International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

Application of Machine Learning and its use in determining Sepsis Prediction

Prashanth Kumar Alay[1]

[1]Department of Health Management and Informatics, MU, School of Medicine.

Abstract - Sepsis is a potentially life-threatening and serious condition that occurs when an infection spreads across the body and triggers a widespread inflammatory response. It has a significantly high death rate, especially for patients in the ICU. Early detection and treatment of sepsis is necessary. Machine learning models for sepsis detection can be trained on a variety of data sources, such as electronic health records (EHRs), vital signs, laboratory test results, and demographic information. Exploratory Data analysis was performed using different pre-processing techniques. To classify the disease, the Random Forest Classifier is used, and comparison also performed among various Classifiers like MLP, KNN, etc.

***Key Words*:** Sepsis Prediction, Random Forest Classifier, vital signs.

1.INTRODUCTION

As a result of an unbalanced body's response to these toxins, sepsis can affect several organ systems. Sepsis is a disease that anyone can get because of infections. People with chronic diseases like cancer, diabetes, renal, lung, and kidney disease, as well as pregnant women, are more likely to catch it because of their compromised immune systems. Also at danger are infants under one year old. This illness poses a significant risk to public health because to its high mortality, high cost of care, and morbidity. The outcomes can be improved with early identification and antibiotic therapy. Pneumonia, stomach infections, kidney infections, and bloodstream infections are all factors that contribute to sepsis. Fast heartbeat, low blood pressure, hypothermia (very low body temperature), hyperventilation (rapid breathing), and severe pain or discomfort are all signs of sepsis. IV antibiotics to Figureht infection, vasoactive medicines to boost blood pressure in individuals with low blood pressure, and IV antibiotics to treat infection make up the treatment for sepsis.

2. LITERATURE REVIEW:

Vital sign-based detection of sepsis in neonates (2023) Sepsis prediction was performed using the Naive Bayes algorithm in a maximum a posteriori framework up to 24 hours before clinical sepsis suspicion. Data on vital signs were continuously and automatically collected for this investigation. Compared to research using manually obtained vital signs, this takes less time and is less prone to data entry mistake. However, due to its therapeutic relevance, this technique restricts the number of positive instances and changes how the cases are distributed among the various subgroups.

Learning representations for the early detection of sepsis with deep neural networks (2021) This study's objective was to contrast the new deep learning methodology's performance and viability with that of the regression approach using traditional temporal feature extraction. Through comparison with hand-crafted, the accuracy, performance, and feature extraction capability of DNNs are improved. Having the goal variable set, access to enough data, and sufficient explanatory power.

A computational approach to sepsis detection (2021) Evaluating the Insight algorithm's sensitivity and specificity three hours before a protracted SIRS event in the prediction of sepsis. The examination of correlations between nine widely used vital sign measures yields this prediction. The sensitivity of 90%, specificity of 81%, and quick prediction of this model are its benefits. SIRS criteria are sensitive to sepsis, but it also has a high proportion of false positive results, which is a drawback.

A Deep Learning-Based Sepsis Estimation Scheme (2020) Of the 34 features in the machine learning implementation, seven values were chosen from the six data quantification channels. Physiochemical prediction models are created using SVM classifiers that are decision-tree based. The proposed plan for the validation procedure of LSTM-RNN and SVM classifiers is validated using the ACNN classifier and two more intelligent classifiers. It is very precise, which shows that the model can accurately forecast the start of shock, severe sepsis, and consequences. Lacks sensitivity; ineffective for any databases that contain a variety of data.

Machine learning for prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy (2020) Three steps were used in the feature selection process: reviewing previous sepsis screening models, speaking with local subject matter experts, and finally using supervised machine learning known as gradient boosting. Important plots and the fundamental simplicity of the individual trees are benefits of this. Having the goal variable set, access to enough data, and sufficient explanatory power was observed from this study.

Prediction of Sepsis in the ICU: A Systematic Review (2019) Three steps were taken to pick features: reviewing current models for sepsis screening, consulting with local subject matter experts, and using supervised machine learning algorithm known as gradient boosting. Key indicators included alert rate, receiver operating characteristic curve area under, sensitivity, specificity, and precision. Because of its great specificity, the model can accurately forecast the development of shock, severe sepsis, and sequelae. Lacks sensitivity; inapplicable to all databases when there is a mix of data.

