Sepsis Prediction Using Machine Learning

1st Rajesh Nalla 2nd Manohar Ronanki 3rd Suresh Kumar Batchu

Abstract—One of the most frequent causes of admission to the critical care unit is sepsis, which is seen as a race to the death between the pathogens and the host immune system (ICU). Despite better healthcare outcomes, sepsis-related mortality is still significant and is the ”second-leading cause of death ”in noncoronary ICUs. The ability to recognize the development of sepsis disease as the outcome is crucial in the patient’s condition. Our main objective is to contrast ”various machine learning algorithms, such as Logistic Regression, Naive Bayes, KNN classifier, and Random Forest classifier”, in order to determine the best classifier that provides the best accuracy and performance using data that is present in the form of ’digital healthcare data’ and to predict the sepsis disease is present in the patient or not. The other goal is to create an interface that is user-friendly. Three steps make up the proposed system. These are model building, feature importance, and pre-processing. We want to perform this on two different datasets for better results. The datasets used here are Physio Net Challenge data and the MIMI III dataset. In this way, models are built using different classifiers to find out the best classifier with high accuracy and detect the sepsis disease in minimal time.

Index Terms—Sepsis,Logistic regression, Naive Bayes, KNN classifier, Random forest classifier

I. Introduction

Sepsis is the outcome of the body’s immune system reacting to an infection, a condition that can be fatal and cause multiple organ failure. In high-income nations, it is estimated that there are 31.5 million cases of sepsis per year, 18.4 millions of severe sepsis, and 4.9 million deaths caused due to sepsis. Research have indicated that prompt antibiotic therapy beginning after early sepsis diagnosis improves patient outcomes, and that every 6 hours of treatment delay increases the risk of fatality by 7.6%. However, because organ failure and deterioration are symptoms of many illnesses, sepsis is frequently misdiagnosed and improperly treated. Treatment for sepsis is difficult because of the variable ’infection source, immunological reactions, and pathophysiology changes’. Furthermore, septic patients’ signs and prognosis are impacted by the variety in age, gender, and comorbidities. A host reaction to an infection that is out of balance causes sepsis, an organ failure that can be fatal. It is a significant worldwide health issue. Sepsis is the leading cause of hospital fatalities despite promising medical advancements over the past few decades. It places a significant load on the health care systems around the world and is linked to unacceptably high mortality and morbidity rates. This can be partly linked to difficulties with early sepsis diagnosis and prompt and effective therapy initiation. The need of prompt diagnosis and treatment commencement is further highlighted by the rising body of research that indicates mortality rises for every hour that antimicrobial intervention is postponed. Machine learning (ML) focuses on creating new prediction models from data by doing a thorough search through a wide range of models and parameters, followed by validation. The viability of developing clinically relevant predictive models was earlier work. In a perfect world, one would be able to create such models using information that is regularly gathered from electronic medical records (EMR). Our goal in the current effort was to detect sepsis from data using ML techniques, which would enable earlier diagnosis and better illness management. Furthermore, we stress that we use normally collected data rather than data that was expressly collected for a hypothesis-driven study process in order to highlight the utility of data mining tools.

II. Motivation

Early diagnosis and timely and appropriate clinical management of sepsis, such as optimal antimicrobial use and fluid resuscitation, are crucial to increase the likelihood of survival. Even though the onset of sepsis can be acute and poses a short-term mortality burden, it can also be the cause of significant long-term morbidity requiring treatment and support. Thus, sepsis requires a multidisciplinary approach.

III. Objectives

This paper is focused on evaluating different Machine Learning algorithms in predicting whether a patient is suffering from sepsis or not. This paper has two main objectives:

- Comparing different classifiers’ performance in sepsis prediction
- Building a web interface using the best classifier.

GitHub Code Link: https://github.com/RajeshUCM/MLProject

IV. Related work

Mahmud et al.[1] aimed to use a logistic regression classifier to detect the start of sepsis using 40 ICU patient features. The large class imbalance between patients with septic and nonseptic infections was reduced by giving the minority class more weight. The data was normalised, and the Pearson correlation coefficient was used to choose pertinent features for predicting the target variable. The model's performance was assessed using the PhysioNet/Computing in Cardiology (CinC) Challenge 2019 scoring system, and it received a score of 32.4 percent on the training dataset and 14 percent on the unseen test dataset.

Using the most recent definition, Sepsis-3, Wang et al.[2] developed sepsis prediction models. Three separate classification techniques—logistic regression (LR), support vector machines (SVM), and logistic model trees—were used to achieve this (LMT). Our algorithms used blood culture findings and vital sign data to forecast the onset of sepsis in adult intensive care unit patients. For patients who did not develop sepsis, predictor values were chosen at random from a 48-hour window; for those who did, values were taken at random from a 48 to 6-hour window prior to commencement using the closest previous time. Comparing the LR and SVM classification methods, it was discovered that the LMT technique was the most successful at predicting sepsis. However, the models did show a high false positive rate, despite similar sensitivity and specificity to previous models using the Sepsis-3 definition.

