Pediatric septic shock estimation using deep learning and electronic medical records

Ji Weon Lee¹, Bongjin Lee²,³, June Dong Park²

¹Integrated and Respite Care Center for Children, Seoul National University, Seoul, Korea
²Department of Pediatrics, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea
³Innovative Medical Technology Research Institute, Seoul National University Hospital, Seoul, Korea

Background: Diagnosing pediatric septic shock is difficult due to the complex and often impractical traditional criteria, such as systemic inflammatory response syndrome (SIRS), which result in delays and higher risks. This study aims to develop a deep learning-based model using SIRS data for early diagnosis in pediatric septic shock cases.

Methods: The study analyzed data from pediatric patients (<18 years old) admitted to a tertiary hospital from January 2010 to July 2023. Vital signs, lab tests, and clinical information were collected. Septic shock cases were identified using SIRS criteria and inotrope use. A deep learning model was trained and evaluated using the area under the receiver operating characteristics curve (AUROC) and area under the precision-recall curve (AUPRC). Variable contributions were analyzed using the Shapley additive explanation value.

Results: The analysis, involving 9,616,115 measurements, identified 34,696 septic shock cases (0.4%). Oxygen supply was crucial for 41.5% of the control group and 20.8% of the septic shock group. The final model showed strong performance, with an AUROC of 0.927 and AUPRC of 0.879. Key influencers were age, oxygen supply, sex, and partial pressure of carbon dioxide, while body temperature had minimal impact on estimation.

Conclusions: The proposed deep learning model simplifies early septic shock diagnosis in pediatric patients, reducing the diagnostic workload. Its high accuracy allows timely treatment, but external validation through prospective studies is needed.

Key Words: clinical decision support systems; early diagnosis; machine learning, pediatrics; septic shock

INTRODUCTION

Septic shock is associated with high morbidity and mortality, and delayed diagnosis often results in poor treatment outcomes [1-5]. Therefore, it is crucial to detect it early and administer appropriate treatment promptly [2,6]. However, detecting septic shock in children presents several challenges. Prior to 2016, both adults and children utilized systemic inflammatory response syndrome (SIRS)-based diagnostic criteria for septic shock [7,8]. Since then, adults have transitioned to a system based on the sepsis-3 definition, employing the Sequential Or-
organ Failure Assessment (SOFA) score [4]. However, due to the impracticality of applying SOFA in children, a diagnostic system based on the sepsis-3 definition could not be implemented for pediatric cases [4,9,10].

As a result, the diagnosis of septic shock in children comprises a blend of existing SIRS-based methods and diagnostic approaches based on pediatric SOFA and age-specific SOFA criteria developed by various researchers [10-12]. Nevertheless, SIRS-based diagnostics remain a significant component of various diagnostic systems [7,9,11]. The SIRS-based diagnostic process, however, is intricate. Parameters such as leukocyte count, body temperature (BT), heart rate (HR), respiratory rate (RR), and systolic blood pressure (SBP), each assessed based on age-specific criteria, must be examined to determine SIRS [7,8,13]. Subsequently, the diagnosis of septic shock depends on whether it is infection-related and accompanied by organ dysfunction, necessitating detailed information such as hourly fluid bolus treatment, capillary refill time, and changes in urine volume [8,13,14]. This results in an increased workload for medical staff and adds complexity to the diagnostic process [15,16]. Nevertheless, even though manually performing this complex and repetitive task can be challenging for an individual, it can be facilitated more efficiently with artificial intelligence. Moreover, it may be possible to diagnose using variables that are easier to collect or more concise than the parameters required by the existing diagnostic system.

Therefore, the purpose of this study was to develop a model for diagnosing septic shock in children using deep learning and information that can be easily obtained from electronic medical records (EMRs).

**MATERIALS AND METHODS**

The Institutional Review Board of Seoul National University Hospital reviewed and approved this study (No. H-1909-043-1062), and as a retrospective observational study that does not use personal identification codes, written consent was exempted.

