Investigating the Role of Machine Learning Algorithms in Predicting Sepsis using Vital Sign Data

Amit Sundas[1], Sumit Badotra[2], Gurpreet Singh[3], Amit Verma[4],
Salil Bharany[5]*, Imtithal A. Saeed[6], Ashraf Osman Ibrahim[7]*
Department of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India[1, 3, 5]
School of Computer Engineering and Technology, Bennett University, Greater Noida, India[2]
University Centre for Research and Development, Department of Computer Science and Engineering,
Chandigarh University, India[4]
Prince Sattam Bin Abdulaziz University, College of Computer Engineering and Sciences,
Department of Information Systems, Saudi Arabia[6]
Creative Advanced Machine Intelligence Research Centre-Faculty of Computing and Informatics,
Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia[7]

***Abstract*—Objective: In hospitals, sepsis is a common and costly condition, but machine learning systems that utilize electronic health records can enhance the timely detection of sepsis. The purpose of this research is to verify the effectiveness of a machine learning tool that makes use of a gradient boosted ensemble for sepsis diagnosis and prediction in relation. San Francisco University of California, (SFUC) Medical Center and the Medical Information Mart for Intensive Care (MIMIC) databases were consulted for historical information. The study encompassed adult patients who were admitted without sepsis and had a minimum single logging of six vital signs (SpO2, temperature, heart rate, respiratory rate, diastolic blood pressure and systolic). Using the area under the receiver operating characteristic (AUROC) curve, the performance of the machine learning algorithm was compared to commonly used scoring systems, and its accuracy was determined. Performance of the MLA (machine learning algorithm) was evaluated at sepsis onset, as well as 24 and 48 hours before sepsis onset. The AUROC for the MLA was 0.88, 0.84, and 0.83 for sepsis onset, 24 hours prior, and 48 hours prior, respectively. At the time of onset, these values were superior to those of SOFA, MEWS, qSOFA, and SIRS. Using UCSF data for training and MIMIC data for testing, the sepsis onset AUROC was 0.89. The MLA can safely predict sepsis up to forty-eight hours before it occurs and the accuracy in detecting the onset of sepsis is higher in comparison to traditional instruments. When trained and evaluated on distinct datasets, the MLA maintains high performance for sepsis detection.**

***Keywords*—Machine learning; sepsis; vital sign; prediction; electronic health records**

I. Introduction

Sepsis, a widespread and economically burdensome syndrome affecting hospitals worldwide, has undergone a transformative shift in its conceptualization. While it was formerly categorized within a three-tier system, the prevailing understanding characterizes it as a two-stage process, spanning from septic conditions to full-blown sepsis. The projected annual cost to the global healthcare system attributed to sepsis is a staggering $24 billion, an alarming financial burden. This is particularly disconcerting when considering the fatality rates associated with sepsis, which range from 25% to 40%. However, emerging evidence underscores the significance of early diagnosis and intervention before the onset of septic shock, as it holds the potential to significantly improve patient outcomes and reduce hospitalization durations [1].

Sepsis, often characterized by organ malfunction due to a systemic inflammatory response to infection, poses a formidable diagnostic challenge due to the intricate and varied origins of infections and the unique responses of individual patients. Consequently, there has been a growing impetus in medical science to advance automated patient monitoring systems tailored for the early identification of sepsis among hospitalized patients [2].

The advent of automated diagnostic decision and prediction technologies, facilitated by the widespread adoption of electronic health records (EHRs) in healthcare facilities, holds substantial promise for revolutionizing the tracking and management of complex medical conditions [3]. These technologies derive their foundations from the comprehensive medical records of patients, utilizing this wealth of data to generate warnings and treatment recommendations [4]. However, it is noteworthy that the existing diagnostic methods for sepsis predominantly lack predictive capabilities [5-7], with the majority relying on rule-based processes to trigger alarms and provide recommendations.

Within clinical settings, the commonly employed sepsis scoring systems encompass the Systemic Inflammatory Response Syndrome (SIRS) criteria [6], the Modified Early Warning Scale (MEWS) [8], and the Sequential Organ Failure Assessment (SOFA) score [9]. While these systems exhibit commendable sensitivity, they often grapple with issues related to specificity and are not explicitly designed for predicting the development of sepsis. Moreover, rule-based scores may struggle to accurately account for the diverse patient populations and the multifaceted sources of infection. Machine learning-based prediction techniques hold the potential to offer superior specificity, broader generalizability, and early sepsis risk identification, thus potentially reducing false alarms and enabling more timely physician responses [12].

