

Application of Machine Learning Techniques in Sepsis Mortality Prediction

Mahyar Mohammadian, Parsa Haghighatgoo, Salar Rahnama, Hosein Khajeh, Somayeh Afrasiabi
Department of CSE&IT, School of ECE, Shiraz University, Shiraz, Iran, mhn.mahyarmohammadian1383@gmail.com
Department of CSE&IT, School of ECE, Shiraz University, Shiraz, Iran, prsahaghighatgoo@gmail.com
Department of CSE&IT, School of ECE, Shiraz University, Shiraz, Iran, salarrahnama0001@gmail.com
Department of Computer Engineering, Yasuj University, Yasuj, Iran, ho3ein.khajeh8282@gmail.com
Department of CSE&IT, School of ECE, Shiraz University, Shiraz, Iran, afrasiabi@shirazu.ac.ir

Abstract—Sepsis, a life-threatening condition resulting from an infection, remains one of the leading causes of mortality in critically ill patients. Early identification of high-risk patients is essential for timely intervention, which can significantly improve survival rates. This study explores the use of machine learning (ML) models, specifically XGBoost, Random Forest, and a Stacked ensemble model, for predicting sepsis-related mortality using the MIMIC-III critical care database. The research compares the performance of these models based on accuracy and AUC-ROC scores. After hyperparameter tuning, XGBoost achieved the highest performance with an AUC of 83.96% and an accuracy of 85.29%. This paper demonstrates the potential of machine learning models as effective tools for supporting clinical decisions in sepsis mortality prediction.

Keywords—Sepsis, Machine Learning Models, MIMIC III, Mortality Rate, Prediction

Introduction

Sepsis, defined as life-threatening organ dysfunction caused by a dysregulated host response to infection, remains a global health burden. Recent estimates attribute approximately 20% of annual global deaths to sepsis, with significant variations in mortality rates across regions [1][2][3]. The World Health Organization (WHO) has emphasized the need for improving sepsis management and diagnosis, making it a global health priority [4][5]. Traditional risk scoring systems such as the Simplified Acute Physiology Score II (SAPS II), Sequential Organ Failure Assessment (SOFA), and Acute Physiology and Chronic Health Evaluation II (APACHE II) are widely used for mortality prediction in sepsis. However, these models often exhibit limited predictive power due to their reliance on linear relationships and a small number of features [6][7][8]. Machine learning (ML) offers a promising alternative for predicting sepsis mortality, as it can analyze complex, non-linear relationships in large datasets. The MIMIC-III database has been instrumental in enabling such research by providing comprehensive electronic health records from ICU patients [9][10]. By leveraging ML models such as Random Forest, XGBoost, and ensemble techniques, researchers can improve the predictive accuracy of sepsis-related outcomes [11][12]. This study explores the use of three machine learning techniques—XGBoost, Random Forest (RF), and a stacked ensemble model of XGBoost, RF, and Logistic Regression—

and Critical clinical features such as urine output, lactate, blood urea nitrogen (BUN), systolic blood pressure, and age were selected based on their established association with sepsis outcomes [13][14] to predict the 30-day mortality of sepsis patients from the MIMIC-III database, a comprehensive dataset of ICU patient records. For example, G. Kong et al. [28] evaluated five models, including Lasso, RF, GBM, LR, and SAPS II, achieving the best AUC of 84.5% with the GBM method. This study's XGBoost model demonstrated superior performance with the best AUC of 84.75%, highlighting its potential for clinical decision support in sepsis management.