**Study Setting and Subjects**

This retrospective observational study was conducted in a tertiary children’s hospital with approximately 350 beds affiliated with the university. Eligible subjects included patients under the age of 18 admitted to the institution between January 2010 and July 2023. Patients admitted to the emergency department or intensive care unit were excluded.

**KEY MESSAGES**

- The study introduces a deep learning-based model for early diagnosis of septic shock in pediatric patients, providing a simpler, more effective alternative to complex traditional criteria.
- Analysis of data from a tertiary university hospital demonstrates the model’s high accuracy (area under the receiver operating characteristics curve of 0.927 and area under the precision-recall curve of 0.879), emphasizing its potential to reduce diagnostic delays and improve treatment outcomes in pediatric septic shock cases.

**Data Source and Preprocessing**

The data used in this study were provided in pseudonymized form from the data warehouse of our medical information system. Demographic information such as age and sex was collected. Vital signs including SBP, diastolic blood pressure (DBP), HR, RR, BT, and pulse oxygen saturation (SpO2) were also recorded. Additionally, laboratory tests and clinical information on blood pH, partial pressure of carbon dioxide (pCO2), C-reactive protein (CRP), oxygen therapy, and the use of inotropes were gathered. As vital signs are typically measured multiple times a day, whereas laboratory tests may be conducted once a day or may not be conducted at all, we matched the measurement times of vital signs with the dates of laboratory test executions. Consequently, we joined each corresponding set of records based on matching measurement dates and laboratory test execution dates. Even if it was a vital sign of a patient admitted to a general ward, measurements taken outside of general wards, such as examination rooms and operating rooms, were excluded from the analysis. Non-physiological anomalies that could be attributed to keystroke errors in the input process were also excluded from the collected vital signs. These anomalies included SBP >300 mm Hg or <30 mm Hg, DBP > SBP, HR >300 beats/min or <5 beats/min, RR >120 breaths/min, BT >42 °C or <30 °C, and SpO2 >100% or <0%. For vital signs of SBP, DBP, HR, and RR, which exhibit age-related variations in normal values, z-scores were calculated based on age rather than using the raw values. Centile data from previous studies were utilized for the z-score conversion [17].

Among all measurements, if SBP, DBP, HR, and RR were all missing, even if there were other values, they were excluded.
from the analysis. However, if only some variables were missing, they were imputed to the values immediately prior to the patient’s same hospital period, and the remaining missing values were imputed as the mean values of the variables. For deep learning, categorical variables were converted to one-hot encoding, and continuous variables were standardized to values between 0 and 1. R software version 4.3.1 (R Foundation for statistical computing; https://www.r-project.org) and R packages such as the generalized additive models for location scale and shape and the sitar model were used for data preprocessing [18-20].

Outcomes
The primary outcome was the septic shock estimation performance of the developed model, evaluated using the area under the receiver operating characteristics curve (AUROC) and the area under the precision-recall curve (AUPRC). Secondary outcomes included the contribution of each variable used in the model to the estimation of septic shock and how the value of each variable affected the estimation.

Septic Shock Labeling
Determining whether a patient had septic shock was crucial for labeling from the collected data. Although there is a method to check for a septic shock diagnosis in the EMR, as mentioned earlier, a challenge arose because it was not possible to discern the criteria used for the diagnosis at the time of entry into the EMR due to the mixed nature of the septic shock diagnosis system. Even if it was a genuine septic shock, it could not be ruled out that the diagnostic input was missing in the EMR due to a mistake by the medical staff, or vice versa.

Therefore, we opted to simulate a SIRS-based diagnosis using retrospective medical records. While the SIRS diagnosis and a potential association with infection were ascertainable through retrospective medical record analysis, there was a limitation in determining organ dysfunction, particularly cardiovascular dysfunction, based on retrospective medical records alone. Hence, we defined a modified diagnostic criterion for cardiovascular system dysfunction with the use of inotropes. In other words, continuous intravenous infusion of inotropes in SIRS with a potential association with infection was defined as septic shock. Inotropes included epinephrine, norepinephrine, dopamine, and dobutamine.

