Predict the Readmission Patterns of Chronic Diseases Using Machine Learning

Karukonda Asha[1], K Manisri[2], A Bhanu Rakesh[3], Ch Rohitha[4], and S.V. Phani Kumar[5]

[1,2,3,4] Students, Department of Computer Science & Engineering, Dhanekula Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, India

[5] Assistant Professor, Department of Computer Science & Engineering, Dhanekula Institute of Engineering & Technology/JNTUK/, Vijayawada, Andhra Pradesh, India

Correspondence should be addressed to Karukonda Asha, ashakarukonda24@gmail.com

ABSTRACT- Machine learning plays an essential role in predicting presence or obsence of locomotor disorders. The diseases like heart, diabetes, cancer, kidney we collect data and such information if predicted well in advance can provide important insights to doctors who can adapt their diagnosis and treatment per patient basis. Here comes to heart disease alternatively known as cardiovascular disease, encases various conditions that impact the heart is the primary basis of death worldwide over the span of the past few decades. when we come to the disease like diabetes which is among critical disease and lot of people are suffering from this disease. People having diabetes, kidney disease, heart stroke and nerve damage. Cancer disease is very dangerous and critical. Kidney disease also recognized as chronic renal disease, is an uncharacteristic functioning of kidney or a failure of renal function expanding over a period of months or years and it is detected during the screening of people who are known to be in threat by kidney problems. So, the early prediction is necessary in combating these diseases and to provide good treatment. For predicting these diseases in people, we are using machine learning techniques.

KEYWORDS- Locomotors disorders, Diagnosis, Heart, Diabetes, Lungs, Cancer and Kidney.

I. INTRODUCTION

Diseases is seen as a common problem with no proper treatment it can't be cured. We can easily identify diseases on the basis of recognizable signs and symptoms that might be damage of tissues like eyes, kidneys, heart, blood vessels and nerves[1][2]. Machine Learning can help people to predict the diabetes and analyze weather the person is readmitting or not. And we have to select a proper dataset and a generic model that actually classifies the problem and it gives better accuracy. In general Machine Learning is playing a major role in solving problems in multiple domains. That including Healthcare, Stock Market, Finance, and others[6][7]. Based on the problem statement we have to select the dataset and make the data to fit to the model then we can predict the data with a better accuracy.

II. LITERATURE SURVEY

Various prediction models have been developed and implemented by various researchers using variants of data mining techniques, machine learning algorithms or also combination of this techniques Hybrid Prediction Model which includes Simple K-means clustering algorithm, followed by application of classification algorithm to the result obtained. In order to build classifiers C4.5 decision tree algorithm is used. Mani Butwall et al. proposed a model using Random Forest Classifier to forecast diabetes behavior[3][4][5].

III. EXISTING SYSTEM

Machine Learning can help people to predict the diseases and analyze whether the person is readmitting or not. And we have to select a proper dataset and a generic model that actually classifies the problem and it gives better accuracy. In general Machine Learning is playing a major role in solving problems in multiple domains[8][9]. That including Healthcare, Stock Market, Finance, and others.Based on the problem statement we have to select the dataset and make the data to fit to the model then we can predict the data with a better accuracy.

IV. PROPOSED SYSTEM

The dataset has been tested with multiple classifiers and those Classifiers classified the patients into two classes with respective having diseases and non-diseases[10]. And not only consider the accuracy metrics but also here in this metric classification we will recall with a good accuracy].To improve better accuracy, we have some of the techniques such as EDA (exploratory data analysis), Model optimization or Fine tuning to analyze the data in a graphical manner. Let us assume we have taken 5 medical parameters for heart disease such as:

- Age
- Gender
- Blood Pressure
- Cholesterol And
- Obesity

After predicting data by other person, he/she got the accuracy which is 80%.We predict those data in a better

way and get more than 80% i.e., some 85% with less error rate i.e., errors are minimized which means it is error free. In such a way we can able to improve better accuracy. Also, we haveothermethods such as treating missing values and outliers, hyper tuning and feature engineering.