I. BACKGROUND

Sepsis is a severe medical condition characterized by a dysregulated immune response to infection, leading to multi-organ dysfunction and high mortality rates [1][3][15]. Despite advances in clinical care, sepsis outcomes remain poor, especially in resource-limited settings [16]. Globally, the incidence of ICU-treated sepsis has increased, emphasizing the need for early detection and effective management strategies [17][18]. Traditional prognostic models like APACHE II, SOFA, and SAPS II rely on clinical and laboratory data to estimate mortality risk. Although these tools are valuable, their accuracy is constrained by linear assumptions and limited variables, reducing their effectiveness in heterogeneous populations [7][9][19]. Recent studies have underscored the limitations of these tools, prompting the exploration of data-driven approaches for mortality prediction [20][21]. Machine learning approaches have demonstrated remarkable potential in addressing the complexities of sepsis. By integrating large datasets and identifying non-linear patterns, ML models can improve predictive performance compared to traditional scoring systems [11][22]. Studies utilizing the MIMIC-III database have shown that advanced ML techniques outperform traditional models in mortality prediction. For example, Mu et al. used ML to predict mortality in sepsis-associated acute respiratory distress syndrome (ARDS), achieving AUC scores exceeding 0.80 [23]. Similarly, Wang et al. explored respiratory support strategies in septic patients, emphasizing the importance of individualized interventions [24]. Gu et al. developed a nomogram incorporating ML for 28-day mortality prediction

in sepsis patients with coronary artery disease, demonstrating its clinical utility [25].

This study evaluates the predictive performance of three ML models—XGBoost, Random Forest, and a stacking ensemble—in predicting 30-day sepsis mortality. By focusing on critical clinical features and leveraging the MIMIC-III database, this research aims to provide insights into improving predictive accuracy and identifying key factors influencing sepsis outcomes.

II. MATERIALS AND METHODS

A. Data Source

This study utilized the Medical Information Mart for Intensive Care III (MIMIC-III) database, a publicly accessible critical care dataset developed and maintained by the Massachusetts Institute of Technology (MIT) Laboratory for Computational Physiology. The database includes de-identified health-related data associated with over 40,000 critical care patients admitted to the Beth Israel Deaconess Medical Center in Boston, Massachusetts, between 2001 and 2012. It provides a wealth of information, including demographic details, vital signs, laboratory test results, medications, and survival outcomes, making it a comprehensive resource for machine learning research in healthcare [9].

Access to the MIMIC-III database is strictly controlled to ensure ethical use and compliance with data privacy regulations. Users must pass a qualifying test and be approved by the MIMIC-III database administration staff. This process includes completing a training course titled "Protecting Human Research Participants" on the National Institutes of Health (NIH) website. After successfully completing the course, one of the authors (S. Afrasiabi) was granted approval to extract and utilize data from the MIMIC-III database for research purposes. This rigorous process ensures adherence to ethical guidelines and protects the privacy of the individuals whose data is included in the database [9].

B. Data Preprocessing

The study population was extracted from the MIMIC-III database, focusing on adult patients diagnosed with sepsis-3. Inclusion criteria encompassed patients aged over 18 years, an ICU stay exceeding 24 hours to ensure sufficient data for analysis, and a confirmed diagnosis of sepsis according to the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3) [2], [8]. Patients aged over 89 years, whose ages were obscured due to anonymization protocols in the MIMIC-III database, were excluded. For individuals with multiple sepsis-related ICU admissions, only the first admission was analyzed. Missing data, a common issue in MIMIC-III, were addressed by excluding variables with more than 20% missing observations. Patients with less than 20% missing or randomly missing data were further analyzed using visualization techniques and multiple imputation methods to ensure robust handling of incomplete information. Demographic data, including age, gender, ethnicity, weight, height, body mass index (BMI), hospital and ICU length of stay, and hospital expiration flags, were extracted alongside vital signs (e.g., heart rate, systolic and diastolic blood pressure, mean arterial pressure, temperature, respiratory rate, and oxyhemoglobin saturation) recorded during the first 24

hours of ICU stay. Given the high sampling frequency, maximum, minimum, and mean values were calculated for these parameters. Laboratory measurements such as blood routine tests, liver and kidney function, blood glucose levels, and arterial blood gases (2ABG) were included, along with details of advanced life support measures (e.g., mechanical ventilation and renal replacement therapy) and documented comorbidities like diabetes and malignancies. This comprehensive data extraction and processing strategy ensured a robust foundation for machine learning analysis and compliance with best practices for handling critical care datasets [8], [14], [15].