Deep Learning
During the hospitalization period, all vital signs measured were preprocessed according to the aforementioned process, and the data were labeled for the presence of septic shock for each timestamped data point. The use of oxygen was coded based on the start and end times of oxygen use. If a vital sign measurement occurred during the period of oxygen use, it was coded as oxygen being administered. Given the expected scarcity of samples in cases of septic shock compared to those without, septic shock and non-septic shock were under-sampled at a ratio of 1:10, respectively, and propensity score matching based on age and sex was employed for effective learning and interpretation. For training and testing the deep learning model, the entire dataset was randomized into a training dataset and a test dataset in an 8:2 ratio, respectively. The deep learning algorithm employed was an artificial neural network, and the network consisted of three hidden layers, each with a dropout rate of 0.3. The learning rate was set at 0.001, and learning was performed over 100 epochs using the stochastic gradient descent optimizer. The input layer comprised 12 variables: age, sex, oxygen supply, CRP, pH, pCO₂, SpO₂, BT, and z-scores of SBP, DBP, HR, and RR. The output layer indicated the presence or absence of the modified SIRS-based septic shock defined earlier. Python version 3.8.10 (Python Software Foundation) and open libraries scikit-learn and PyTorch were employed for deep learning. Shapley additive explanation (SHAP) and matplotlib libraries were used to analyze and visualize the contribution of each variable to the estimation process [21-23].

RESULTS
Baseline Characteristics
During the study period, 10,255,757 measurements were collected, and after applying exclusion criteria, 9,616,115 measurements from 73,911 patients were included. Among the total measurements, 34,696 (0.4%) corresponded to septic shock. After the 1:10 under-sampling process, a total of 381,656 measurements were used for deep learning (Figure 1). The median (interquartile range) age was 4 years (3–7 years), and 53% were male. The baseline characteristics of the datasets used for learning are summarized in Table 1.

Main Outcomes
The evaluation of the septic shock estimation performance of the developed deep learning model yielded excellent results with an AUROC of 0.927 (95% confidence interval [CI], 0.924–0.931) and AUPRC of 0.879 (95% CI, 0.871–0.887), indicating a
**DISCUSSION**

Through this study, we developed a deep learning model to estimate septic shock in children, and the model exhibited high estimation accuracy. However, there are several important implications to consider, such as the absence of a newly established diagnostic system after the SIRS-based septic shock diagnosis system. This study utilized a modified diagnostic system based on SIRS, and these factors should be taken into account.

One key aspect of this study is its operational simplicity and reduced burden on healthcare personnel. The model utilizes a few parameters from the SIRS-based septic shock diagnosis system, providing a less complex yet effective method for the estimation process that does not require extensive efforts from medical staff. Importantly, by inputting a few parameters into the EMR, our model can efficiently estimate septic shock in children with high accuracy. This aligns our study with previous research attempting to use deep learning to estimate pediatric septic shock, especially in the absence of a fully established definition. For example, Liu et al. [24], conducted an analysis using the following four sets of diagnostic criteria for early prediction and risk stratification of septic shock in pediatric sepsis patients: SIRS-based criteria, SOFA-based criteria, pediatric logistic organ dysfunction (PELOD)-2 (2 points), and PELOD-2 (6 points) [25]. Le et al.'s study [26] demonstrated a machine learning-based algorithm's high potential for early...
Figure 2. Septic shock estimation performance of the developed model. (A) The area under the receiver operating characteristics curve (AUROC) graph of the estimated model. (B) The area under the precision-recall curve (AUPRC) graph. CI: confidence interval.

Figure 3. Impact on the model output for each feature of the developed model. pCO$_2$: partial pressure of carbon dioxide; RR: respiratory rate; CRP: C-reactive protein; SBP: systolic blood pressure; DBP: diastolic blood pressure; HR: heart rate; SpO$_2$: pulse oxygen saturation; SHAP: Shapley additive explanation.

detection of severe sepsis in pediatric patients with the definition of severe sepsis based on Goldstein’s criteria [7]. Our study differs by aiming for screening and estimation in hospital general wards, while previous studies focused on early detection in emergency settings [27]. Additionally, while studies on deep learning and septic shock or severe sepsis are prevalent in adults, fewer have been conducted in pediatrics [28]. Fleuren et al.’s research [29], the first to systematically review
Machine learning’s role in predicting adult sepsis, analyzed 28 papers detailing 130 models, demonstrating the accuracy of individual machine learning models in predicting sepsis onset in advance with excellent performance on retrospective data. Similarly, in pediatrics, there is a need for a systematic review that includes machine learning models from various studies.