V. DESIGN

Flowchart is a diagram of the sequence of movements or actions of people or things involved in a complex system or activity. Flowcharts are used in analyzing, designing, documenting or managing a process or program in various fields. It is a graphical representation of a computer program in relation to its sequence of functions. A flow chart can also be represented as diagrammed representation of an algorithm, a step-by-step approach to solve a task[11][12]. It represents a work flow or a process as shown in figure 1.
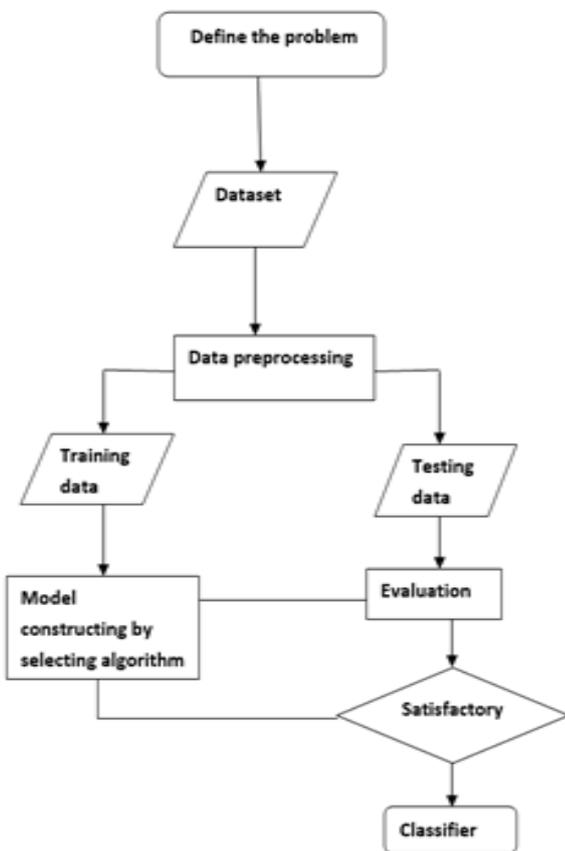

Fig. 1: Flow chart Diagram

VI. RESULTS/OUTPUT

In this work different steps were taken. The proposed approach uses different classification and ensemble methods and implemented using python. These methods are standard Machine Learning methods used to obtain the best accuracy from data. In this work we see that random forest classifier achieves better compared to others. Here we got different diseases (Diabetes, Heart, Kidney and cancer) outputs as shown in figure 2, 3, 4, 5. Overall we have used some of the best machine learning algorithms for the prediction and to get a better accuracy.
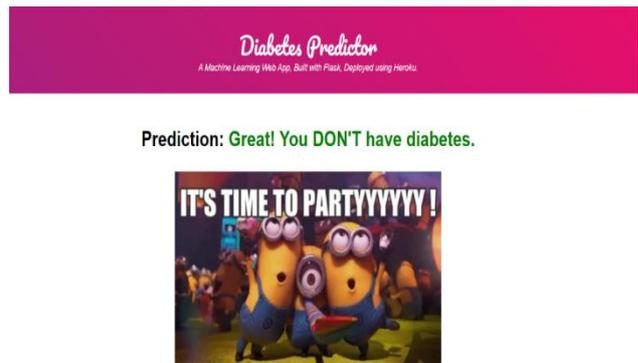
A. Diabetes (Output)

Fig. 2: Diabetes disease

B. Heart (Output)

Fig. 3: Heart disease

C. Kidney (Output)

Fig. 4: Kidney disease
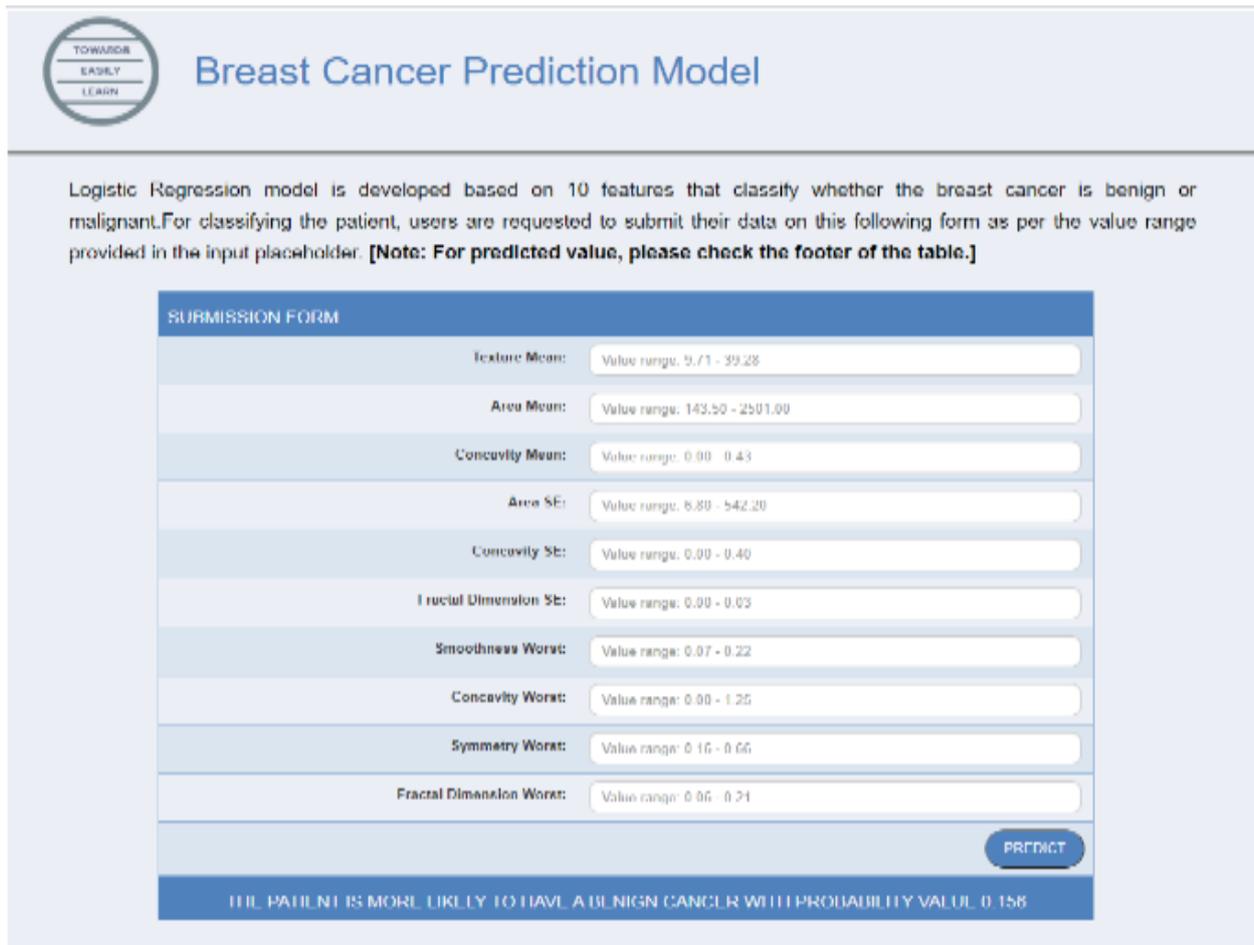
D. Breast Cancer (Output)

