Improving sepsis classification performance with artificial intelligence algorithms: A comprehensive overview of healthcare applications

Anjana G. [a], Nisha K.L. [a,*], Arun Sankar M.S. [b,c]

[a] Department of Electronics and Communication Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, Kerala, India
[b] School of CS&IT, University College Cork, Cork, Ireland
[c] Electronic and Communications Department, South East Technological University, Carlow, Ireland

ABSTRACT

Purpose: This study investigates the potential of machine learning (ML) algorithms in improving sepsis diagnosis and prediction, focusing on their relevance in healthcare decision-making. The primary objective is to contribute to healthcare decision-making by evaluating the performance of various supervised and unsupervised models.
Materials and methods: Through an extensive literature review, optimal ML models used in sepsis research were identified. Diverse datasets from relevant sources were employed, and rigorous evaluation metrics, including accuracy, specificity, and sensitivity, were applied. Innovative techniques were introduced, such as a Stacked Blended Ensemble Model and Skopt Optimization with Blended Ensemble, incorporating Bayesian optimization for hyperparameter tuning.
Results: ML algorithms demonstrate efficacy in sepsis diagnosis, presenting an improved balance between specificity and sensitivity, critical for effective clinical decision-making. Classifier ensemble models show enhanced accuracy and efficiency, with novel optimization techniques contributing to improved adaptability.
Conclusion: The study emphasizes the potential benefits of ML algorithms in sepsis management, advocating for ongoing research to optimize performance and ensure ethical utilization in healthcare decision-making. Ethical considerations, interpretability, and transparency are crucial factors in implementing these algorithms in clinical practice.

1. Introduction

Sepsis [1,2] is a critical and time-sensitive medical condition characterized by organ dysfunction resulting from the body's dysregulated response to infection, demanding prompt interventions to mitigate mortality risks. The incidence of sepsis has steadily risen since the introduction of the first consensus definition (Sepsis-1) in 1991, reaching approximately 49 million reported cases and 11 million sepsis-related deaths worldwide [3] by 2017. The World Health Organization (WHO) declared sepsis a global health priority to address this. The journey toward a more accurate definition [4,5,7,9] can be traced through three pivotal stages: Sepsis-1 (1999), Sepsis-2 (2001), and Sepsis-3 (2016) (Appendix A).

Sepsis 1, in 1991, introduced sepsis as an "ongoing process." linked to the combination of Systemic Inflammatory Response Syndrome [5,6,9] (SIRS) criteria and known or suspected infection. Subsequent classifications, such as Sepsis-2 in 2001, sought to refine diagnostic criteria by incorporating various signs and symptoms, emphasizing biochemical indicators [8]. Sepsis 3 in 2016, recognizing organ dysfunction as central and introducing the Sequential Organ Failure Assessment (SOFA) [9,10] score, developed in 1994, as a tool for assessing organ failure. This classification addressed the limitations of its predecessors by focusing on organ dysfunction [11] rather than SIRS criteria. Recent studies emphasize the importance of recognizing distinct dysfunction patterns influenced by pathogen type, exposure duration, and host characteristics. Additionally, introducing the qSOFA (quick Sepsis-related Organ Failure Assessment) score was introduced for bedside screening outside the intensive care unit (ICU) for identifying organ dysfunction in sepsis cases.

Early diagnosis and management of sepsis are crucial, prompting exploration into artificial intelligence (AI) as a solution. This study provides a comprehensive overview of AI algorithms in sepsis classification, discussing their potential in healthcare decisions, recent developments, performance metrics, and ethical considerations. The main

contributions of this work are summarized in Table 1.

2. Artificial intelligence and machine learning

Recent studies underscore the potential of AI algorithms in sepsis diagnosis and prediction, utilizing supervised, unsupervised, and reinforcement learning approaches [12-16]. Despite challenges in data quality, privacy, security, and ethics, these algorithms excel in processing extensive patient datasets, facilitating early detection of sepsis patterns. AI has emerged as a significant player in healthcare decision-making [17-19] making substantial contributions to various domains such as Diagnosis and Treatment [20,21], Risk Prediction and Prevention [22-24], Patient Monitoring [25], Clinical Decision Support [14,18], and Workflow Optimization [26].The role of AI in healthcare decision-making can be summarized in Fig. 1.

Supervised learning [27], a task-driven approach, includes algorithms like logistic regression, random forest, naïve Bayes, and decision trees. These algorithms excel in predicting disease development based on clinical features and require labelled data for effective training. Logistic regression [28] and decision trees [27], known for their simplicity and robustness. Random Forest, an ensemble learning algorithm [29,30], excels in disease classification and identifying high-risk patients, while the probabilistic Naïve Bayes algorithm [16,27] finds applications in disease diagnosis, risk prediction, and treatment outcome analysis.

Conversely, unsupervised learning algorithms [31] are data-driven approaches to uncover patterns and relationships within healthcare datasets without labelled data. Clustering algorithms [32], such as K-means, group similar patients, diseases, or symptoms, aiding in the identification of patient subgroups and disease subtypes for sepsis prediction.

Reinforcement learning (RL) algorithms [33,34] offer a framework for sequential healthcare decision-making based on learning through trial and error. Q-Learning optimizes healthcare decision-making by learning from patient responses to treatments [35]. Deep Q-Networks (DQNs) use neural networks to estimate the Q-function, enabling complex decision-making in tasks like drug dosing and patient risk stratification.