Evaluation and Development of a Machine Learning Model for Early Identification of Patients at Risk for Sepsis (2019) Three steps were used in the feature selection process: reviewing previous sepsis screening models, speaking with local subject matter experts, and finally using supervised machine learning known as gradient boosting. Alert rate, area under the receiver operating characteristic curve, sensitivity, specificity, and precision were some important performance indicators. Advantages outperform available standards in terms of discrimination, sensitivity, precision, and timeliness. The disadvantages include the ongoing difficulty in developing criteria-based diagnostic procedures for sepsis.

Machine Learning Algorithm to Predict Severe Sepsis and Septic Shock: Implementation and Impact on Clinical Practice (2019) To predict severe sepsis and septic shock, create a machine learning algorithm, put it into practice, and assess how it affects clinical practice and patient outcomes. Advantage: The precision has improved. The requirement for a substantial amount of excellently annotated medical data is a drawback.
Using machine learning algorithms to predict in-hospital mortality of sepsis patients in the ICU (2019) To create prediction models, they used the least absolute shrinkage and selection operator (LASSO), random forest, gradient boosting machine, and the conventional logistic regression (LR) approach. Better resource use is an advantage. The ethical issues around data privacy and the restricted availability in low-resource environments are disadvantages.
Evaluation of machine learning algorithm for advance prediction of sepsis using six vital signs (2019) In this paper, a gradient-boosted ensemble machine learning tool for sepsis detection and prediction is validated and its performance compared to other approaches. Continuous patient monitoring is achievable with quicker and more precise detection than with conventional techniques and danger of bias in the models require significant amounts of annotated data.

3. EXISTING SYSTEM

Sepsis is a potentially fatal organ failure brought on by an improperly controlled host response to infection. It is necessary to improve the early detection of suspected sepsis in prehospital and emergency circumstances in order to reduce the high case fatality rates and morbidity for sepsis and septic shock. ANN models like Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), Radial Basis Function Neural Networks (RBFNN), etc. are used in the current methodology. They are a kind of recurrent neural network that can pick up order dependence in situations involving sequence prediction. In this case, training the model is exceedingly challenging since it cannot handle the lengthy sequences. They are created with the ability to identify the sequential properties of data and then use patterns to foresee future events. The model's slow computational capacity affects accuracy. They also have the issues with gradient fading and exploding. With the current systems, the computation is slow, and the classification will take longer and have an accuracy of only 82–87 percent.

4. PROPOSED SYSTEM

Designing and creating a method for sepsis early detection using Random Forest Classifier is the main goal of this study. Pre-processing, determining the value of features, and classifying data are the three main processes in the suggested methodology. The data will first go through the resampling method of pre-processing. Using log plots and Yeo Johnson, the feature importance is calculated. The suggested classifier, known as the Random Forest Classifier, provides results for the sepsis detection. The architecture, or block view, of the proposed system's modules, which represents the technique we suggested. Python will be used to put the suggested method into practice. In order to demonstrate how well the technique works, the system is assessed in terms of Accuracy and Log Loss. The measured performance of the method suggested in this research will be contrasted with that of the previous work.

5. SYSTEM ARCHITECHTURE

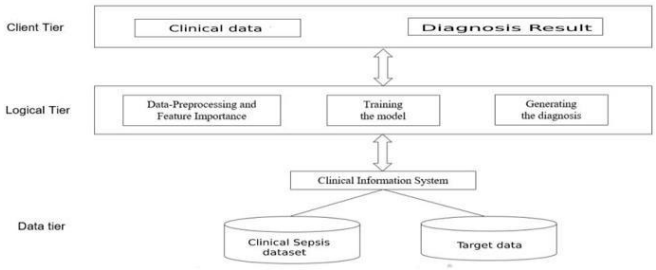

Figure 1: System Architecture

6. MODULES

A. Data Collection: Data from electronic health records (EHR), information on vital signs, and findings from laboratory tests. HER improves patient care in all areas, including safety, effectiveness, patient-centeredness, communication, timeliness, efficiency, and equity.
B. Tools Used: The libraries for machine learning, including Sklearn, Numpy, Pandas, and Matplotlib. The most widely used library for model validation and meaningful feature extraction is called Sklearn. SMOTE analysis is one of the numerous resampling techniques offered by the Imbalanced- learn library.
C. Data Preprocessing: The dataset's significant number of missing values required to be imputed for better prediction results utilizing several methods, including the mean, median, mode, and missforest methods. It uses Random Forest as its foundation, handling every missing value in accordance with its specifications.
D. Machine Learning Algorithms: One of the finest algorithms for classifying objects and displaying accurate performance is Random Forest. The best prediction is made after all the findings from weak decision trees have been combined and shown to exhibit iterative phenomena. Vital indicators such HR, temp, O2Sat, MAP, and Resp taken six hours prior to prediction, laboratory results, and demographic data were used to construct the prediction of sepsis.

