Improving sepsis mortality prediction with machine learning: A comparative study of advanced classifiers and performance metrics

*Puyu Zhou[1,A–F], *Jiazheng Duan[2,A–F], Jianqing Li[1,A–F]

[1] Macau University of Science and Technology, China
[2] Dongguan Yumei Photoelectric Co., Ltd., China

A – research concept and design; B – collection and/or assembly of data; C – data analysis and interpretation;
D – writing the article; E – critical revision of the article; F – final approval of the article

Advances in Clinical and Experimental Medicine, ISSN 1899–5276 (print), ISSN 2451–2680 (online) *Adv Clin Exp Med*. 2025;34(8):1393–1402

Address for correspondence
Jianqing Li
E-mail: jqli@must.edu.mo

Funding sources
None declared

Conflict of interest
None declared

* Puyu Zhou and Jiazheng Duan contributed equally to this work.

Received on January 22, 2024
Reviewed on August 9, 2024
Accepted on October 15, 2024

Published online on February 11, 2025

Abstract

Background. High sepsis mortality rates pose a serious global health problem. Machine learning is a promising technique with the potential to improve mortality prediction for this disease in an accurate and timely manner.

Objectives. This study aimed to develop a model capable of rapidly and accurately predicting sepsis mortality using data that can be quickly obtained in an ambulance, with a focus on practical application during ambulance transport.

Materials and methods. Data from the Medical Information Mart for Intensive Care-IV (MIMIC-IV) dataset were used to compare the performance of 11 machine learning algorithms against the widely utilized quick Sequential Organ Failure Assessment (qSOFA) score. A dynamic updating model was implemented. Performance was evaluated using area under the curve (AUC) and precision-recall area under the curve (PRAUC) scores, and feature importance was assessed with SHapley Additive exPlanations (SHAP) values.

Results. The light gradient boosting machine (LightGBM) model achieved the highest AUC (0.79) and PRAUC (0.44) scores, outperforming the qSOFA score (AUC = 0.76, PRAUC = 0.40). The LightGBM also achieved the highest PRAUC (0.44), followed by Optuna_LightGBM (0.43) and random forest (0.42). The dynamically updated and tuned model further improved performance metrics (AUC = 0.79, PRAUC = 0.44) compared to the base model (AUC = 0.76, PRAUC = 0.39). Feature importance analysis offers clinicians insights for prioritizing patient assessments and interventions.

Conclusions. The LightGBM-based model demonstrated superior performance in predicting sepsis-related mortality in an ambulance setting. This study underscores the practical applicability of machine learning models, addressing the limitations of previous research, and highlights the importance of real-time updates and hyperparameter tuning in optimizing model performance.

Key words: machine learning, LightGBM, hyperparameter tuning, sepsis-related mortality, dynamic updating

Cite as
Zhou P, Duan J, Li J. Improving sepsis mortality prediction with machine learning: A comparative study of advanced classifiers and performance metrics. *Adv Clin Exp Med*. 2025;34(8):1393–1402. doi:10.17219/acem/194660

DOI
10.17219/acem/194660

Introduction

Sepsis is a critical global health issue, contributing significantly to worldwide morbidity and mortality. Rapid identification and intervention are crucial for improving patient outcomes because early interventions have been shown to increase survival rates. Developing accurate and efficient diagnostic tools for sepsis is, therefore, essential. Due to the limitations of current medical technologies in diagnosing sepsis from a biomarker perspective, statistics-driven machine learning techniques have gained increasing attention from medical diagnostic researchers. A growing volume of data, including laboratory results, vital signs, genetic, molecular, and clinical data, as well as patient health histories, is available in high resolution for high-risk individuals and sepsis patients.[1–5] Gradient boosting trees (GBT) are among the most widely used machine learning methods,[6] followed by logistic regression,[6,7] random forest (RF),[8] ridge regression,[9] lasso regression (LR),[10] naïve Bayes (NB),[11] K-nearest neighbors (KNN),[12–14] gated recurrent units,[15,16] and long short-term memory.[17] For hyperparameter tuning, 2 studies used grid search methods,[18,19] and 2 use Bayesian optimization methods.[20,21] Amrollahi et al.[22] and Zhang et al.[23] compared their algorithms to scoring systems used in clinical practice, such as systemic inflammatory response syndrome (SIRS), Sequential Organ Failure Assessment (SOFA), quick SOFA (qSOFA), Modified Early Warning System (MEWS), or targeted real-time Early Warning System (TREWScore).

However, the aforementioned papers have some shortcomings. Some studies have relatively small sample sizes, which may affect the generalizability of the results, such as the studies by Delahanty et al.[24] and Hammoud et al.[25] In some studies, the proposed models were not adequately validated. For example, the study by Culliton et al.[26] did not provide detailed information about the model's performance on an independent test set. Some studies used different feature selection methods, making it difficult to fairly compare the performance differences between methods, such as the studies by Goh et al.[3] and Qin et al.[27] In some studies, multiple algorithms were compared without explaining why these particular algorithms were chosen, such as in the studies by Horng et al.[28] and Apostolova and Velez.[29] When using machine learning methods, several papers did not provide detailed information about hyperparameter tuning, such as the study by Amrollahi et al.[22] Other publications compared their algorithms to existing clinical scoring systems but did not elaborate on the limitations of these scoring systems. For example, the study by Delahanty et al.[24] compared their algorithm to SIRS, SOFA, qSOFA, MEWS, and TREWScore, without discussing the shortcomings of these scoring systems in detail. Additionally, racial, age and sex differences in the study participants were not adequately considered in some research, which may affect the predictive ability of the model.