Previous research endeavors have revealed the capacity to forecast the onset of sepsis, septic shock, and severe sepsis with a lead time of up to four hours before the condition manifests, employing machine learning-based systems trained on patient EHR data [13-14]. However, these studies were primarily conducted within the confines of a single institution's critical care group. In this study, we align with the contemporary definition of sepsis proposed by Singer et al. [1] to evaluate the historical performance of an algorithm employing a mixed-ward dataset, predicting sepsis up to two days in advance, solely relying on vital sign inputs. Moreover, our research aims to assess the algorithm's effectiveness by benchmarking its performance against prevailing rule-based scoring systems and scrutinizing its reliability through cross-population validation, as elucidated in study [15].

II. Related Work

In the realm of predicting sepsis using vital sign data, extensive research has been conducted to explore the role of machine learning algorithms. This section provides an overview of existing studies and their contributions, offering insights into the progress made in this critical domain and highlighting the gaps and areas requiring further investigation.

Numerous researchers have delved into the development of sepsis prediction models, aiming to enhance early detection and intervention. Studies by [1-3] and [10-14] have primarily focused on utilizing machine learning algorithms to analyze vital sign data for sepsis prediction. These studies have demonstrated promising results in terms of accuracy and timeliness, providing a foundation for further exploration.

In contrast, [16] and [17] have employed alternative approaches, such as rule-based scoring systems, to predict sepsis. While these methods have proven valuable in clinical settings, they raise questions about the potential advantages of machine learning algorithms in terms of predictive power and adaptability.

While substantial progress has been made in the field of sepsis prediction, there are still various challenges that demand attention. These include addressing the interpretability of machine learning models, optimizing feature selection, and ensuring generalizability across diverse patient populations and healthcare settings. The research presented in this study seeks to contribute to this ongoing discourse by:

- Our unique approach employs a gradient boosted ensemble for sepsis diagnosis, leveraging SFUC and MIMIC electronic health records.
- We include adult patients without sepsis who have recorded six vital signs. We evaluate the algorithm's performance with AUROC and compare it to traditional scoring systems.
- Also in Results, MLA has an AUROC of 0.88 at sepsis onset, 0.84 and 0.83 for predictions 24 and 48 hours earlier. Outperforms SOFA, MEWS, qSOFA, and SIRS. MLA trained on UCSF data, tested on MIMIC, reaches AUROC of 0.89.
- Our research shows MLA may predict sepsis up to 48 hours earlier with high accuracy, contributing to improved early sepsis management.

By reviewing the existing solutions and identifying areas that warrant further exploration, this research aims to position itself within the broader landscape of sepsis prediction, ultimately striving to enhance the effectiveness of early intervention in critical healthcare scenarios.

III. Components

A. Ethics Certification and Informed Consent

As mandated by the Health Insurance Portability and Accountability Act (HIPAA), we removed all personally identifying information from patient records before collecting the datasets. There was no compromise in patient well-being due to the data gathering procedure [16].

B. Measurements

Six vital signs (systolic BP, heart rate, temperature, respiration rate, diastolic BP and peripheral oxygen saturation (SpO2)) were examined to establish sepsis risk ratings. To be included in the research, it was required that every patient encounter had a minimum of one record for each vital sign. These are the sole vitals we engage in the act of generating or producing characteristics for assessing sepsis threat scores since they are directly related to sepsis development and are evaluated often even in the absence of a clinical concern for septic shock [17].

C. Sources Data

The datasets utilised in this study came from the Medical Center at the University of California, San Francisco (SFUC) and the Intensive Care Unit section of the Medical Information Market (MIMIC). Patients who visited the Parnassus Heights, Mission Bay, or Mount Zion facilities between June 2016 and March 2023 accounted for 17,467,987 of the total contacts in the SFUC dataset. Our final group consisted of 91,445 patients after excluding those with hospital stays less than seven hours and more than 2000 hours from the original 96,646 inpatients (95,869 of whom had at least one recording of each vital sign). We employed subsets of this final sample, differentiated by patients' lengths of stay, to conduct our 24- and 48-hour lookahead analyses. Different frequencies of data collection and types of care provided were documented in the SFUC data from the ICU, the ED, and the floor units [18-19]. Due to missing unit transfer timestamps, it was impossible to determine where a patient was located at any given moment. The MIMIC information was culled from the 61,532 ICU interactions recorded in the Medical Information Mart for Intensive Care III (MIMIC-III) v1.3 database between the years of 2012 and 2023. Patients 18 and older had 52,902 visits to the hospital, but only 21,507 had at least one recording of each vital sign, qualifying them for inclusion in the final cohort. Missing measurements of any vital sign were grounds for excluding encounters. Patient safety was not jeopardised by the data collecting process, and all patient information was de-identified in accordance with HIPAA regulations. SFUC's IRB (Institutional Review Board) gave its clearance to this project [20].

