International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

Deep Learning-Based Sepsis Prediction Using Electronic Health Records

D. Priyadharshini[1], Dr. S. Subbaiah[2]

[1]Research Scholar, Department of Computer Science, SRM Institute of Science and Technology, Ramapuram, Chennai, India

[2]Associate Professor, Head of the Department (AI and ML), SRM Institute of Science and Technology, Ramapuram, Chennai, India

***Abstract*— Sepsis is a life-threatening medical condition caused by the body's dysregulated response to infection, leading to organ dysfunction, septic shock, and potentially death if not treated promptly. According to global healthcare studies, sepsis remains one of the leading causes of mortality among hospitalized patients, particularly in intensive care units (ICUs). Early diagnosis is essential because every hour of delayed treatment significantly increases mortality risk. Traditional clinical scoring systems such as the Sequential Organ Failure Assessment (SOFA) and quick SOFA (qSOFA) are widely used but often identify sepsis after significant physiological deterioration has already occurred. Recent advances in Artificial Intelligence (AI), especially deep learning, provide new opportunities for earlier and more accurate sepsis prediction using Electronic Health Records (EHRs). EHRs contain longitudinal patient information such as demographics, laboratory results, medications, vital signs, diagnoses, and clinical notes, enabling deep learning models to recognize subtle temporal patterns preceding sepsis onset. Systematic reviews report that machine learning and deep learning models can often predict sepsis 2–6 hours before clinical diagnosis, with strong discriminative performance, although challenges remain around generalization and clinical deployment.**

***Keywords*—Sepsis Prediction, Deep Learning, Electronic Health Records (EHR), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Transformer Networks, Healthcare Analytics, Artificial Intelligence, Clinical Decision Support System, Early Disease Detection, Time-Series Analysis, Predictive Healthcare.**

I. Introduction

Healthcare organizations generate enormous volumes of patient information through Electronic Health Records. These records include laboratory reports, vital signs, medication histories, physician notes, diagnostic imaging summaries, and demographic information. Unlike traditional paper-based medical records, EHRs provide continuous patient monitoring and enable data-driven healthcare.

Sepsis develops when an infection triggers an excessive immune response, resulting in widespread inflammation, tissue damage, organ failure, and shock. It affects millions of patients annually and is associated with high mortality rates and substantial healthcare costs. The challenge in managing sepsis lies in its rapid progression and nonspecific early symptoms, which often resemble less severe illnesses.

Conventional diagnostic methods rely heavily on physician experience and predefined clinical criteria. However, these approaches may fail to identify complex interactions among physiological variables.

Artificial Intelligence has transformed predictive healthcare by learning intricate relationships from large datasets. Deep learning models are especially suited for EHR analysis because they can process large, high-dimensional, time-dependent clinical data without requiring extensive manual feature engineering.

This research explores how deep learning techniques can predict sepsis before clinical manifestation, allowing physicians to initiate timely treatment and improve patient survival.

II. Background

Sepsis is defined as a life-threatening organ dysfunction caused by an abnormal host response to infection.

It generally progresses through three stages:

- Infection
- Sepsis
- Septic Shock

Common symptoms include:

- Fever
- Rapid heart rate
- Low blood pressure
- Increased respiratory rate
- Confusion
- Elevated white blood cell count
- High lactate concentration

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

If untreated, sepsis can rapidly progress to multiple organ failure.

2.2 Electronic Health Records (EHR)

Electronic Health Records are digital repositories of patient medical information.

Typical EHR components include:

Patient demographics
Medical history
Laboratory test reports
Medication prescriptions
Vital signs
Imaging reports
Clinical notes
ICU monitoring data
Diagnoses and procedures

These records form sequential time-series data, making them well suited for recurrent neural networks, LSTMs, GRUs, and Transformer architectures.

III. Problem Statement

Current sepsis detection methods often identify patients after organ dysfunction has already begun. Traditional scoring systems may not fully exploit the complex temporal relationships within EHR data.

Hospitals therefore require predictive systems that can:

- Detect sepsis several hours before clinical diagnosis.
- Reduce mortality.
- Minimize ICU admissions.
- Support clinicians with real-time decision making.
- Decrease healthcare costs.

IV. Literature Review

Recent research demonstrates significant progress in AI-based sepsis prediction.

A systematic review of 42 studies found that machine learning and deep learning models achieved AUROC values ranging roughly from 0.80 to 0.97 and frequently predicted sepsis 2–6 hours before onset. However, studies differed in datasets, sepsis definitions, preprocessing methods, and evaluation metrics, making direct comparison difficult.

Researchers have explored several deep learning architectures:

Model	Advantages	Limitations
CNN	Learns local feature patterns	Limited temporal memory
RNN	Handles sequential data	Vanishing gradient problem
LSTM	Captures long-term dependencies	Higher computational cost
GRU	Faster than LSTM	Slightly less expressive
Transformer	Models long-range dependencies with attention	Require more data and computation

Newer reviews emphasize the growing role of multimodal learning, explainability, and standardized reporting frameworks such as TRIPOD-AI for clinical translation.

V. Proposed Methodology

The proposed framework consists of six phases to enable early sepsis prediction using Electronic Health Records (EHRs). Initially, patient data are collected from publicly available ICU databases such as **MIMIC-IV**, **PhysioNet Challenge**, and the **eICU Collaborative Research Database**, which contain extensive clinical information on critically ill patients.

The collected data undergo preprocessing to improve quality by removing duplicate records, handling missing values, normalizing features, aligning time-series observations, encoding categorical variables, detecting outliers, and generating patient-level sequences suitable for deep learning models.

