International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 03 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

Heart Health Classification and Prediction Using Machine Learning

Hardi Rathod[1], Pratik Kumar Singh[2], Nikita Tikone[3], Suresh Babu[4]

[1,2,3]BE student, Information Technology, Pillai College of Engineering, Navi Mumbai, Maharashtra, India
[4]Professor, Dept. of Information Technology, Pillai College of Engineering, Navi Mumbai, Maharashtra, India

---

Abstract - *According to a survey [1], one in every four deaths occurs due to cardiovascular diseases. Early diagnosis of these diseases can help prevent deaths by alerting the state of heart. It has been found that medical professionals working in the field of heart diseases can predict the chances of heart attack with around 67% accuracy. Our attempt is to create a machine learning model which can be more accurate than the medical practitioners. Machine learning algorithms like support vector machines, logistic regression, random forest and ensemble learning can provide logical reasoning to base our predictions. In order to increase accurate diagnosis on the medical practioner's part, a heart sound classification model based on heart sounds can help any physician, radiologist or patient determine the current health of their heart. It is based on S1 i.e. lub and S2 i.e. dub sounds of the heart and the model classifies the heart sound to determine how much of the sound can be classified into normal, murmur and extrasystole. To identify the significant heart sound beats, a heart sound segmentation model is created. This can help physicians to provide early advice to their patients without any exclusive and time exhaustive tests, saving money and any delay. Further, a sound recorder placed inside the stethoscope can make the whole process more efficient and easier.*

Key Words: Heart sound segmentation, Support Vector Machines, Regression, Random Forest, Deep Learning, Convolutional Neural Network

1. INTRODUCTION

Machine learning can be defined as the identification of patterns, functions or inferences from a set of varied data in order to compute similar tasks automatically and solve statistical problems. The importance of machine learning in healthcare is its ability to process huge datasets, and then provide analysis of the data that can help doctors in planning and providing care, ultimately leading to lower costs of care, and increased patient life expectancy.

The paper consists of two major sections: Heart disease prediction based on medical records and Heart Sound Classification based on Heart Sounds. The project takes data sets from two sources. The first dataset has been taken from the University of California Irvine repository's Cleveland database. PASCAL Challenge includes a dataset that was obtained from the general public using an iPhone app called iStethoscope and a clinic trial using the DigiScope digital stethoscope.

The proposed system is intended to be used by doctors, physicians as well as normal users. The users can give input to the web portal. Using the input provided by the users, the Machine Learning model can easily predict whether the user is diagnosed with heart disease or not. The doctors can use the sound recorder present in the stethoscope to give the input to the classification model which will classify the patient's heart sound as healthy or murmur.

2. METHODOLOGY

2.1 Heart Disease Prediction

In the proposed system, Support Vector Machine, Logistic regression and Random forest algorithms were used. System's workflow is discussed below:

The dataset initially consists of 76 attributes, out of which we selected 14 most relevant attributes.

During our exhaustive literature survey, we found out that Support Vector Machine along with other Machine learning algorithms such as Logistic Regression and Random Forest algorithms were excellent for predicting heart diseases. Along with these algorithms, we also made use of the ensemble learning. Hence, these algorithms were used to create our ML Model.

In the proposed application, users (doctor, patient, physician, etc.) will be able to input the attribute values and send it to the ML Model. The ML Model, upon receiving the input, will predict the heart disease.

The proposed system model is shown in Figure 1.