Differences exist among various populations. For example, in the study by Liu et al.,[30] natural language processing methods were used to predict sepsis, but the demographic characteristics of the study participants were not discussed in detail. Some studies do not fully address the feasibility and practicality of the models in clinical practice. For example, in the studies by Goh et al.[3] and Qin et al.,[27] although their models demonstrate high accuracy in predicting sepsis, challenges related to data acquisition, data processing and model deployment in actual clinical applications were not addressed. Additionally, some studies did not thoroughly discuss the handling methods and potential issues associated with different data types. For example, in the study by Johnson et al.,[31] various types of data (such as laboratory, vital signs, genetic, molecular, and clinical data) were mentioned, but detailed integration methods and potential problems were not discussed. Furthermore, some studies did not explicitly point out the advantages and areas for improvement of machine learning methods in sepsis prediction. For example, Hammoud et al.[25] used LR for prediction but did not discuss the advantages and limitations of this method compared to others. The choice of evaluation metrics may also affect the interpretation of results. For example, Horng et al.[28] used multiple evaluation metrics but did not discuss the relationships between these metrics and their applicability in assessing model performance. Finally, some studies did not consider the baseline risk characteristics of patients, which may influence the accuracy and applicability of model predictions. For example, Apostolova and Velez[29] did not discuss the patient's baseline risk characteristics and their impact on model predictions.

In this study, we aimed to develop a model that can quickly and accurately predict sepsis mortality using data that can be rapidly obtained during ambulance transport. Using the Medical Information Mart for Intensive Care (MIMIC)-IV dataset, we compared the performance of 11 machine learning algorithms and benchmarked our results against the widely used qSOFA score. To enhance the clinical applicability of our model, we carefully selected features that are feasible to collect during ambulance transport, addressed demographic differences among patients, and considered baseline risk characteristics that could impact predictive performance. Additionally, we emphasized the importance of the precision-recall area under the curve (PRAUC) metric for evaluating models on imbalanced datasets, a common challenge in sepsis research. To further improve the predictive capabilities of our clinical decision support system, we implemented a real-time updating model, leveraging both online and incremental learning approaches to dynamically incorporate new patient data.

Objectives

Our research contributes to the practical application of machine learning methods for predicting sepsis-related mortality.

Materials and methods

The data used in this study were derived from the MIMIC-IV database, a comprehensive de-identified database that provides intensive care data from Beth Israel Deaconess Medical Center (BIDMC; Boston, USA).[31] The database contains information on over 40,000 patients who were admitted to the intensive care units (ICUs) at BIDMC between 2008 and 2019. It adopts a modular approach to data organization, emphasizing data sources and enabling the separation and combination of different data types. The dataset includes 'hosp' modules and 'icu' modules. The 'hosp' module contains data from electronic health records throughout the hospital. These measurements are mainly recorded during hospitalization, though some tables also include data from outpatient sources (e.g., outpatient laboratory tests). Patient demographics (patients), hospitalizations (admissions) and within-hospital transfers (transfers) are recorded in the 'hosp' module. The 'icu' module contains data from BIDMC's clinical information system, MetaVision (iMDSoft).[31] Necessary training data samples have been completed, with a record ID of 49953233.

Ambulance measurable features

The features in the MIMIC-IV dataset (Table 1) can be rapidly measured during ambulance transport. Early assessment and intervention by ambulance teams are crucial for improving patient outcomes and minimizing complications. Real-time monitoring of vital signs and laboratory indicators allows emergency medical providers to assess the patient's condition promptly and administer appropriate treatment. This approach can improve patient survival rates, shorten hospital stays and reduce medical costs.[32,33]

Methods

Features

The features in the dataset are mainly divided into 2 types: numerical and categorical. Numerical features are represented by real numbers, such as age, heart rate (HR) and blood pressure, and usually require scaling (normalization or standardization). Categorical features consist of fixed categories, such as sex, ethnicity and marital status, and must usually be encoded (e.g., one-hot encoding) to convert them into numerical form. To distinguish between numerical and categorical features in the dataset, we defined 2 separate lists and applied different preprocessing steps to them. Missing values in numerical features were filled with the median, whereas missing values in categorical features were filled with the mode. For outlier handling, HR values were limited to the range of 30–200 bpm.