Recent studies have explored diverse AI approaches for sepsis prediction. Scherpf et al. [36] employed a recurrent neural network (RNN) achieving an 87% accuracy but raising concerns about its black-box nature. XGBoost by Hou et al. [37,38] and probabilistic modelling by Yao et al. [39], showcased promising results. Li et al. [40] achieved an AUC of 0.85 with gradient-boosting decision trees (GBDT). However, challenges persist due to uncertainties in model performance, limited algorithmic information in studies by Mao et al. [41], Barton et al. [42], Delahanty et al. [43], Bedoya et al. [44], Su et al. [45], Zhao et al. [46],
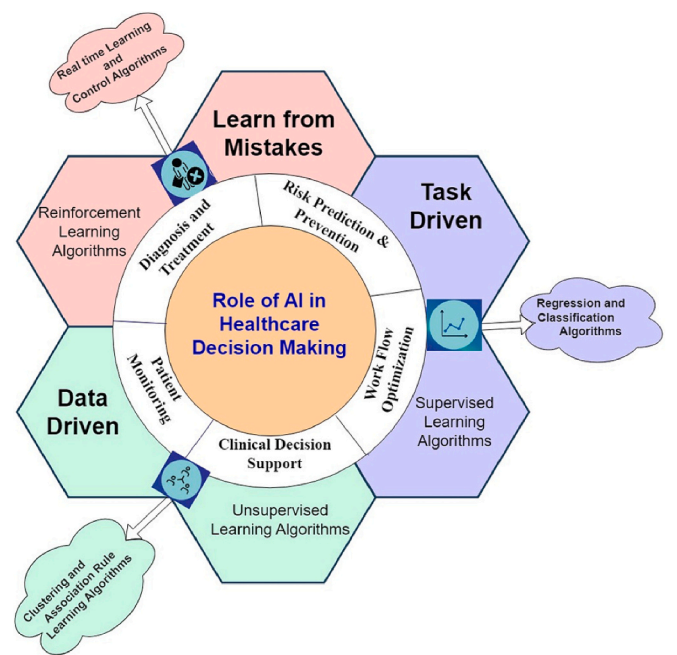
Table 1
Main contributions of the work.

Contribution	Description
Critical state-of-the-art review	Initiates a thorough exploration of the latest AI algorithms, spanning supervised, unsupervised, and reinforcement learning, to enhance overall classification performance.
Introduction of novel approaches	Introduces innovative methodologies like Stacked Blended Ensemble Approach, Skopt Optimization with Blended Ensemble Approach, and Skopt Optimization Stacked Blended Ensemble Approaches, aiming to elevate predictive performance through optimized model integration.
Performance evaluation	Systematically assesses the performance of diverse AI algorithms in sepsis classification, focusing on achieving a balanced trade-off between sensitivity and specificity.
Discussion on future directions	Discuss AI's potential future directions and applications in sepsis classification, emphasizing their impact on healthcare decision-making.

Fig. 1. Healthcare decision-making and role of AI algorithms.

and Kijpaisalratana et al. [47] hindering reproducibility. Incorporating human-in-the-loop approaches, such as Gupta et al. [48] human-in-the-loop artificial intelligence system for sepsis treatment, recognizes the need for clinical expertise in decision-making. However, its real-world performance remains uncertain without clinical trials, and the system lacks interpretability, making the reasons behind its predictions unclear.

3. Simulation background and methodology

In this section, we delve into the methodological approaches of our simulation, providing a comprehensive background for understanding its application in predicting sepsis and evaluating machine learning models. The workflow of this research work is summarized in Fig.2.

3.1. Data collection and pre-processing

In this work, we processed a sepsis dataset from the Physio Net/Computing in Cardiology Challenge [49,50]. The dataset was publicly available for download, which included de-identified and labelled data for 40,336 patients (non-sepsis-36,786 and sepsis-3550). The dataset comprised 40 clinical factors, combining hourly details of eight vital sign parameters, 26 laboratory metrics, and six demographic attributes (Appendix B).

Challenges within the dataset included missing values, varied units of measurement, and categorical variables. Representativeness was limited as the data came from specific sources, raising concerns about generalizability. The dataset lacked comprehensive patient histories, detailed insights into antibiotic administration, and a broader clinical context. Additional information on pre-existing conditions, comorbidities, antibiotic treatment impact, and sepsis onset context would be beneficial to improve interpretability.