Thakur et al.[3] wanted reduce the delay, prediction models or screening tests are used to initiate antimicrobial therapy. Unfortunately, most of the prediction models require invasive parameters, making them unsuitable for rural areas of low-income countries lacking laboratory facilities. In this retrospective study, two prediction models were developed and compared using invasive and non-invasive parameters, respectively, through binary logistic regression analysis. The model developed from non-invasive parameters performed just as well as the one made from invasive parameters, as determined by the area under the receiver operating characteristic (AUROC) values of 0.824 and 0.777 for the non-invasive and invasive models, respectively. The AUROC values in the validation dataset were 0.824 and 0.830 for the non-invasive and invasive models, respectively, indicating their significance with p ¡ 0.001.

Siren et al.[4] identified neonatal sepsis (NSD) early in order to lower mortality, morbidity, and antibiotic use in premature infants. Nonetheless, in case-control setups, NSD models are frequently constructed and tested only on patient electrocardiogram(ECG) data. In this paper, we suggest a retrospective cohort research setting that is more realistic, utilising data from many modalities, including ECG, chest impedance, pulse oximetry, demographic characteristics, and repeated measures of body weight. In a framework for sequence-to-sequence mapping, we compare the vanilla and Long-Short-Term Memory (LSTM) recurrent neural networks (RNN) architectures for NSD. In a leave-one-out cross-validation framework, we compare the performance of these models to that of logistic regression (LR) using a number of classification measures.

Pawar et al.[5] aimed to create and test a straightforward algorithm for early sepsis detection. Unfortunately, there were missing data in the dataset, which we addressed by imputation, filling in the gaps with population means. Clinically relevant variables were combined, and transformation features were included, such as dichotomization, z-scores, derivative, and changes from baseline. Although the model showed promising results, further optimization is necessary to decrease the false positive rate. Adding features capturing change over time is expected to provide scope for further investigations.

Shankar et al.[6] identified that sepsis is a fatal condition brought on by infection. Particularly among patients in intensive care units, it has a notably high death rate. Early and accurate detection of sepsis is essential since postponing treatment significantly increases mortality. Using the patient's EMR, vital signs, and demographic data, the proposed research aims to develop a categorizer that accurately predicts sepsis up to six hours before the clinical diagnosis of the illness. The study presents numerous imputation methods and proposes a brand-new filling algorithm called Mixed Filling. The critical elements that influence the categorizer's predictions have been outlined, making the model easier for medical professionals to understand.

Shanthi et al.[7] proposed to employ machine learning algorithms for early prediction of sepsis, a severe medical condition that can lead to tissue damage, organ failure, and even death if not detected early. As sepsis is a complex disease and its symptoms can resemble those of other illnesses, early diagnosis is crucial. The study focuses on using physiological data to identify sepsis patients, and the algorithms utilized include Extreme Gradient Boost, Logistic Regression, Support Vector Machine, and Decision Tree classifiers. Visual Studio was used to create the study, while Python was applied to access it dynamically. In high-dimensional spaces, Logistic Regression and Support Vector Machine algorithms were found to be more effective, while Boosting algorithms were employed to increase prediction accuracy. The results indicate that Logistic Regression and Support Vector Machine algorithms can produce a prediction accuracy rate of about 98.

Mitchell et al.[8] proposed techniques for diagnosing and forecasting sepsis have centered on ICU patients and depended on an out-of-date definition based on SIRS criteria, which lacks clinical validation and misses at least 1 in 8 cases. The potential for enhanced early warning systems (EWS) and prediction models is presented by recently updated consensus guidelines on sepsis. Our goal was to create a more precise EWS using the updated criteria and improved models to detect sepsis in patients in non-

ICU wards early. We created an Early Warning System (EWS) to predict the onset of sepsis 12 to 24 hours in advance using multivariate logistic regression analysis of physiological and laboratory data from electronic health records (EHRs). The EWS produced an area under the receiver operating characteristic curve (AUC) result.

Moore et al.[9] observed that due to potential difficulties throughout their hospital stay, intensive care unit (ICU) patients are at a high risk of morbidity and mortality, with severe sepsis being a major cause of death. The creation of a new framework is necessary due to the bias and insufficiency of identification and prediction approaches now in use. Predictive models have the potential to help with early detection of severe sepsis and rapid intervention. Of the 3,446 patients in a retrospective cohort of ICU patients, the SVM model using laboratory and vital signs effectively predicted the development of severe sepsis in 339 (65%). The generated models offer suggestions for clinical decision support in situations other than ICUs. lower output in this situation.