D. Statistical Analysis

We extracted the sensitivity, specificity, and AUROC value with 95% CI for predicting sepsis patients in the ICU from each of the included studies. The ROC curve compares different thresholds by plotting the true positive rate (TPR) against the false positive rate (FPR). Excellent, good, fair, poor, and fail are defined by AUROC curve values of 0.9-1, 0.8-0.9, 0.7-0.8, 0.6-0.7, and 0.5-0.6, respectively. In the end, we calculated the ROC, sensitivity, and specificity with 95% CI. To gauge the degree of statistical heterogeneity among the included trials, we calculated an I^2 value. Heterogeneity is classified as extremely low (l^2 ~25%), low (l^2 ~50%), medium (l^2 ~75%), or high (l^2 > 75%), respectively. There was less variation in impact sizes across trials because data from all included research were combined using a random effect model. The proportion of overall study variance that can be attributed to factors other than chance is measured by the I2 statistic [21]. To determine I^2, we used the formula:

$$l^2 = 100\ \% \ (Q - df)/Q \quad (1)$$

Q = dfQ, where df = number of observations and Q = Cochrane's heterogeneity statistic. As a consequence, the I^2 findings range from 0% (no observed heterogeneity) to 100% (highest heterogeneity), with all negative values adjusted to zero.

We selected the symmetric approach in our meta-analysis because we hypothesized that the included papers would be of varying quality. The pooled estimate of AUROC, sensitivity, specificity, and diagnostic odds ratio was calculated using MetaDiSc (version 1.4). It's useful for doing things like (a) summarizing data from each research, (b) analyzing the graphical and statistical similarity of studies, (c) computing the pooled estimate, and (d) examining heterogeneity. The likelihood ratio was calculated to illustrate the extent to which a given outcome was more common in studies including patients with sepsis illness compared to those involving subjects without sepsis disease.

$$LR+ = (Sensitivity/1 - Specificity) \quad (2)$$

$$LR- = (1 - Sensitivity/Specificity) \quad (3)$$

Additionally, diagnostic odds ratio (DOR) was calculated to reveal how much higher the chances are for persons with a positive test result to have the sepsis illness compared to those with a negative test result. The formula for DOR is LR + /LR-. Each technique's efficacy was measured using a number of different metrics (Supplementary Table 2), including area under the receiver operating curve, sensitivity, specificity, diagnostic odd ratio, and probability ratio.

The precise confidence bounds for the binomial percentage were calculated using the F distribution technique, and the confidence ranges for overall sensitivity and specificity were analyzed as well [19]. However, excess dispersion correction was applied in the computations, and Meta-DiSc was the tool of choice. The typical approximation to binomial was used here.

$$SE\ (SenT\ ) = \sqrt{\frac{SenT(1 - SenT\ )}{\sum_{i\ Di}}} \quad (4)$$

$$SE\ (Sper\ ) = \sqrt{\frac{Senr(1 - Senr\ )}{\sum_{i\ Di}}} \quad (5)$$

$$SenT \pm Z_{\frac{\alpha}{2\varphi Sen}} SE(Senr) \quad (6)$$

$$Sper \pm Z_{\frac{\alpha}{2\varphi Spe}} SE(Spen) \quad (7)$$

IV. Methodologies

A. Results and Methods

The capability of the algorithm to detect individuals that are septic at start and in the preceding 24 and 48 hours was the primary focus of this study. We evaluated the efficacy of the method by calculating the AUROC, or area under the receiver operating characteristic curve.

The data was collected through queries built for the PostgreSQL (PostgreSQL Global Development Group) database and then saved as CSV files [19]. Features for predicting sepsis risk were created using just six vital signs: heart rate, respiration rate, systolic blood pressure, SpO2, diastolic blood pressure and temperature. If there wasn't a fresh reading for each hour leading up to the patient's designated onset time, the previous reading was used to estimate the value. When several readings were obtained within the same hour, an average was calculated and utilized. This cut down on the classification system's exposure to measurement frequency data that wasn't relevant to physiology [21].