Summary statistics reveal significant imbalances in the dataset, particularly in the distribution between sepsis and non-sepsis classes. This imbalance can lead to biased models favoring the majority class and poor performance in the minority class. To address this, we utilize SMOTE [51] (Synthetic Minority Over-sampling Technique. It generates synthetic samples for the minority class by interpolating between each minority instance and its k-nearest neighbours, effectively balancing the class distribution and addressing the issue of data imbalance. Fig. 3
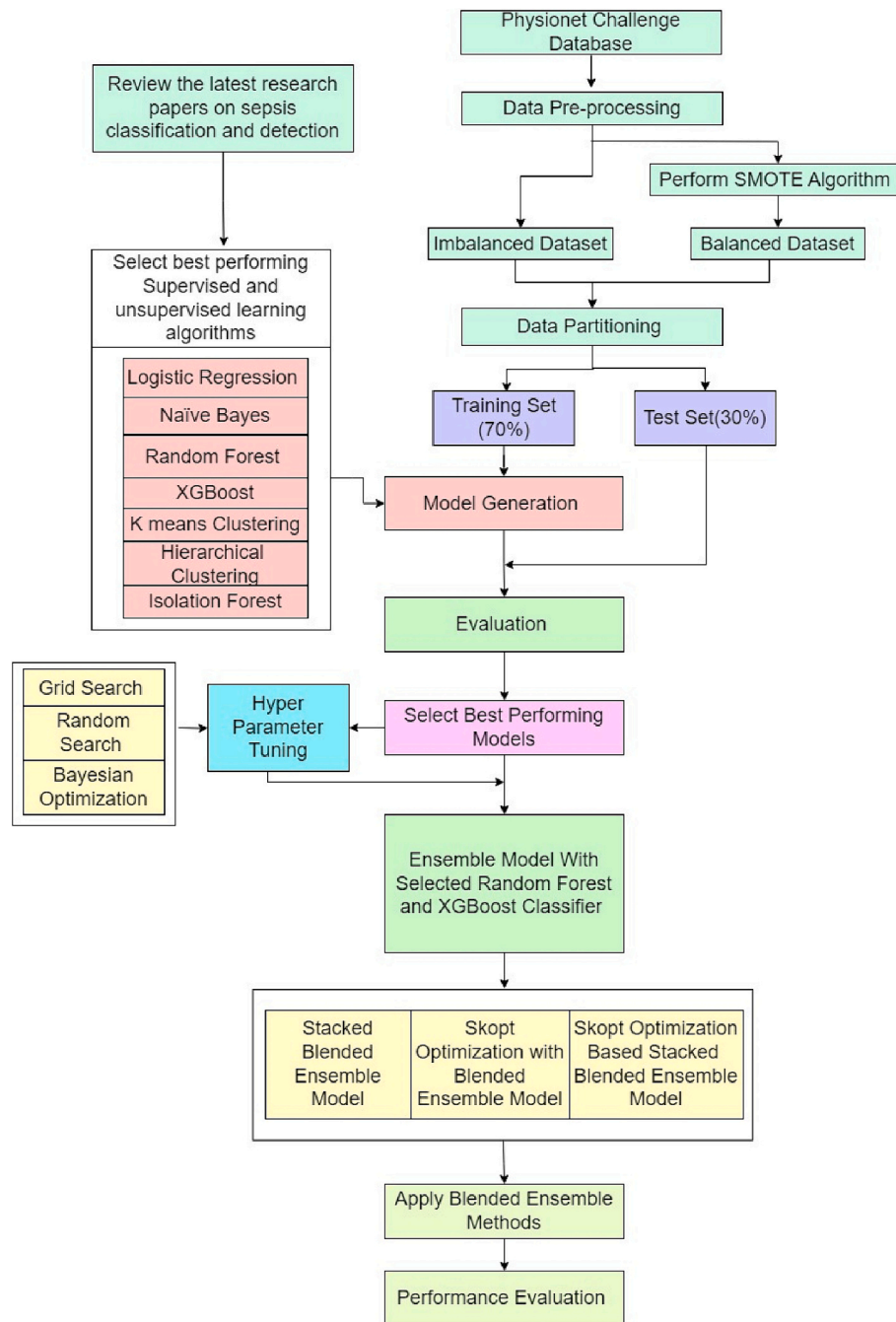
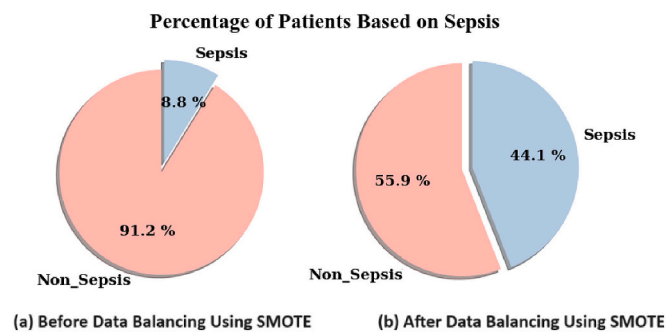**Fig. 2.** Proposed workflow for efficient sepsis classification.

Fig. 3. Distribution of Sepsis and Non- sepsis classes in the dataset before and after data balancing using SMOTE.

shows the dataset class distribution before and after data balancing.

Finally, the pre-processed data was split into an 80% training set and a 20% test set to train and evaluate the performance of the classification models. In this work, two methods were simulated. The first used the original dataset for Data partitioning, while the second used the SMOTE algorithm to balance the dataset before splitting it into corresponding training and testing sets.

3.2. Algorithm selection, model generation, and implementation

The initial step in this phase involved considering well-established machine learning algorithms commonly used in classification tasks, including supervised and unsupervised techniques (Appendix C). The algorithms were rigorously evaluated based on a comprehensive set of performance metrics, including specificity, sensitivity, accuracy,

precision, recall, F1 score, AUC-ROC, mean absolute error, and root mean squared error. Subsequently, the algorithms demonstrating superior performance—characterized by high accuracy, precision, recall, optimal sensitivity and specificity, and adeptness in handling imbalanced data—were identified for further analysis and implementation. Hyperparameter tuning was executed using various established techniques, including Grid Search [52], Random Search [52,53], and Bayesian Optimization [52,54]. These methods, while diverse in their approach, shared the common goal of systematically exploring and optimizing the hyperparameter space to ensure the models' optimal performance.