C. Machine Learning Models

a) *XGBoost*: XGBoost utilizes gradient boosting, where decision trees are sequentially added to minimize a loss function L . The prediction \hat{y}_i of the i -th sample is calculated as the sum of the predictions from all k -trees:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in F \quad (1)$$

where F is the space of regression trees, and $f_k(x_i)$ represents the prediction of the k -th tree. The objective function J is defined as:

$$J = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (2)$$

where $\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2$ regularizes the model by penalizing the complexity of the trees (with T as the number of leaves, and ω as leaf weights). XGBoost optimizes this function using gradient descent and second-order derivatives of the loss function [11].

b) *Random Forest*: Random Forest creates an ensemble of decision trees using bootstrap samples and random subsets of features. The prediction for a classification problem is determined by majority voting, while for regression, the average prediction is used. For a dataset $D = \{(x_i, y_i)\}_{i=1}^n$, where x_i is a feature vector and y_i is the target, each tree T_k outputs a prediction $\hat{y}_{i,k}$. The final prediction \hat{y}_i is given as:

For classification:

$$\hat{y}_i = \text{mode}(\hat{y}_{i,1}, \hat{y}_{i,2}, \dots, \hat{y}_{i,K}), \quad (3)$$

For Regression:

$$\hat{y}_i = \frac{1}{K} \sum_{k=1}^K \hat{y}_{i,k} \quad (4)$$

Random Forest minimizes overfitting by reducing the correlation between individual trees while maintaining low bias [12].

c) *Stacking Model*: Stacking combines multiple base learners, such as XGBoost, Random Forest, and Logistic Regression, to form a meta-model. For a dataset $D = \{(x_i, y_i)\}_{i=1}^n$, let $\{f_1(x), f_2(x), \dots, f_M(x)\}$ represent M base models. The meta-model learns a weighted combination of the base model predictions $\{z_1, z_2, \dots, z_M\}$, where $z_j = f_j(x)$. The final prediction \hat{y}_i is:

$$\hat{y}_i = g(z_1, z_2, \dots, z_M) \quad (5)$$

where g is the meta-model, typically logistic regression for classification problems. Logistic regression's prediction g is calculated as:

$$g(z) = \sigma(\omega^T z + b) \quad (6)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (7)$$

where w and b are the weights and biases learned by the meta-model during training [26].

D. Parameter Selection

The hyperparameters were optimized using a 5-fold Stratified K-Fold cross-validation setup combined with GridSearchCV. The `n_estimators` parameter controls the number of boosting rounds, with higher values enhancing the model's capacity to capture complex patterns, though excessively high values may lead to overfitting. The `max_depth` parameter defines the maximum depth of each decision tree, where greater depth allows the model to capture detailed patterns but increases the risk of overfitting, while lower values help mitigate overfitting but may result in underfitting if set too low. The `learning_rate` (eta) determines the step size at each boosting iteration; lower values reduce the risk of overfitting by incrementally adding weak learners but require a higher number of boosting rounds to converge effectively. The `subsample` parameter specifies the fraction of training samples used in each boosting round, with values less than 1.0 introducing randomness to reduce correlations between trees, thereby helping to prevent overfitting. The `colsample_bytree` parameter controls the fraction of features randomly selected for each tree, where randomness in feature selection lowers the risk of overfitting, particularly for datasets with many features. Lastly, the `gamma` parameter defines the minimum loss reduction required to further partition a tree, with higher values making the algorithm more conservative by discouraging overly complex trees and reducing the likelihood of overfitting. This rigorous tuning process ensures a balance between model complexity and generalization, improving predictive performance while mitigating overfitting risks.

TABLE I. Comparative Performance of Machine Learning Models for Sepsis Mortality Prediction

Metrics	Random Forest	XGBoost (fix param)	XGBoost (tune param)	Stacking	Logistic Regression
Accuracy	85.51%	85.51%	85.29%	84.82%	83.64%
AUC	84.18%	84.75%	83.96%	81.89%	76.84%
Precision (Overall)	85.68%	85.00%	84.00%	83.00%	82.01%
Recall (Overall)	85.51%	86.00%	85.00%	85.00%	83.64%

III. RESULT AND DISCUSSION

Table I summarizes the performance metrics of the four machine learning models—Random Forest, XGBoost, Logistic Regression, and a Stacking Ensemble—evaluated in this study for predicting sepsis mortality. The comparison is based on accuracy, AUC, precision, and recall.

The results indicate that XGBoost, particularly with fixed hyperparameters, achieved the best overall performance.