7. UML DIAGRAMS

In the class diagram, the major class is Sepsis model which is connected to UI and Dataset. It contains the attributes like user data, preprocessed training data and preprocessed testing data. The are two functions for model creation as well as model generation. The predict function present in the model is used to analyze the parameters and predict the sepsis. The user class is used to provide input and display.

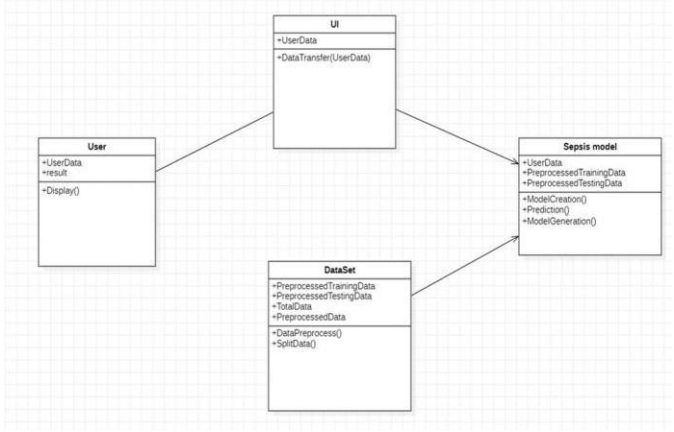

Figure 2: Class Diagram

In the sequence diagram below the lifelines are:
• User
• API
• Dataset
• Machine learning model

Firstly, the datasets are fetched into the Notebook. Then the datasets are trained and tested using various algorithms. When the user gives vital signs, demographics and clinical reports as input, the model analyses parameters and predicts the sepsis.

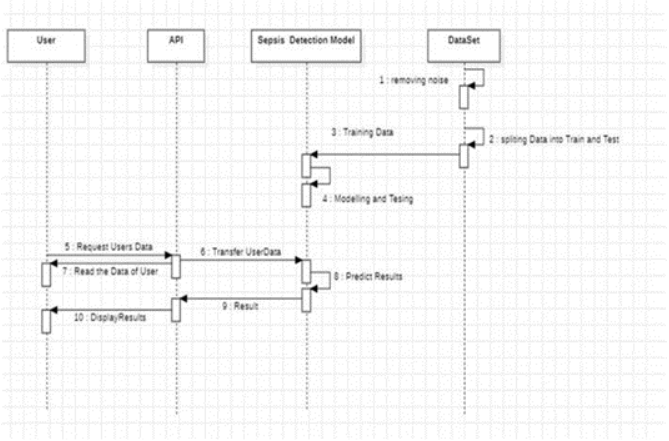

Figure 3: Sequence diagram

In the use case diagram shown below, a request is sent by the system, requesting user's data for sepsis classification. User communicates with the website to fill up the form fields. The pre-processing of data is performed on the sepsis dataset. The machine learning model to detect sepsis will classify sepsis and calculate the sepsis results. The results are displayed to the user using the Sepsis API