Following individual model refinement, an ensemble learning approach [55] was adopted to synergize the strengths of multiple classifiers. The utilization of ensemble models aims to fortify predictive capabilities by amalgamating diverse base models, capturing nuanced aspects of the data for improved generalization and performance. Three distinctive ensemble approaches were explored in-depth:

Stacked Blended Ensemble Model [54,55]: The Stacked Blended Ensemble Model combines logistic regression as a meta-model with Random Forest [29] and XGBoost classifiers [57] as base models. Base model predictions are used as features for the meta-model, enabling optimal combination through stacking and blending. Evaluation involves Stratified K-Fold cross-validation and comprehensive metric assessment.

Skopt Optimization with Blended Ensemble Approach [56,58-60]: This approch employs sequential model-based optimization (SMBO) [59] through Scikit-Optimize [54,60] to iteratively optimize hyperparameters for Random Forest and XGBoost models. It utilizes a probabilistic Bayesian optimization strategy, incorporating a Gaussian Process as a surrogate model. The optimization iterates through fitting the surrogate model, selecting hyperparameters, and updating the model based on an acquisition function like Expected Improvement. Once optimized individually, the models' predictions are combined using a weighted average, determined automatically by Skopt, potentially enhancing overall predictive performance.

Skopt Optimization based Stacked Blended Ensemble Approach [54,58,60]: This approach combines the two powerful concepts of Bayesian optimization using scikit-optimize [60] (skopt) and stacked blended ensemble. Like the previous approach, Skopt is used to optimize the hyperparameters of individual models. However, instead of directly blending the predictions of these models by weighted average, a meta-model (a logistic regression model) is trained with predictions from optimized Random Forest and XGBoost models, capturing higher-order interactions between models through stacking.

4. Simulation performance for predicting sepsis

This section presents the performance evaluation of various classification models for predicting sepsis. Table 2 shows the performance metrics of various classification algorithms on two different datasets, one balanced and one imbalanced.

The performance evaluation results from Table 2 highlight distinct characteristics of various classifiers in sepsis detection. Random Forest Classifier exhibits exceptional accuracy (0.956) on balanced datasets, demonstrating balanced sensitivity (0.957) and specificity (0.956). However, its high accuracy (0.982) on imbalanced datasets comes at the cost of low sensitivity (0.010), indicating challenges in identifying the minority class. Conversely, XGBoost demonstrates balanced precision (0.857) and recall (0.719) on balanced datasets, with relatively high specificity (0.925) and sensitivity (0.719). Yet, it struggles with sensitivity (0.006) on imbalanced datasets despite maintaining high accuracy (0.982). Naive Bayes Classifier performs well on imbalanced datasets (0.929) but poorly on balanced datasets (0.134), excelling in accuracy but struggling with sensitivity. Logistic Regression achieves reasonable accuracy (0.753) on balanced datasets, with balanced precision and recall, but on imbalanced datasets, it maintains high accuracy (0.982) while exhibiting very low sensitivity (0.001). K-means and Hierarchical Clustering show relatively poor sensitivity on both dataset types, highlighting limitations in correctly classifying the minority class. Isolation Forest demonstrates moderate sensitivity (0.191) and high specificity (0.945) on balanced datasets, suitable for classifying both classes. However, on imbalanced datasets, it faces challenges in sensitivity (0.188) despite maintaining relatively high specificity (0.903). In summary, Random Forest exhibits high sensitivity, indicating its ability to capture intricate patterns in imbalanced datasets, while XGBoost contributes to maintaining high accuracy with balanced precision and recall. Combining Random Forest and XGBoost into an ensemble model could leverage their complementary strengths, potentially enhancing classifier reliability and accuracy.

The ROC curves of the classifiers shown in Fig.4 reveal varying performance levels in the sepsis detection task.

The Random Forest and XGBoost classifiers effectively balance true positive rate (TPR) and false positive rate (FPR), showcasing their robust performance in classifying instances. Conversely, K-means Clustering

Table 2
Performance evaluation on balanced and imbalanced dataset (the best results are highlighted).