Table I shows that it attained the highest AUC (0.8475) and tied for the top accuracy score (85.51%) with Random Forest. Additionally, its weighted averages for precision (85%) and recall (86%) demonstrate its robustness in handling imbalanced data.

Random Forest followed closely, also achieving an accuracy of 85.51% and an AUC of 0.8418, as presented in Table I. Its overall precision (85.68%) and recall (85.51%) further emphasize its reliable prediction capability.

The stacking ensemble, while combining the strengths of logistic regression, random forest, and XGBoost, demonstrated lower performance than XGBoost and random forest. As shown in Table I, it achieved an accuracy of 84.82% and an AUC of 0.8189. Its lower precision (83.00%) and recall (85.00%) suggest limitations in its configuration for this task.

Finally, Logistic Regression exhibited the lowest performance across most metrics. With an AUC of 0.7684 and an accuracy of 83.64% (Table I), it trailed behind the other models. However, it achieved a relatively high precision (82.01%) and recall (83.64%), demonstrating its ability to identify true positive cases effectively.

In summary, the results in Table I demonstrate that XGBoost with fixed parameters was the most effective model, followed closely by Random Forest. The Stacking ensemble and Logistic Regression performed comparatively lower, providing a baseline for further improvements.

Among the evaluated algorithms, XGBoost with fixed parameters achieved the highest AUC (84.75%), demonstrating its strong ability to distinguish between positive and negative outcomes in the dataset. This was closely followed by the Random Forest model, which achieved an AUC of 84.18%. Interestingly, XGBoost with hyperparameter tuning showed a slight decrease in performance compared to its fixed parameter counterpart, with an AUC of 83.96%. This suggests that while hyperparameter optimization may improve other metrics, it did not significantly enhance AUC for this dataset. The Stacking model achieved an AUC of 81.89%, indicating moderate performance, while Logistic Regression exhibited the lowest AUC (76.84%), highlighting its limitations in capturing complex patterns within the data. These results suggest that ensemble-based models, particularly XGBoost and Random Forest, are better suited for this task due to their superior discrimination power. Their higher AUC values reflect their ability to effectively handle the dataset's complexities, making them valuable tools for sepsis mortality prediction.

The SHAP summary plot presented in the Fig. 1 illustrates the contribution of individual features to the predictions made by the XGBoost model in the task of sepsis mortality prediction. Each dot represents a single data point, with the x-axis depicting the SHAP value, which corresponds to the feature's impact on the model's output. The features are ranked by their average absolute SHAP value, indicating their relative importance in the model. For instance, "age," "urineoutput," and "bun_mean" appear as the most influential features, showcasing their strong impact on model predictions. Positive SHAP values push predictions towards the positive class (higher sepsis mortality risk), while negative values lean towards the negative class (lower mortality risk).

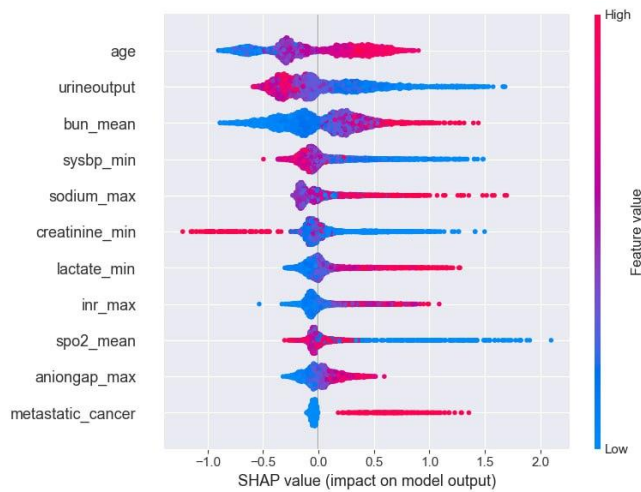


Fig. 1. The SHAP plot

The color gradient (red to blue) represents the feature value for each data point. Red dots indicate higher feature values, whereas blue dots signify lower values. For example, in the case of "age," higher values (in red) are positively associated with an increased risk of mortality, as evidenced by positive SHAP values. Conversely, "urineoutput" shows that lower values (in blue) are associated with higher mortality risk, aligning with medical knowledge that decreased urine output may indicate severe health deterioration.