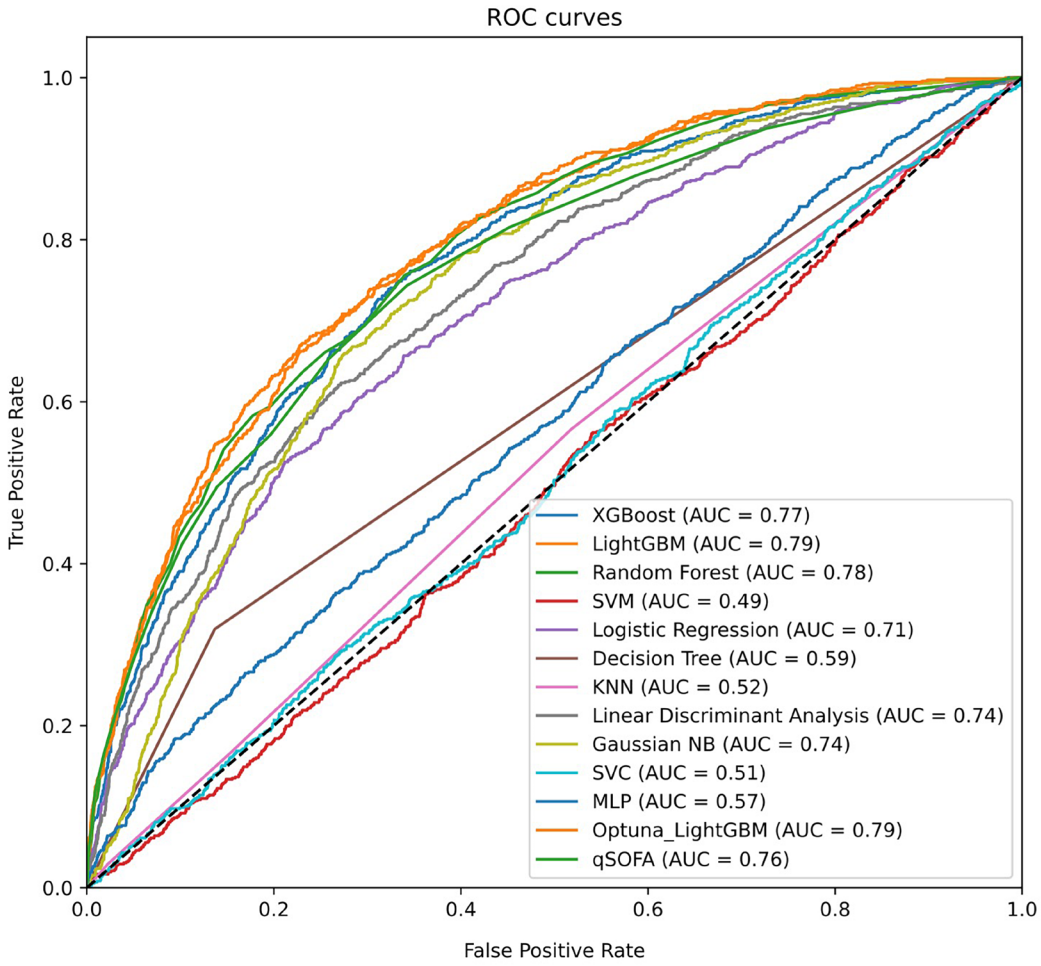
Although both feature and label values after one-hot encoding were represented as 0/1, their meanings differed. Features after one-hot encoding indicated the presence of a particular category, whereas label values represented the survival status of the patient. During model training, the algorithm attempts to combine all features to predict the target variable. If concerns arise about the potential negative impact of one-hot encoding on predictions, alternative encoding methods, such as ordinal encoding or target encoding, may be considered. However, in most cases, one-hot encoding is an effective method for encoding categorical features.

Statistical analyses

When the dataset contains many categorical features, each with multiple categories, one-hot encoding can result in a significant increase in data dimensionality. In such cases, dimensionality reduction methods (e.g., principal

Table 1. Features that can be measured quickly in an ambulance from the Medical Information Mart for Intensive Care-IV (MIMIC-IV) dataset

Category	Feature	Meaning
Vital signs	heart rate (HR)	An abnormal HR may indicate deterioration of the patient's condition.
	systolic blood pressure (SBP)	Abnormal blood pressure may be a sign of septic shock.
	diastolic blood pressure (DBP)	?
	respiratory rate (RR)	An increased RR may be a sign of complications such as hypoxemia.
	temperature (temp)	Abnormal temperature may be a sign of infection.
	oxygen saturation (SpO_2)	Low oxygen saturation may indicate hypoxemia.
Laboratory indicators	white blood cell count (WBC)	Abnormal WBC count may indicate infection.
	lactic acid	Elevated lactic acid levels may be associated with septic shock.
	C-reactive protein (CRP)	Elevated CRP levels during the acute-phase response may indicate infection.
	creatinine	Elevated creatinine levels may be a sign of renal impairment.
	liver function indicators (e.g., ALT, AST)	Abnormal liver function indicators may indicate liver damage.
Patient demographics	age	Age may be an important factor in determining the patient's health status.
	sex	Different sexes may have different risks for certain diseases, such as cardiovascular diseases or cancers.