|		Accuracy	Precision	Recall	F1 Score	Specificity	Sensitivity	AUC-ROC	Mean absolute error	Root mean squared error
Random forest classifier	Balanced dataset	0.956	0.913	0.957	0.935	**0.956**	**0.957**	0.956	0.044	0.209
	Imbalanced dataset	0.982	0.744	0.010	0.010	0.978	0.010	0.505	0.018	0.134
XGBoost	Balanced dataset	0.897	0.857	0.719	0.768	**0.933**	**0.820**	0.822	0.143	0.378
	Imbalanced dataset	0.982	0.531	0.006	0.012	0.989	0.006	0.503	0.018	0.134
Logistic regression	Balanced dataset	0.753	0.753	0.389	0.508	0.931	0.389	0.660	0.247	0.497
	Imbalanced dataset	0.982	0.182	0.001	0.003	0.988	0.001	0.501	0.018	0.135
Naive Bayes classifier	Balanced dataset	0.134	0.704	0.432	0.536	0.911	0.496	0.672	0.246	0.496
	Imbalanced dataset	0.929	0.052	0.171	0.080	0.943	0.171	0.557	0.071	0.267
K-means clustering	Balanced dataset	0.668	0.518	0.038	0.070	0.913	0.662	0.510	0.332	0.576
	Imbalanced dataset	0.974	0.037	0.018	0.025	0.991	0.018	0.505	0.026	0.162
Hierarchical clustering	Balanced dataset	0.668	0.519	0.035	0.066	0.984	0.035	0.509	0.332	0.576
	Imbalanced dataset	0.965	0.026	0.018	0.023	0.982	0.016	0.510	0.028	0.162
Isolation forest	Balanced dataset	0.694	0.636	0.191	0.294	0.945	0.191	0.568	0.306	0.553
	Imbalanced dataset	0.890	0.034	0.188	0.058	0.903	0.188	0.545	0.110	0.332
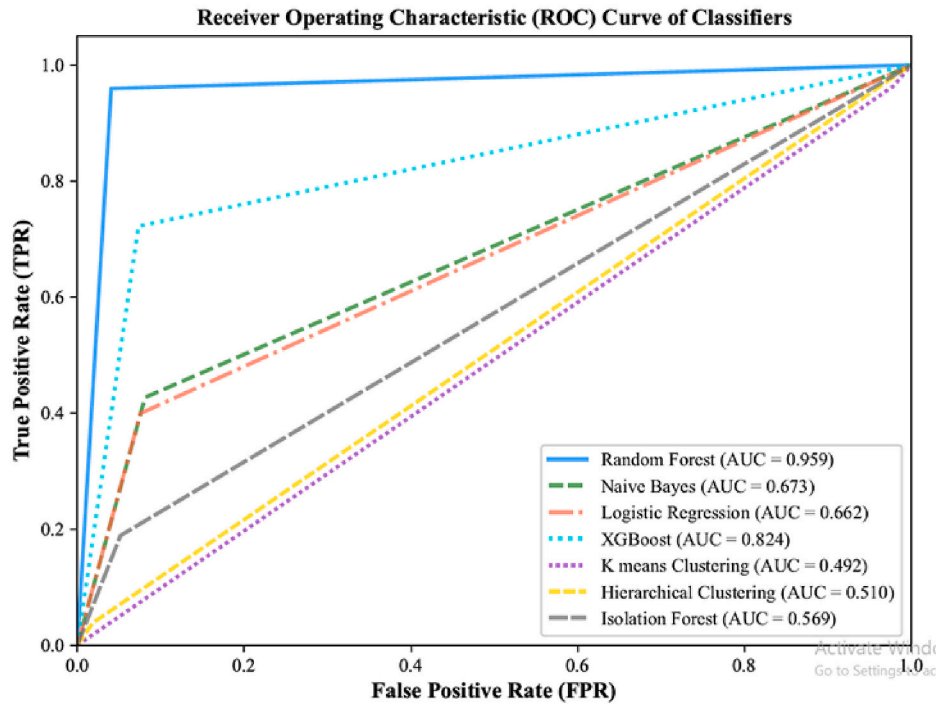
Fig. 4. Comparative ROC curves of sepsis detection classifiers.

and Hierarchical Clustering classifiers exhibit inadequate discrimination between positive and negative instances. Naive Bayes, Logistic Regression, and Isolation Forest classifiers demonstrate moderate performance with varying AUC values. Naive Bayes shows moderate discriminative power but lags the Random Forest classifier in accuracy. Logistic Regression performs moderately better than Naive Bayes but falls short compared to Random Forest. Isolation Forest displays moderate discriminative power but does not match the accuracy of Random Forest or XGBoost classifiers.

The sensitivity and specificity trade-off of hyper parameter optimizations obtained through the simulations is tabulated in Table 3. This hyperparameter tuning focuses on finding the optimal values for a single model.

The results in Table 3 indicate that, considering sensitivity, specificity, and the trade-off, the Bayesian Optimization process is the preferred choice among the provided hyperparameter tuning methods for achieving a balanced classifier in sepsis scenarios. A blended ensemble method combining multiple models was considered to enhance its predictive performance. Implementation results of these blended ensemble models in terms of specificity and sensitivity are tabulated in Table 4.

In terms of sensitivity and specificity the Skopt Optimization-based Stacked Blended Ensemble Approach offers a compelling trade-off, maintaining a sensitivity of 0.965 and specificity of 0.965. Considering the need for a balanced trade-off between sensitivity and specificity, the Skopt optimization-based stacked Blended Ensemble

Table 3
Performance evaluation of hyperparameter tuning methods.

Hyper parameter tuning methods	Sensitivity		Specificity	
	Random forest classifier	XGBoost classifier	Random forest classifier	XGBoost classifier
Grid search	0.546	0.871	0.961	0.950
Random search	0.521	0.982	0.941	0.962
Bayesian optimization	0.567	0.978	0.940	0.963

Table 4
Performance evaluation of blended ensemble methods.

Ensemble models	Sensitivity	Specificity
Stacked blended ensemble model	0.983	0.961
Skopt optimization with blended ensemble approach	0.975	0.956
Skopt optimization based stacked blended ensemble approach	0.965	0.965

Approach appears to be the best performer and excellent choice, demonstrating strong predictive capabilities for both positive and negative instances in the case of sepsis classification.

5. Discussion

Spanning supervised learning, unsupervised learning, and reinforcement learning establishes a thorough understanding of the diverse landscape of sepsis classification. The evaluation of AI algorithms for sepsis classification underscores the effectiveness of ensemble methods, notably the Random Forest and XGBoost classifiers. These models exhibit robust performance across various metrics, demonstrating their superiority over unsupervised algorithms like K-means, Hierarchical Clustering, and Isolation Forest. The decision to blend Random Forest and XGBoost in an ensemble model is justified by their complementary strengths, as evidenced by their balanced performance. Bayesian Optimization emerges as a promising hyperparameter tuning method, achieving a balanced sensitivity and specificity trade-off suitable for the complex nature of sepsis.

