International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 10 Issue 4, Jul-Aug 2026 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Early Prediction of Sepsis Using Machine Learning

Er. Yasmeen Viqar[1], Furqaan[2], Mehak Ayoub[3], Toiba Tahir[4]

[1]Associate Professor, Department of Computer Science and Engineering,
[2,3,4]Student, Department of Computer Science and Engineering,
[1,2,3,4]SSM College of Engineering, Pattan, Jammu and Kashmir

ABSTRACT

Sepsis is a deadly condition where human body's reaction to an infection causes organ damage and inflammation throughout the body. Quick detection of sepsis is vital because early detection of sepsis helps reduce mortality risk and makes treatment easier. Existing machine learning solutions for sepsis prediction are designed mostly for ICUs and employ non-disease-specific prediction models. In this work, we present a Machine Learning framework for early prediction of sepsis for non-ICU patients. Individual Random Forests are trained for each of the four infection types – Pneumonia, Urinary Tract Infection(UTI), Intra-Abdominal Infection, and Skin and Soft Tissue Infection(SSTI). Prediction is done based on structured information obtained from patients' symptoms, vital sign measurements, and lab test results. Trained models are deployed as a Flask based web application which facilitates features like authentication of users, selection of disease type, prediction generation, confidence scoring, risk- stratification and history management of predictions. Our system serves as an example for the potential application of ML driven systems for clinical decision support and demonstrates benefits of disease-specific prediction approaches for early detection of sepsis.

How to cite this paper: Er. Yasmeen Viqar | Furqaan | Mehak Ayoub | Toiba Tahir "Early Prediction of Sepsis Using Machine Learning" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-10 | Issue-4, August 2026, pp.43-47, URL: www.ijtsrd.com/papers/ijtsrd141903.pdf

IJTSRD141903

KEYWORDS: *Prediction of Sepsis, Machine Learning, Random Forest, Healthcare Analytics, Flask, Disease- specific Prediction, Clinical Decision Support.*

1. INTRODUCTION

Sepsis is one of the leading causes of hospital mortality across the globe. It is a condition when the body's response to an infection is deregulated, causing tissue damage, multi-organ failure, and death. Timely diagnosis and treatment have been found to improve patient survival rates in clinical literature. Diagnosing sepsis in the early stage is difficult because many of its symptoms resemble symptoms commonly seen in infections.

Existing solutions use clinical scoring systems such as SIRS and qSOFA to determine sepsis. These systems follow predefined rules and do not take into consideration interactions between variables associated with patient health. The accumulation of healthcare big data enabled application of machine learning algorithms to diagnose illnesses earlier and more precisely.

Upon reviewing related work, it was found that most sepsis prediction systems target patients in ICU and trained their models with general prediction systems. These models may overlook clinical features unique to certain diseases that contribute to deterioration of sepsis. Pneumonia, UTI, Intra-Abdominal Infection, and SSTI each have distinct symptoms and physiological characteristics. Applying one model to detect sepsis from all these infections can decrease the predictivity of the model.

To address this problem, we present a disease-specific machine learning based sepsis prediction system. Individual Random Forest classifiers are trained and deployed for each infection type using a Flask interactive web app. The intention is to assist with early clinical decision-making and offer a usable predictive model outside of the ICU.

2. RELATED WORK

Machine learning models have been applied to predicting sepsis. Early efforts focused on hand-engineered features from structured clinical datasets paired with traditional machine learning techniques such as Logistic Regression, Support Vector Machines, Decision Trees, etc. While such approaches were promising in terms of data-driven predictions, they struggled to model complex clinical relationships.

Nemati et al. introduce an interpretable machine learning based approach to predict sepsis in ICU patients given physiological time-series data. Results indicate their model can outperform traditional clinical scoring methods. One major limitation of this approach is its requirement of continuous ICU monitoring and clinical hardware/software infrastructure.

The PhysioNet Sepsis Challenge allowed researchers and developers to design and train advanced prediction models given publicly available ICU datasets. While several high performing systems were developed as part of the challenge, most focused on benchmarking results versus operational deployment.

Futoma et al. evaluated various machine learning and deep learning models for sepsis prediction and highlighted that simple techniques like Random Forest can achieve strong performance while also maintaining model interpretability. Other approaches have leveraged recurrent neural networks and long short-term memory architectures to understand temporal patterns present in patient data. While these techniques have shown promise, most require large datasets, are computationally expensive, and/or have restrictive deployment requirements.

