International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 02 | Feb 2021 www.irjet.net p-ISSN: 2395-0072

PREMATURE DETECTION OF LIFE-THREATENING SEPSIS DISEASE USING MACHINE LEARNING ALGORITHMS

Dr.P.Durgadevi
Assistant Professor (Sr.G),Department of Computer Science and Engineering
SRM Institute Of Science and Technology,Vadapalani,Chennai

Abstract-*Sepsis is a Life-Threatening disease which arises when our body gives response to any infection. When our body starts responding to external infection it starts releasing chemicals in our body to fight against them if that chemical gets excessively released, it causes inflammation. Sepsis is a dangerous disease which needs to be detected and cured well before time to save the patient's life because the effect of Sepsis will increase with each hour of delay and can cause severe damage in the patient. There are many therapies as well as model been built to save patient life but are a bit slower in finding the cure for them but we are building a model which will predict the Sepsis well before time using only vital sign of the patients like Heart Rate, Age, WBC Count, PaCO2 level and then the patient can go for further treatment. Sepsis is a major public health concern with significant mortality, and healthcare expenses. We have proposed such a system which will predict the occurrence of sepsis well before time to save a patient's life. We will be using a sepsis dataset and training our model to predict the desired output. To check the accuracy by giving a sample test of patient record we will predict whether the patient is suffering from sepsis or not well before 5 hour because it is really important for a person to know whether he/she is suffering from a life threatening disease namely sepsis cause it's effect in human body will be increasing by 3-9% with each hour of delay.*

Key Words: Sepsis, inflammation, Dataset, prediction, accuracy

1. INTRODUCTION

According to the recent survey done by WHO, It is estimated that every year around 3 million new-borns and 1.2 million children suffer from Sepsis. Three out of ten deaths due to neonatal Sepsis are thought to be caused by resistant pathogens. Any kind of infectious pathogens can potentially cause Sepsis. It can lead to Septic Shock. The major Symptoms of Sepsis are patches of discoloured skin, decreased urination, changes in mental ability, low platelet count, problem breathing, abnormal heart functions, chills due to fall in body temperature, extreme weakness. Any type of infection can trigger Sepsis, but following types of infections are more likely to cause Pneumonia, abdominal infection, kidney infection, bloodstream infection. Moreover, young children and seniors, people with weaker immune systems, such as those with HIV are more likely to suffer with Sepsis. Prior models have been built but with less accuracy but we are using different machine learning algorithms like KNN, Logistic Regression, Gaussian Bayes. We compared all the algorithm and came up with different accuracies but among them KNN stands out and gives us higher accuracy.

2. LITERATURE REVIEW

Allison Sutherland et.al, 2011 introduced a four tertiary critical care setting for a multi-centre, prospective clinical trial all around Australia for collecting data. They applied an expression biomarkers to MT PCR(multiplex tandem-polymerase chain reaction) data to create a diagnostic rule. The MT PCR data were randomly partitioned into training and testing data set were applying different bioinformatics and statistical analysis predictions were done.

Chen Lin aimed to develop a sepsis detection model using LSTM concept. They used a CNN model which was added before LSTM to obtain local characteristic of EHR. The second component they used was a fully neural network introducing static information. Their framework was evaluated for two experimental tasks: visit level early diagnosis(left) and event level early prediction. Their result shows that for visit level early diagnosis by incorporating both CNN and static information.

Hye Jie Kam aimed to develop a sepsis detection model with deep learning methodologies. They were using feed-forward networks and for the model they were using a combined feature set, the AUC they produced was the same as that of the basic feature set. For the long short-term memory model they were using only the basic feature set and they improved the AUC to 0.929 compared with existing 0.887 of the In-Sight model.

Jacob aimed to develop high-performance early sepsis prediction technology for the general patient population. Their algorithm was developed and applied to the prediction of sepsis up to three hours prior to a patient's first five hour systemic inflammatory Response Syndrome(SIRS). When applied to a never before seen set of test patients. Their analysis showed the sepsis risk level in patient from the co-evolution of multiple risk factors were more important than the contributions from isolated individual risk factors when making prediction further in advance.

Senthil K.Nachimuthu et.al collected the data of around 3100 patients admitted to the emergency dept and measured the accuracy in the first 3 hours, 6 hours, 12 hours, 24 hours to understand the effect of sepsis with increasing time. Using Dynamic Bayesian Networks they have prepared their data using a sequence of steps and using the resulting data set to train and test the DBN model. Their main aim was early detection of sepsis even before many laboratory tests become available, ideally within the first few hours after admission.