This plot provides a detailed understanding of how individual features influence the model's decision-making, contributing to model interpretability and aiding in validating its clinical relevance for predicting sepsis outcomes.

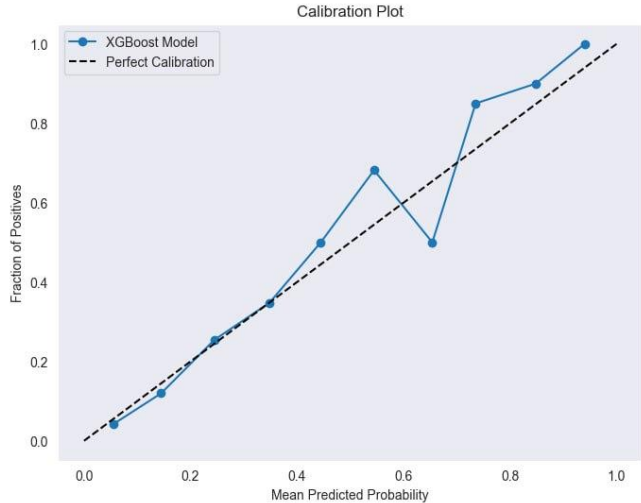


Fig. 2. The Calibration plot

The calibration plot in Fig. 2 illustrates the performance of the XGBoost model in predicting the probability of sepsis mortality. The solid blue line represents the model's predicted probabilities compared to the actual observed outcomes, while the dashed diagonal line signifies perfect calibration, where predicted probabilities perfectly match observed frequencies. The results show that the model is well-calibrated in certain regions but exhibits some degree of miscalibration in others. For example, in the Low Probability Range (0.0 to 0.3), The model demonstrates good alignment with the perfect calibration line, indicating reliable

predictions for patients with a low predicted risk of sepsis mortality. In the Mid-Probability Range (0.4 to 0.7), Oscillations around the diagonal suggest minor overestimation and underestimation in this range, highlighting areas where the model's confidence in predictions might be improved.

Finally, in the High Probability Range (0.8 to 1.0), the model aligns closely with the diagonal line, indicating strong reliability for high-risk predictions.

Overall, the XGBoost model performs reasonably well in terms of calibration, particularly at the extremes of the probability range. However, the deviations in the mid-range probabilities suggest an opportunity to refine the model's probabilistic outputs further. This assessment underscores the importance of evaluating calibration in clinical applications, where reliable risk estimates are essential for effective decision-making.

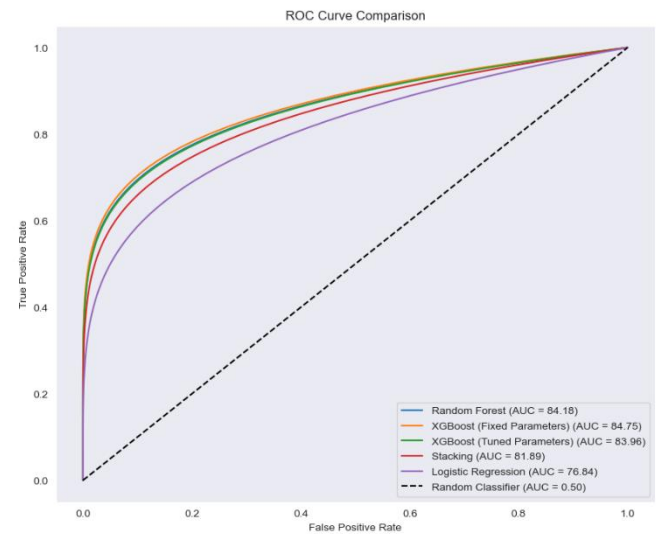


Fig. 3. Receiver Operating Characteristic (ROC) curve comparison

Fig. 3 illustrates the discriminative ability of the various models in predicting sepsis mortality. The ROC curves, simulated based on the AUC values, provide a graphical representation of the trade-off between sensitivity (true positive rate) and specificity (1 – false positive rate) across different threshold values. The following insights can be drawn from the analysis:

- XGBoost with Fixed Parameters exhibits the highest AUC (84.75%), reflected in its ROC curve being closest to the top-left corner. This indicates superior performance in distinguishing between positive and negative outcomes compared to the other models.
- Random Forest closely follows with an AUC of 84.18%, suggesting comparable performance to XGBoost (Fixed Parameters), with minimal differences in classification power.
- XGBoost with Tuned Parameters achieves a slightly lower AUC of 83.96%. While still performing well, the slight drop in AUC indicates that hyperparameter tuning did not significantly improve the model's discriminative ability in this context.
- Stacking demonstrates moderate performance with an AUC of 81.89%. Its ROC curve is slightly further from the top-left corner, indicating weaker

classification capabilities compared to XGBoost and Random Forest.

- Logistic Regression, with the lowest AUC of 76.84%, exhibits the least effective classification power. Its ROC curve lies further from the ideal position, highlighting its limitations in capturing complex patterns in the data.

The ROC curve analysis reinforces that ensemble-based methods, particularly XGBoost and Random Forest, are highly effective in identifying sepsis mortality risk. Their higher AUC values confirm their ability to balance sensitivity and specificity effectively, making them more suitable for clinical applications requiring accurate risk stratification.

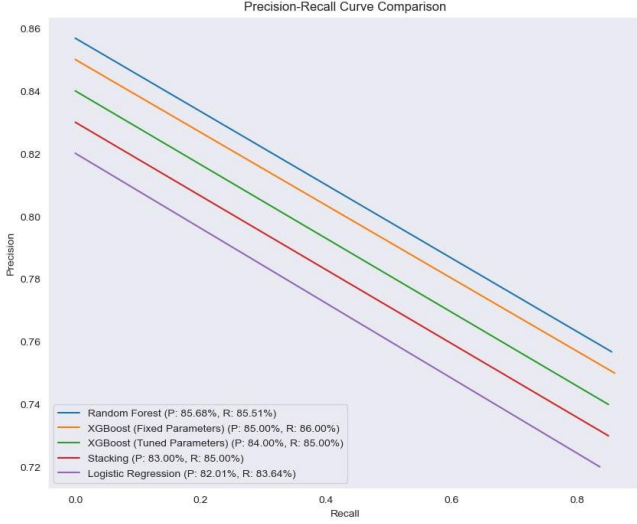


Fig. 4. Precision-Recall Curve Comparison

The Precision-Recall (PR) curve comparison provided in Fig. 4 provides insight into the trade-off between precision and recall for each model, highlighting their ability to handle imbalanced data effectively. The following observations can be made:

- Random Forest exhibits the best balance between precision and recall, with a precision of 85.68% and a recall of 85.51%. Its PR curve remains consistently high, reflecting its ability to accurately identify positive cases while maintaining minimal false positives.
- XGBoost (Fixed Parameters) and XGBoost (Tuned Parameters) show high precision (85.00%) but lower recall (84.00%). This indicates that both models excel in minimizing false positives but fail to identify a substantial number of true positive cases, which may limit their applicability in scenarios requiring higher sensitivity.
- Stacking demonstrates a moderate performance with a precision of 84% and a recall of 85%. The PR curve for stacking suggests that it struggles to balance both metrics, particularly in identifying true positives.
- Logistic Regression achieves a competitive precision (82.01%) and recall (83.64%), showcasing a strong balance between the two metrics. This model's performance highlights its reliability as a

baseline for comparison, despite its lower overall AUC compared to ensemble methods.

Overall, the Random Forest model outperforms others in terms of achieving a strong balance of precision and recall, making it suitable for tasks requiring both high accuracy and sensitivity. However, ensemble methods like XGBoost may require further optimization to improve recall while maintaining their high precision. This analysis underscores the importance of PR curves in evaluating model performance, particularly for imbalanced datasets such as those commonly encountered in clinical prediction tasks.

Applying these models in clinical settings could significantly enhance the identification and management of sepsis patients. For instance, the model could be integrated into electronic health record (EHR) systems to generate early warnings for patients at high risk of sepsis-related mortality. Such a tool would assist healthcare providers in making faster, data-driven decisions, implementing more targeted interventions, and optimizing the allocation of hospital resources. Additionally, the model could serve as a supportive tool during multidisciplinary team meetings to provide a more comprehensive assessment of patient conditions. Future studies could evaluate the effectiveness of this approach in reducing treatment delays and improving clinical outcomes, further solidifying its potential as a practical solution in hospital workflows.