The relevance of our study is further emphasized by the current landscape of pediatric septic shock diagnostics. Prior research, including initiatives like the Surviving Sepsis Campaign, has explored various diagnostic frameworks beyond SIRS. These studies collectively highlight the heterogeneity of existing diagnostic approaches and the critical need for a more standardized and efficient system, addressing a gap that our study begins to fill.

One intriguing finding in our study pertains to oxygen supply and its impact on the diagnosis of septic shock. Oxygen administration showed a significant impact on diagnosis (high SHAP value) (Figure 3). Notably, within the oxygen supply variable, cases where oxygen was not administered (blue) exhibited high positive SHAP values, while cases where oxygen was administered (red) showed negative SHAP values (Figure 4). In other words, the use of oxygen appeared to decrease the likelihood of being diagnosed with septic shock. While it is plausible for oxygen demand to increase during septic shock, we considered other vital signs, such as RR, HR, and BT, which may fluctuate in different conditions. For instance, during pneumonia, all these vital signs can vary, along with increased demand for oxygen. However, during septic shock, although these vital signs may also fluctuate, the increase in oxygen demand is less pronounced compared to pneumonia, especially in the early stages and in compensated states. Therefore, we decided that oxygen supply alone was not sufficient for diagnosis, but that when integrated with other relevant factors, the need for oxygen supply decreases the likelihood of being diagnosed with septic shock.

Despite these promising aspects, our study has limitations that must be acknowledged for guiding future research. The study was conducted in a single-center setting, potentially

Figure 4. Impact changes on the model output according to the value of each feature of the developed model. It shows that the Shapley additive explanation (SHAP) value varies depending on whether the value of each feature is high (red) or low (blue). In the sex division, women are coded as 0 and men are coded as 1, thus blue means female and red means male. Similarly, for oxygen supply, it is coded as 1 (red) if supplied, and 0 (blue) if not supplied. pCO₂: partial pressure of carbon dioxide; RR: respiratory rate; CRP: C-reactive protein; SBP: systolic blood pressure; DBP: diastolic blood pressure; HR: heart rate; SpO₂: pulse oxygen saturation.
limiting the generalizability of the results to different hospital environments with unique characteristics. Caution is advised when applying these findings to other clinical settings. Our study substitutes certain criteria for organ dysfunction, particularly cardiovascular dysfunction, with the continuous infusion status of inotropes. This substitution introduces the possibility of uncertainty in confirming whether a patient truly had septic shock. As a retrospective observational study, it faces inherent limitations in accurately determining the septic shock status of patients at the time. Furthermore, the lack of a gold standard for pediatric septic shock diagnosis complicates the matter, especially for prospective studies in this area. In our study, the dataset was split in an 8:2 ratio to ensure that the training set did not overlap with the test set. However, this approach did not involve multiple-fold cross-validation or external validation. Additionally, our data are limited to patients from a single center, indicating a need for future well-designed multicenter studies for broader validation and generalization of these findings.

In conclusion, our study represents a step forward in developing a deep learning model capable of efficiently estimating septic shock in pediatric populations, demonstrating good performance. While the model shows strong potential, we recognize the importance of interpreting and applying these findings with caution, considering the inherent limitations of the study.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

FUNDING

This study was conducted with the support of the Seoul National University Industry-University Cooperation Foundation (grant no. 800-20190281).

ACKNOWLEDGMENTS

None.

ORCID

Ji Weon Lee https://orcid.org/0000-0003-4012-0843
Bongjin Lee https://orcid.org/0000-0001-7878-9644

REFERENCES


