Heart Sound Analysis & Cardiovascular Prediction

Prof.Archana Chaudhri
DEPARTMENT OF INSTRUMENTATION ENGINEERING, VISHWAKARMA INSTITTUE OF TECHNOLOGY, PUNE INDIA

Ms.Nikita Kotwal[2]
DEPARTMENT OF INSTRUMENTATION ENGINEERING, VISHWAKARMA INSTITTUE OF TECHNOLOGY, PUNE INDIA

Ms.Gauri Unnithan[3]
DEPARTMENT OF INSTRUMENTATION ENGINEERING, VISHWAKARMA INSTITTUE OF TECHNOLOGY, PUNE INDIA

Ms. Anaya Pawar[4]
DEPARTMENT OF INSTRUMENTATION ENGINEERING, VISHWAKARMA INSTITTUE OF TECHNOLOGY, PUNE INDIA

***Abstract*—** Cardiovascular diseases are life-threatening and must be identified early on with a high-precision auscultation examination. Cardiac auscultation is a technique for analysing and listening to heart sounds using an electronic stethoscope. An electronic stethoscope is a device that records heart sounds digitally and is known as a phonocardiogram. The strength, tone, position, quality, and timing of heart sounds in the cardiac cycle may all be used to characterise them.. Using this sound frequency this heart sound is analyzed. The heart sound predicting model, when considered as a system with inputs of historical and current data and outputs of future dataA specific cardiac event, such as the closing of a valve or the tensing of the chordae tendineae, causes heart sounds. Auscultation of the heart sounds can be used to diagnose a variety of pathologic cardiac diseases. Heart sounds, unlike a heart murmur, are discrete, brief auditory occurrences caused by a specific source. A murmur is caused by blood flow turbulence and can sometimes span the whole systole or diastole cycle. The S1 and S2 heart sounds are the most common normal heart sounds. At times, the S3 might be normal, but it can also be pathogenic. A S4 heart sound is usually always a sign of something wrong. The goal of this research is to create an automated heart sound signal analysis system that can assist doctors in identifying heart murmurs early in life. The earliest indications of cardiac illness are heart murmurs. The strength, tone, position, quality, and timing of heart sounds in the cardiac cycle may all be used to characterize them.

Keywords—Data Analytics, Heart Sound Analysis, Machine Learning, Random forest classifier, Multiple Linear Regression,

I. Introduction

A variety of diseases that affect your heart are referred to as heart disease. The terms "Heart disease" and "Cardiovascular disease" are frequently used interchangeably. Cardiovascular disease is a term used to describe diseases in which blood arteries are constricted or obstructed, which can result in a heart attack, chest discomfort (angina), or stroke. Other cardiac diseases, such as those that affect the muscle, valves, or rhythm of your heart, are also classified as heart disease. Heart disease is the top cause of mortality for both women and men, according to a news story. Machine learning (ML) has shown to be useful in aiding in the decision-making and prediction of huge amounts of data generated by the healthcare industry. The scope of this paper is to investigate a multiple regression, Random forest classification approach for the cardiovascular disease prediction. This paper is especially used Random Forest Classifier to classify the Heart sound, and Multiple Linear regression used for prediction of S4. Accuracy was calculated by confusion matrix

II. LITREATUREREVIEW

Machine learning is a field of computer science that is used to create algorithms with self-learning properties, i.e., learning that is done by the machine itself, thus the name. It is considered one of the most important topics in Artificial Intelligence. To show intelligence machine needs to interpret and analyze the input. After analyzing it the result data apart from simply following the instructions on that data. In this paper we have focused on two main algorithms mainly 1. Multiple linear regression .2. Random forest classifier.

Numerous researches have been done in this field and achieved different result and conclusions.

By analysing HRV (Heart Rate Variability) from ECG, data pre-processing, and heart disease pattern, Kiyong Noh et al. [1] utilise a classification technique for the extraction of multiparametric features. The associative classifier experiment was carried out using a dataset of 670 persons divided into two groups: normal people and patients with heart disease[2]. Wiharto and Hari Kusnanto presented a paper titled Intelligence System for Coronary Heart Disease Diagnosis Level using K-Star Algorithm. They present an expectation framework for heart infection in this study. Learning how to quantify vectors computation of the neural system This framework's neural system recognises 13 clinical features as input and forecasts the

presence or absence of cardiac disease in the patient, as well as different execution measures.

