed in [Equation 2](#) as follows:

$$z = \text{Dense}(CLS_0 \parallel x) \in \mathbb{R}^{(col+1) \times d}, \quad (2)$$

where d represents the embedding dimension and \parallel denotes the concatenate function.

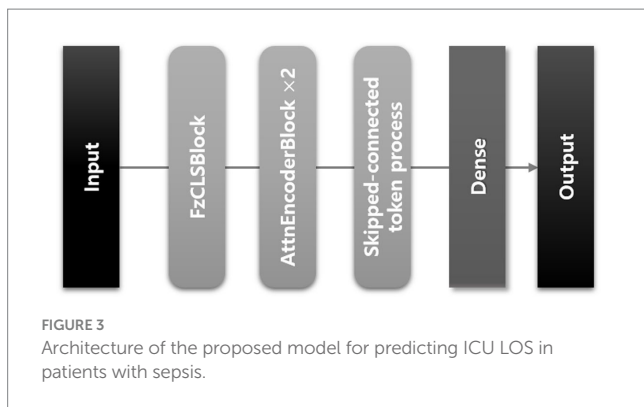
2.2.2 Module of multi-head self-attention

Inspired by networks in several studies using transformers ([25, 34–36](#)), we employed the self-attention mechanism of the transformer encoder. The self-attention mechanism calculates model weights to assess the relevance of each feature and captures interactions between features or instances. Recent research have demonstrated superior prediction accuracy by incorporating self-attention mechanisms in new networks such as TabNet ([35](#)) and FT-transformers ([26](#)). These findings suggest the efficacy of self-attention mechanisms for analyzing tabular datasets.

The projected vector $z \in \mathbb{R}^{(col+1) \times d}$ (as defined above) is analyzed in the “AttnEncoderBlock” illustrated in [Figure 4B](#). The input vector z is linearly transformed into query (Q), key (K), and value (V) matrices within the single attention head of multihead self-attention (MSA) ([37](#)). The attention weight is calculated by taking the dot product of Q and K, normalizing it by the square root of the dimension of K, and applying a softmax function. After that, the attention head outputs the dot product of the attention score and V, which are computed in parallel five times. Besides the MSA, the AttnEncoderBlock includes a fully connected feedforward network (FFN) composed of two linear transformations with a rectified linear unit (ReLU) activation in between ([38](#)).

2.2.3 Module of analyzing global and local information

Previous research indicates that CLS tokens often fail to adequately capture the semantic content of the input because they focus more on global information than local and low-level features ([39](#)). We designed a skip-connected token process, which comprehensively analyzes global and local data information, to address this issue. A



skip-connected token process adds token values representing both global and local information.

As presented in Figure 4C, the output of the AttnEncoderBlock is batch normalized and divided into the CLS $\in \mathbb{R}^{1 \times d}$ token, summarizing all features, and the feature-wise token $f' \in \mathbb{R}^{col \times d}$, maintaining the unique information of each feature. The f' , containing local information, passes through a dense layer to convert the local information into more abstract and high-level features. This layer captures complex dependencies and correlations between local features by identifying the interactions between various feature dimensions and learning appropriate weights. Additionally, the CLS token, containing global information, is added to f' , enabling a comprehensive analysis of global and local information. These computations can be expressed in Equations 3 and 4 as follows:

$$f = \text{Dense}(\text{Flatten}(f')) \in \mathbb{R}^d, \quad (3)$$

$$t_{sc} = \text{Dense}(\text{CLS} + f) \in \mathbb{R}^d. \quad (4)$$

The t_{sc} token is used for the final prediction of the proposed model, predicting the ICU LOS via a dense layer with one unit.

2.3 Implementation details

The proposed model was implemented using Python 3.9 on a workstation with an 11th Gen Intel(R) Core(TM) i7-11700K processor at 3.60 GHz and 64 GB of RAM. We applied exponential decay to control the learning rate during training, gradually reducing it to ensure stable convergence. The proposed model was configured with a batch size of 32 and the Adam optimizer at a learning rate of $1e-3$. Learning rate decay was applied every 10 steps at a rate of 0.96. Furthermore, we compared the prediction accuracy of the proposed model with that of conventional ML and DL models. Hyperparameters for random forest (RF) (40), extreme gradient boosting (XGBoost) (40), support vector regression (SVR) (41), multiple linear regression (MLR) (42), and TabNet (35) were set to their respective default values.