Fleuron et al. [10] analysed that every supervised machine learning model that attempted to anticipate these situations in real time failed the index test. The Grading of Recommendations was used to evaluate the quality of the evidence. Individual machine learning models can precisely forecast the beginning of sepsis in the future, according to this comprehensive review and meta-analysis of the literature on the subject. Between-study heterogeneity restricts the evaluation of the combined results even though they offer alternatives to conventional grading systems. To close the information gap between the bytes and the bedside, systematic reporting and clinical implementation studies are required.

Kong et al. [11] wanted to create machine learning-based tools to forecast the likelihood of hospital death for sepsis patients in intensive care units. The prediction performance of the machine learning-based models created for this investigation was good. The GBM model outperformed the others in terms of accurately forecasting the probability of in-hospital death. It might help ICU doctors treat critically ill sepsis patients with the proper clinical therapies, thereby enhancing the prognoses of sepsis patients there.

Kausch et al. [12] wanted to assess the modeling strategy and statistical approach used in the sepsis prediction models developed for the adult hospital population. The objective of this review was to review the literature on machine learning models for sepsis prediction, summarise the results, and identify potential directions for further study. In emergency rooms, intensive care units, and acute care floors of hospitals, twelve research assessed the development of machine learning models. There is a significant obstacle in the translation from model development to clinical translation and implementation, as just two further research addressed prospective patient outcomes in the ICU setting.

V. Data Preparation

Data Preparation is the process of collecting the data, cleaning, and preparing the data to train the model. It is crucial for any Machine Learning model to have clean data to give accurate predictions and hence this process also consists of data cleaning procedure.

A. Dataset Collection

To check and analyse the capability of various machine learning algorithms on predicting sepsis disease, we have considered two datasets. The patients in the ICUs of three separate hospitals made up the initial dataset, which was taken from the Physionet Challenge. Clinical information on about 30,247 patients from three different hospitals has been collected. Each person's clinical data includes 44 measures of important indicators, demographic data, and laboratory results. Heart rate, body temperature, and mean arterial pressure are important indicators. Among the laboratory readings are calcium, glucose, and platelet count. There are other factors as well, such as age and gender. The labels 0 (Non-sepsis) and 1 indicate sepsis (Sepsis). The MIMIC-III data set, which is likewise gathered via the Physionet website, is another dataset. This database includes data on demographics, vital signs, and the outcomes of tests and drugs. It is remarkable for its diversity and includes a sizable number of ICU patients. When the data has been cleaned, this dataset is utilized to train and test the machine learning models under consideration.

B. Data Cleaning

It is crucial to clean the data before submitting it to the model for training so that any incorrect, corrupted, improperly formatted, duplicated, or missing data can be fixed or eliminated. Even though results and algorithms seem to be accurate, faulty data renders them unreliable. Here let us consider the first dataset. By looking into it we can understand that it is time series data. The dataframe has 44 columns. After the first look at data, it can be observed that the data seems quite sparse. For the dataset considered in this process, there are cells that are null and hence the following data-cleaning steps have been considered:

- Removing null values
- Dropping the redundant and unwanted features
- Feature Importance

- Imputation
- Encoding
- Scaling
- Data Sampling

1) Removing Null Values: By looking at the dataset, we observe that there are many null values. By finding out the percentage of null values for each column, we understand that these should be removed. The features have been removed based on the number of null values and redundancy

2) Dropping the redundant and unwanted features: This stage carefully removes any features or undesirable information from the dataset, such as redundant or pointless information. Discoveries that you make are seen as irrelevant when they are unrelated to the specific problem you are aiming to research. This can improve analytical effectiveness, reduce deviance from your main objective, and produce a dataset that is easier to handle and performs better. In the dataset considered, these features are removed because of redundancy and null values. The features are ’Unnamed: 0’, ’SBP’, ’DBP’, ’EtCO2’,’BaseExcess’, ’HCO3’,’pH’,’PaCO2’,’Alkalinephos’, ’Calcium’, ’Magnesium’, ’Phosphate’, ’Potassium’, ’PTT’, ’Fibrinogen’, ’Unit1’, ’Unit2’. So, the updated dataset is considered for further analysis. The columns are dropped using the drop() function with its attribute column being the list of columns to be removed, inplace been set to true, and axis to 1. The result is the dataset without the above-mentioned attributes.

3) Feature Importance: As there are many features, let us find out the correlation between the features by using the corr matrix() function with its attribute column being the list of columns. As can be seen in this correlation heat map almost all of the features do not have a high correlation.