Niti Guru et al. [3] advocated using neural networks to predict heart disease, blood pressure, and sugar levels. There are 13 characteristics in each entry in the collection. For data training and testing, supervised networks, such as neural networks with back propagation algorithms, are utilised.

Noura Ajam [4] suggested using an artificial neural network to diagnose cardiac problems. Feed forward Back propagation learning techniques were employed to test the model based on their capabilities. Classification accuracy increased to 88 percent when proper function was taken into account, using 20 neurons in the buried layer. For the prediction of cardiac disease, ANN provides a substantial outcome. Seema et al.,[5] focuses on approaches that employ Nave Bayes, Decision Trees, Support Vector Machines (SVM), and Artificial Neural Networks to predict chronic illness by mining data from past health records (ANN). A comparison analysis of classifiers is conducted to determine which performs better on an accuracy rate. SVM had the best accuracy rate in this trial, but Nave Bayes has the highest accuracy rate for diabetes.

III. Proposed System Architecture

Input: Many variables influence the risk of heart disease, including blood pressure, diet, artery thickness, vessel shrinking, and so on.

Data Acquisition: Depending on the required section, user needs to enter the input data.

Data Processing: A disturbed or normal sound is predicted accurately only if entered data is valid and correct.

Output: what exactly is the condition of the heart due to change in heart beat is predicted

Computational statistics, which likewise focuses on generating predictions using computers, is closely connected to machine learning. The algorithms that we used were also highly efficient for this particular task; hence the usage. The Random Forest Classifier is a statistical technique for grouping data points into functional categories. When the dataset is huge and/or there are numerous variables, clustering becomes more challenging since not all of the variables can be included. As a result, the algorithm can predict if a data point belongs to a specific category.

Fig. 1 represents purposed system architecture for heart disease

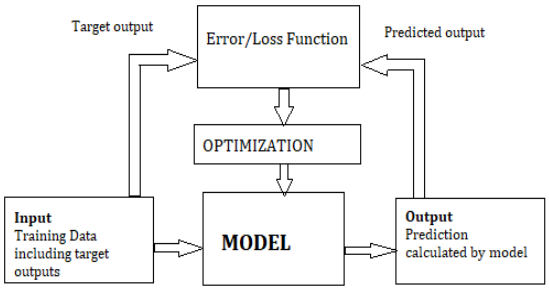

Figure 1. Proposed system architecture

IV. ALGORITHMIC REVIEW

Regression: Random forest is a supervised learning method that may be used for both classification and regression. However, it is mostly employed for categorization issues. As we all know, a forest is made up of trees, and more trees equals a more strong forest. Similarly, the random forest algorithm constructs decision trees from data samples, then receives predictions from each of them, and ultimately votes to choose the best option. It is an ensemble approach that is superior than a single decision tree because it eliminates over-fitting by averaging the results.

Fig. 2 represents the classification using regression model.

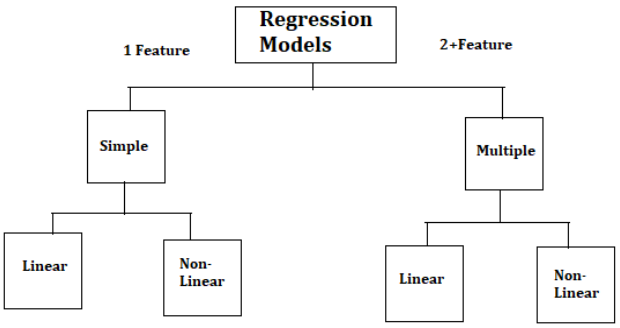

Figure 2. Regression Models for Classification

Multiple Linear Regression: A Multiple Linear Regression model is one that incorporates more than one predictor variable.

Equation for Multiple-Linear Regression.