2.4 Model performance evaluation

We conducted a four-fold cross-validation to verify the reliability and consistency of the predictions of the proposed model. Twenty

percent of the dataset was allocated for testing, while the remaining dataset was divided into four folds. Each iteration of the four-fold validation consisted of one fold used for validation and the remaining folds used for training. We adopted the following three key metrics to quantitatively evaluate the performance of the proposed model because it performed a regression task: coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE). Detailed descriptions of each metric are as follows:

The R^2 value measures the proportion of variance in the dependent feature that can be predicted from the independent features. The R^2 value ranges from 0 to 1, where 0 indicates that the model does not explain the variability in the response data around its mean, and 1 indicates that the model explains all the variability of the response data around its mean (43). R^2 value for an ideal model is close to 1 and is computed using Equation 5 as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (5)$$

where n denotes the number of patients, y_i corresponds to the observed value, \hat{y}_i represents the predicted value, and \bar{y} is the average ICU LOS. The R^2 metric is crucial as it directly correlates with the proportion of the total variation in the target feature explained by the model. A high R^2 indicates that the model captures a significant portion of the variance, vital for predictive accuracy (44).

The MAE represents the average absolute difference between the predicted and observed values of the model (45). It provides a straightforward and interpretable measure of the average prediction error (46). Ideally, the MAE value approaches zero and is computed using Equation 6 as follows:

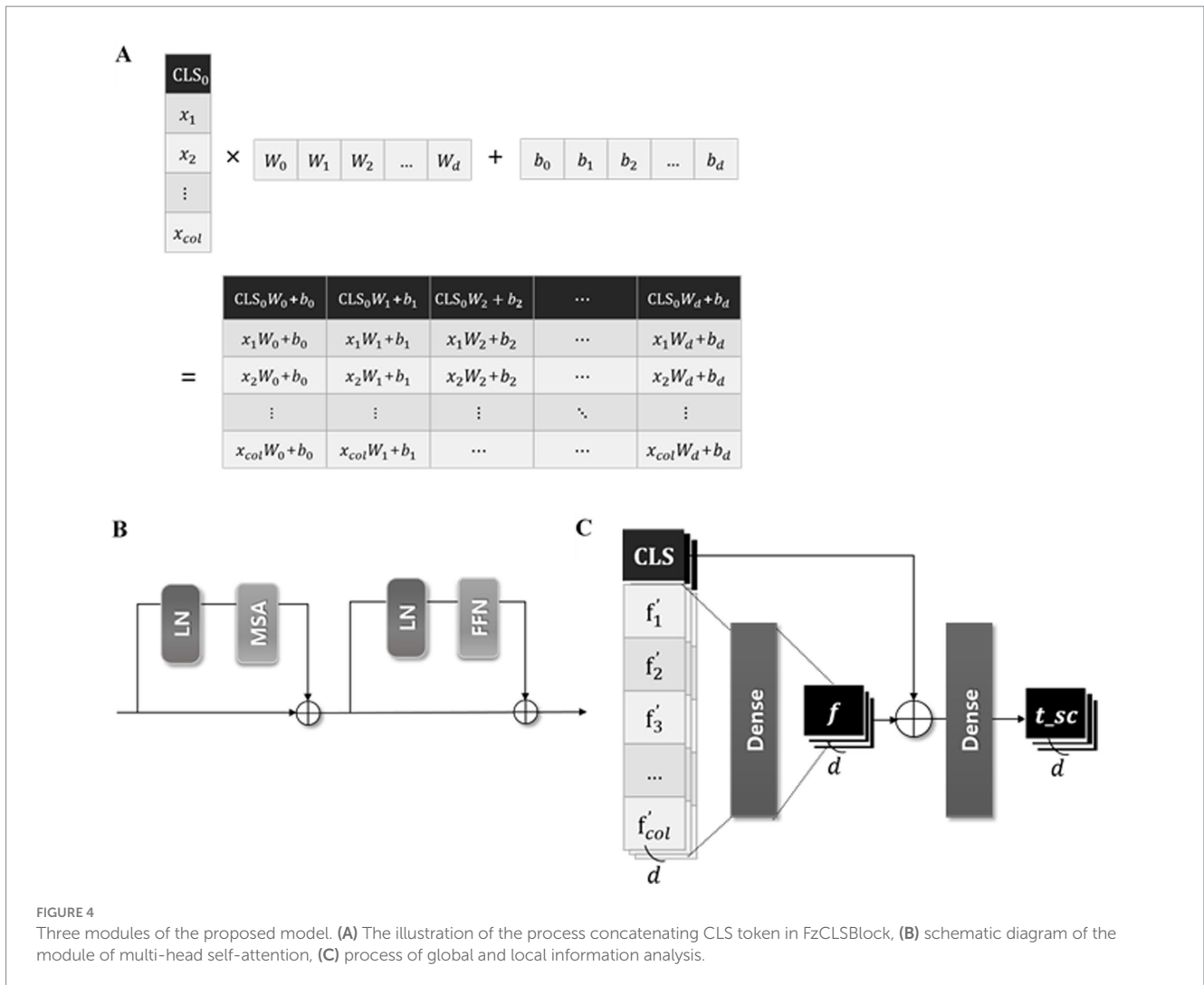
$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (6)$$

where n represents the number of data points used for model testing, \hat{y}_i corresponds to the value predicted by the model for the i -th sample, and y_i denotes the corresponding observed value (47). The MAE is advantageous due to its reduced sensitivity to outliers compared to metrics such as RMSE, making it a more reliable indicator of the average performance of a model, particularly when handling datasets with noisy or extreme values (48).