Moreover, blended ensemble models were implemented to achieve a balanced sensitivity and specificity trade-off. The stacked Blended Ensemble Model achieves exceptional sensitivity, implying its efficacy in correctly identifying true positive cases, however occasionally misclassifying true negative cases. The Skopt Optimization with Blended Ensemble Approach maintains a commendable balance between sensitivity and specificity, favoring a more conservative approach in classifying negative cases. The Skopt Optimization-based Stacked Blended Ensemble Approach achieves a remarkable balance between sensitivity

and specificity, making it ideal for minimizing both false negatives and false positives in sepsis classification.

However, limitations exist, particularly in handling imbalanced datasets and performance variability across different datasets due to the dynamic nature of clinical data. Advanced machine learning techniques are necessary to enhance accuracy and robustness. Reinforcement learning (RL)-based methods show promise, as they can adapt decision-making strategies based on feedback and effectively handle the dynamic nature of clinical data.

Future research directions could explore the integration of reinforcement learning techniques in sepsis classification to improve classifier performance and patient outcomes. These techniques can optimize treatment decisions tailored to individual patients, considering real-time changes in patient data. Further validation and analysis across diverse datasets and real-world scenarios are essential to establish the practical applicability of these approaches in clinical settings.

In summary, the study highlights the superiority of ensemble methods, particularly Random Forest and XGBoost classifiers, in sepsis classification. The introduced ensemble learning techniques show promise in enhancing classifier performance and overall resilience. Sequential decision-making models like reinforcement learning-based methods offer new avenues for advancements in sepsis classification, potentially benefiting patient outcomes with better trade-offs between sensitivity and specificity.

6. Conclusion

The paper provides a comprehensive exploration of AI algorithms in sepsis classification, highlighting their pivotal role in healthcare decision-making for this complex medical condition. It emphasizes the inefficiencies of current approaches reliant on manual review and clinical judgment, advocating for innovative solutions.

Through rigorous evaluation, Random Forest and XGBoost emerge as superior models for sepsis detection, with three novel ensemble approaches proposed for further enhancement. Among these, the Skopt Optimization-based Stacked Blended Ensemble achieves an optimal balance between sensitivity and specificity, offering promise in minimizing false negatives and false positives.

Acknowledging model limitations, especially in handling imbalanced datasets and variability across clinical scenarios, underscores the necessity for ongoing research and advanced machine-learning techniques. Reinforcement learning stands out as a promising avenue, given its adaptability to dynamic clinical data.

In conclusion, this paper underscores the transformative potential of AI in sepsis classification, emphasizing the need for continued innovation, ethical awareness, and rigorous exploration of advanced techniques to enhance patient outcomes.

Author contribution

Anjana G: Contributed to the conceptualization and design of the study, acquisition, and data analysis and participated in the writing, review, and editing process. **Nisha K L**: Conceptualized the study, validated the study, supervised the analysis, participated in the review and editing process, and provided final approval for the version to be submitted. **Arun Sankar M S**: Supervised the analysis, participated in the review and editing process, and provided final approval for the version to be submitted.

Funding and Financial Disclosure

CRediT authorship contribution statement

Anjana G.: Writing – original draft, Visualization, Methodology,

Investigation, Formal analysis, Data curation, Conceptualization. **Nisha K.L.:** Writing – review & editing, Validation, Supervision, Conceptualization. **Arun Sankar M.S.:** Writing – review & editing, Validation, Supervision.

Declaration of competing interest

The authors declare that they have no conflicts of interest.

Data Availability

Data will be available upon request to Anjana G (anjanag2255@gmail.com).

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jcrc.2024.154815.

References

[1] Gauer R, Forbes D, Boyer N. Sepsis: diagnosis and management. Am Fam Physician 2020;101(7):409–18.

[2] Kamath S, Hammad Altaq H, Abdo T. Management of sepsis and septic shock: what have we learned in the last two decades? Microorganisms 2023;11(9):2231.

[3] Guarino M, Perna B, Cesaro AE, Maritati M, Spampinato MD, Contini C, et al. 2023 update on sepsis and septic shock in adult patients: management in the emergency department. J Clin Med 2023;12(9):3188. https://doi.org/10.3390/jcm12093188.

[4] Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA 2016;315(8):801–10. https://doi.org/10.1001/jama.2016.0287.

[5] Marik PE, Taeb AM. SIRS, qSOFA and new sepsis definition. J Thorac Dis 2017;9 (4):943–5. https://doi.org/10.21037/jtd.2017.03.125.

[6] Bone RC, Balk RA, Cerra FB, Dellinger RP, Fein AM, Knaus WA, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. Chest 1992;101(6):1644–55. https://doi.org/10.1378/chest.101.6.1644.

[7] Gül F, Arslantaş MK, Cinel İ, Kumar A. Changing definitions of sepsis. Turk J Anaesthesiol Reanim 2017;45(3):129.

[8] Linde-Zwirble WT, Angus DC. Severe sepsis epidemiology: sampling, selection, and society. Crit Care 2004;8(4):1–5.

[9] Churpek MM, Snyder A, Han X, Sokol S, Pettit N, Howell MD, et al. Quick sepsis-related organ failure assessment, systemic inflammatory response syndrome, and early warning scores for detecting clinical deterioration in infected patients outside the intensive care unit. Am J Respir Crit Care Med 2017;195(7):906–11. https://doi.org/10.1164/rccm.201604-0854OC.