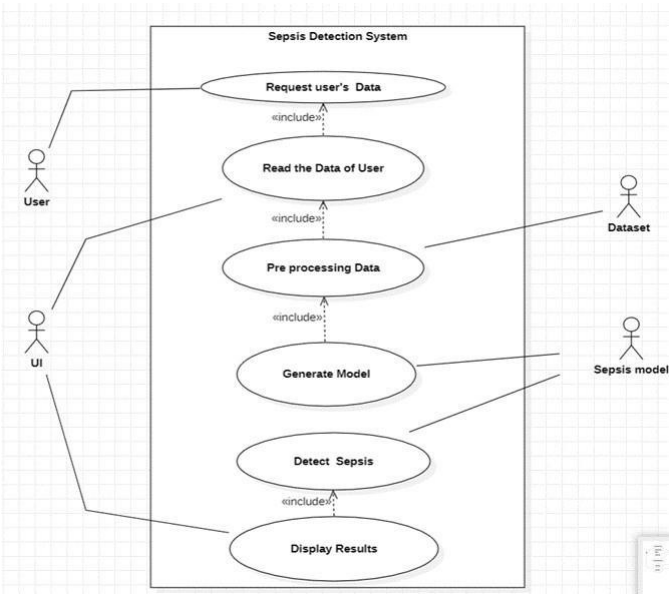

Figure 4: Use case diagram

8. RESULTS

Exploratory Data Analysis performed on the collected data set by checking on the basic parametric requirements.
Probability plots were plotted between some attributes and the ordered values. Plots were plotted for O2Sat, Temp, MAP, BUN, Creatinine, Glucose, WBC, Platelets. Dataset was fit into various algorithms and the best was found based on various attributes such as accuracy, precision, recall, f1 score, confusion matrix, etc. Classifier Accuracy and Log Loss were plotted.