The RMSE is the square root of the average squared difference between the predicted and actual observations. It is widely used due to its ability to penalize larger errors more heavily than MAE, highlighting significant deviations (48). The formula for RMSE is calculated using Equation 7:

$$\text{RMSE} = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (7)$$

This metric provides an aggregate measure of model accuracy, encompassing both bias and variance components of error. RMSE is



valuable in applications where larger errors are more significant and must be minimized. Its sensitivity to large errors makes it essential for ensuring robustness and precision (49).

3 Results

3.1 Model performance comparison

We conducted a performance comparison of the proposed model using a four-fold cross-validation of the datasets. In Table 2, the proposed model demonstrated promising predictive performance, achieving an average R², MAE, and RMSE of 0.29 ± 0.01, 2.05 ± 0.03, and 2.72 ± 0.02, respectively. The average R² indicates the proposed model could explain approximately 29% of the variability in ICU LOS. Notably, the R², MAE and RMSE values showed minimal variation across folds, demonstrating the stability.

Figure 5 presents the calibration plot for each model, illustrating the agreement between predicted and observed ICU LOS. Darker points in these plots represent a better fit with actual values. The plots indicate that while the proposed model accurately predicts shorter ICU stays, it exhibits noticeable deviations for longer stays. This

finding suggests that the proposed model demonstrated strong performance for shorter ICU stays; however, it may require further refinement to improve accuracy for longer stays, which are often associated with more complex and variable patient conditions.

Additionally, we compared the performance of the proposed model with conventional ML and DL models, as shown in Table 3. The proposed model leveraged the skip-connected token process to enhance its predictive power by capturing interactions within tabular data. Comparisons were made with other DL models using MSA, such as TabNet and FT-Transformer, as well as traditional models known for their strong performance on tabular data, including RF, XGBoost, and MLR. The proposed model demonstrated superior performance compared to the other models, indicating that it provides more accurate predictions.

3.2 Ablation study

We conducted an ablation study to evaluate the effectiveness of the proposed skip-connected token process. This study compared the information delivered to the ICU LOS output layer by altering specific components. The performance of models was compared across three

categories: models that used only local information analysis, only global information analysis, or a combination of both. Detailed configurations and corresponding performance indicators are provided in Table 4.

The first model, which utilized only global information analysis, demonstrated poor performance. In contrast, the second model, relying solely on local information analysis, exhibited improved results. Notably, the proposed model, integrating global and local information analysis, achieved an R^2 , MAE, and RMSE of 0.29 ± 0.01 , 2.05 ± 0.03 , and 2.72 ± 0.02 , respectively, outperforming the other two models.

These results indicate that the proposed model, which uses skip-connected token process, has the highest explanatory power and lowest prediction error, demonstrating a significant enhancement in overall model performance. This underscores the necessity of skip-connected token process in integrating local and global information for improved predictions.

TABLE 2 Performance of the proposed model evaluated using four-fold cross-validation.