Information was also gathered to create the Sepsis-3 reference standard and the rules-based grading system. Often used measures such as the Sequential Organ Failure Assessment (SOFA), Modified Rankin Scale (MERS), and qSOFA (quick SOFA) were compared to the prediction algorithm. Similar to Jaimes et al. [3], we searched for SIRS criteria. To determine each patient's MEWS score [14], we used the same procedure as Fullerton et al [2]. The formula for calculating a qSOFA score may be found in Singer et al [1]. While the SOFA score is included in the widely accepted definition of sepsis, we investigated its ability to identify the onset of sepsis independently of other factors. CSV files were needed for bilirubin levels, FiO2, PaO2, the Glasgow Coma Scale, white blood cell counts, vasopressor dosages, and platelet counts in order to calculate these scores [22].

Sepsis is "life-threatening organ dysfunction induced by a dysregulated host response to infection," according to the 2016 consensus definition, which served as the basis for the Sepsis-3 gold standard. A 2-point shift in the Sequential Organ Failure Assessment (SOFA) score was considered indicative of organ failure [1]. To determine when the SOFA score shifted, we relied on the criteria established by Seymour et al. [4]. Antibiotics were administered and culture collected within 24 hours or within 72 hours if we suspected there was an infection. Seymour et al. [4] discovered the same thing when they tried testing the approach in reverse. When both the SOFA score and infection requirements were reached for the first time, we diagnosed sepsis [26-27].

There were 2,649 Sepsis-3 positive SFUC encounters out of 91,445 total that were included (a prevalence of 2.9%) in Fig. 1. The Sepsis-3 criteria were satisfied in 1024 out of

21,507 contacts at the MIMIC, yielding a frequency of 4.8%. There was a Sepsis-3 prevalence of 3.3% across all 112,952 patient interactions. The last stage of inclusion criteria for Sepsis-3 eliminated many potentially eligible encounters since the timing of sepsis onset was during the first seven hours of admission [23].

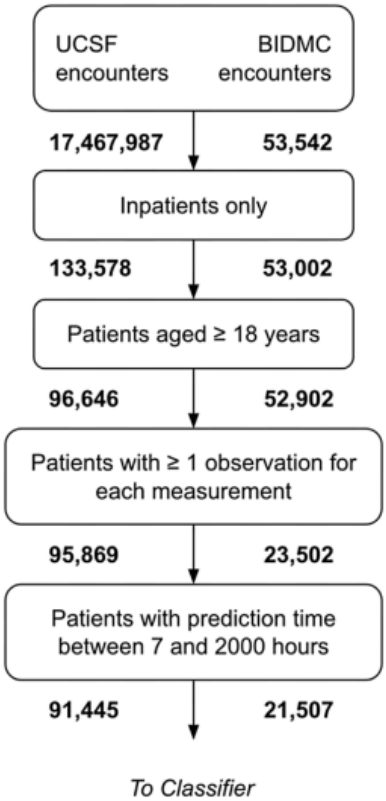

Fig. 1. Graphic illustrating patient inclusion and duration-based subsets in SFUC/MIMIC datasets for training and testing.

The "onset time" for individuals who never acquired sepsis was chosen at random from a continuous, uniform probability distribution so that they might serve as negative examples. The algorithm's risk ratings were derived from data collected both at and before the patient's start time. Patients who were diagnosed with sepsis either at the time of admission or within seven hours of admission were removed from the analysis to make room for prediction windows [24].

Patient contacts were first categorized by duration of stay before training the classifier. Example: a patient who had been hospitalized for 25 hours before contracting sepsis would be included in a 24-hour prediction experiment but not a 48-hour prediction trial. There were 107 cases of Sepsis-3 among the 20,590 MIMIC interactions and 267 cases among the 89,000 SFUC encounters with at least 24 hours of stay data. After at least 48 hours in the hospital, Sepsis-3 was identified in 50 of the 20,533 MIMIC interactions and 97 of the 88,887 SFUC encounters. To keep the calculation matrices manageable, hospital encounters with onset times more than 2000 hours were omitted [25].

B. The Algorithm for Prediction in Machine Learning

An ensemble of trees was used to generate scores for use in the algorithm's classifier, which was then used to get an overall score [28]. The system utilised one-hour, two-hour, and pre-prediction vital indicators, as well as the hourly changes between them, to make its predictions. A feature vector x containing 30 components, with five values derived from each of the six measurement sources was formed by summing these numbers in a causal fashion. The trees were built using the Python XGBoost module, with each branch being divided into two feature groups [29]. We used a five-fold cross-validation grid search on the training set to determine that a maximum of four, three, and six branches should be used for 0-hour, 24-hour, and 48-hour predictions, respectively [30]. Based on this grid search, we settled on the values 0.05, 0.12, and 0.12 for the XGBoost learning rate parameter. As we employed early stopping to avoid model overfit, we did not need to restrict the maximum number of trees in each ensemble. These risk ratings were then utilised by the algorithm to classify patients as having sepsis or at risk for developing it [31-33].