A major limitation we observe in previous work is that models are generalizable predictors. Additionally, most models focus on providing prediction scores for ICU patients and are not deployed in an easy-to-use interface. This inspired us to develop our disease specific prediction framework with interpretable machine learning and a web-application frontend.

3. SYSTEM ANALYSIS AND DESIGN

The prototype system that we propose for the development of early prediction of sepsis based on customized machine learning models. Technical and operational feasibility studies were performed before proceeding with the development to confirm that the system could be developed to enable predictions based on healthcare information.

Modules included in the system are User login page, Disease selection page, Dynamic form page, Machine Learning Prediction page, Prediction Visualization page and History page. The structured clinical information from users are taken as input using forms created based on disease category. Depending on the disease category selected by the user, the appropriate machine learning model is loaded.

Datasets used for model training were preprocessed (cleaned, encoded, normalized and made ready for model training) before building the machine learning model. Individual datasets for Pneumonia, UTI, Intra-Abdominal infection and SSTI were prepared to understand each infection's clinical characteristics.

Random Forest algorithm was chosen as it provided high accuracy, is not susceptible to overfitting, helps in understanding the importance of each feature and works efficiently with structured format of healthcare data. The prototype system also provides us with confidence score of prediction and risk categories (Low/Medium/High risk of sepsis) to make it easier for users to understand the output of prediction.

Frontend of the web application was designed using HTML, CSS and Javascript. Flask framework is used as backend and SQLite database is used to store user details as well as prediction and history data.

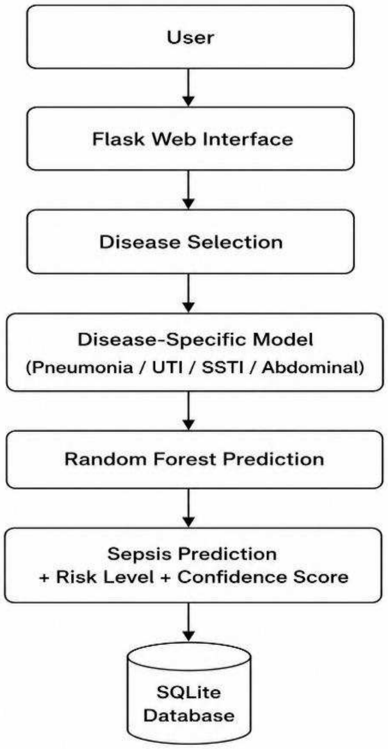


Figure 3.1: Proposed Architecture of the Early Prediction of Sepsis System

4. IMPLEMENTATION

Implementation involved turning the design into an actual working system. Building began with obtaining and preprocessing the datasets needed. Disease specific datasets were prepared and formatted into machine- learning-ready form with techniques like data cleaning, feature encoding, scaling, and normalization etc.

Individual Random Forest models were trained for each disease categories with clinical features such as age, temperature, heart rate, respiratory rate, systolic blood pressure etc. based on disease. Disease-specific symptomatic features were also added to the training data accordingly like temperature and white blood cell count for pneumonia, platelets for covid, oxygen saturation for covid etc.

The trained classifier models were saved using Joblib and added to the Flask app. Once users register and create an account on the website they can login through email verification(using OTP based authentication) and a password. After logging in they are prompted to select a disease category and fill in symptoms and details related to This data then gets processed in the backend where the data is preprocessed if needed and fed into the disease specific model to generate a prediction. The prediction along with confidence values to determine the risk category are returned. Prediction records are kept in an SQLite database file so that users can review their previous predictions.

This project's architecture allows frontend/frontend elements, backend services, machine learning models, and database to all talk to each other nicely. You now have a full-stack end-to-end prediction machine.