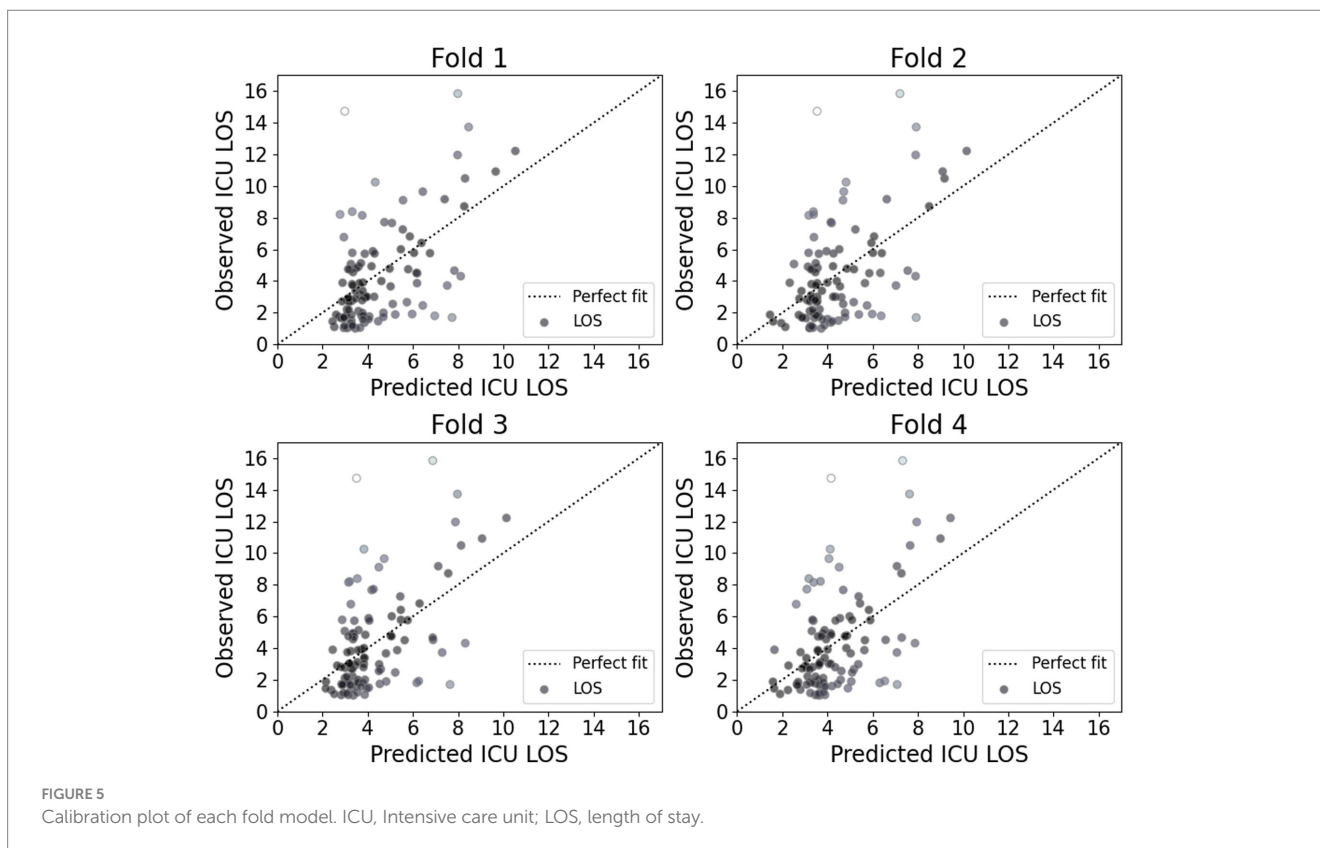
| Four-fold cross-validation | CBNUH (521 patients) | | |
|----------------------------|----------------------|-----------------|-----------------|
| | R^2 | MAE | RMSE |
| Fold1 | 0.30 | 2.03 | 2.69 |
| Fold2 | 0.28 | 2.06 | 2.73 |
| Fold3 | 0.29 | 2.01 | 2.71 |
| Fold4 | 0.27 | 2.08 | 2.74 |
| Average | 0.29 ± 0.01 | 2.05 ± 0.03 | 2.72 ± 0.02 |

3.3 Model interpretation

We employed Shapley additive descriptions (SHAP) to assess the impact of each feature on the model predictions. Figure 6A displays the mean absolute SHAP values for each feature, highlighting their importance in the model predictions. The top three most influential features were GCS, MV and PF. This ranking elucidates the primary factors that drive the predictions of the proposed model, offering valuable insights into which features most significantly affect ICU length of stay predictions.

Figure 6B presents the SHAP summary plots for four different percentiles (20th, 40th, 60th, and 80th) of ICU LOS. These plots visualize the SHAP values for individual predictions, indicating how each feature affects the predicted ICU LOS. The color scale shows the direction of the impact, with red and blue indicating an increase and decrease in the predicted LOS, respectively.

Sequentially across percentiles, the high GCS score and application of MV reduced predicted LOS, primarily at the 20th and 40th percentiles of ICU LOS (see blue section). Conversely, the low PF ratio increased LOS. The GCS represents a level of consciousness rating of 3–15 that assesses neurological status, with a lower score indicating worse status (50). The MV and PF are important indicators of respiratory function, reflecting the need for mechanical ventilation and the oxygen exchange capacity of the lungs, respectively (49). The impact of MV application and low PF was also evident at the 60th percentile, significantly increasing ICU LOS. On the other hand, an average level of Plt indicates a properly functioning coagulation system and reduces ICU LOS. The low GCS score, high Bil level, and MV application played a significant role in 80th percentile ICU LOS, with severe GCS score significantly increasing expected ICU LOS. Bil is another important predictor of ICU



LOS, with elevated levels indicating liver dysfunction or hemolysis. However, contrarily in our study, elevated Bil was shown to reduce ICU LOS.

4 Discussion

This study demonstrated that the transformer-based DL model outperformed traditional ML and DL models in predicting ICU LOS

for patients with sepsis using SOFA-based tabular data. The proposed model, leveraging a skip-connected token process to integrate global and local information, achieved an average R², MAE, and RMSE of 0.29, 2.05 days, and 2.72 days, respectively. Reliable predictions of ICU LOS are clinically and operationally impactful, as they enable better resource allocation and improve patient outcomes, particularly for critical conditions like sepsis (51, 52). The proposed model builds on these insights by providing an efficient tool that uses limited SOFA-based data to achieve practical predictions.