4) Imputation: Since there are a lot of missing values in the dataset, therefore imputation was done to fill the missing values. While imputing, it is important to note that imputation should be done on per patient basis, otherwise the data from one patient will leak into the data of the other patient. Also another point that should be taken into consideration is that mean, median, mode can not directly be used to impute as it will result in uneven distribution of the parameters with respect to time. So, on the ’Patient ID’ column we perform bfill() and ffill() functions for imputing. By checking the remaining proportion of missing values, ’TroponinI’, ’Bilirubin direct’, ’AST’, ’Bilirubin total’, ’Lactate’, ’SaO2’, ’FiO2’, ’Unit’, ’Patient ID’ have more than 25 percent of null values and hence are dropped from the dataset using drop() function.

5) Encoding Categorical Variables: Encoding categorical variables is one of the important data preparation tasks. The real-life data may consist of attributes that are categorical string values. Most machine learning models perform several mathematical computations to reach their goal and hence they work only on numerical data and on other data types which are considered by that algorithm.It is required to convert these string values into integer values because the model will not function on them. Encoding categorical variables is all about this process. Out of the many different types of encoding, we used the one hot encoding technique used to represent categorical variables as numerical values in a machine learning model. In this dataset there is 1 attribute that is categorical. The gender attribute is categorical consisting two categories Male(M), Female(F)

6) Scaling: The concept of standardization emerges when continuous independent variables are measured at several scales. This suggests that these variables did not affect the analysis equally. Therefore the attributes that take large values will be given more weightage by the model irrespective of whether or not those attributes contribute to such weightage towards the goal of the model. Therefore, it is required to transform the data to comparable scales. Generally models tend to give a better result for a normal distribution. So we explored different techniques to plot histograms and QQ plots of all the features and then we applied different transformations on it to see which were giving good results. The ones giving the best results were then adopted in the dataframe. Here we applied different transformations like yeojohnson plot, exponential transformation, inverse transformation and logarithmic transformation to plot guassian transformation. By observing different plots, it can be concluded that only log was somewhat effective and that too for MAP, BUN, Creatinine, Glucose, WBC and Platelets. So we apply log transformations on the columns and apply standard normalization using StandardScaler() function.

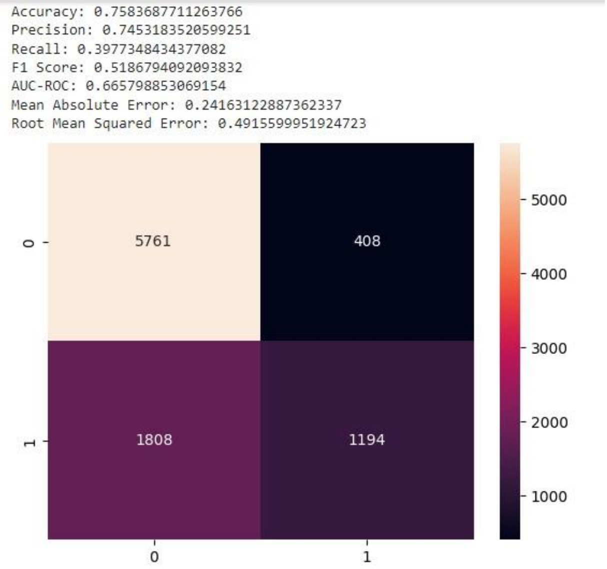
7) Data Sampling: By checking the distribution of data points between the two classes we obtained that the number of sepsis label 1 is 15284 while number of sepsis label 0 is 750935. So, this shows a clear imbalance between sepsis label 1 and label 0, to deal with this, we did undersampling. It is a technique for balancing unequal datasets by keeping all of the data in the minority class and reducing the size of the majority class.

VI. Comparison of different classifiers

The physionet dataset is split into training and testing data in 80:20 ratio. To obtain a better understanding, we considered various evaluation metrics like accuracy, precision, recall,f1 score, AUC-ROC, Mean Absolute Error, Root Mean Squared Error and Confusion Matrix. Here the different machine learning algorithms performance is compared.