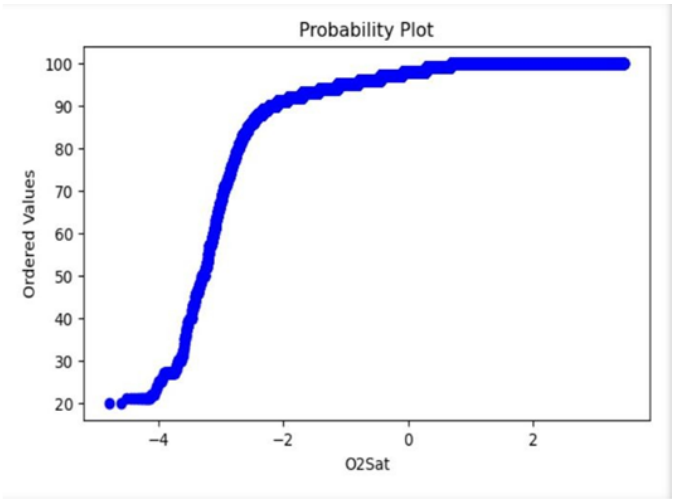

Figure 5: O_2Sat vs Ordered Values Probability plot

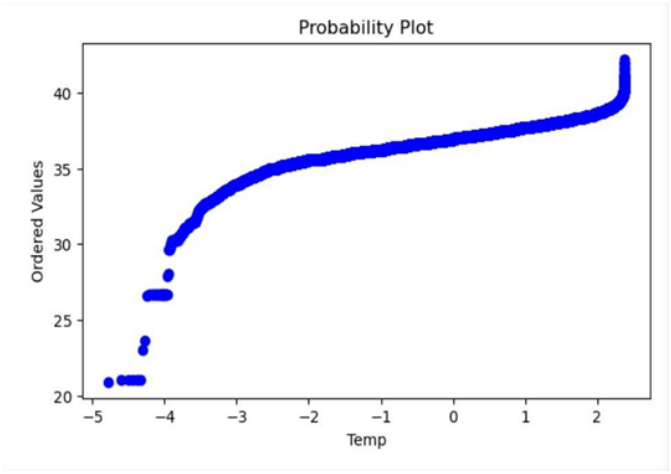

Figure 6: Temp vs Ordered Values Probability plot

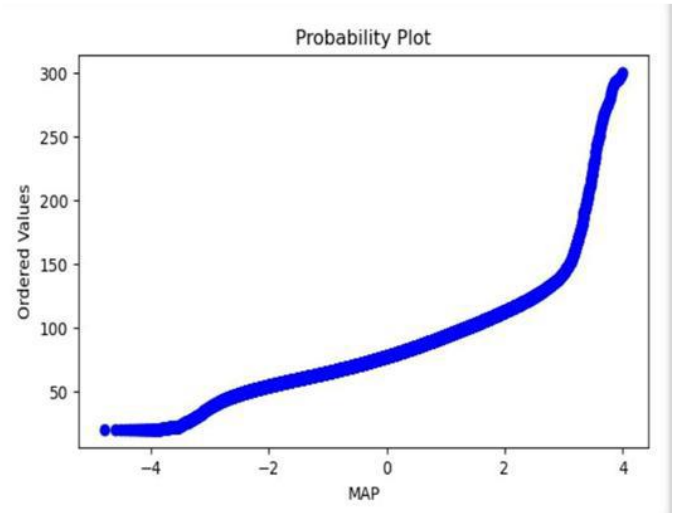

Figure 7: MAP vs Ordered Values Probability plot

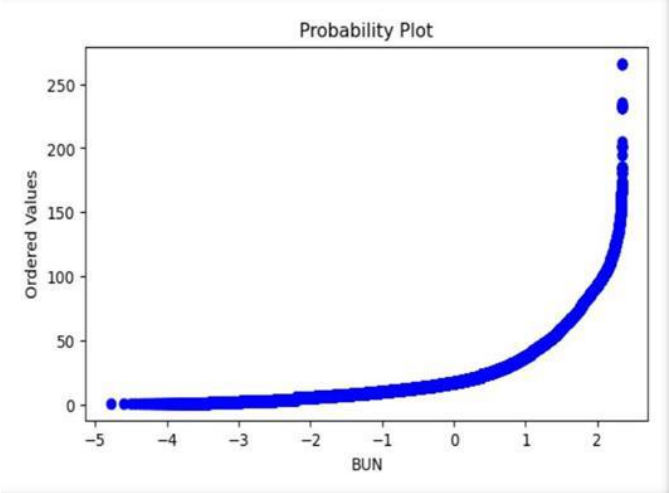

Figure 8: BUN vs Ordered Values Probability plot

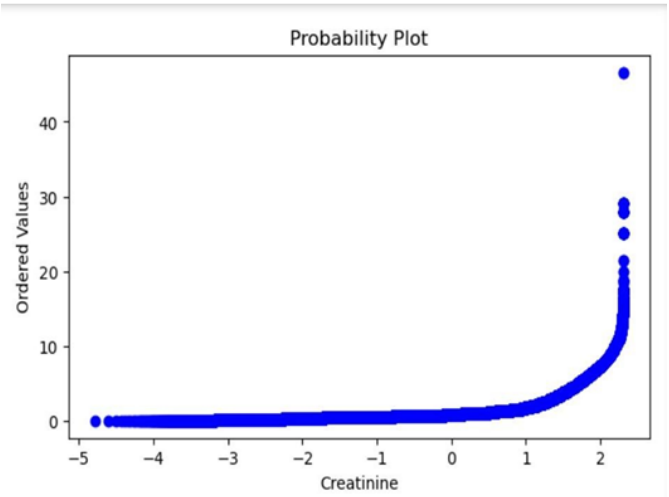

Figure 9(e): Creatinine vs Ordered Values Probability plot

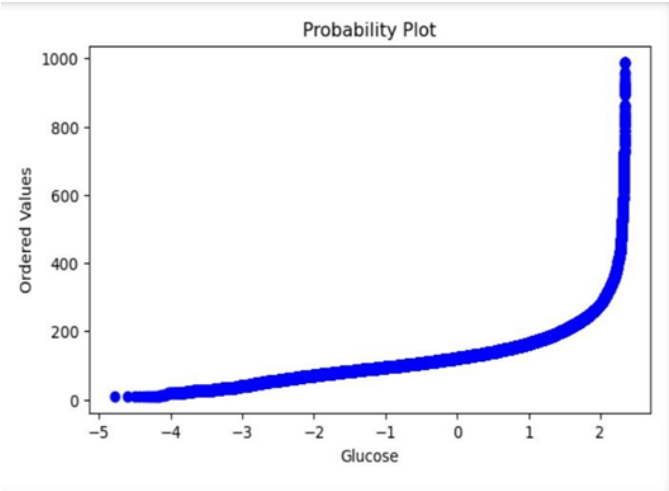

Figure 10: Glucose vs Ordered Values Probability plot

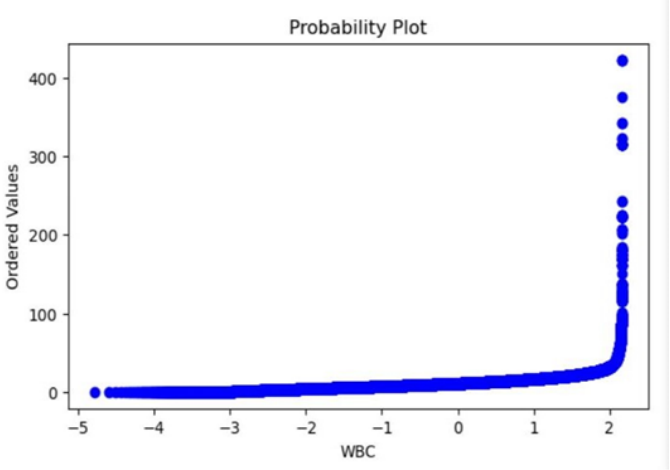

Figure 11: WBC vs Ordered Values Probability plot

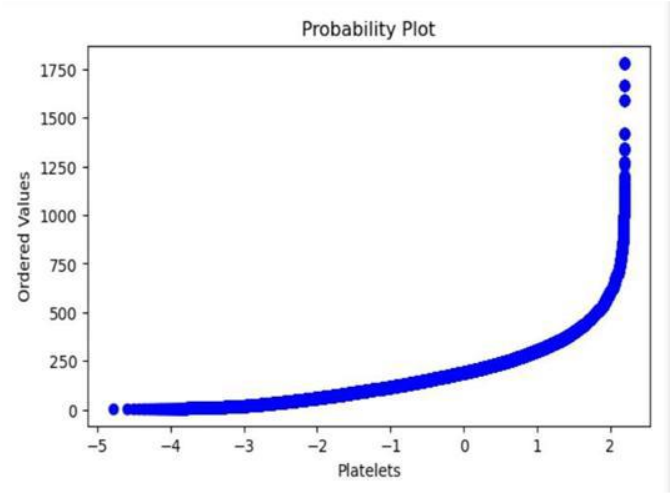

Figure 12 Platelets vs Ordered Values Probability plot

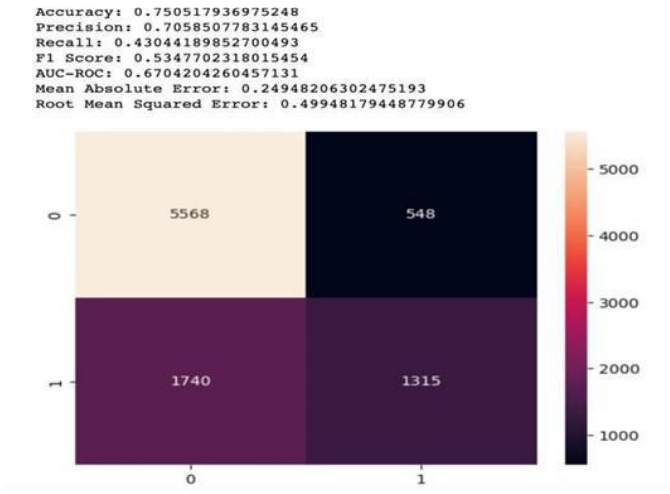

Figure 13 Results for Naïve Bayes Classifier

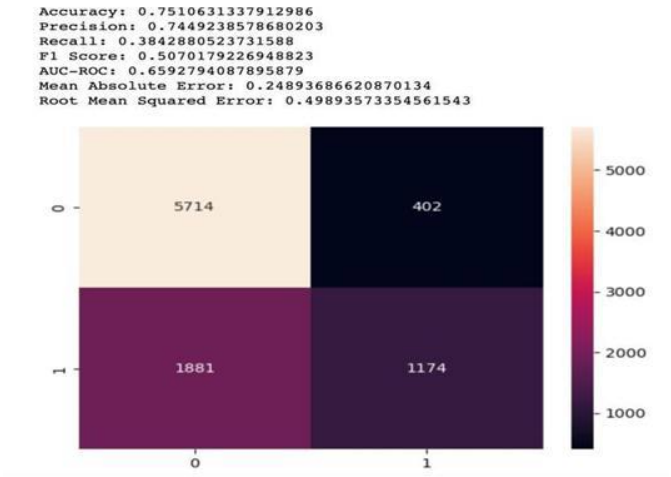

Figure 14 Results for Logistic Regression

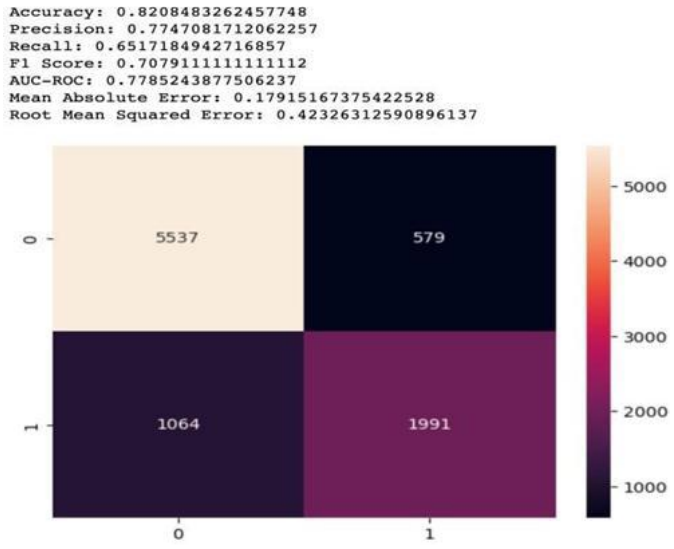