[10] Song J-U, Sin CK, Park HK, Shim SR, Lee J. Performance of the quick sequential (sepsis-related) organ failure assessment score as a prognostic tool in infected patients outside the intensive care unit: A systematic review and meta-analysis. Critic Care (London, England) 2018;22(1). https://doi.org/10.1186/s13054-018-1952-x.

[11] Finkelsztein EJ, Jones DS, Ma KC, Pabón MA, Delgado T, Nakahira K, et al. Comparison of qSOFA and SIRS for predicting adverse outcomes of patients with suspicion of sepsis outside the intensive care unit. Crit Care 2017;21(1):73. https://doi.org/10.1186/s13054-017-1658-5.

[12] Duncan CF, Youngstein T, Kirrane MD, Lonsdale DO. Diagnostic challenges in sepsis. Curr Infect Dis Rep 2021;23(12):22. https://doi.org/10.1007/s11908-021-00765-y.

[13] Ramlakhan S, Saatchi R, Sabir L, Singh Y, Hughes R, Shobayo O, et al. Understanding and interpreting artificial intelligence, machine learning and deep learning in emergency medicine. Emerg Med J EMJ 2022;39(5):380–5. https://doi.org/10.1136/emermed-2021-212068.

[14] Wu M, Du X, Gu R, Wei J. Artificial intelligence for clinical decision support in sepsis. Front Med 2021;8:665464. https://doi.org/10.3389/fmed.2021.665464.

[15] Alloghani M, Al-Jumeily D, Mustafina J, Hussain A, Aljaaf AJ. A systematic review on supervised and unsupervised machine learning algorithms for data science. In: Supervised and Unsupervised Learning for Data Science; 2020. p. 3–21.

[16] Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 2019;19 (1):1–16.

[17] Masic I. Medical decision making - an overview. Acta Inform Med AIM J Soc Med Inform Bosnia & Herzegovina: casopis Drustva za medicinsku informatiku BiH 2022;30(3):230–5. https://doi.org/10.5455/aim.2022.30.230-235.

[18] Giordano C, Brennan M, Mohamed B, Rashidi P, Modave F, Tighe P. Accessing artificial intelligence for clinical decision-making. Front Digital Health 2021;3: 645232. https://doi.org/10.3389/fdgth.2021.645232.

[19] Sankaran R, Kumar A, Parasuram H. Role of artificial intelligence and machine learning in the prediction of the pain: a scoping systematic review. Proc Inst Mech Eng Part H J Eng Med 2022;236(10):1478–91. https://doi.org/10.1177/ 09544119221122012.

[20] Kiran Kumar M, Divya Udayan J. A study on machine learning techniques in cancer disease prediction and diagnosis. Ind J Public Health Res Dev 2019;10(4):157–62.

[21] Al-Antari MA. Artificial intelligence for medical diagnostics-existing and future AI technology! Diagnostics (Basel, Switzerland) 2023;13(4):688. https://doi.org/ 10.3390/diagnostics13040688.

[22] Moor M, Rieck B, Horn M, Jutzeler CR, Borgwardt K. Early prediction of sepsis in the ICU using machine learning: a systematic review. Front Med 2021;8:607952. https://doi.org/10.3389/fmed.2021.607952.

[23] Jadi A, Zedan H, Alghamdi T. Risk management based early warning system for healthcare industry. In: In 2013 International Conference on Computer Medical Applications (ICCMA). IEEE; 2013. p. 1–6.

[24] Mohamed AKS, Mehta AA, James P. Predictors of mortality of severe sepsis among adult patients in the medical intensive care unit. Lung India Off Organ Indian Chest Soc 2017;34(4):330–5. https://doi.org/10.4103/lungindia.lungindia_54_16.

[25] Shriram KV, Mural N, Subashri V, Sivaraman R. Design and development of an embedded system for monitoring the health status of a patient. Int J Intel Syst Appl 2013;5:64–71.

[26] Zhai, K., Yousef, M. S., Mohammed, S., Al-Dewik, N. I., & Qoronfleh, M. W. Optimizing clinical workflow using precision medicine and advanced data analytics. Processes, 11(3), 939.

[27] Osisanwo FY, Akinsola JET, Awodele O, Hinmikaiye JO, Olakanmi O, Akinjobi J. Supervised machine learning algorithms: classification and comparison. Int J Comput Trends Technol (IJCTT) 2017;48(3):128–38.

[28] Tolles J, Meurer WJ. Logistic regression: relating patient characteristics to outcomes. Jama 2016;316(5):533–4.

[29] Breiman L. Random forests. In: Machine Learning. 45; 2001. p. 5–32.

[30] Kuzhali JV, Vengataasalam S. A novel ensemble classifier based classification on large datasets with hybrid feature selection approach. Res J Appl Sci Eng Technol 2014;7(17):3633–42.

[31] Amruthnath N, Gupta T. A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. In: In 2018 5th International Conference on Industrial Engineering and Applications (ICIEA). IEEE; 2018. p. 355–61.

[32] Papin G, Bailly S, Dupuis C, Ruckly S, Gainnier M, Argaud L, et al. … a study group clinical and biological clusters of sepsis patients using hierarchical clustering. PloS One 2021;16(8):e0252793. https://doi.org/10.1371/journal.pone.0252793.

[33] Liu S, See KC, Ngiam KY, Celi LA, Sun X, Feng M. Reinforcement learning for clinical decision support in critical care: comprehensive review. J Med Internet Res 2020;22(7):e18477.