A. Logistic Regression

In contrast to its name, logistic regression is more of a classification model. Logistic regression is a simple and superior method for scenarios involving binary and linear classification. This classification model works brilliantly with linearly separable classes and is exceedingly easy to use. When Logistic Regression is used, the accuracy obtained is 75.8% for Physionet dataset and 66.4% for MIMIC-III dataset. The evaluation metrics obtained are shown in the figures 1 and 2

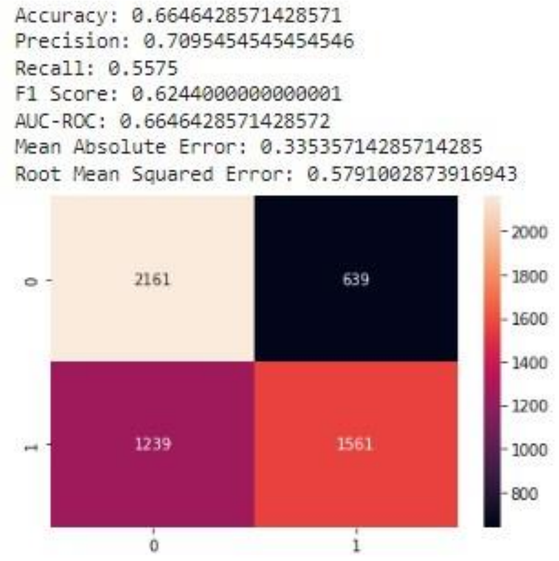


```
Accuracy: 0.7583687711263766
Precision: 0.7453183520599251
Recall: 0.3977348434377082
F1 Score: 0.5186794092093832
AUC-ROC: 0.665798853069154
Mean Absolute Error: 0.24163122887362337
Root Mean Squared Error: 0.4915599951924723
```

Fig. 1.

B. Random Forest

Random Forest is the most popular machine learning algorithm and a part of the supervised learning strategy. Random Forest, as the name suggests, is a classifier that increases the projected accuracy of the dataset by averaging numerous decision trees applied to different subsets of the provided data. Instead of relying exclusively on one decision tree, the random forest uses estimates from each decision tree and predicts the result based on the votes of the majority of projections. The greater number of trees prevents higher accuracy and overfitting. The purpose of a criterion is to assess a split's quality. Here we considered both gini and entropy as criterion to obtain best outcome[14]. When a dataset piece is randomly labelled, the gini impurity counts the number of times it will

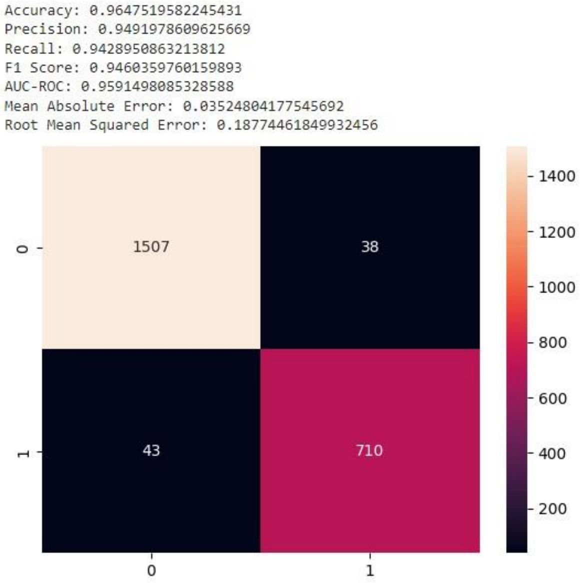


```
Accuracy: 0.6646428571428571
Precision: 0.7095454545454546
Recall: 0.5575
F1 Score: 0.6244000000000001
AUC-ROC: 0.6646428571428572
Mean Absolute Error: 0.33535714285714285
Root Mean Squared Error: 0.5791002873916943
```

Fig. 2.

GitHub Code Link: https://github.com/RajeshUCM/MLProject

be erroneously identified. The minimal value of the Gini Index is 0.

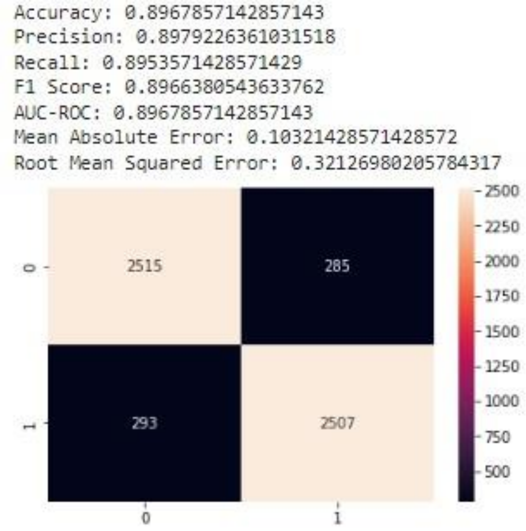


```
Accuracy: 0.9647519582245431
Precision: 0.9491978609625669
Recall: 0.9428950863213812
F1 Score: 0.9460359760159893
AUC-ROC: 0.9591498085328588
Mean Absolute Error: 0.03524804177545692
Root Mean Squared Error: 0.18774461849932456
```

Fig. 3.