Two sets of SFUC interactions were created. The first group, made up of 80% of all interactions, was arbitrarily divided into a test set and a training set. Twenty percent of the second set was put aside as a control group for further examination. Using just the training data, we conducted a five-fold cross-validation to find the optimal hyperparameters for the grid search. For this prediction job, we looked into a parameter space that was similar to that of previously described hyperparameters [34-36]. We looked at learning rates between 0.05 and 0.12, in 0.01 increments, and explored numbers between 3 and 8 for the maximum number of branchings. For each look ahead, we settled on the optimal combination of branching level and learning rate based on the average area under the receiver operating characteristic (AUROC) curve [37]. For ten-fold cross-validation, we randomly distributed encounters over ten groups of similar size, each containing 20% of the training set. Nine of these groups were used for training the algorithm for each fold, while the other was used for testing. Each of the ten potential permutations of training and test sets was put through the algorithm and put to the test on the independent test set [38]. We produced machine learning algorithm performance measures by averaging the metrics from ten cross-validation models, including tabular and graphical representations of the findings. In addition, the averaged feature significance scores from XGBoost were presented; these values show how often a feature was utilised to partition the data across the trees. In addition, we calculated the AUROC standard deviation using the cross-validation outcomes [39].

Patient encounter cohorts utilised for 24- and 48-hour prediction were limited to those with sufficient stay data, as previously indicated [40]. As a consequence, there is an inequity in the distribution of socioeconomic classes since fewer septic patients were seen in these cohorts. We took use of XGBoost's built-in capacity to deal with unbalanced classes [40] rather than using minority oversampling to artificially inflate the number of septic patients, which may not be typical of the real-world situation in which such an MLA is implemented.

C. Methods for Validation in Cross Populations

Cross-population validation studies were undertaken to evaluate the algorithm's sepsis detection ability after being trained on a data set from an individual institution and then evaluated on another with demographic and clinical disparities. We evaluated the algorithm on MIMIC patient measurements after training it on SFUC data but before retraining it on the target dataset. The whole dataset was put through its paces during testing, and the algorithm was trained and validated in the same manner as detailed above. Only at the outset was testing done so that it could be compared to rule-based approaches.

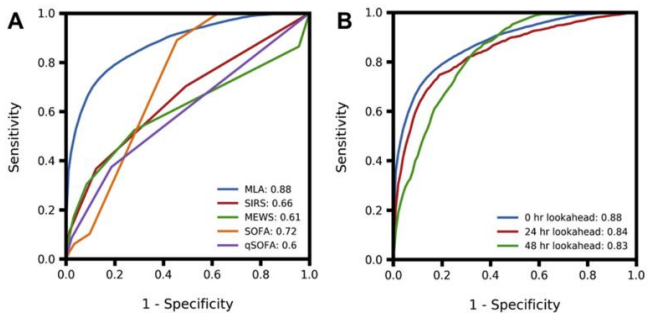

Fig. 2. MLA's efficiency in the hours before onset, and in comparison, to that of rival methods. A) Comparing ROC and AUROC of MLA to rival scoring systems at sepsis onset using SFUC data. B) ROC and AUROC for MLA at 0, 24, and 48 hours prior to sepsis onset using SFUC patient data.

V. Results

The research involved analyzing 91,445 patient interactions from SFUC and 21,507 patient encounters from MIMIC based on the collected data. Demographic characteristics of the two patient groups are compared in Table I, revealing significant differences in various aspects such as healthcare units visited, sepsis rates, in-hospital mortality, and age distributions. Importantly, the MIMIC database exclusively included ICU admissions, while the SFUC dataset encompassed all inpatient contacts. This deliberate selection of disparate data sets was aimed at evaluating the potential generalizability of the prediction system across a diverse range of patient groups.

In contrast to the qSOFA (0.60), MEWS (0.61), SIRS (0.66), and SOFA (0.72) scoring systems applied to the same dataset, the machine learning method developed and evaluated on the SFUC dataset exhibited a superior AUROC (0.88) for sepsis prediction (see Fig. 2). Additional performance indicators are detailed in Table II. Notably, the false alarm rates generated by the MLA, SIRS, and SOFA models were 0.22, 0.49, and 0.41, respectively, at specified operating points. It is significant to note that SIRS and SOFA generated 2.22 and 1.86 times as many false warnings as the MLA, respectively.