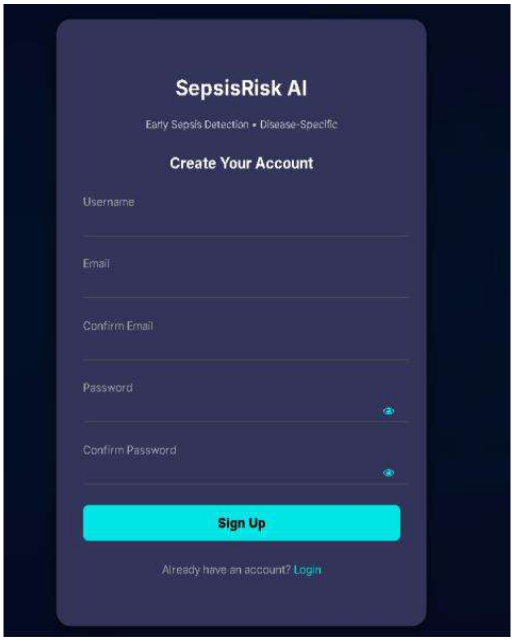


Figure 4.1: User Registration Interface

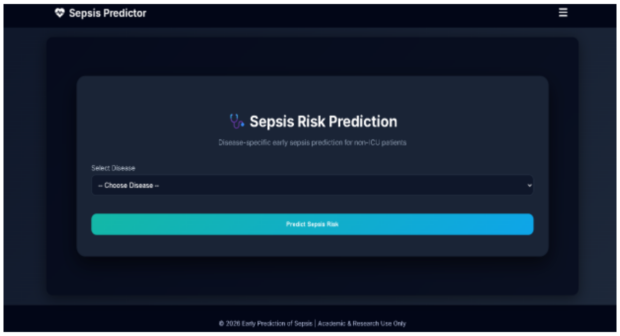
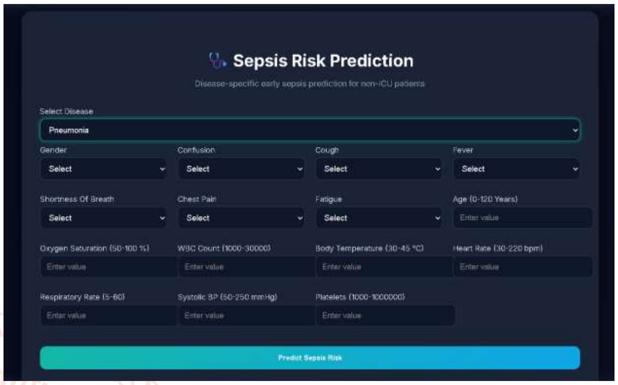


Figure 4.2: Disease Selection Dashboard



Select Disease: Pneumonia

Figure 4.3: Disease-Specific Patient Data Entry Form

5. RESULT AND DISCUSSION

The prototype system was integrated and tested on disease-specific datasets acquired for this task. All the major functionalities including registration/login with email verification, authentication, disease selection, prediction and history functionality are implemented as expected.

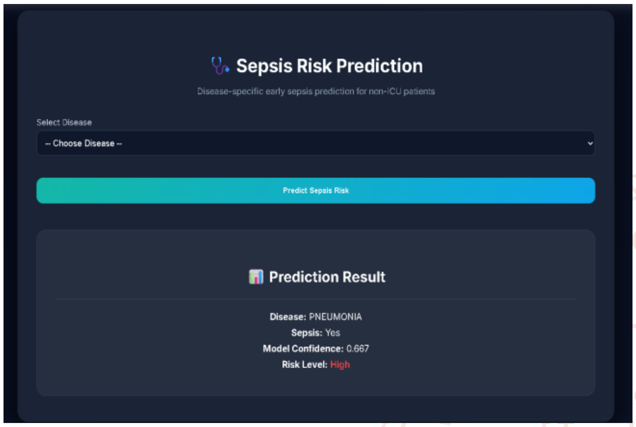
Random Forest classifiers demonstrate decent accuracy on each of the disease categories supported by our framework. Each model is evaluated on standard metrics such as Accuracy, Precision, Recall and F1-Score. Predictions are made on disease specific models which learn features pertaining to one infection type rather than a generalized form of the infection.

Accuracy scores from the prediction task suggest disease specific models have more focused representation of a patient's medical data. Since only features associated with each infection category affect prediction output, the model's decision-making process is clearer and not affected by unrelated features. Incorporating disease- specific models also provides a better predictive framework as progression of sepsis is determined by the site of infection.

Interface for our Flask application was also found to be user friendly. Forms were populated based on disease selection, requiring users to only input features associated with their specific disease. Prediction took seconds to display and data was stored/retrieved efficiently using our database.