TABLE 3 Comparison of conventional model performance.

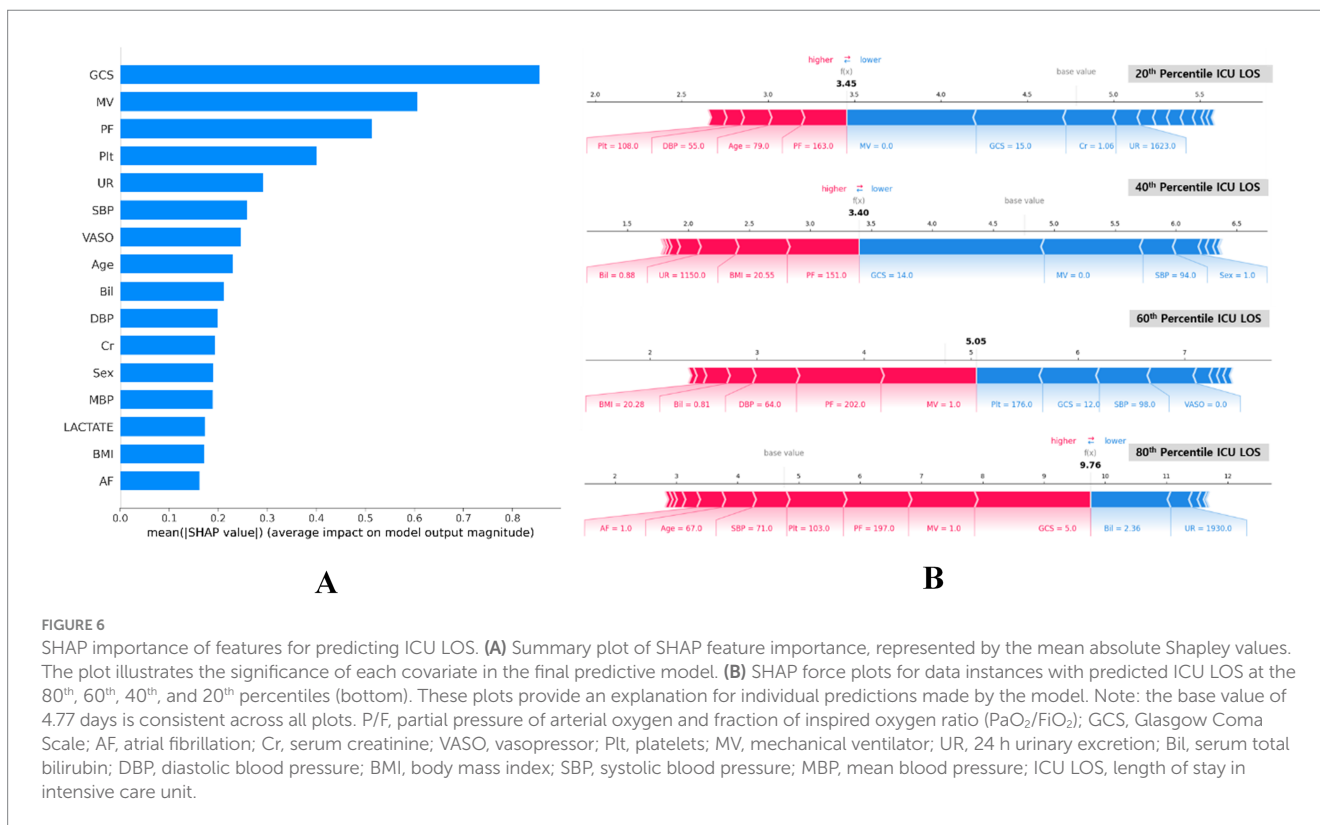
| Model | CBNUH | | |
|----------------|----------------|-------------|-------------|
| | R ² | MAE | RMSE |
| RF | 0.09 ± 0.01 | 2.25 ± 0.01 | 3.07 ± 0.01 |
| XGBoost | 0.10 ± 0.02 | 2.23 ± 0.02 | 3.05 ± 0.04 |
| MLR | 0.22 ± 0.00 | 2.13 ± 0.01 | 2.84 ± 0.01 |
| TabNet | -0.39 ± 0.21 | 2.77 ± 0.17 | 3.79 ± 0.29 |
| FT-Transformer | 0.26 ± 0.03 | 2.12 ± 0.02 | 2.76 ± 0.05 |
| Proposed model | 0.29 ± 0.01 | 2.05 ± 0.03 | 2.72 ± 0.02 |

TABLE 4 Ablation study on the proposed model.

| Method | R ² | MAE | RMSE |
|---------------------------------------|----------------|-------------|-------------|
| Local information analysis | 0.22 ± 0.02 | 2.11 ± 0.06 | 2.84 ± 0.04 |
| Global information analysis | 0.15 ± 0.02 | 2.27 ± 0.04 | 2.97 ± 0.03 |
| Local and Global information analysis | 0.29 ± 0.01 | 2.05 ± 0.03 | 2.72 ± 0.02 |