Furthermore, the study evaluated the algorithm's performance in predicting sepsis 24 and 48 hours before its onset, achieving AUROC values of 0.84 and 0.83, respectively (see Fig. 2 and Table II). During the 24-hour prediction, a dataset of 89,000 SFUC patients was analyzed, including 267 septic cases, and for the 48-hour prediction, a dataset of 88,887 SFUC patients was analyzed, including 97 septic cases. Notably, both predictions yielded higher diagnostic odds ratios (DOR) than initially predicted by rules-based approaches (see Table II).

Additionally, the study included data from MIMIC, which comprised 21,507 patients and 1024 septic cases. Remarkably, the algorithm, trained on the SFUC database without retraining, achieved an AUROC of 0.890 when applied to the MIMIC dataset.

The critical question arising from these results pertains to whether the MLA method can generate similar results when applied to infections other than sepsis. It prompts further consideration regarding whether the methodology's applicability is confined exclusively to sepsis or if it can be generalized to other medical contexts.

TABLE I. Cohort Demographics between SFUC and MIMIC

Summary of the Population	Characteristic	MIMIC (%)	SFUC (%)
Gender	Male	56.13	45.38
	Female	43.87	54.62
Ages[1]	70+	40.26	20.44
	60–69	22.79	21.22
	50–59	17.89	18.62
	40–49	9.73	13.02
	30–39	4.99	15.20
	18–29	4.34	11.50
Length of Stay (days)[2]	9+	10.05	17.02
	6–8	8.24	14.21
	3–5	32.38	38.42
	0–2	49.34	30.35
Death During Hospital Stay	No	72.92	97.81
	Yes	27.08	2.19
ICD-9 Code	Septic Shock	7.00	1.85
	Severe Sepsis	10.19	3.69
	Sepsis	3.48	5.83

Note - 91,445 patients make up the SFUC cohort. There are 21,507 patients in the MIMIC cohort.

1. Median age at SFUC was 55 (interquartile range [IQR]: 38-67), whereas at MIMIC it was 65 (IQR: (53, 77).

2. Median (SFUC): 4 (interquartile range [IQR]: (2, 6.32); median (MIMIC): 3 (2, 4)

Every one of the prediction windows was given a feature significance value (Supplementary Table I). The five most highly rated characteristics are shown in Table II. Age was the single most influential factor across all prediction intervals. After taking into account age, the greatest total score came from taking the patient's temperature, systolic blood pressure and heart rate simultaneously.

TABLE II. PRE-SEPSIS MLA PERFORMANCE ASSESSMENTS AND COMPARISON SCORES DURING THE START OF SEPSIS

	MLA (t = −48)	MLA (t = −24)	MLA (t = 0)	MEWS (t = 0)	SIRS (t = 0)	SOFA (t = 0)	qSOFA[1] (t = 0)
AUROC (SD)	0.83 (0.04)	0.84 (0.04)	0.88 (0.008)	0.61	0.66	0.72	0.60
LR+	2.86	2.57	3.76	1.86	1.43	1.92	2.00
Sensitivity	0.84	0.80	0.80	0.52	0.70	0.78	0.37
LR-	0.21	0.28	0.26	0.66	0.56	0.37	0.77
Specificity	0.66	0.72	0.78	0.72	0.51	0.59	0.81
DOR	13.15	13.69	14.79	2.81	2.44	5.20	2.60

Note. On the basis of patient data from SFUC, we evaluate MLA performance at 0, 24, and 48 hours before commencement, as well as competing scores at the time of onset. The setpoints were selected with sensitivities around 0.80 in mind. Just the MLA's cross-validation standard deviation in AUROC was computed.

[a.] **Abbreviation**. SD, standard deviation; LR+, positive likelihood ratio; LR-, negative likelihood ratio.

[1] All settings for qSOFA generated sensitivities much outside of the 0.80 range.

VI. CONCLUSION

The machine learning system evaluated in this research has the potential to revolutionize the way sepsis is diagnosed and treated. With an impressive AUROC of 0.83, the system has shown a high degree of accuracy in predicting sepsis up to 48 hours before the onset of symptoms. This early warning capability is crucial in ensuring that patients receive timely and appropriate treatment, leading to improved health outcomes.

Future work could focus on the implementation of the algorithm in clinical practice to assess its practical utility and to validate its performance across multiple healthcare systems. Additionally, further exploration could be done to identify additional predictive variables that may enhance the algorithm's performance, as well as potential applications of the algorithm for other clinical conditions. Overall, the potential for machine learning algorithms to improve sepsis detection and patient outcomes warrants further investigation and development in the field of healthcare.