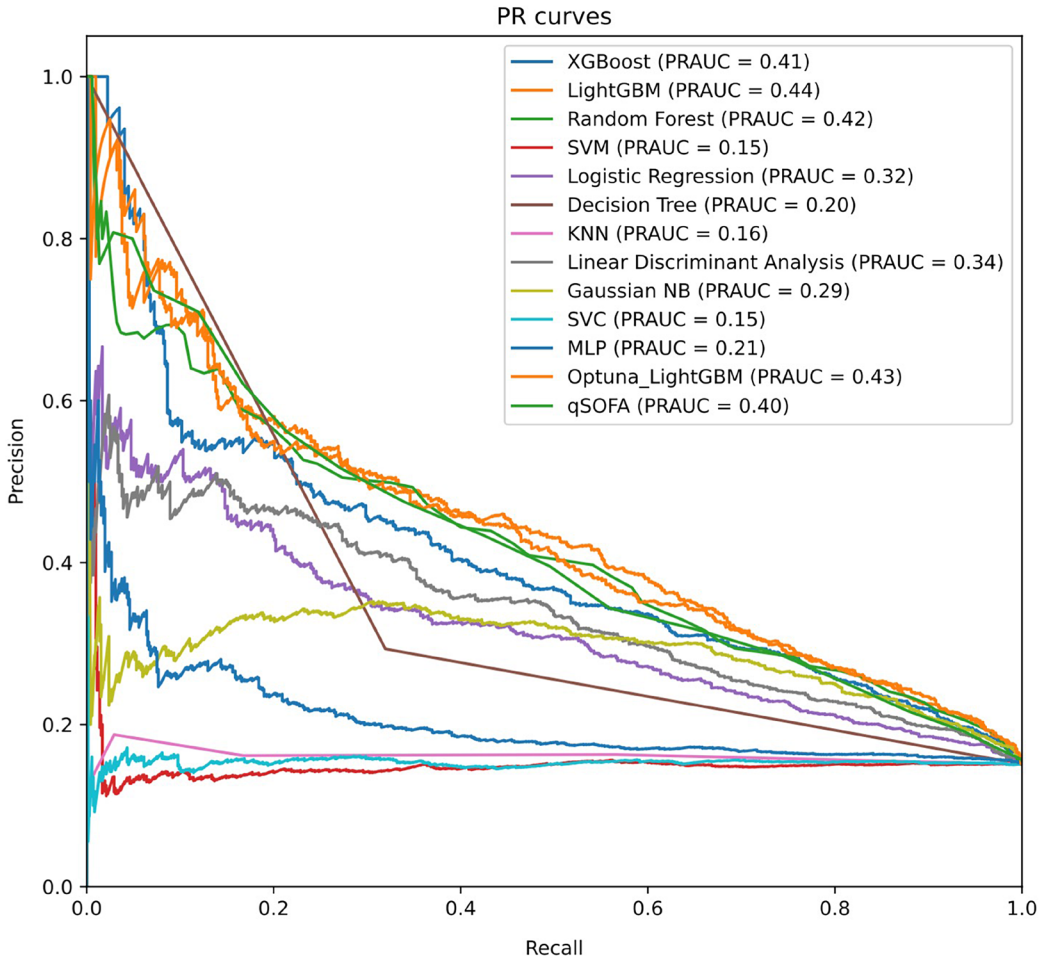


Fig. 1. Performance comparison of machine learning algorithms for sepsis-related mortality prediction. This figure presents the area under the curve (AUC) and precision-recall area under the curve (PRAUC) scores for 11 machine learning algorithms and the quick Sequential Organ Failure Assessment (qSOFA) score evaluated in our study. The LightGBM algorithm achieved the highest AUC score (0.79) and PRAUC score (0.44), outperforming both the qSOFA score (AUC = 0.76, PRAUC = 0.40) and other machine learning models. These results highlight the superior performance of the LightGBM-based model for predicting sepsis-related mortality in an ambulance setting

component analysis) or feature selection methods can be used to reduce the number of features.

In clinical applications, identifying minority samples is often more valuable than identifying majority samples, which is the focus of classifier construction. However, current machine learning models for predicting sepsis mortality are mainly designed to maximize overall classification accuracy, which limits their ability to effectively identify minority samples. Therefore, we included both PRAUC and area under the curve (AUC) as indicators to select the optimal algorithm.

We used several algorithms to classify the dataset, including XGBoost, light gradient boosting machine (LightGBM) (https://github.com/microsoft/LightGBM), RF, support vector machine, logistic regression, decision tree, KNN, linear discriminant analysis, Gaussian NB, support vector classification, and multilayer perceptron.

The purpose of this study was to analyze specific datasets. The chosen algorithm may not always be the one considered best for general use, but rather the most appropriate one for the specific data at hand. In addition to the aforementioned algorithms, we introduced an updating model strategy to improve the predictive performance of our clinical decision support system. We explored both online learning and incremental learning approaches to dynamically update the model as new patient data became available. This updating strategy enables the model to adapt to changes in patient physiological parameters, providing more accurate predictions for clinicians.

The updating model involves retraining the classifier with new data, which can be achieved by either retraining the model from scratch or updating the existing model with the new data. For algorithms that support incremental learning (e.g., XGBoost, LightGBM), we employed an incremental learning approach. For algorithms that do not support incremental learning, we used an online learning approach, which involves retraining the model from scratch using the updated dataset.

In summary, our methodology involved preprocessing the dataset, selecting the optimal algorithm based on PRAUC and AUC indicators, and implementing an updating model strategy to ensure the model's performance remains accurate and relevant in the face of changing patient data.

The underlying hypothesis for our methodology is the variance contribution hypothesis, which posits that the variance of the data is mainly contributed by a few principal components (i.e., most of the information can be summarized by a smaller number of composite variables). This hypothesis is tested by evaluating the cumulative variance contribution ratio.

Results

The LightGBM algorithm achieved the highest AUC score of 0.79, followed closely by RF (AUC = 0.78) and XGBoost (AUC = 0.77), as shown in Fig. 1, which provides a comprehensive comparison of the AUC and PRAUC scores for all evaluated models. The qSOFA score had an AUC of 0.76, demonstrating that the LightGBM-based model outperformed traditional methods in terms of discriminatory ability. In terms of PRAUC, LightGBM also achieved the highest score (0.44), followed by Optuna_LightGBM (0.43) and RF (0.42). The qSOFA score had a PRAUC of 0.40, demonstrating the advantage of our proposed model in detecting the positive class in the presence of class imbalance.

Despite involving hyperparameter tuning, the Optuna_LightGBM model did not outperform the default LightGBM model in terms of AUC and PRAUC scores.

By calculating the SHAP values of the LightGBM model, as shown in Fig. 2, we ranked the features by importance in descending order: maximum blood urea nitrogen, patient age at admission, maximum HR, minimum mean arterial pressure, minimum blood glucose, patient ethnicity, maximum blood sodium concentration, minimum respiratory rate, maximum respiratory rate, maximum blood creatinine, minimum blood urea nitrogen, minimum HR, minimum blood sodium concentration, minimum blood creatinine, minimum white blood cell count, minimum hematocrit, maximum blood glucose, maximum mean arterial pressure, maximum white blood cell count, and maximum hematocrit. These features are ranked according to their contribution to the model's predictions.

In this study, we implemented a real-time updating model to improve the predictive performance of our clinical decision support system. We explored both online learning and incremental learning approaches to dynamically update the model as new patient data became available. The incremental GBT, which was compatible with our current LightGBM model, was used as the incremental learning method.

To evaluate the performance of our dynamically updated model, we used sliding window validation and other time series validation methods. The results show that the dynamically updated model better adapts to changes in patient physiological parameters, providing more accurate predictions for clinicians.

We also addressed the issue of overfitting by incorporating regularization, reducing model complexity, employing early stopping, and utilizing additional data when available. These techniques help prevent overfitting while ensuring optimal model performance with the available data.