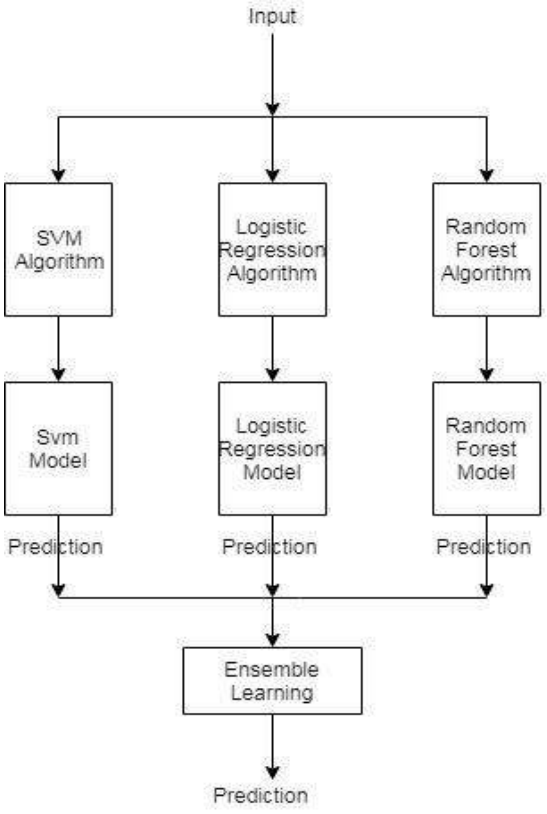

Fig. 1. System Architecture

1. Support Vector Machines (SVM): Support Vector Machine algorithm is most widely used for solving classification problems. It is based on statistical learning theory. In this, every single data-item is drawn as coordinates in the n-dimensional space. The SVM contains decision-hyperplanes which divides different classes of data points using maximum margin. Data points near hyperplanes are called support vectors. They affect hyperplane position and their orientation. The classification process creates a non-linear decision boundary and classifies data points not represented in vector space.
2. Logistic Regression Algorithm: Logistic regression is yet another popular machine learning algorithm. It is used for binary classification and is based on the concept of probability. Logistic Regression uses a complex cost function called the 'Sigmoid function'. This function is used for mapping any real value into another value between 0 and 1. Therefore, Logistic regression algorithm helps in predicting whether something is True or False.
3. Random Forest Algorithm: Random forest is mostly used for classification and regression problems. The random forest is composed of multiple decision trees. By averaging out the impact of several decision trees, random forests tend to improve prediction. It uses a modified tree learning algorithm that inspects, at each split in the learning process, a random subset of the features.
4. Ensemble Learning: Ensemble learning is a technique wherein different models trained over the same data are bought together and combined to solve a particular machine learning problem. It is used to increase the accuracy of the individual models since one model may predict correctly for one trend of data, while for another trend, another model may predict right. It also increases the performance of the program.

2.2 Heart Sound Classification

The idea of using heart sounds to classify heart health was highly popularized by Dr. Andrew Ng in 2017. It tries to mimic how a real physician would try to determine whether a certain person has heart disease or not.

The main motivation behind this technique was to enable the use of our system to detect genetic and hereditary heart problems as well that may not have been influenced by lifestyle choices. This section consists of Sound Segmentation and Classification. Both models are primarily based on the Convolutional Neural Network (CNN) algorithm.

Convolutional Neural Network Algorithm/CNN: CNN is a deep learning-based algorithm that has capability to understand features in complex media like images, audio, etc. It understands the features by finding differences between two objects. CNN is composed of neurons and layers. These neurons are defined by weights. These layers are arranged in order and receive input as the previous layer's output.

The heart sounds need to be segmented in order to determine the location of the two important beats - S1 and S2. This audio processing step will also divide the incoming wave into short segments and reduce noises. The training dataset consists of labelled locations of S1 and S2. Other files are unlabeled and need to be segmented into small sections after figuring out the locations of S1 and S2. S1 is the first sound and caused by the closure of mitral and tricuspid valves at the start of systole. The second sound, S2 is caused by the closure of the aortic and pulmonic valves. The time between S1 and S2 defines systole and the time between the S2 and S1 defines diastole. The rising phase of the wave corresponds to the beginning of systole(S1) and the declining phase of the wave is S2. Before starting with classification, exhaustive visualization of each of the classes was done. Heart Sound Classification algorithm takes in segmented data in order to understand the duration between each lub and dub and vice versa. The order of the lub and dub and the time duration between each of these beats are very significant for classification. For example, if a lub is followed by another lub, then it denotes that the heart sound belongs to extrasystole and has some irregularity.