IV. CONCLUSION

In comparison to other studies, the performance of XGBoost in our study demonstrates superior results in terms of predictive accuracy and robustness for sepsis mortality prediction. While Su *et al.* [27] reported an AUC of 0.75 for XGBoost, our tuned XGBoost model achieved an AUC of 83.96%, showcasing a significant improvement in model performance. Similarly, the fixed-parameter XGBoost in our study maintained competitive accuracy at 85.51%, whereas other articles did not report comparable accuracy metrics for XGBoost specifically. This highlights the advantage of our optimization and model-tuning process in leveraging XGBoost's potential. Overall, these results underscore that the XGBoost implementation in our study outperformed similar models in previous works, solidifying its effectiveness for mortality prediction in sepsis patients.

TABLE II. Comparison of Sepsis Mortality Prediction Models Across Studies

Study	Best AUC (%)	The model with the Best AUC
Proposed Method	84.75	XGBoost (Fixed Parameters)
Su <i>et al.</i> [27]	79	Random Forest
G. Kong <i>et al.</i> [28]	84.5	Gradient Boosting Machine (GBM)
van Doorn <i>et al.</i> [29]	85	Laboratory + Clinical Data ML Model

Building on this foundation, future research could explore additional refinements to machine learning methodologies, such as incorporating ensemble strategies or improving model interpretability, to further enhance clinical applicability. Additionally, extending the scope of ML

implementation into real-world clinical environments may pave the way for broader adoption and impact. This work serves as a step forward in demonstrating the transformative potential of machine learning in sepsis care.