Additionally, we conducted hyperparameter tuning using grid search, focusing on key parameters such as "num_leaves", "feature_fraction", "bagging_fraction", "bagging_freq", and "learning_rate". We then compared the tuned model to the base model in terms of AUC and PRAUC.

The results showed that the tuned model with dynamic updating outperformed the base model, achieving an AUC of 0.79 (compared to 0.76 for the base model) and a PRAUC

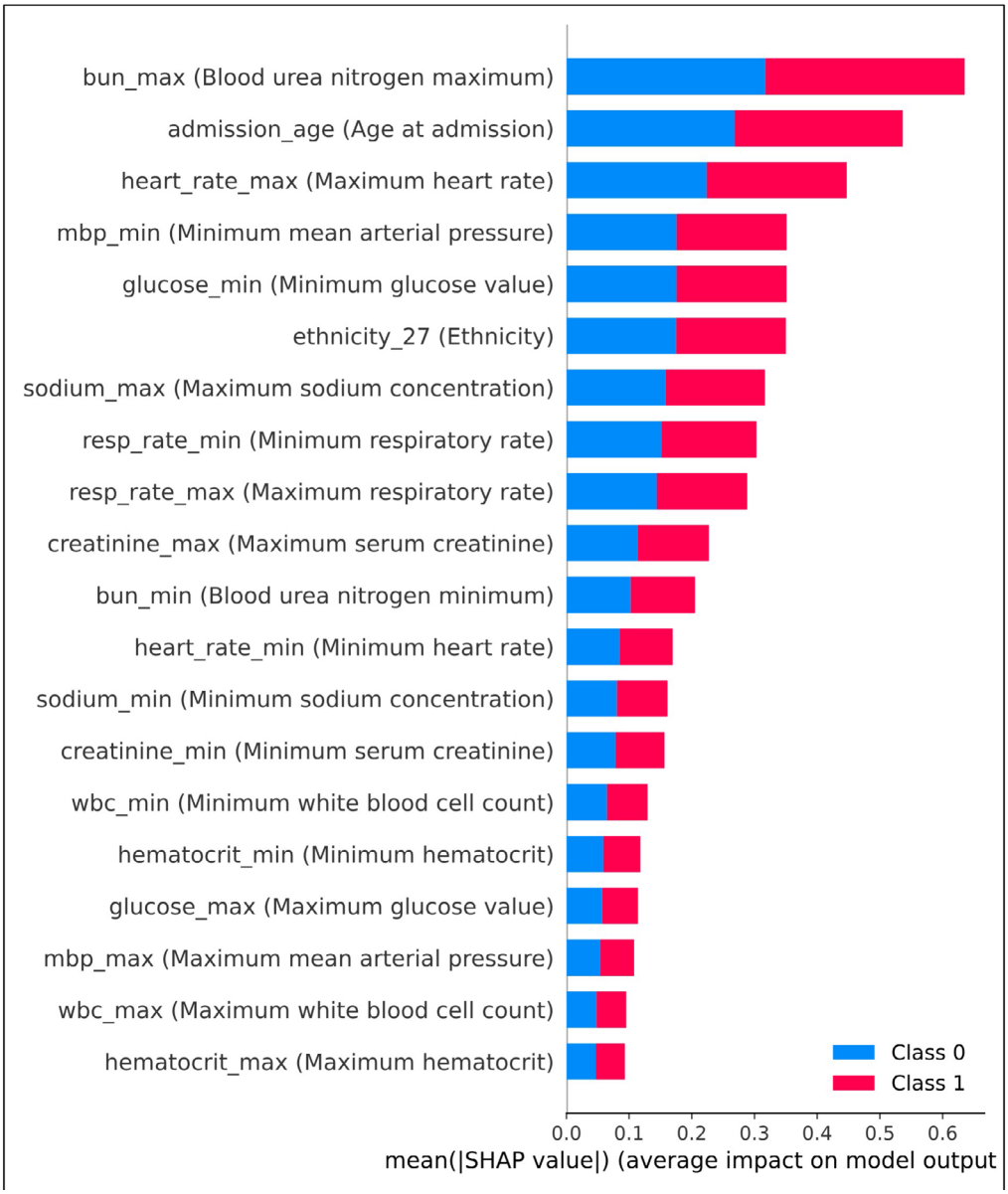


Fig. 2. Feature importance ranking for the LightGBM model using SHAP values. This figure presents a horizontal bar chart illustrating the relative importance of the top 20 features, ranked in descending order, for predicting sepsis-related mortality using the LightGBM model. The x-axis represents the SHAP (SHapley Additive exPlanations) values, reflecting each feature's contribution to the model's prediction. The y-axis lists the features, with the most important feature (maximum blood urea nitrogen) at the top and the least important feature (maximum hematocrit) at the bottom. This ranking helps healthcare professionals prioritize physiological indicators and focus on critical factors when assessing a patient's condition during ambulance transport

of 0.44 (compared to 0.39 for the base model), as illustrated in Fig. 3, which highlights the advantages of real-time updates and hyperparameter tuning.

Discussion

Addressing the challenges posed by imbalanced classification problems, this study applies machine learning algorithms to improve sepsis mortality prediction. Our results show that the LightGBM algorithm outperformed other classifiers in this context, largely due to its ability to handle imbalanced datasets and its efficient computational properties. Furthermore, we found that the PRAUC was a more appropriate evaluation metric for imbalanced classification problems because it better reflected the performance changes of the classifier at different thresholds and provided insight into the trade-off between precision and recall.

The Optuna_LightGBM did not outperform the default LightGBM model in terms of AUC and PRAUC scores. This may be because the default hyperparameters of LightGBM were already well optimized for this specific problem, and additional tuning did not result in a significant improvement. Moreover, hyperparameter tuning may introduce the risk of overfitting, which could potentially limit the generalizability of the model.