C. KNN

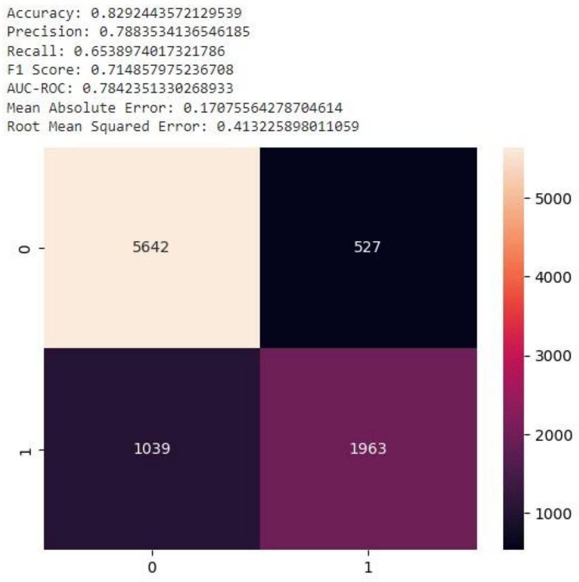
The K-NN algorithm groups the new instance, assuming that it is equivalent to the prior instances, into the category that most closely resembles the present categories. A new data point is categorised using the K-NN algorithm based on similarity after storing all the prior data. This suggests that employing the K-NN method, new data can be consistently and quickly classified[15]. K-NN is a non-parametric method that makes no assumptions about the underlying data. This



```
Accuracy: 0.8967857142857143
Precision: 0.8979226361031518
Recall: 0.8953571428571429
F1 Score: 0.8966380543633762
AUC-ROC: 0.8967857142857143
Mean Absolute Error: 0.10321428571428572
Root Mean Squared Error: 0.32126980205784317
```

Fig. 4.

method is commonly referred to as a lazy learner since it saves the training dataset rather than learning from it right away. As an alternative, it does a task while classifying data using the dataset. The KNN method merely stores the data from the training phase and classifies newly acquired data into a subset that is highly similar to the training data. When KNN is used, the accuracy obtained is 82.9% for Physionet dataset and 80% for MIMIC-III dataset. The evaluation metrics obtained are shown in the figures 5 and 6

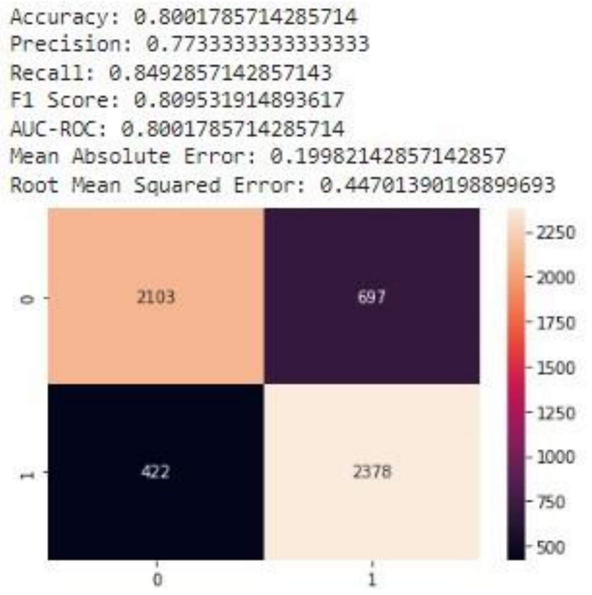


```
Accuracy: 0.8292443572129539
Precision: 0.7883534136546185
Recall: 0.6538974017321786
F1 Score: 0.714857975236708
AUC-ROC: 0.7842351330268933
Mean Absolute Error: 0.17075564278704614
Root Mean Squared Error: 0.413225898011059
```

Fig. 5.

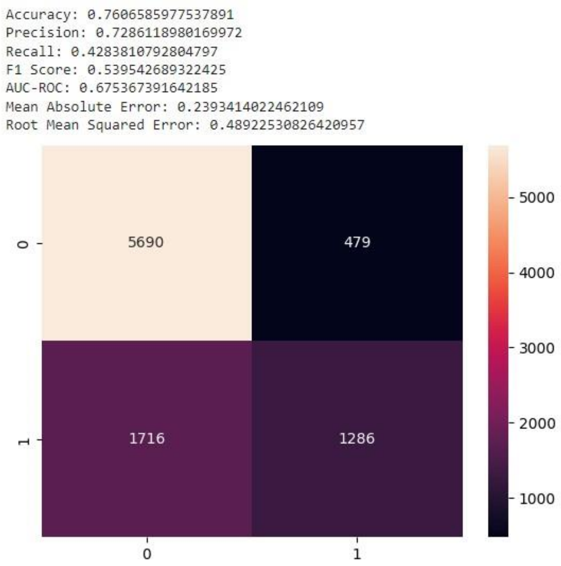
D. Naive Bayes

For classification tasks, the Naive Bayes classifier, a probabilistic machine learning model, is used. The classifier's underlying principle is the Bayes theorem



```
Accuracy: 0.8001785714285714
Precision: 0.7733333333333333
Recall: 0.8492857142857143
F1 Score: 0.809531914893617
AUC-ROC: 0.8001785714285714
Mean Absolute Error: 0.19982142857142857
Root Mean Squared Error: 0.44701390198899693
```

Fig. 6.