Observed data: $y = b_0 + b_1x_1 + b_2x_2 + \ldots\ldots + b_px_p + \varepsilon$

Predicted data: $y' = b_0 + b_1x_1 + b_2x_2 + \ldots\ldots + bpxp$

Error: $\varepsilon = y - y'$

Multiple linear regression is performed by generating a model with existing dataset or a real time data set which is used for future prediction

Step 1: Load the data set

Step 2: Feature selection

Step 3: Select the predictor and targeted variables

Step 4: Normalize the data

Step 5: Training/model fitting

Step 6: Model parameters study

Random Forest Classifier: - Random forest is a supervised learning technique that may be used for classification and regression. However, it is mostly employed to solve categorization issues. A forest, as we all know, is made up of trees, and more trees equals a more strong forest. Similarly, the random forest algorithm constructs decision trees from data samples, then receives predictions from each of them, and ultimately votes to choose the best option. It is an ensemble approach that outperforms a single decision tree because it eliminates over-fitting by averaging the results.

Step 1 – First, start with the selection of random samples from a given dataset.

Step 2 – Next, this algorithm will construct a decision tree for every sample. Then it will get the prediction result from every decision tree.

Step 3 – In this step, voting will be performed for every predicted result.

Step 4 – At last, select the most voted prediction result as the final prediction result.

Fig 4 Represents random forest classifier model.

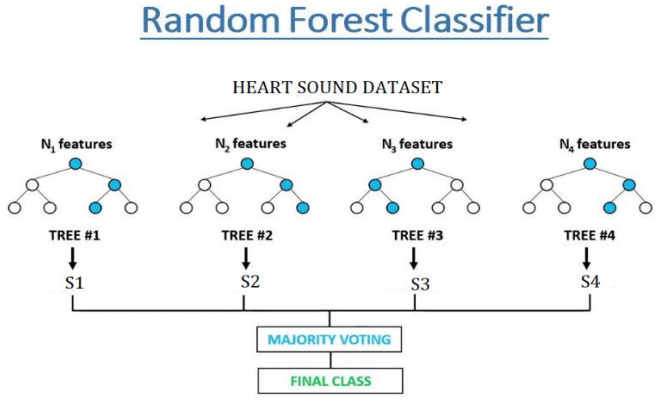

Figure 4. Random Forest Classifier Model

V. ALGORITHM IMPLEMENTATION

Table1 Data Description for traffic flow prediction

S. No	Attribute Name	Possible Range (Hz)
1	S1	50-180
2	S2	50-200
3	S3	50-90
4	S4	50-80
5	Murmur	200-600

Above table1 contains attributes which has to be considered for the better prediction and their range.

A. Dataset used for traffic flow Prediction

Fig. 5 Represents dataset used for prediction & analysis which contains parameters sound waves S1, S2, S3, S4, Murmur in Hz