The superior performance of LightGBM over other classifiers can be attributed to its ability to handle large-scale data, high-dimensional features and class imbalance more effectively. As a gradient boosting framework, LightGBM is known for its efficiency and scalability, making it well-suited for the complex nature of sepsis-related mortality prediction.

These results are of great value for doctors using the LightGBM model to predict sepsis mortality in patients transported using ambulances. First, doctors can

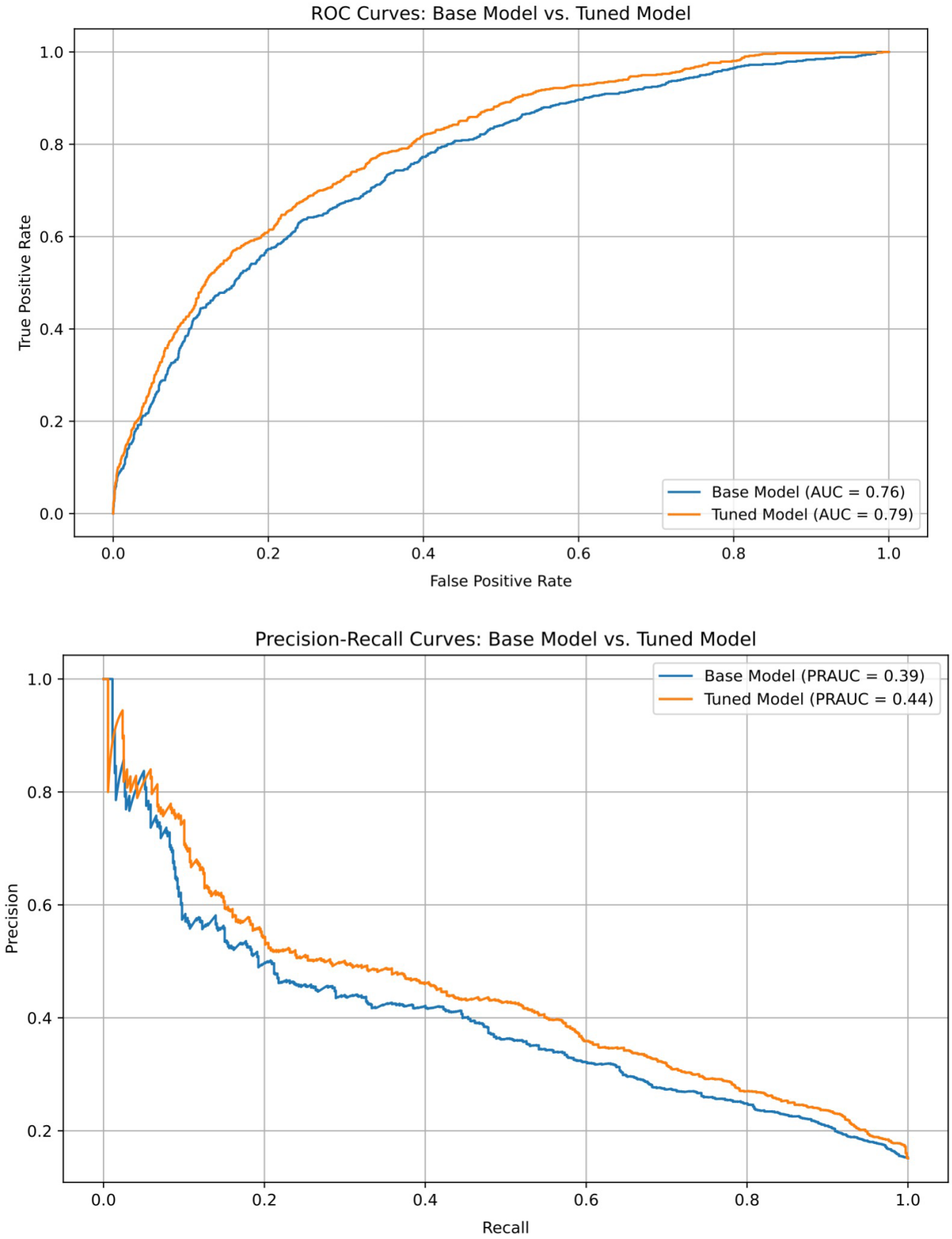


Fig. 3. Performance comparison of the dynamically updated model and the base model. The dynamically updated model achieves higher area under the curve (AUC) (0.79) and precision-recall area under the curve (PRAUC) (0.44) scores compared to the base model (AUC = 0.76, PRAUC = 0.39), demonstrating the advantages of real-time updates and hyperparameter tuning in improving the model's predictive performance for sepsis-related mortality in an ambulance setting

understand which physiological indicators play a more significant role in mortality prediction, based on feature rankings, allowing them to prioritize attention to these indicators. Second, the feature rankings help doctors quickly assess the patient's condition in the ambulance and take appropriate intervention measures based on the model's predictions.

The LightGBM demonstrates superior performance in our experiment for several reasons:

Data structure: As a gradient boosted trees-based algorithm, LightGBM employs an optimized feature histogram method that handles large-scale datasets and high-dimensional features more efficiently. Additionally, it incorporates gradient-based one-side sampling and exclusive feature bundling techniques to reduce memory consumption and computational complexity, enabling faster convergence while maintaining high accuracy.

Model structure: LightGBM uses a leaf-wise growth strategy that reduces the risk of overfitting compared to traditional level-wise growth strategies. This approach focuses on fitting the training data by splitting the leaf with the highest gain.

Regularization strategy: LightGBM implements effective regularization strategies, including L1 and L2 regularization, along with parameters for maximum tree depth, minimum leaf node weight and minimum split gain. These strategies control model complexity, prevent overfitting and enhance the classifier's generalization ability.