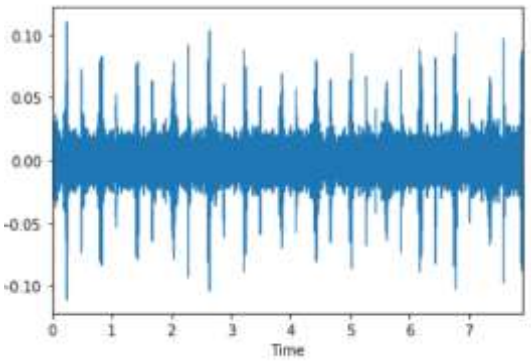

Fig. 2. Plotted waveform of an audio sample

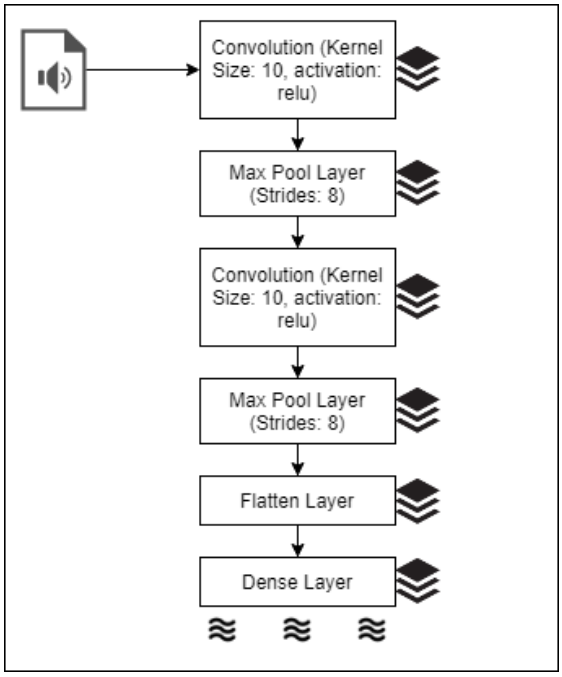

Fig. 3. CNN Model for Heart Sound Segmentation

After the sounds have been segmented and the locations of S1 and S2 beats identified, classification can be done. For this, we applied SVM algorithm and CNN Algorithm. The deep learning method helps in better audio or any media analysis. Transfer learning approach is used to obtain maximum efficiency and provide the model with more data trained model's inferences. The pretrained model is trained on Audio set Data which has around 6000 hours of audio data and is trained over 567 categories. The structure of the pretrained model is VGG CNN. It starts with spectrograms at the base and convolution layers with full connected layers form the top. Group of Convolution Layer, Max Pool Layer and Batch Normalization Layer needs to be applied seven times for the final model. The convolution layer provides with a matrix formed by extracted features from the input audio, the max pool layer helps in reducing the dimensions of the matrix so that assumptions about the down-sampled image's features can be made and batch normalization layer helps in scaling and independent learning between layers. The classes of the classification are presented in Input Details. The model takes about an hour and half to train for the first time. After saving it as an 'H5' file, it took about a minute to see results for a new patient. The model is scalable as new results are combined with the 'H5' model file.

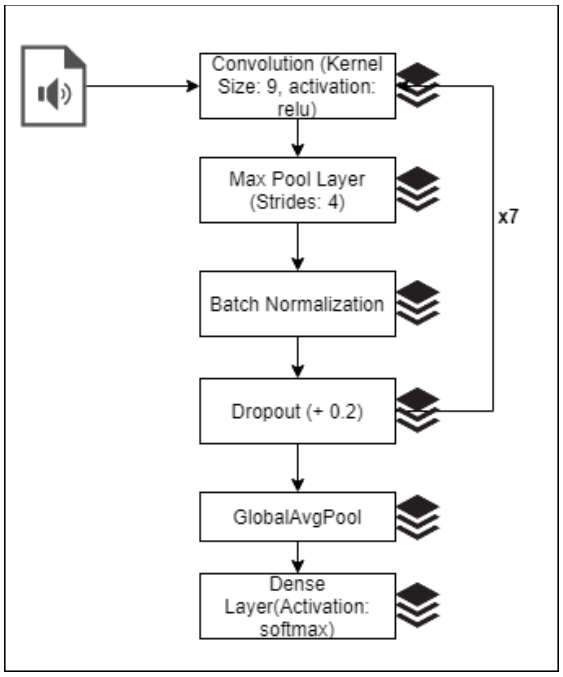

Fig. 4. CNN Model for Heart Sound Classification

3. CONCLUSIONS

3.1 Input details

The dataset that we have used has been taken from University of California Irvine machine learning repository which includes 14 attributes and have been summarized in table 1.