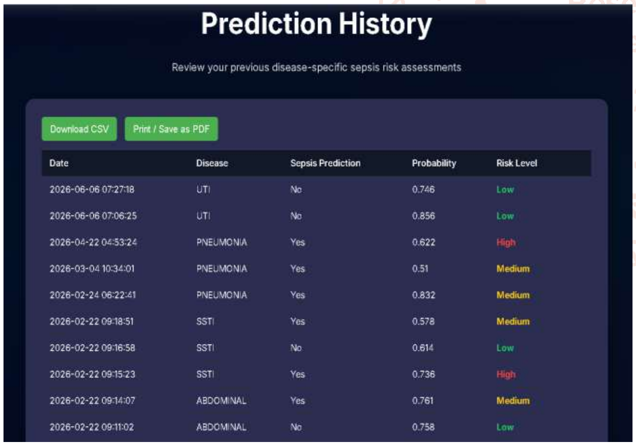
There are some clear limitations to our framework. Each model is only limited by the data it's trained on. In a real- world scenario our application has no direct access to hospital monitors and requires users to manually input clinical features. Our implementation currently supports only 4 disease categories. Our framework can be improved by training models on more diverse datasets and including more infection categories..

In conclusion, our experiments have shown that machine learning has the potential to aid in early prediction of sepsis. Training predictors on disease specific datasets has proven to have tangible benefits over general models.



Disease: PNEUMONIA
Sepsis: Yes
Model Confidence: 0.667
Risk Level: High

Figure 5.1: Sepsis Risk Prediction Result Page



Date	Disease	Sepsis Prediction	Probability	Risk Level
2026-06-06 07:27:18	UTI	No	0.746	Low
2026-06-06 07:06:25	UTI	No	0.856	Low
2026-04-22 04:53:24	PNEUMONIA	Yes	0.622	High
2026-03-04 10:34:01	PNEUMONIA	Yes	0.51	Medium
2026-02-24 06:22:41	PNEUMONIA	Yes	0.832	Medium
2026-02-22 09:18:51	SSTI	Yes	0.578	Medium
2026-02-22 09:16:58	SSTI	No	0.614	Low
2026-02-22 09:15:23	SSTI	Yes	0.736	High
2026-02-22 09:14:07	ABDOMINAL	Yes	0.761	Medium
2026-02-22 09:11:02	ABDOMINAL	No	0.758	Low

Figure 5.2: Prediction History Management Interface

6. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a machine learning based system that predicts sepsis early on using structured clinical data. Models trained for Random Forest were built individually for Pneumonia, Urinary Tract infection, Intra- Abdominal infection and Skin and Soft tissue infection. We were able to deploy the trained models into a flask based web application. This application came with capabilities like prediction/risk assessment, risk categorization, confidence scoring and user management/history of predictions.

Compared to other sepsis prediction models out there our system works well in several aspects. It operates on non- ICU patients and has disease specific models, something that can be deployed in the real world as a healthcare application. We tested it's capabilities through experiments and found that it can aid clinicians in identifying early risk and is interpretable/easy to use.

Prospective developments could involve the integration of bigger and broader real-world clinical datasets, linking hospital information systems in real-time, developing mobile apps, and applying Explainable Artificial Intelligence (XAI) methods to enhance interpretability. We can also consider other types of infections and more complex machine learning models such as XGBoost and LightGBM to enhance the predictive capabilities further. These future directions would enable us to build more accurate and clinically pertinent sepsis prediction tools.

REFERENCES

[1] M. Singer 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. W. Seymour et al., "Assessment of Clinical Criteria for Sepsis," JAMA, vol. 315, no. 8, pp. 762–774, 2016.

[3] S. Nemati et al.,"AnInterpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU,"Critical Care Medicine, vol. 46, no. 4, pp. 547–553, 2018.

[4] M. A. Reyna et al., "Early Prediction of Sepsis from Clinical Data: The PhysioNet Challenge," Critical Care Medicine, vol. 48, no. 2, pp. 210–217, 20 20.

[5] F. Fleuren et al., "Machine Learning for the Prediction of Sepsis: A Systematic Review and Meta-Analysis,"Intensive Care Medicine, vol. 46, no. 3, pp. 383–400, 20 20.

[6] R. Delahanty et al., "Development and Evaluation of a Machine Learning Model for Early Identification of Patients at Risk for Sepsis," Annals of Emergency Medicine, vol. 73, no. 4, pp. 334–344, 2019.

[7] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[8] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 28 25–28 30, 2011.

[9] R. Evans et al., “Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 20 21,” Intensive Care Medicine, vol. 47, no. 11, pp. 11 81–12 47, 20 21.

[10] T. van der Poll et al., “The Immunopathology of Sepsis and Potential Therapeutic Targets,” Nature Reviews Immunology, vol. 17, no. 7, pp. 407–420, 2017.