Fig. 5: Breast Cancer disease

VII. CONCLUSION

In this study, we employed five ML algorithms: Support Vector Machine, Naïve Bayes, k-Nearest neighbor, decision trees and random forest in one of the four diseases such as Diabetes, Heart, Breast Cancer, and Kidney (original) datasets. We tried to compare a better accuracy by using various machine learning algorithms. Also we have tried to compare its precision, sensitivity and specifity to find the best classification accuracy. After comparison of all the algorithms Random Forest reaches the accuracy of 99% and outperforms, therefore, all other algorithms. In conclusion, RF has proven its efficiency in one of the 4 diseases (Diabetes, heart, kidney and breast cancer) and achieves the best performance in terms of precision and low error rate.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] Ryd´en L, Standl E, et al. Guidelines on diabetes, pre-diabetes and cardiovascular diseases: executive summary: The task force on diabetes and cardiovascular diseases of the European Society of Cardiology ESC) and of the European Association for the Study of Diabetes (EASD).

[2] International Diabetes Federation. IDF diabetes atlas, 8th edition. http://diabetesatlas.org/IDF Diabetes Atlas 8e interactive , Belgium.

[3] Diabetes Prediction using Machine Learning Algorithms, Aishwarya Mujumdara et al., International Conference On Recent Trends In Advanced Computing 2019, ICRTAC 2019, ScienceDirect, www.sciencedirect.com

[4] Mani Butwall and Shraddha Kumar," A Data Mining Approach for the Diagnosis of Diabetes Mellitus using Random Forest Classifier",International Journal of Computer Applications, Volume 120 - Number 8,2015.

[5] Carol M. Ashton, David H. Kuykendall, et al. The Association between the Quality of Inpatient Care and Early Readmission. Annals of Internal Medicine.

[6] Sara J Healy, Dawn Black, et al. Inpatient diabetes education is associated with less frequent hospital readmission among patients with poor glycemic control. Diabetes Care, page DC 130108, 2013.

[7] Reena Duggal, Suren Shukla, et al. Predictive risk modelling for early hospital readmission of patients with diabetes in India. International Journal of Diabetes in Developing Countries , 2016.

[8] B. Strack, J. DeShazo, C. Gennings, J. Olmo, S. Ventura, K. Ciosand J. Clore, "Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records," BioMed Research International, vol. 2014.

[9]   Seckeler MD, Hoke TR. The worldwide epidemiology of acute rheumatic fever and rheumatic heart disease.

[10] Gaziano TA, Bitton A, Anand S, Abrahams-Gessel S, Murphy A. Growing epidemic of coronary heart disease in low-and middle-income countries. Curr Probl Cardiol.

[11] Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS ONE. 2017.

[12] Ramalingam VV, Dandapath A, Raja MK. Heart disease prediction using machine learning techniques: a survey. Int J Eng Technol.

ABOUT THE AUTHORS

Karukonda Asha, A UG Final Year student seeking his degree in Computer Science Engineering at Dhanekula Institute of engineering and Technology, Vijayawada, Andhra Pradesh

K Manisri, A UG Final Year student seeking her degree in Computer Science Engineering at Dhanekula Institute of engineering and Technology, Vijayawada, Andhra Pradesh

A Bhanu Rakesh, A UG Final Year student seeking her degree in Computer Science Engineering at Dhanekula Institute of engineering and Technology, Vijayawada, Andhra Pradesh

Ch Rohitha, A UG Final Year student seeking her degree in Computer Science Engineering at Dhanekula Institute of engineering and Technology, Vijayawada, Andhra Pradesh

S.V. Phani Kumar, Assistant professor in Computer Science Engineering at Dhanekula Institute of engineering and Technology, Vijayawada, Andhra Pradesh.