REFERENCES

[1] M. Singer, C.S. Deutschman, C.W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, et al., The third international consensus definitions for sepsis and septic shock (Sepsis-3), J. Am. Med. Assoc. 315 (8) (2016) 801–810, https://doi.org/10.1001/jama.2016.0287.

[2] J.N. Fullerton, C.L. Price, N.E. Silvey, S.J. Brace, G.D. Perkins, Is the Modified Early Warning Score (MEWS) superior to clinician judgement in detecting critical illness in the pre-hospital environment? Resuscitation 83 (2012) 557–562.

[3] F. Jaimes, J. Garcés, J. Cuervo, F. Ramirez, J. Ramirez, A. Vargas, et al. The systemic inflammatory response syndrome (SIRS) to identify infected patients in the emergency room, Intensive Care Med. 29 (2003) 1368–1371, https://doi.org/10.1007/s00134-003-1874-0.

[4] C.W. Seymour, V.X. Liu, T.J. Iwashyna, F.M. Brunkhorst, T.D. Rea, A. Scherag, et al., Assessment of clinical criteria for sepsis: for the third international consensus definitions for sepsis and septic shock (Sepsis-3), J. Am. Med. Assoc. 315 (8) (2016 Feb 23) 762–774.

[5] Sundas, A., & Panda, S. (2019). Iot and wsn based smart surveillance system for patients with closed-loop alarm. International Journal of Scientific & Technology Research, 8, 508-511.

[6] Nashif, S., Raihan, M. R., Islam, M. R., & Imam, M. H. (2018). Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system. World Journal of Engineering and Technology, 6(4), 854-873.

[7] Luo, J., Lan, L., Huang, S., Zeng, X., Xiang, Q., Li, M., ... & Zhou, X. (2023). Real-time prediction of organ failures in patients with acute pancreatitis using longitudinal irregular data. Journal of Biomedical Informatics, 139, 104310.

[8] Sundas, A., Badotra, S., Rani, S., & Gyaang, R. (2023). Evaluation of Autism Spectrum Disorder Based on the Healthcare by Using Artificial Intelligence Strategies. Journal of Sensors, 2023.

[9] Dahan, F., Alroobaea, R., Alghamdi, W., Mohammed, M. K., Hajjej, F., & Raahemifar, K. (2023). A smart IoMT based architecture for E-healthcare patient monitoring system using artificial intelligence algorithms. Frontiers in Physiology, 14, 40.

[10] Forkan, A. R. M., Khalil, I., & Atiquzzaman, M. (2017). ViSiBiD: A learning model for early discovery and real-time prediction of severe clinical events using vital signs as big data. Computer Networks, 113, 244-257.

[11] Alfian, G., Syafrudin, M., Ijaz, M. F., Syaekhoni, M. A., Fitriyani, N. L., & Rhee, J. (2018). A personalized healthcare monitoring system for diabetic patients by utilizing BLE-based sensors and real-time data processing. Sensors, 18(7), 2183.

[12] Alghatani, K., Ammar, N., Rezgui, A., & Shaban-Nejad, A. (2021). Predicting intensive care unit length of stay and mortality using patient vital signs: machine learning model development and validation. JMIR medical informatics, 9(5), e21347.

[13] Sundas, A., Badotra, S., Bharany, S., Almogren, A., Tag-ElDin, E. M., & Rehman, A. U. (2022). HealthGuard: An Intelligent Healthcare System Security Framework Based on Machine Learning. Sustainability, 14(19), 11934.

[14] da Silva, D. B., Schmidt, D., da Costa, C. A., da Rosa Righi, R., & Eskofier, B. (2021). DeepSigns: A predictive model based on Deep Learning for the early detection of patient health deterioration. Expert Systems with Applications, 165, 113905.

[15] Kim, S. Y., Kim, S., Cho, J., Kim, Y. S., Sol, I. S., Sung, Y., ... & Sohn, M. H. (2019). A deep learning model for real-time mortality prediction in critically ill children. Critical care, 23(1), 1-10.

[16] Ahmed, U., Lin, J. C. W., & Srivastava, G. (2023). Multivariate time-series sensor vital sign forecasting of cardiovascular and chronic respiratory diseases. Sustainable Computing: Informatics and Systems, 100868.

[17] Peng, H. T., Siddiqui, M. M., Rhind, S. G., Zhang, J., Teodoro da Luz, L., & Beckett, A. (2023). Artificial intelligence and machine learning for hemorrhagic trauma care. Military Medical Research, 10(1), 1-20.

[18] Staszak, K., Tylkowski, B., & Staszak, M. (2023). From Data to Diagnosis: How Machine Learning Is Changing Heart Health Monitoring. International Journal of Environmental Research and Public Health, 20(5), 4605.