[34] Shortreed SM, Laber E, Lizotte DJ, Stroup TS, Pineau J, Murphy SA. Informing sequential clinical decision-making through reinforcement learning: an empirical study. Mach Learn 2011;84:109–36.

[35] Raghu A, Komorowski M, Singh S. Model-based reinforcement learning for sepsis treatment. In: arXiv preprint; 2018. arXiv:1811.09602.

[36] Scherpf M, Gräßer F, Malberg H, Zaunseder S. Predicting sepsis with a recurrent neural network using the MIMIC III database. Comput Biol Med 2019;113:103395.

[37] Johnson AE, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Scient Data 2016;3(1):1–9.

[38] Hou N, Li M, He L, Xie B, Wang L, Zhang R, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. J Transl Med 2020;18(1):1–14.

[39] Yao S, Yang JB, Xu DL, Dark P. Probabilistic modeling approach for interpretable inference and prediction with data for sepsis diagnosis. Expert Syst Appl 2021;183: 115333.

[40] Li K, Shi Q, Liu S, Xie Y, Liu J. Predicting in-hospital mortality in ICU patients with sepsis using gradient boosting decision tree. Medicine 2021;100(19).

[41] Mao Q, Jay M, Hoffman JL, Calvert J, Barton C, Shimabukuro D, 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.

[42] Barton C, Chettipally U, Zhou Y, Jiang Z, Lynn-Palevsky A, Le S, et al. Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. Comput Biol Med 2019;109:79–84.

[43] 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):334–44.

[44] Bedoya AD, Futoma J, Clement ME, Corey K, Brajer N, Lin A, et al. Machine learning for early detection of sepsis: an internal and temporal validation study. JAMIA Open 2020;3(2):252–60.

[45] Su L, Xu Z, Chang F, Ma Y, Liu S, Jiang H, 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 2021;8:664966.

[46] Zhao X, Shen W, Wang G. Early prediction of sepsis based on machine learning algorithm. Comput Intell Neurosci 2021;13:6522633. https://doi.org/10.1155/ 2021/6522633.

[47] Kijpaisalratana N, Sanglertsinlapachai D, Techaratsami S, Musikatavorn K, Saoraya J. Machine learning algorithms for early sepsis detection in the emergency department: a retrospective study. Int J Med Inform 2022;160:104689.

[48] Gupta A, Lash MT, Nachimuthu SK. Optimal sepsis patient treatment using human-in-the-loop artificial intelligence. Expert Syst Appl 2021;169:114476.

[49] Reyna MA, Josef CS, Jeter R, Shashikumar SP, Westover MB, Nemati S, et al. Early prediction of sepsis from clinical data: the PhysioNet/computing in cardiology challenge. Crit Care Med 2019;48(2):210–7. https://doi.org/10.1097/ CCM.0000000000004145.

[50] Reyna M, Josef C, Jeter R, Shashikumar S, Moody B, Westover MB, et al. Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019 (version 1.0.0). PhysioNet 2019. https://doi.org/10.13026/v64v-d857.

[51] Yi X, Xu Y, Hu Q, Krishnamoorthy S, Li W, Tang Z. ASN-SMOTE: a synthetic minority oversampling method with adaptive qualified synthesizer selection. Complex Intel Syst 2022;8(3):2247–72. https://doi.org/10.1007/s40747-021-00638-w.

[52] Elgeldawi E, Sayed A, Galal AR, Zaki AM. Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis. Inform MDPI 2021;8(4):79. https://doi.org/10.3390/informatics8040079.

[53] Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res 2012;13(2).

[54] Yao J, Zhang X, Luo W, Liu C, Ren L. Applications of stacking/blending ensemble learning approaches for evaluating flash flood susceptibility. Int J Appl Earth Observ Geoinform ITC J 2022;112(102932):102932. https://doi.org/10.1016/j. jag.2022.102932.

[55] Mahajan P, Uddin S, Hajati F, Moni MA. Ensemble learning for disease prediction: a review. Healthcare (Basel, Switzerland) 2023;11(12). https://doi.org/10.3390/ healthcare11121808.

[56] Mahesh TR, Kumar D, Vinoth Kumar V, Asghar J, Mekcha Bazezew B, Natarajan R, et al. Blended ensemble learning prediction model for strengthening diagnosis and treatment of chronic diabetes disease. Comput Intell Neurosci 2022:1–9. https:// doi.org/10.1155/2022/4451792.

[57] Rangan ES, Pathinarupothi RK, Anand KJS, Snyder MP. Performance effectiveness of vital parameter combinations for early warning of sepsis-an exhaustive study using machine learning. JAMIA Open 2022;5(4):ooac080. https://doi.org/ 10.1093/jamiaopen/ooac080.

[58] Lim Y-F, Ng CK, Vaitesswar US, Hippalgaonkar K. Extrapolative Bayesian optimization with Gaussian process and neural network ensemble surrogate models. Adv Intel Syst (Weinheim an Der Bergstrasse, Germany) 2021;3(11): 2100101. https://doi.org/10.1002/aisy.202100101.

[59] Lacoste A, Larochelle H, Laviolette F, Marchand M. Sequential Model-Based Ensemble Optimization. 2014. https://doi.org/10.48550/ARXIV.1402.0796.

[60] Head T, MechCoder Louppe G, Shcherbatyi I, Fcharras Vinícius Z, Fabisch. A scikit-optimize/scikit-optimizev0.5.2; 2018. https://doi.org/10.5281/ ZENODO.1207017.