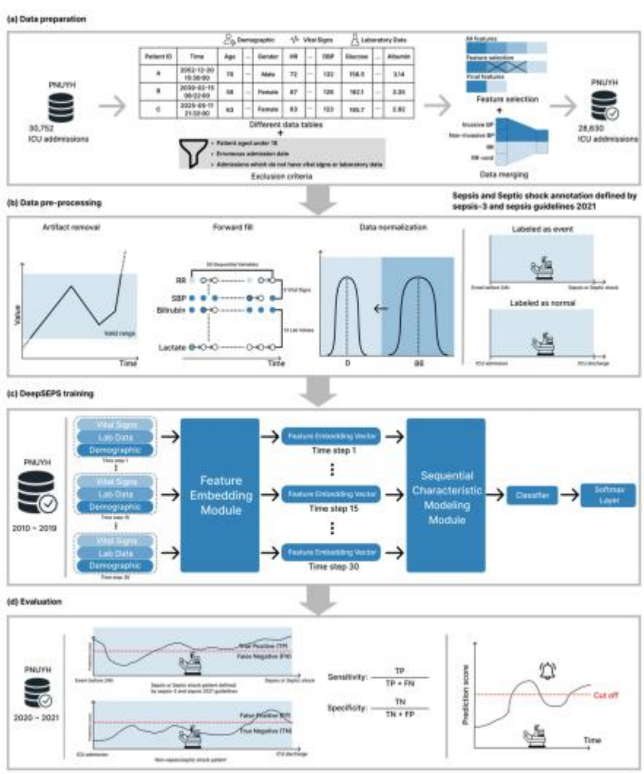
In the feature selection phase, important clinical predictors are identified, including demographic information (age, gender, weight), vital signs (heart rate, respiratory rate, blood pressure, temperature, oxygen saturation), laboratory values (white blood cell count, lactate, platelets, creatinine, bilirubin, glucose), and clinical variables such as previous infections, antibiotic administration, ICU stay duration, and comorbidities.

An **LSTM-based deep learning model** is employed to analyze sequential EHR data. The architecture includes an embedding layer, LSTM layer, dropout layer, dense layer, and a sigmoid output layer that predicts the probability of sepsis. LSTM networks effectively capture temporal dependencies in patient health records, making them well suited for early disease prediction.

The model is trained using the **Adam optimizer** and **Binary Cross-Entropy loss function**, with a batch size of **32–128**, a learning rate of **0.001**, and **50–100 training epochs**. Performance is validated using **5-fold cross-validation** to ensure robustness and reliability.

Finally, the trained model estimates the probability of sepsis for each patient.

If the predicted probability exceeds a predefined threshold, an early warning alert is generated, enabling clinicians to initiate timely interventions and improve patient outcomes.



VI. CONCLUSION

Deep learning-based sepsis prediction using Electronic Health Records represents a significant advancement in precision medicine. By leveraging longitudinal patient data, models such as LSTMs, GRUs, and Transformers can identify subtle physiological changes that precede clinical recognition of sepsis, enabling earlier intervention and potentially improving survival. Although challenges such as data quality, model interpretability, and external validation remain, the growing availability of large-scale EHR datasets and advances in explainable and multimodal AI position deep learning as a promising tool for future clinical decision support. With rigorous validation and thoughtful integration into healthcare workflows, these systems have the potential to reduce mortality, optimize resource utilization, and enhance patient care.

REFERENCES

[1] K. R. Islam, J. Prithula, J. Kumar, T. L. Tan, M. Murugappan, and M. J. Uddin, "Machine Learning-Based Early Prediction of Sepsis Using Electronic Health Records: A Systematic Review," *Journal of Clinical Medicine*, vol. 12, no. 17, Art. no. 5658, Aug. 2023, doi:10.3390/jcm12175658.

[2] M. Murugappan *et al.*, "A Novel Classical Machine Learning Framework for Early Sepsis Prediction Using Electronic Health Record Data From ICU Patients," *Computers in Biology and Medicine*, vol. 182, 109284, 2024, doi:10.1016/j.compbiomed.2024.109284.

[3] H. Cai, D. Wu, J. Xu, X. Liu, Y. Zhu, X. Shu, Y. Li, and B. Yi, "End-to-End Autoencoder MLP Framework for Sepsis Prediction," *arXiv preprint*, arXiv:2508.18688, 2025.

[4] C. Yin, S. Fu, B. Yao, T.-H. Pham, W. Cao, D. Wang, J. Caterino, and P. Zhang, "SepsisCalc: Integrating Clinical Calculators into Early Sepsis Prediction via Dynamic Temporal Graph Construction," *arXiv preprint*, arXiv:2501.00190, 2025.

[5] A. Taparugssanagorn, M. Särestöniemi, and M. Hämäläinen, "Lightweight, Explainable, and Efficiency-Aware Deep Learning for Early Sepsis Prediction from Structured Electronic Health Record Time-Series," in *Digital Health and Wireless Solutions: Integrating AI, LLMs and Multimodal Health Data for Next-Generation Decision Support*, 2026, pp. 225–244.

[6] B. P. R. Rella, A. K. Chaudhary, R. B. Patel, B. Nikolic, M. Janjic, and N. Bacanin, "Privacy-Preserving Federated Multimodal Deep Learning for Sepsis Prediction Using Vision Transformers and Secure Aggregation," *Internet Technology Letters*, vol. 9, no. 4, e70324, 2026.

[7] J. Wang, J. Luo, M. Ye, X. Wang, Y. Zhong, A. Chang, G. Huang, Z. Yin, C. Xiao, J. Sun, and F. Ma, "Recent Advances in Predictive Modeling with Electronic Health Records," *arXiv preprint*, arXiv:2402.01077, 2024.