[19] Sundas, A., & Badotra, S. (2022, October). Recurring Threats to Smart Healthcare Systems Based on Machine Learning. In 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 1-8). IEEE.

[20] Elhaj, H., Achour, N., Tania, M. H., & Aciksari, K. (2023). A comparative study of supervised machine learning approaches to predict patient triage outcomes in hospital emergency departments. Array, 17, 100281.

[21] Zhang, H., Wang, L. C., Chaudhuri, S., Pickering, A., Usvyat, L., Larkin, J., ... & Kotanko, P. (2023). Real-time prediction of intradialytic

hypotension using machine learning and cloud computing infrastructure. Nephrology Dialysis Transplantation, gfad070.

[22] Gupta, S., Shabaz, M., Gupta, A., Alqahtani, A., Alsubai, S., & Ofori, I. (2023). Personal HealthCare of Things: A novel paradigm and futuristic approach. CAAI Transactions on Intelligence Technology.

[23] Manjula, S., Vanitha, R., & Geetha, K. Prediction of Health Conditions Of Patient Using Iot Based Machine Learning.

[24] Im, J. E., Yoon, S. A., Shin, Y. M., & Park, S. (2023). Real-Time Prediction for Neonatal Endotracheal Intubation Using Multimodal Transformer Network. IEEE Journal of Biomedical and Health Informatics.

[25] Sundas, A., Badotra, S., & Singh, G. (2023, January). Sensor Data Transforming into Real-Time Healthcare Evaluation: A Review of Internet of Things Healthcare Monitoring Applications. In 2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE) (pp. 559-567). IEEE.

[26] Churpek MM, Yuen TC, Winslow C, et al. Multicenter development and validation of a risk stratification tool for ward patients. Am J Respir Crit Care Med 2014; 190(6): 649–655.

[27] Churpek MM, Yuen TC, Winslow C, et al. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. Crit Care Med 2016; 44(2):368–374.

[28] Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system. Chest 1991; 100(6):1619–1636.

[29] Subbe CP, Slater A, Menon D, et al. Validation of physiological scoring systems in the accident and emergency department. Emerg Med J 2006; 23(11): 841–845.

[30] Ferreira FL. Serial evaluation of the SOFA score to predict outcome in critically ill patients. JAMA 2001;286(14): 1754–1758.

[31] Marafino BJ, Park M, Davies JM, et al. Validation of prediction models for critical care outcomes using natural language processing of electronic health record data. JAMA Netw Open 2018; 1(8):e185097.

[32] Siontis GCM. Predicting death: an empirical evaluation of predictive tools for mortality. Arch Intern Med 2011; 171(19): 1721–1726.

[33] Mao Q, Jay M, Hoffman JL, et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ Open 2018; 8(1): e017833.

[34] Lehman L-WH, Adams RP, Mayaud L, et al. A physiological time series dynamics-based approach to patient monitoring and outcome prediction. IEEE J Biomed Health 2015; 19(3): 1068–1076.

[35] Johnson AE, Pollard TJ, Mark RG, et al. Reproducibility in critical care: a mortality prediction case study. In: Proceedings of the 2nd machine learning for healthcare conference (vol. 68), Boston, MA, 18–19 August 2017.

[36] Song, W., Jung, S. Y., Baek, H., Choi, C. W., Jung, Y. H., & Yoo, S. (2020). A predictive model based on machine learning for the early detection of late-onset neonatal sepsis: development and observational study. JMIR medical informatics, 8(7), e15965.

[37] Cheng, C. Y., Kung, C. T., Chen, F. C., Chiu, I. M., Lin, C. H. R., Chu, C. C., ... & Su, C. M. (2022). Machine learning models for predicting in-hospital mortality in patient with sepsis: Analysis of vital sign dynamics. Frontiers in Medicine, 9, 964667.

[38] Moor, M., Rieck, B., Horn, M., Jutzeler, C. R., & Borgwardt, K. (2021). Early prediction of sepsis in the ICU using machine learning: a systematic review. Frontiers in medicine, 8, 607952.

[39] Zhao, Y., Jia, L., Jia, R., Han, H., Feng, C., Li, X., ... & Li, T. (2022). A new time-window prediction model for traumatic hemorrhagic shock based on interpretable machine learning. Shock (Augusta, Ga.), 57(1), 48.

[40] Wardi, G., Carlile, M., Holder, A., Shashikumar, S., Hayden, S. R., & Nemati, S. (2021). Predicting progression to septic shock in the emergency department using an externally generalizable machine-learning algorithm. Annals of emergency medicine, 77(4), 395-406.