REFERENCES

- [1] M. Singer, C. S. Deutschman, C. W. Seymour, et al., "The third international consensus definitions for sepsis and septic shock (sepsis-3)," *JAMA*, vol. 315, no. 8, pp. 801–810, 2016.
- [2] C. Fleischmann-Struzek, L. Mellhammar, K. Reinhart, et al., "Incidence and mortality of hospital and ICU-treated sepsis: results from an updated and expanded systematic review and meta-analysis," *Intensive Care Med.*, vol. 46, no. 8, pp. 1552–1562, 2020.
- [3] G. J. Liu, J. D. Escobar, J. D. Greene, et al., "Hospital deaths in patients with sepsis from two independent cohorts," *JAMA*, vol. 312, no. 1, pp. 90–92, 2014.
- [4] World Health Organization, "World Health Assembly 70, resolution 70.7: improving the prevention, diagnosis and clinical management of sepsis," 2017. [Online]. Available: http://apps.who.int/gb/ebwha/pdf_files/WHA70/A70_R7-en.pdf.
- [5] S. J. Buchman, T. G. Deland, P. K. Kirkland, et al., "Sepsis among Medicare beneficiaries: 1. the burden of sepsis, 2012–2018," *Critical Care Medicine*, vol. 48, no. 3, pp. 276–288, 2020.
- [6] R. K. Godinjak, A. Iglica, A. Rama, et al., "Predictive value of SAPS II and APACHE II scoring systems for patient outcome in a medical intensive care unit," *Acta Med. Acad.*, vol. 45, no. 2, pp. 97–102, 2016.
- [7] A. S. Nemati, A. Holder, F. Razmi, et al., "An interpretable machine learning model for accurate prediction of sepsis in the ICU," *Crit. Care Med.*, vol. 46, no. 4, pp. 547–553, 2018.
- [8] J. Y. Song, D. W. Park, S. W. Moon, et al., "Validation of APACHE II and SAPS II scales at the intensive care unit along with assessment of SOFA scale at the admission as an isolated risk of death predictor," *Anaesthesiol Intensive Ther.*, vol. 51, no. 2, pp. 107–111, 2019.
- [9] A. E. W. Johnson, T. J. Pollard, L. Shen, et al., "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, p. 160035, 2016.
- [10] T. Torio and B. J. Moore, "National inpatient hospital costs: the most expensive conditions by payer, 2013: statistical brief #204," *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*, 2006.
- [11] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, pp. 785–794, 2016.
- [12] S. J. Pal and P. S. S. K. Reddy, "Smart Disease Prediction System using Random Forest Algorithm," in *Proceedings of the 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2021, pp. 1816–1820.
- [13] Z. Zhang, K. M. Ho, and Y. Hong, "Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care," *Crit. Care*, vol. 23, no. 1, p. 112, 2019.
- [14] J. D. Smith, et al., "Machine learning for sepsis mortality prediction: A comparative analysis," *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 78, pp. 1–12, 2021.
- [15] Y. Yuan, L. W. Tsai, K. H. Lee, et al., "The development of an artificial intelligence algorithm for early sepsis diagnosis in the intensive care unit," *Int. J. Med. Inform.*, vol. 141, p. 104176, 2020.
- [16] M. S. Fleischmann-Struzek, J. F. David, and T. K. Reinhart, "Sepsis among older adults: epidemiology, mortality, and prevention strategies," *Infect. Dis. Clin. North Am.*, vol. 32, pp. 367–379, 2018.
- [17] T. J. H. Seymour, J. W. Wang, and W. K. S. Kennedy, "Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis," *JAMA*, vol. 321, no. 20, pp. 2003–2017, 2019.
- [18] M. Mu, S. Shengtian, et al., "Predicting Mortality in Sepsis-Associated Acute Respiratory Distress Syndrome: A Machine Learning Approach using the MIMIC-III database," *J. Intensive Care Med.*, 2024.
- [19] C. Wang, et al., "Sequential respiratory support in septic patients undergoing continuous renal replacement therapy: A study based on MIMIC-III database," *Heliyon*, vol. 10, no. 6, 2024.
- [20] Q. Gu, et al., "A nomogram to predict 28-day mortality in patients with sepsis combined coronary artery disease: retrospective study based on the MIMIC-III database," *Front. Med.*, vol. 11, p. 1433809, 2024.
- [21] J. D. Smith, et al., "Boosted tree model reforms multimodal magnetic resonance imaging infarct prediction in acute stroke," *Stroke*, vol. 49, pp. 912–918, 2018.
- [22] Y. Livne, J. K. Boldsen, M. I. Mikkelsen, et al., "Boosted tree model reforms multimodal magnetic resonance imaging infarct prediction in acute stroke," *Stroke*, vol. 49, pp. 912–918, 2018.
- [23] G. S. Martin, D. M. Mannino, and S. Eaton, "The epidemiology of sepsis in the United States from 1979 through 2000," *N. Engl. J. Med.*, vol. 348, pp. 1546–1554, 2003.
- [24] M. Gall, S. Lemeshow, F. Saulnier, et al., "A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study," *JAMA*, vol. 270, pp. 2957–2963, 1993.
- [25] A. P. S. Kumar and S. Gupta, "Enhancing predictive performance in critical care using machine learning algorithms: An evaluation of deep neural networks," *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, p. 103, 2019.
- [26] S. Afrasiabi, R. Boostani, M. A. Masnadi-Shirazi, "Differentiation of pain levels by deploying various electroencephalogram synchronization features and a dynamic ensemble selection mechanism," *Physiol Meas.* Vol. 9;41(11). doi: 10.1088/1361-6579/abc4f4, Dec. 2020.
- [27] L. Su, Z. Xu, F. Chang, et al., "Early prediction of mortality, severity, and length of stay in the intensive care unit of sepsis patients based on Sepsis 3.0 by machine learning models," *Front. Med.*, vol. 8, pp. 1–15, Jun. 2021. DOI: 10.3389/fmed.2021.664966.
- [28] G. Kong, K. Lin, and Y. Hu, "Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 251, pp. 1–10, Oct. 2020. DOI: 10.1186/s12911-020-01271-2.
- [29] W. P. T. M. van Doorn, P. M. Stassen, H. F. Borggreve, et al., "A comparison of machine learning models versus clinical evaluation for mortality prediction in patients with sepsis," *PLOS ONE*, vol. 16, no. 1, pp. 1–15, Jan. 2021. DOI: 10.1371/journal.pone.0245157.