3. METHODOLOGIES EMPLOYED

Methodologies 1: Dataset were collected from "Early Prediction of Sepsis from Clinical Data -- the Physio Net Computing in Cardiology Challenge".This dataset contains the patient admitted within an hour to within 10 hours and with each hour of delay the risk of a patient being losing his life will increase and till now many models have been produced but didn't meet the expected accuracy. The dataset contains 90 % patient without Sepsis and 10 % with Sepsis where many patients having common symptoms like discoloured skin, unconsciousness. Whenever a patient is admitted in the hospital with being accused of Sepsis he/she need to go for intense check-ups which may take 4-5 hours and in that time gap, the patient may die too so using only vital signs like HR, Temp, O2Sat, SBP, MAP, DBP, Rasp, EtCO2.These vital signs can easily be obtained within 1st hour of admission of the patient so our using these values our model will predict whether the patient is suffering from corona or not. Vital Signs are very important data which immensely help us to predict the happening of Sepsis

Methodologies 2:

Machine learning is a framework that allows the system to automatically learn from its experience without being explicitly programmed so using machine learning algorithms like Gaussian Bayes, KNN, Logistic Regression we can create our model to predict the Sepsis.

Gaussian Bayes: Gaussian Bayes is a classification algorithm (probabilistic classifier) with a strong independence assumption between features.Using Gaussian Bayes we got an accuracy of around 95 % with 85:15 ratio as training is to testing data and the classification report for Naive Bayes is shown in Table 1.

Table-1: Classification Report of Gaussian Bayes

	Precision	Recall	F1-Score	Support
0	0.99	0.97	0.98	28317
1	0.09	0.21	0.12	442
accuracy			0.95	28759
Macro Avg	0.54	0.59	0.55	28759

K-Nearest Neighbours: KNN is a classification or regression algorithm also called a Lazy Learner. KNN is best suited for this kind of prediction as Sepsis Prediction is also a classification problem as we based on vital signs, we will classify between whether the patient is Sepsis positive or not. For distance metrics, we will use a Euclidean metric

d(x, x') = √(x-x')^2 – (y – y')^2-------(1)

sing KNN we got an accuracy of around 98% with 85:15 ratio as training is to testing data and the classification report for KNN is shown in Table 2.

Table- 2: Classification report for KNN

	Precision	Recall	F1-Score	Support
0	0.99	1.00	1.00	47185
1	0.83	0.60	0.70	747
accuracy			0.98	47932
Macro Avg	0.91	0.80	0.85	47932

Logistic Regression: Logistic Regression is used to describe relationship between the dependent variable and the independent variable. Since Logistic regression is a binary classification algorithm it gives output between 0 and 1. The basic formula Logistic Regression uses is:

$$ln\left(\frac{p}{1-p}\right) = b_0 + b_1 x$$

Fig -1: Logistic Regression Formula1

where b0 is constant, b1 is the amount of change with one unit change in x

$$p = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p)}}$$

Fig -2: Logistic Regression Formula2

Logistic Regression can handle any number of categorical variables so

$$\frac{p}{1-p} = \exp\left(b_0 + b_1 x\right)$$

Fig -3: Logistic Regression Formula3

Using Logistic Regression we got an accuracy of around 99 % with 85:15 ratio as training is to testing data and the classification report for Logistic Regression is shown in Table 3.

Table- 3: Classification report for Logistic Regression Technique

	Precision	Recall	F1-Score	Support
0	0.98	1.00	0.99	47185
1	0.78	0.01	0.02	747
accuracy			0.99	47932
Macro Avg	0.88	0.50	0.51	47932

4. RESULTS AND DISCUSSION

Sepsis is an inflammatory disease that was predicted by our model in the form of 0 or 1. If 1 means a person is suffering from Sepsis and if 0 means a person is not suffering from Sepsis. The goal of our model was to predict Sepsis with higher accuracy and we did, Comparing all the algorithms which we use as shown in fig 4.we found out that Logistic Regression gave the best accuracy and is best suited for our model to predict Sepsis. This makes our model more suitable to be used in different health posts, Hospitals to predict the Sepsis as soon as the patient is admitted to the hospital. The following figure 1 shows the effect of all vital sign on the occurrence of Sepsis.