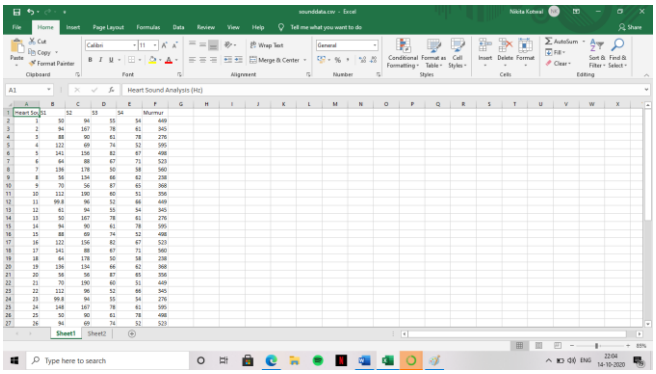

Figure 5. Features used for Sound prediction & Analysis

Fig. 6 Represents prediction using Random Forest Classifier

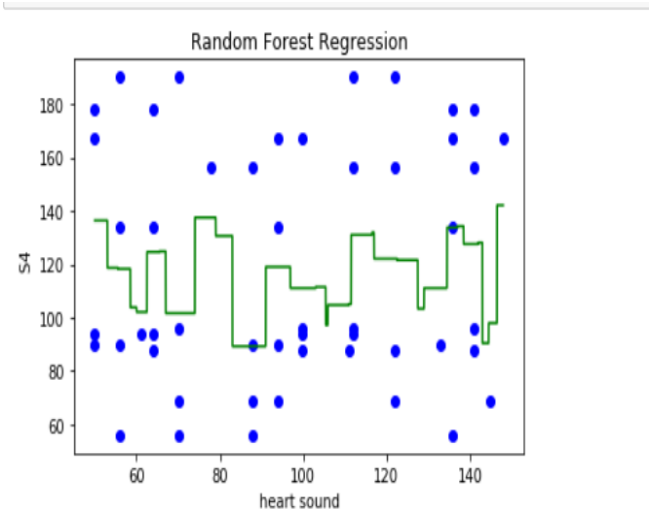

Figure 6. Convergence of the proposed method for prediction

Fig 7. Represents Prediction of S4 using S1 & S2 and its classification using Multiple regression model.

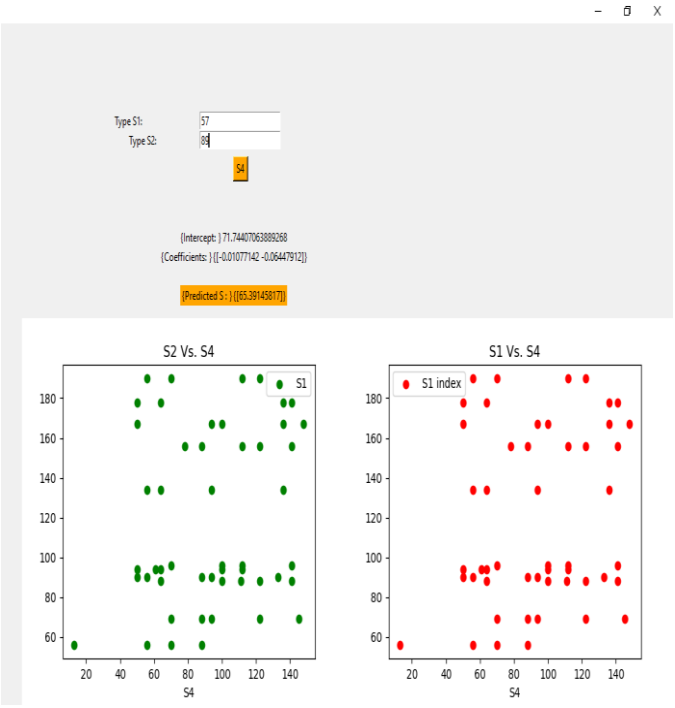

Figure 7. S4 prediction & classification of sound waves

B. Validation Metrics

The validation of the proposed method is computed using Accuracy. The ratio of correct predictions to the total number of input samples is known as accuracy.

For prediction using multiple linear regression:

Accuracy using accuracy prediction function
Accuracy score is 2.
Table 5 demonstrates the accuracy as the evaluation metrics for both heart sound analysis and its classification.

Table5 Evaluation Metrics

Accuracy	76.6666%

VI. CONCLUSION AND FUTURE SCOPE

In Heart sound analysis using the heart sound heartbeat is classified in abnormal & normal heartbeat. Depending on given input of S1 & S2 possible range of S4 is predicted.

The primary limitation of the study is the lack of sufficiently large data to train the prediction model. Our next steps will be to collect more data from a variety of sources and to increase the suggested machine learning classifier's accuracy.

REFERENCES

[1] Kiyong Noh, Heon Gyu Lee, Ho-Sun Shon, Bum Ju Lee, and Keun Ho Ryu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", Springer, Vol:345, pp: 721- 727, 2006

[2] Diagnosing Heart Disease Patients Using k-Nearest Neighbour Tim Turner, Mai Shouman, and Rob Stocker Vol. 2, No. 3, June 2012, International Journal of Information and Education Technology

[3] Niti Guru, Anil Dahiya, Navin Rajpal, "Decision Support System for Heart Disease Diagnosis Using Neural Network", Delhi Business Review, Vol. 8, No. 1 (January - June 2007).

[4] Noura Ajam, "Heart Disease Diagnoses using Artificial Neural Network", The International Insitute of Science, Technology and Education, vol.5, No.4, 2015, pp.7-11.

[5] Dr.S.Seema Shedole, Kumari Deepika, "Predictive analytics to prevent and control chronic disease", https://www.researchgate.net/punlication/316530782, January 2016.