Figure 15 Results for KNN Classifier

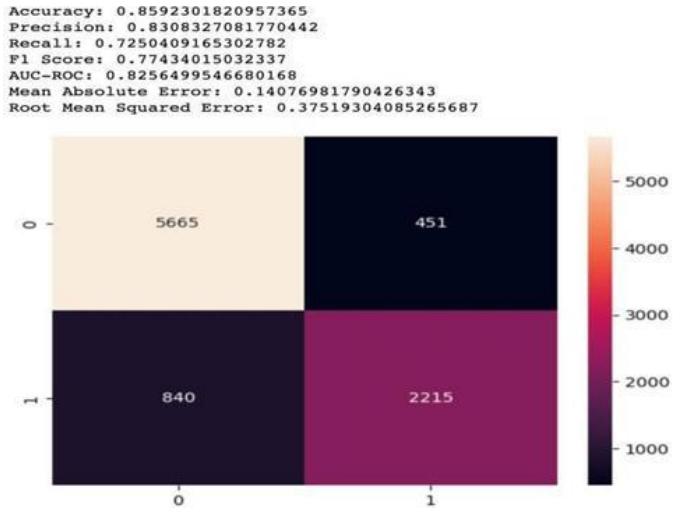

Figure 16 Results for XGBoost Classifier

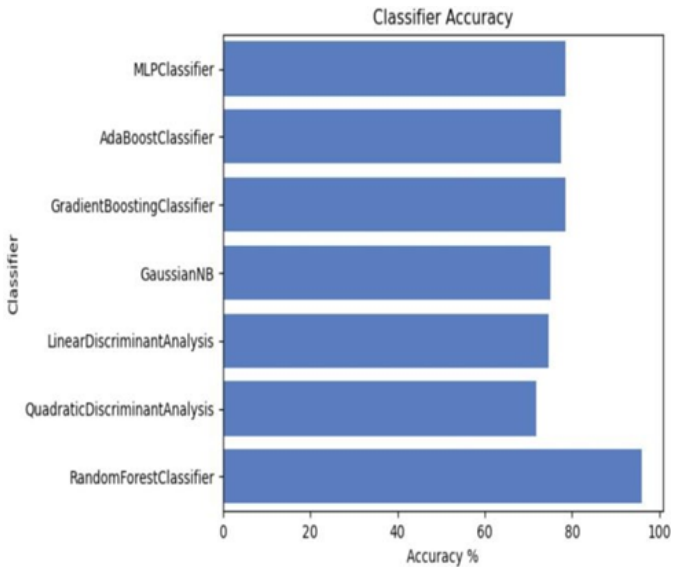

Figure 17 Classifier Accuracy

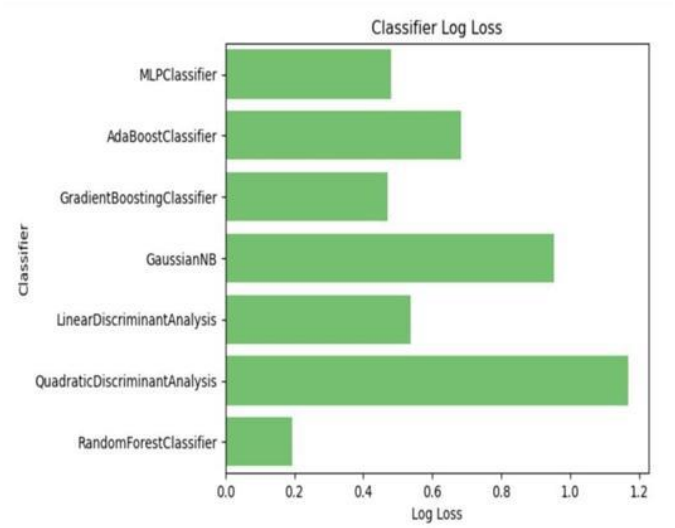

Figure 18 Classifier Log Loss

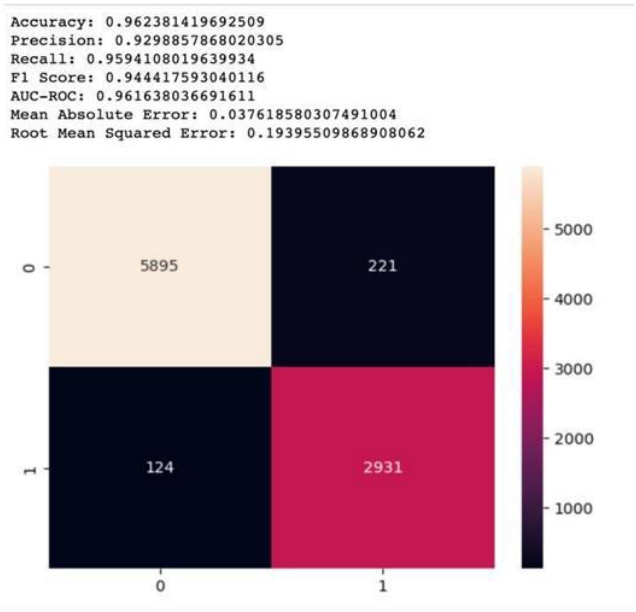

Figure 19 Results for Random Forest Classifier

9. CONCLUSION

Sepsis is a serious condition that causes tissue harm, organ failure, or even demise of the person. The main goal is to identify sepsis as soon as a patient arrives at the emergency room for care. Few sepsis detection systems currently in use are LSTM and RNN. However, the main disadvantage of RNN is that, because of its recurrent nature, an RNN model's computation is slow and only provides 82% accuracy. To address these issues, a method for early sepsis diagnosis employing stacking classifiers—including KNN Classifier, Naive Bayes Classifier, and Random Forest