TABLE I. ATTRIBUTES WITH DESCRIPTION

Sr No.	Attributes	Description
1	Age	Age of patient in years
2	Sex	1=Male, 0=Female
3	Chest Pain(cp)	1: typical angina 2: atypical angina 3: non-anginal pain 4: asymptomatic
4	Resting blood pressure	Resting blood pressure of the patient
5	Serum cholesterol	Serum cholesterol of the patient
6	Fasting blood pressure	1 = true; 0 = false

7	Resting electrocardiographic results	0: normal 1: ST-T wave abnormality 2: probable hypertrophy
8	Thalach	Maximum heart rate of the patient
9	Exercise induced angina	1= Yes, 0= No
10	Oldpeak	ST depression induced by exercise relative to rest
11	Slope	1: upsloping, 2: flat, 3: downsloping
12	CA	number of major vessels (0-3) colored by fluoroscopy
13	Thal	3 = normal; 6 = fixed defects; 7 = reversible defects
14	Num	0: < 50% diameter 1: > 50% diameter

The other dataset that we have used is from the PASCAL Challenge that obtained data from clinical trials. The sound recordings were in the .wav format and gathered from the general public using an iPhone app called iStethoscope and DigiScope stethoscope [8]. The dataset group 1 is used for heart sound segmentation algorithm. It consists of 176 files. The training files had S1(lub) and S2(dub) locations. The dataset group 2 which is used for classification consisted of 656 audio files of about 30 seconds. Files provided are noisy and have many background noises. The categories were:

1. Normal: These are the normal. healthy heart sounds. A normal heart sound has a clear "lub dub, lub dub" pattern, and the time from "lub" to "dub" is shorter than the time from "dub" to the next "lub". Example of Temporal Description provided in the dataset website [8]:

   …lub……..dub……..lub……..dub……..lub ……..dub…….. lub…….dub…

2. Murmur: Heart Murmurs are found between "lub" and "dub" or between "dub" and "lub". They are precursors or indicators of many serious heart diseases. An example of temporal description of Heart Murmur from the website [8] is:

   *** defines murmur.

   …lub..****...dub……..lub..****..dub……..lub ..****..dub…….. lub..****..dub …

3. Extrasystole: These are the heart sounds consisting of some irregularities like skipped or extra beats. An extrasystole may or may not be a sign of disease. The temporal description from the website [8] is

   ……..lub…….dub……..lub…….dub ……..lub.lub……..dub…….

3.2 Output details

The accuracy of the algorithms we have used have been summarized in Table 2.

TABLE II. ACCURACY OF ALGORITHMS

Sr. No.	Algorithm	Accuracy
Heart Disease Prediction		
1	Support Vector Machine (SVM)	90.11%
2	Logistic Regression	86.90%
3	Random Forest	73.69%
4	Ensemble learning	83.56%
Heart Sound Classification		
5	Support Vector Machine	70%
6	Convolutional Neural Network (CNN)	85%