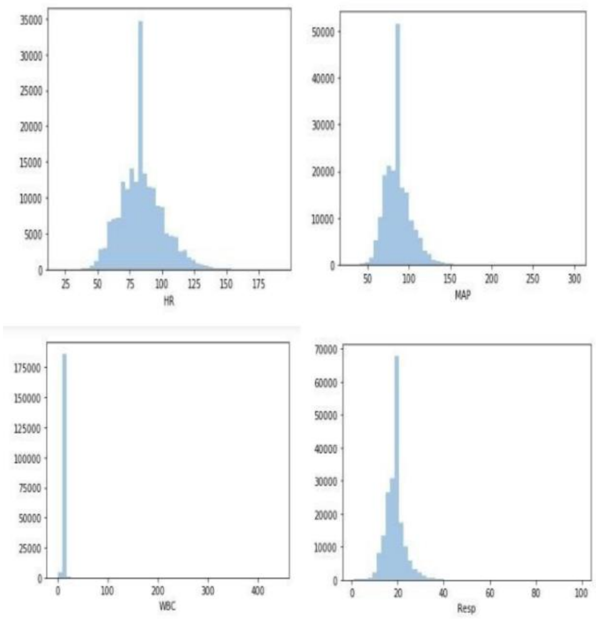

Fig-4: Level of Vital Sign-on Patient Admitted

5. CONCLUSION

This paper focuses on a detailed description of multiple sepsis modelling experiments using different machine learning algorithms like KNN, Gaussian Bayes, logistic regression. The data was collected, well pre-processed using python libraries and on that data different algorithms were applied, and based on accuracy given by them the final output was predicted to complete our model. This Model helps to save people's life by predicting sepsis inpatient using only normal vital signs like Heart Rate, Pressure, Temperature, Age, respiratory rate to save their lives. This model predicts the occurrence of sepsis in 1st hour of admission of the patient and saves their life.

REFERENCES

[1] Allison Sutherland, Mervyn Thomas, Richard B Brandon, June 2011"Development and validation of a novel molecular biomarker diagnostic test for the early detection of sepsis", 15:R149.

[2] Chen Lin, Yuan Zhang, Julie Ivy, July 2018“Early Diagnosis and Prediction of Sepsis Shock by Combining Static and Dynamic Information Using Convolutional-LSTM,” IEEE International Conference on Healthcare Informatics (ICHI).

[3] Hye jin Kam, Ha Young Kim,2016, "Learning representations for the early detection of sepsis with deep neural networks", vol. 89, pp. 248-255.

[4] Jacob S.Calvert, Daniel A.Price, and Ritankar Das, July 2016"A computational approach to early sepsis detection," vol 74, pp. 69-73.

[5] Matthew Reyna, Chris Josef, Russell Jeter, Aug 2019”Early Prediction of Sepsis from Clinical Data -- the PhysioNet Computing in Cardiology Challenge”.

[6] E. Slade, P. S. Tamber and J.-L. Vincent, "The Surviving Sepsis Campaign: raising awareness to reduce mortality", Critical Care journal article, vol. 7, no. 1, pp. 1, Jan. 2003.

[7] D. De Backer and T. Dorman, "Surviving sepsis guidelines: A continuous move toward better care of patients with sepsis", JAMA, vol. 317, no. 8, pp. 807-808, 2017.

[8] V. X. Liu et al., "The Timing of Early Antibiotics and Hospital Mortality in Sepsis", American Journal of Respiratory and Critical Care Medicine, vol. 196, no. 7, pp. 856-863, Oct. 2017.

[9] F. Lupu, "”Crossroads in Sepsis Research” Review Series Overview of the pathophysiology of sepsis", Journal Of Cellular And Molecular Medicine, vol. 12, no. 4, pp. 1072-1073, May 2008.

[10] T. C. Minderhoud, C. Spruyt, S. Huisman, E. Oskam, S. C. E. Schuit and M. D. Levin, "Microbiological outcomes and antibiotic overuse in Emergency Department patients with suspected sepsis", Neth J Med, vol. 75, no. 5, pp. 196-203, June 2017.

[11] C. Pierrakos and J.-L. Vincent, "Sepsis biomarkers: a review", Critical Care, vol. 14, no. 1, pp. R15, Feb. 2010.

[12] P. Y. Wu, C. W. Cheng, C. D. Kaddi, J. Venugopalan, R. Hoffman and M. D. Wang, "Omic and Electronic Health Record Big Data Analytics for Precision Medicine", IEEE Transactions on Biomedical Engineering, vol. 64, no. 2, pp. 263-273, 2017.