The strengths of the proposed model are manifold: First, the input features are based on the SOFA criteria, widely used in ICUs to assess organ dysfunction severity in critically ill patients. The model requires only 16 SOFA-related clinical features collected within 24 h of ICU admission, making it a convenient tool for predicting ICU LOS in patients with sepsis due to the accessibility of SOFA criteria data. Second, the proposed model was designed to work with tabular data, the most common structured data format, which requires less computational power than other data types and does not necessitate high-end hardware. Third, the model effectively captures comprehensive information from the features utilizing CLS and feature-wise tokens, analyzing global and local information. The proposed model employs MSA to capture global interactions between features, further analyze local information through dense layers and then integrates both in the final prediction to enhance performance.

Furthermore, the proposed model was interpreted using SHAP, providing valuable insights into the relative importance of various features in predicting ICU LOS. The top three influential features in this study were GCS, MV, and PF. The GCS score, underscored for its critical role in assessing neurological status, showed a positive correlation with ICU LOS. This finding is consistent with a previous



study (53), and highlights the importance of GCS as the most significant predictor. Similarly, MV and PF are respiratory indices associated with ICU LOS prediction in this study. The results in our study are consistent with previous studies showing that MV use and lower PF increased ICU LOS (53, 54). Conversely, this study found that elevated bilirubin levels were associated with a shorter ICU LOS, which contrasts with a previous study where higher bilirubin levels prolonged the length of hospital stay (55). The correlation of these factors indicates that these may assist in determining ICU LOS.

However, this study has several limitations. The dataset was derived from a single institution, potentially limiting the generalizability of the findings. Future research should aim to validate the proposed model across diverse healthcare settings and larger multicenter datasets. Additionally, while the transformer-based model outperformed others in predicting ICU LOS, it showed an opportunity for improvement, particularly in predicting stays longer than 8 days. This result suggests the need for additional data on longer durations to improve the prediction of extended ICU LOS in real medical scenarios.

5 Conclusion

We developed a transformer-based DL model to predict ICU LOS in patients with sepsis using data collected within the first 24 h of ICU admission. The proposed model achieved an MAE of 2.05 days. The proposed model effectively captures complex feature interactions by integrating global and local information through a novel skip-connected token process. Additionally, the proposed model utilizes a set of SOFA-related features that are widely used to assess the severity of organ dysfunction in clinical practice. Such an approach ensures simplicity of data collection and wide applicability, making the proposed model practical for use in a variety of healthcare settings.

Data availability statement

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

Ethics statement

Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article because the review board waived the

requirement for informed consent, owing to the retrospective design of this study.

Author contributions

JK: Methodology, Software, Validation, Visualization, Writing – original draft. G-HK: Formal analysis, Methodology, Software, Writing – review & editing. J-WK: Data curation, Formal analysis, Methodology, Software, Writing – review & editing. KK: Data curation, Formal analysis, Validation, Writing – original draft. J-YM: Conceptualization, Data curation, Formal analysis, Writing – original draft. Y-GS: Methodology, Project administration, Supervision, Validation, Writing – review & editing. SP: Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was supported by a research grant from Chungbuk National University in 2024, and by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (Grant Number: RS-2023-00267328).

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.

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.

References

- Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*. (2016) 315:801–10. doi: 10.1001/jama.2016.0287
- Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Crit Care Med*. (2021) 49:e1063–143. doi: 10.1097/CCM.0000000000005337
- Khwannimit B, Bhurayanontachai R. The direct costs of intensive care management and risk factors for financial burden of patients with severe sepsis and septic shock. *J Crit Care*. (2015) 30:929–34. doi: 10.1016/j.jcrc.2015.05.011
- Kim HI, Park S. Sepsis: early recognition and optimized treatment. *Tuberc Respir Dis*. (2019) 82:6–14. doi: 10.4046/trd.2018.0041
- Fortenberry JD, Paden ML. Extracorporeal therapies in the treatment of sepsis: experience and promise. *Seminars in pediatric infectious diseases*. Amsterdam: Elsevier (2006).
- Dombrovskiy VY, Martin AA, Sunderram J, Paz HL. Facing the challenge: decreasing case fatality rates in severe sepsis despite increasing hospitalizations. *Crit Care Med*. (2005) 33:2555–62. doi: 10.1097/01.CCM.0000186748.64438.7B
- van Gestel A, Bakker J, Verraart CP, van Hout BA. Prevalence and incidence of severe sepsis in Dutch intensive care units. *Crit Care*. (2004) 8:1–10. doi: 10.1186/cc2858
- Stoller J, Halpin L, Weis M, Aplin B, Qu W, Georgescu C, et al. Epidemiology of severe sepsis: 2008–2012. *J Crit Care*. (2016) 31:58–62. doi: 10.1016/j.jcrc.2015.09.034