4. RESULTS AND DISCUSSION

1. Home Page

Fig. 5. Home Page

2. Input from user

Fig. 6. Input from User

3. Result (Heart Disease Prediction based on Lab Test Parameters)

Fig. 7 (a). Result

Fig. 7 (b). Result

4. Results (Heart Sound Classification)

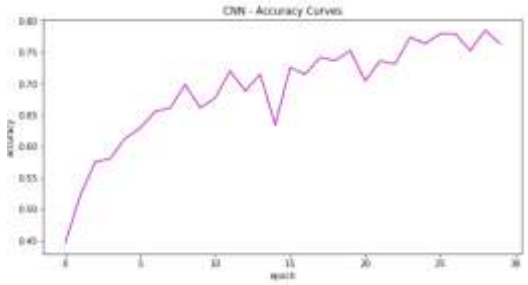

Fig. 8 (a). Accuracy Curves for CNN

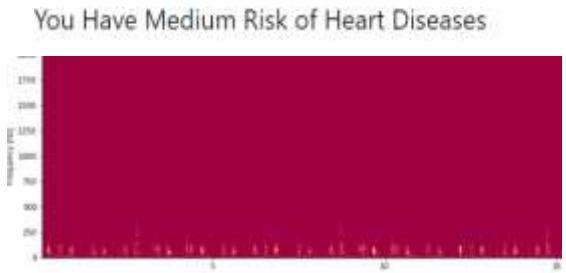

Fig. 8 (b). Wave Diagram of a Patient and Analysis/ Result

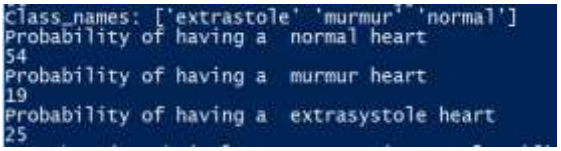

Fig. 8 (c). Probability Statistics of Another Patient

5. CONCLUSIONS

The system is GUI-based, user-friendly and scalable. The proposed working model can help in reducing treatment costs by providing early diagnostics in time. The model can serve the purpose of training tools for medical students and will be a soft diagnostic tool available for physician and cardiologist.

This system can easily be incorporated with doctors by incorporating heart sound recorders into their stethoscope so that they can record heart beats and the model can give them some insight into health issues. Since these recorders are cheap, even normal people can use them and check their heart health status at our website.

ACKNOWLEDGEMENT

We would like to acknowledge and render our warmest thanks to our Guide Prof. Suresh Babu who gave us the opportunity to do this Artificial Intelligence project on the topic 'Heart health classification and prediction using

machine learning'. His friendly guidance and expert advice have been invaluable throughout all stages of the project.

We thank our H.O.D. of the Information Technology department, Dr. Satishkumar L. Varma and Principal Dr. Sandeep Joshi for extended discussions and valuable suggestions which have contributed greatly to the improvement of the project.

REFERENCES

[1] Ahmad A. Abdul Aziz, "Tackling the burden of Cardiovascular diseases in India", Cardiovascular Quality and Comparison, 2018.

[2] K. Aravinthan, Dr. M. Vanitha, "A Comparative Study on Prediction of Heart Disease using Cluster and Rank based Approach", International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 2, February 2016.

[3] Rajesh N, T Maneesha, Shaik Hafeez, Hari Krishna, "Prediction of Heart Disease Using Machine Learning Algorithms", International Journal of Engineering & Technology, 7 (2.32) (2018) 363-366, May 2018.

[4] Mirpouya Mirmozaffari, Alireza Alinezhad, and Azadeh Gilanpour, "Heart Disease Prediction with Data Mining Clustering Algorithms" , Int'l Journal of Computing, Communications & Instrumentation Engg. (ECCIE) Vol. 4, Issue 1 (2017).

[5] Shashikant Ghumbre, Chetan Patil, and Ashok Ghatol, "Heart Disease Diagnosis using Support Vector Machine", International Conference on Computer Science and Information Technology (ICCSIT 2011) Pattaya, Dec. 2011.

[6] Mudasir Manzoor Kirmani,Syed Immamul Ansarullah,"Prediction of Heart Disease using Decision Tree a Data Mining Technique", International Journal of Computer Science and Network, Volume 5, Issue 6, December 2016.

[7] Amin Ul Haq, Jian Ping Li, Muhammad H Memon, Shah Nazir, and Ruinan Sun, "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms", School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China, December 2018.

[8] Classifying Heart Sounds Challenge, Peter Bentley, PASCAL.