Ability to handle class imbalance: LightGBM includes built-in mechanisms for addressing class imbalance, such as automatic class weight adjustment using the class_weight parameter and adjusting the weights of positive and negative samples via the scale_pos_weight parameter. These mechanisms help the training process focus on improving the prediction performance of minority classes, resulting in better classification results.

The PRAUC is considered a more appropriate evaluation metric for imbalanced classification problems because it emphasizes the classifier's performance in predicting negative samples, such as deaths in sepsis mortality prediction. Compared to other metrics, such as AUC and F1 scores, PRAUC provides a more accurate and reliable assessment of the classifier's performance in imbalanced datasets. This is due to the following reasons:

Unaffected by the number of negative samples: PRAUC is not influenced by the number of negative samples, unlike the AUC of the receiver operating characteristic (ROC) curve, which considers both the true positive rate (recall) and false positive rate. The latter is heavily affected by the number of negative samples.

Focus on small probability events: PRAUC is more suited for cases where the main concern is correctly detecting positive samples, especially those with a small probability. Precision and recall rates provide a better reflection of model performance in these situations compared to other evaluation metrics.

Trade-off between precision and recall: PRAUC emphasizes both precision and recall, allowing for the identification of an optimal balance point to achieve the best trade-off between these 2 metrics.

In our study, we also explored the performance of updated models by incorporating additional features, fine-tuning hyperparameters and employing ensemble techniques. These updated models aim to further enhance prediction accuracy and generalization ability for sepsis mortality.

Additional features: By integrating relevant clinical data, such as laboratory test results, vital signs and comorbidities, our updated models can capture a more comprehensive view of the patient's condition, which may contribute to a better understanding of the underlying risk factors associated with sepsis mortality.

Hyperparameter tuning: We performed a systematic search for optimal hyperparameters using techniques such as grid search and random search. These methods help our updated models achieve better performance by optimizing their configurations.

Ensemble techniques: By combining predictions from multiple base models, we employed ensemble techniques such as bagging, boosting and stacking. These techniques aim to reduce overfitting, increase model stability and improve overall performance by leveraging the strengths of different base models.

Limitations

The updated models showed improved performance compared to the initial models, indicating that incorporating additional features, fine-tuning hyperparameters and using ensemble techniques can enhance the prediction accuracy of sepsis mortality. However, further research is needed to explore other potential factors influencing sepsis mortality predictions and to investigate novel machine learning algorithms and techniques for continuously improving model performance.

Conclusions

This study addressed the limitations and gaps in the existing literature on predicting sepsis-related mortality using machine learning models, focusing on their practical application in an ambulance setting. We demonstrated that the LightGBM-based model outperformed other classifiers and the qSOFA score. By implementing a dynamic updating model and fine-tuning hyperparameters, we further enhanced the model's performance, resulting in more accurate and reliable predictions for clinicians.

Our findings significantly contribute to the practical application of machine learning models in the medical field, particularly for predicting sepsis-related mortality

in an ambulance setting. The feature importance analysis provides valuable insights that help doctors prioritize patient assessment and interventions, ultimately improving patient outcomes. This research demonstrates the potential of real-time updating and hyperparameter tuning to further optimize the performance and clinical utility of sepsis-related mortality prediction models in real-world ambulance settings.

Supplementary data

The Supplementary materials are available at https://doi.org/10.5281/zenodo.14644757. The package includes the following files:

Supplementary Table 1. Training data for ASPII.
Supplementary Table 2. Training data for SOFA.
Supplementary Table 3. Testing data for sepsis.

Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Use of AI and AI-assisted technologies

Not applicable.

ORCID iDs

Puyu Zhou https://orcid.org//0000-0002-2361-2101
Jiazheng Duan https://orcid.org/0009-0009-3805-3937
Jianqing Li https://orcid.org/0000-0002-6768-1483