9. Martin GS, Mannino DM, Eaton S, Moss M. The epidemiology of sepsis in the United States from 1979 through 2000. *N Engl J Med.* (2003) 348:1546–54. doi: 10.1056/NEJMoa022139
10. Martin GS. Sepsis, severe sepsis and septic shock: changes in incidence, pathogens and outcomes. *Expert Rev Anti-Infect Ther.* (2012) 10:701–6. doi: 10.1586/eri.12.50
11. Patel JJ, Taneja A, Niccum D, Kumar G, Jacobs E, Nanchal R. The association of serum bilirubin levels on the outcomes of severe sepsis. *J Intensive Care Med.* (2015) 30:23–9. doi: 10.1177/0885066613488739
12. Paoli CJ, Reynolds MA, Sinha M, Gitlin M, Crouser E. Epidemiology and costs of sepsis in the United States—an analysis based on timing of diagnosis and severity level. *Crit Care Med.* (2018) 46:1889–97. doi: 10.1097/CCM.0000000000003342
13. Torio CM, Moore BJ. (2016). National inpatient hospital costs: The most expensive conditions by payer, 2013.
14. Raman V, Laupland KB. Challenges to reporting the global trends in the epidemiology of ICU-treated sepsis and septic shock. *Curr Infect Dis Rep.* (2021) 23:1–8. doi: 10.1007/s11908-021-00749-y
15. Fingar K, Washington R. (2016). Trends in hospital readmissions for four high-volume conditions, 2009–2013.
16. Yinusa A, Faezipour M. Optimizing healthcare delivery: a model for staffing, patient assignment, and resource allocation. *Appl Syst Innov.* (2023) 6:78. doi: 10.3390/asi6050078
17. Cosgrove SE. The relationship between antimicrobial resistance and patient outcomes: mortality, length of hospital stay, and health care costs. *Clin Infect Dis.* (2006) 42:S82–9. doi: 10.1086/499406
18. Wu J, Lin Y, Li P, Hu Y, Zhang L, Kong G. Predicting prolonged length of ICU stay through machine learning. *Diagnostics.* (2021) 11:2242. doi: 10.3390/diagnostics11122242
19. Deng Y, Liu S, Wang Z, Wang Y, Jiang Y, Liu B. Explainable time-series deep learning models for the prediction of mortality, prolonged length of stay and 30-day readmission in intensive care patients. *Front Med.* (2022) 9:933037. doi: 10.3389/fmed.2022.933037
20. Zangmo K, Khwannimit B. Validating the APACHE IV score in predicting length of stay in the intensive care unit among patients with sepsis. *Sci Rep.* (2023) 13:5899. doi: 10.1038/s41598-023-33173-4
21. Farimani RM, Amini S, Bahaadini K, Eslami S. (2024). Predicting length of stay in intensive care units for cardiovascular surgery patients using APACHE II, APACHE IV, SAPS II and SOFA scores.
22. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: a highly efficient gradient boosting decision tree. In: Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. *Advances in Neural Information Processing Systems*. Vol. 30. Long Beach, CA, USA: Curran Associates Inc. (2017). p. 3146–3154. Available at: <https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde84869bdd9eb6b76fa-Paper.pdf>
23. Hope TM. Linear regression. Machine learning. Amsterdam: Elsevier; (2020). p. 67–81.
24. Devlin J, Chang M-W, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805.* (2018)
25. Somepalli G, Goldblum M, Schwarzschild A, Bruss CB, Goldstein T. Saint: improved neural networks for tabular data via row attention and contrastive pre-training. *arXiv preprint arXiv:2106.01342.* (2021)
26. Gorishniy Y, Rubachev I, Khrullkov V, Babenko A. Revisiting deep learning models for tabular data. In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates Inc. (2021). p. 18932–43. Available at: <https://proceedings.neurips.cc/paper/2021/file/01a1a84df7a0baaa88a1a9fa7848d42c-Paper.pdf>
27. Chen C-FR, Fan Q, Panda R, editors. (2021). “Crossvit: cross-attention multi-scale vision transformer for image classification.” in *Proceedings of the IEEE/CVF international conference on computer vision.*
28. Yuan L, Chen Y, Wang T, Yu W, Shi Y, Jiang Z-H, et al., editors. (2021). “Tokens-to-token vit: training vision transformers from scratch on imagenet.” in *Proceedings of the IEEE/CVF international conference on computer vision.*
29. Wang J, Yu X, Gao Y. Feature fusion vision transformer for fine-grained visual categorization. *arXiv preprint arXiv:2107.02341.* (2021)
30. Mansoori A, Zeinalnezhad M, Nazarianesh L. Optimization of tree-based machine learning models to predict the length of hospital stay using genetic algorithm. *J Healthc Eng.* (2023) 2023:3395. doi: 10.1155/2023/9673395
31. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol.* (2014) 14:1–13. doi: 10.1186/1471-2288-14-135
32. Walfish S. A review of statistical outlier methods. *Pharm Technol.* (2006) 30:82.
33. Raju VG, Lakshmi KP, Jain VM, Kalidindi A, Padma V, editors. (2020). “Study the influence of normalization/transformation process on the accuracy of supervised classification.” in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT): IEEE.*
34. Hwang Y, Song J. Recent deep learning methods for tabular data. *Commun Appl Methods.* (2023) 30:215–26. doi: 10.29220/CSAM.2023.30.2.215
35. Arik SÖ, Pfister T. Tabnet: attentive interpretable tabular learning. *Proc AAAI Conf Artif Intel.* (2021) 35:6679–87. doi: 10.1609/aaai.v35i8.16826
36. Song W, Shi C, Xiao Z, Duan Z, Xu Y, Zhang M, et al., editors. (2019). “AutoInt: automatic feature interaction learning via self-attentive neural networks.” in *Proceedings of the 28th ACM international conference on information and knowledge management.*
37. Choi Y, Lee Y, Cho J, Baek J, Kim B, Cha Y, et al., editors. (2020). “Towards an appropriate query, key, and value computation for knowledge tracing.” in *Proceedings of the seventh ACM conference on learning@scale.*
38. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, CA, USA: Curran Associates Inc. (2017). p. 6000–6010. Available at: <https://proceedings.neurips.cc/paper/2017/file/3f5ec243547dee91fbd053c1c4a845aa-Paper.pdf>
39. Choi J-W, Yang M, Kim J-W, Shin YM, Shin Y-G, Park S. Prognostic prediction of sepsis patient using transformer with skip connected token for tabular data. *Artif Intell Med.* (2024) 149:102804. doi: 10.1016/j.artmed.2024.102804
40. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn.* (2006) 63:3–42. doi: 10.1007/s10994-006-6226-1
41. Dou B, Zhu Z, Merkurjev E, Ke L, Chen L, Jiang J, et al. Machine learning methods for small data challenges in molecular science. *Chem Rev.* (2023) 123:8736–80. doi: 10.1021/acs.chemrev.3c00189
42. Tranmer M, Elliot M. Multiple linear regression. *Cathie Marsh Centre Census Surv Res.* (2008) 5:1–5.
43. Nagelkerke NJ. A note on a general definition of the coefficient of determination. *Biometrika.* (1991) 78:691–2. doi: 10.1093/biomet/78.3.691
44. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. Heidelberg: Springer (2013).
45. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature. *Geosci Model Dev.* (2014) 7:1247–50. doi: 10.5194/gmd-7-1247-2014
46. Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim Res.* (2005) 30:79–82. doi: 10.3354/cr030079
47. Li C, Chen L, Feng J, Wu D, Wang Z, Liu J, et al. Prediction of length of stay on the intensive care unit based on least absolute shrinkage and selection operator. *IEEE Access.* (2019) 7:110710–21. doi: 10.1109/ACCESS.2019.2934166
48. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE). *Geosci Model Dev Discuss.* (2014) 7:1525–34. doi: 10.5194/gmd-7-1525-2014
49. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. *Int J Forecast.* (2006) 22:679–88. doi: 10.1016/j.ijforecast.2006.03.001
50. Bastos PG, Sun X, Wagner DP, Wu AW, Knaus WA. Glasgow coma scale score in the evaluation of outcome in the intensive care unit: findings from the acute physiology and chronic health evaluation III study. *Crit Care Med.* (1993) 21:1459–65. doi: 10.1097/00003246-199310000-00012
51. Stone K, Zwiggelaar R, Jones P, Mac PN. A systematic review of the prediction of hospital length of stay: towards a unified framework. *PLoS Dig Health.* (2022) 1:e000017. doi: 10.1371/journal.pdig.0000017
52. Yu Z, Ashrafi N, Li H, Alaei K, Pishgar M. Prediction of 30-day mortality for ICU patients with Sepsis-3. *BMC Med Inform Decis Mak.* (2024) 24:223. doi: 10.1186/s12911-024-02629-6
53. Peres IT, Hamacher S, Oliveira FLC, Thomé AMT, Bozza FA. What factors predict length of stay in the intensive care unit? Systematic review and meta-analysis. *J Crit Care.* (2020) 60:183–94. doi: 10.1016/j.jcrc.2020.08.003
54. Tobi K, Ekwere I, Ochukpe C. Mechanical ventilation in the intensive care unit: a prospective study of indications and factors that affect outcome in a tertiary hospital in Nigeria. *J Anesth Clin Res.* (2017) 8:2. doi: 10.4172/2155-6148.1000718
55. Yang Z-X, Lv X-L, Yan J. Serum total bilirubin level is associated with hospital mortality rate in adult critically ill patients: a retrospective study. *Front Med.* (2021) 8:697027. doi: 10.3389/fmed.2021.697027