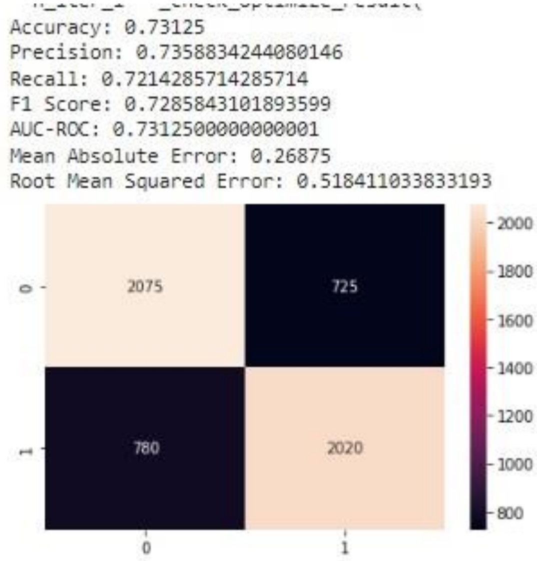


```
Accuracy: 0.7606585977537891
Precision: 0.7286118980169972
Recall: 0.4283810792804797
F1 Score: 0.539542689322425
AUC-ROC: 0.675367391642185
Mean Absolute Error: 0.2393414022462109
Root Mean Squared Error: 0.48922530826420957
```

Fig. 7.

VII. Output

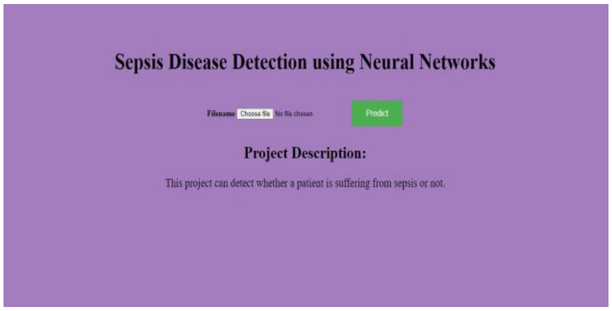
From the results obtained, we understood that Random forest gave better accuracy compared to other machine learning algorithms. So, to build web application, we integrated the Random forest model to detect sepsis disease. Using flask, we built a website as shown in figure 9. A text file which contains the 16 parameter values is given as input.Later we upload the text file as shown in figure 10. When we click on the predict button, then the output is displayed as shown in figure 11. In this way we get the output. If the prediction result is 0 it means sepsis is absent. If the result is 1 it means that sepsis is present.



```
Accuracy: 0.73125
Precision: 0.7358834244080146
Recall: 0.7214285714285714
F1 Score: 0.7285843101893599
AUC-ROC: 0.7312500000000001
Mean Absolute Error: 0.26875
Root Mean Squared Error: 0.518411033833193
```

Fig. 8.

GitHub Code Link: https://github.com/RajeshUCM/MLProject



Sepsis Disease Detection using Neural Networks

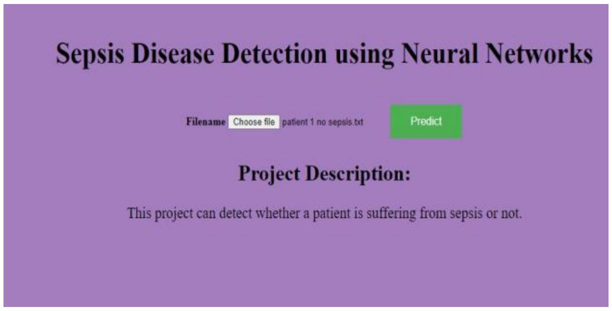
Project Description:

This project can detect whether a patient is suffering from sepsis or not.

Fig. 9.

CONCLUSION

In this study, various classifiers were utilised to compare the sepsis detection accuracy. The Random Forest classifier has the best accuracy of all of them, scoring 96.47% on the 'Physionet Challenge Sepsis dataset and 89.6% on the MIMIC-III dataset'. This model aids in the highly accurate early detection of sepsis disease. As a result, the practicality of sepsis detection has been suggested. Also, a web interface that can be installed on the hospital website is being developed so that clinicians may diagnose diseases quickly and accurately.