Classifier that involves feature importance, pre-processing, and classification has been created. This method is quick, less likely to produce incorrect findings, and provided accuracy of 96.23%.
This project also demonstrates the early and more accurate detection of this disease using a variety of classifiers, with a focus on stacking classifiers to create accurate prediction models and reduce the need for time-consuming lab tests. In the future, we hope to improve the application by making it customer-specific, implementing this model in a hospital website, and assisting the medical staff in spotting any early indications of disease.

10. REFERENCES

1. Nachimuthu S.K., Huag P.J. Early Detection of Sepsis in the Emergency Department using Dynamic Bayesian Networks; Proceedings of the 20 12 AMIA Annual Symposium; Chicago, IL, USA. 3–7 November 2012; pp. 653–662.
2. Sutherland, A., Thomas, M., Brandon, R.A. et al. Development and validation of a novel molecular biomarker diagnostic test for the early detection of sepsis: https://doi.org/10.1186/cc10274
3. Karen K. Giuliano; Physiological Monitoring for Critically Ill Patients: Testing a Predictive Model for the Early Detection of Sepsis: https: // doi.org/ 10.4037/ ajcc 2007.16.2.122.
4. M. Fu, J. Yuan, M. Lu, P. Hong and M. Zeng, "An Ensemble Machine Learning Model For the Early Detection of Sepsis From Clinical Data," 2019 Computing in Cardiology (CinC), Singapore, Singapore, 2019, pp. Page 1-Page 4
5. Kam, Hye Jin, and Ha Young Kim. "Learning representations for the early detection of sepsis with deep neural networks." Computers in biology and medicine 89 (2017): 248-255.
6. Kaylor, R. Andrew, et al. "Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data–driven, machine learning approach." Academic emergency medicine 23.3 (2016): 269-278
7. https://www.healthline.com/health/sepsis
8. https://physionet.org/content/challenge-2019/1.0.0