References

1. Fleischmann C, Scherag A, Adhikari NKJ, et al. Assessment of global incidence and mortality of hospital-treated sepsis: Current estimates and limitations. *Am J Respir Crit Care Med*. 2016;193(3):259–272. doi:10.1164/rccm.201504-0781OC
2. Fleuren LM, Klausch TLT, Zwager CL, et al. Machine learning for the prediction of sepsis: A systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med*. 2020;46(3):383–400. doi:10.1007/s00134-019-05872-y
3. Goh KH, Wang L, Yeow AYK, et al. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nat Commun*. 2021;12(1):711. doi:10.1038/s41467-021-20910-4
4. Valik JK, Ward L, Tanushi H, et al. Predicting sepsis onset using a machine learned causal probabilistic network algorithm based on electronic health records data. *Sci Rep*. 2023;13(1):11760. doi:10.1038/s41598-023-38858-4
5. Yong L, Zhenzhou L. Deep learning-based prediction of in-hospital mortality for sepsis. *Sci Rep*. 2024;14(1):372. doi:10.1038/s41598-023-49890-9
6. Brownlee J. A gentle introduction to the gradient boosting algorithm for machine learning. San Juan, Puerto Rico: Machine Learning Mastery; 2020. https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning. Accessed August 21, 2024.
7. Bonte C, Vercauteren F. Privacy-preserving logistic regression training. *BMC Med Genomics*. 2018;11(Suppl 4):86. doi:10.1186/s12920-018-0398-y
8. Wang L, ed. *Support Vector Machines: Theory and Applications*. Vol. 177. Studies in Fuzziness and Soft Computing. Berlin–Heidelberg, Germany: Springer Berlin Heidelberg; 2005. doi:10.1007/b95439
9. Rokem A, Kay K. Fractional ridge regression: A fast, interpretable reparameterization of ridge regression. *GigaScience*. 2020;9(12):giaa133. doi:10.1093/gigascience/giaa133
10. Lee JH, Shi Z, Gao Z. On LASSO for predictive regression. *J Econometrics*. 2022;229(2):322–349. doi:10.1016/j.jeconom.2021.02.002
11. Anand MV, KiranBala B, Srividhya SR, Kavitha C, Younus M, Rahman MH. Gaussian naïve Bayes algorithm: A reliable technique involved in the assortment of the segregation in cancer. *Mobile Information Systems*. 2022;2022:1–7. doi:10.1155/2022/2436946
12. Borgohain O, Dasgupta M, Kumar P, Talukdar G. Performance analysis of nearest neighbor, K-nearest neighbor and weighted K-nearest neighbor for the classification of Alzheimer disease. In: Borah S, Pradhan R, Dey N, Gupta P, eds. *Soft Computing Techniques and Applications*. Vol. 1248. Advances in Intelligent Systems and Computing. Singapore: Springer Singapore; 2021:295–304. doi:10.1007/978-981-15-7394-1_28
13. Negi A, Hajati F. Analysis of variants of KNN for disease risk prediction. In: Barolli L, Hussain F, Enokido T, eds. *Advanced Information Networking and Applications*. Vol. 451. Lecture Notes in Networks and Systems. Cham, Switzerland: Springer International Publishing; 2022:531–545. doi:10.1007/978-3-030-99619-2_50
14. Uddin S, Haque I, Lu H, Moni MA, Gide E. Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Sci Rep*. 2022;12(1):6256. doi:10.1038/s41598-022-10358-x
15. Agarap AF. A neural network architecture combining gated recurrent unit (GRU) and support vector machine (SVM) for intrusion detection in network traffic data [preprint posted online September 10, 2017]. *arXiv*. doi:10.48550/ARXIV.1709.03082
16. Erichson NB, Lim SH, Mahoney MW. Gated recurrent neural networks with weighted time-delay feedback [preprint posted online December 1, 2022]. arXiv. doi:10.48550/arXiv.2212.00228. Accessed October 15, 2024.
17. Li Q, Kamaruddin N, Yuhaniz SS, Al-Jaifi HAA. Forecasting stock prices changes using long-short term memory neural network with symbolic genetic programming. *Sci Rep*. 2024;14(1):422. doi:10.1038/s41598-023-50783-0
18. Jiang X, Xu C. Deep learning and machine learning with grid search to predict later occurrence of breast cancer metastasis using clinical data. *J Clin Med*. 2022;11(19):5772. doi:10.3390/jcm11195772
19. Magalhães MMDC. Hyperparameter fine tuning for a time series forecasting model [doctoral thesis]. Carcavelos, Portugal: Nova School of Business and Economics. 2022.
20. Tomlinson G, Al-Khafaji A, Conrad SA, et al. Bayesian methods: A potential path forward for sepsis trials. *Crit Care*. 2023;27(1):432. doi:10.1186/s13054-023-04717-x
21. Zhao QY, Liu LP, Luo JC, et al. A machine-learning approach for dynamic prediction of sepsis-induced coagulopathy in critically ill patients with sepsis. *Front Med*. 2021;7:637434. doi:10.3389/fmed.2020.637434
22. Amrollahi F, Shashikumar SP, Razmi F, Nemati S. Contextual embeddings from clinical notes improves prediction of sepsis. *AMIA Annu Symp Proc*. 2020;2020:197–202. PMID:33936391. PMCID:PMC8075484.
23. Zhang Y, Xu W, Yang P, Zhang A. Machine learning for the prediction of sepsis-related death: A systematic review and meta-analysis. *BMC Med Inform Decis Mak*. 2023;23(1):283. doi:10.1186/s12911-023-02383-1
24. Delahanty RJ, Alvarez J, Flynn LM, Sherwin RL, Jones SS. Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis. *Ann Emerg Med*. 2019;73(4):334–344. doi:10.1016/j.annemergmed.2018.11.036
25. Hammoud I, Ramakrishnan I, Henry M, Morley E. Multimodal early septic shock prediction model using LASSO regression with decaying response. In: *2020 IEEE International Conference on Healthcare Informatics (ICHI)*. Oldenburg, Germany: IEEE; 2020:1–3. doi:10.1109/ICHI48887.2020.9374377
26. Culliton P, Levinson M, Ehresman A, Wherry J, Steingrub JS, Gallant SI. Predicting severe sepsis using text from the electronic health record [preprint posted online November 30, 2017]. *arXiv*. doi:10.48550/ARXIV.1711.11536

27. Qin F, Madan V, Ratan U, et al. Improving early sepsis prediction with multimodal learning [preprint posted online July 23, 2021]. arXiv. doi:10.48550/arXiv.2107.11094. Accessed October 15, 2024.
28. Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PLoS One*. 2017;12(4):e0174708. doi:10.1371/journal.pone.0174708
29. Apostolova E, Velez T. Toward automated early sepsis alerting: Identifying infection patients from nursing notes [preprint posted online September 11, 2018] *arXiv*. doi:10.48550/arXiv.1809.03995
30. Liu R, Greenstein JL, Sarma SV, Winslow RL. Natural language processing of clinical notes for improved early prediction of septic shock in the ICU. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Berlin, Germany: IEEE; 2019:6103–6108. doi:10.1109/EMBC.2019.8857819
31. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3(1):160035. doi:10.1038/sdata.2016.35
32. Báez AA, Hanudel P, Perez MT, Giraldez EM, Wilcox SR. Prehospital Sepsis Project (PSP): Knowledge and attitudes of United States advanced out-of-hospital care providers. *Prehosp Disaster Med*. 2013;28(2):104–106. doi:10.1017/S1049023X12001744
33. Studnek JR, Artho MR, Garner CL, Jones AE. The impact of emergency medical services on the ED care of severe sepsis. *Am J Emerg Med*. 2012;30(1):51–56. doi:10.1016/j.ajem.2010.09.015