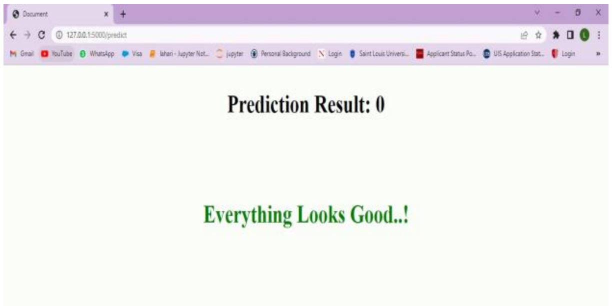


Sepsis Disease Detection using Neural Networks

Project Description:

This project can detect whether a patient is suffering from sepsis or not.

Fig. 10.



Prediction Result: 0

Everything Looks Good..!

Fig. 11.

REFERENCES

[1] Mahmud, F., Pathan, N. S., Quamruzzaman, M. (2019, December). Early detection of sepsis in ICU patients using logistic regression. In 2019 3rd International Conference on Electrical, Computer Telecommunication Engineering (ICECTE) (pp. 173-176). IEEE.

[2] Wang, R. Z., Sun, C. H., Schroeder, P. H., Ameko, M. K., Moore, C. C., Barnes, L. E. (2018, June). Predictive models of sepsis in adult ICU patients. In 2018 IEEE International Conference on Healthcare Informatics (ICHI) (pp. 390-391). IEEE.

GitHub Code Link: https://github.com/RajeshUCM/MLProject

[3] Thakur, J., Pahuja, S. K., Pahuja, R. (2018, May). Neonatal sepsis prediction model for resource-poor developing countries. In 2018 2nd International Conference on Electronics, Materials Engineering NanoTechnology (IEMENTech) (pp. 1-5). IEEE.

[4] Honor´e, A., Siren, H., Vinuesa, R., Chatterjee, S., Herlenius, E. (2022, December). An LSTM-based Recurrent Neural Network for Neonatal Sepsis Detection in Preterm Infants. In 2022 IEEE Signal Processing in Medicine and Biology Symposium (SPMB) (pp. 1-6). IEEE.

[5] Pawar, R., Bone, J., Ansermino, J. M., G¨orges, M. (2019, September). An algorithm for early detection of sepsis using traditional statistical regression modeling. In 2019 Computing in Cardiology (CinC) (pp. Page-1). IEEE.

[6] Shankar, A., Diwan, M., Singh, S., Nahrpurawala, H., Bhowmick, T. (2021, January). Early prediction of sepsis using machine learning. In 2021 11th International Conference on Cloud Computing, Data Science Engineering (Confluence) (pp. 837-842). IEEE.

[7] Shanthi, N. (2022, January). A novel machine learning approach to predict sepsis at an early stage. In 2022 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-7). IEEE.

[8] Mitchell, S., Schinkel, K., Song, Y., Wang, Y., Ainsworth, J., Halbert, T., ... Barnes, L. E. (2016, April). Optimization of sepsis risk assessment for ward patients. In 2016 IEEE Systems and Information Engineering Design Symposium (SIEDS) (pp. 107-112). IEEE.

[9] Guill´en, J., Liu, J., Furr, M., Wang, T., Strong, S., Moore, C. C., ... Barnes, L. E. (2015, April). Predictive models for severe sepsis in adult ICU patients. In 2015 Systems and Information Engineering Design Symposium (pp. 182-187). IEEE.

[10] Guill´en, J., Liu, J., Furr, M., Wang, T., Strong, S., Moore, C. C., ... Barnes, L. E. (2015, April). Predictive models for severe sepsis in adult ICU patients. In 2015 Systems and Information Engineering Design Symposium (pp. 182-187). IEEE.

[11] Guill´en, J., Liu, J., Furr, M., Wang, T., Strong, S., Moore, C. C., ... Barnes, L. E. (2015, April). Predictive models for severe sepsis in adult ICU patients. In 2015 Systems and Information Engineering Design Symposium (pp. 182-187). IEEE.

[12] Kausch, S. L., Moorman, J. R., Lake, D. E., Keim-Malpass, J. (2021). Physiological machine learning models for prediction of sepsis in hospitalized adults: an integrative review. Intensive and critical care nursing, 65, 103035.

[13] DeMaris, A., Selman, S. H. (2013). Logistic regression. In Converting Data into Evidence (pp. 115-136). Springer, New York, NY.

[14] DeMaris, A., Selman, S. H. (2013). Logistic regression. In Converting Data into Evidence (pp. 115-136). Springer, New York, NY.

GitHub Code Link: https://github.com/RajeshUCM/MLProject