INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 3 - V7I3-2222)
Available online at: https://www.ijariit.com

Human heart condition prediction using Machine Learning implementation paper

Akshay Baban Pathare
akshaypathare2977@gmail.com
JSPM'S Imperial College of Engineering and Research, Pune, Maharashtra

Abhijit Yuvraj Sonawane
sonawane.abhijit58@gmail.com
JSPM'S Imperial College of Engineering and Research, Pune, Maharashtra

Santosh Machhindra Punde
santoshpunde1999@gmail.com
JSPM'S Imperial College of Engineering and Research, Pune, Maharashtra

Akash Rajendra Shete
akashshete9615@gmail.com
JSPM'S Imperial College of Engineering and Research, Pune, Maharashtra

Santosh Waghmode
stwaghmode@gmail.com
JSPM'S Imperial College of Engineering and Research, Pune, Maharashtra

ABSTRACT

Heart disease is one of the complex diseases and globally many people suffered from this disease. On time and efficient identification of heart disease plays a key role in healthcare, particularly in the field of cardiology. In this article, we proposed an efficient and accurate system to diagnosis heart disease and the system is based on machine learning techniques. The system is developed based on classification algorithms includes Support vector machine, Logistic regression, Artificial neural network, K-nearest neighbour, Naïve bays, and Decision tree while standard features selection algorithms have been used such as Relief, Minimal redundancy maximal relevance, Least absolute shrinkage selection operator and Local learning for removing irrelevant and redundant features. We also proposed novel fast conditional mutual information feature selection algorithm to solve feature selection problem. The features selection algorithms are used for features selection to increase the classification accuracy and reduce the execution time of classification system. Furthermore, the leave one subject out cross-validation method has been used for learning the best practices of model assessment and for hyper meter tuning. The performance measuring metrics are used for assessment of the performance of the classifiers. The performances of the classifiers have been checked on the selected features as selected by features selection algorithms. The experimental results show that the proposed feature selection algorithm (FCMIM)is feasible with classifier Random Forest Classifier for designing a high-level intelligent system to identify heart disease. The suggested diagnosis system achieved good accuracy as compared to previously proposed methods. Additionally, the proposed system can easily be implemented in healthcare for the identification of heart disease.

Keywords: *Heart Classification, Support Vector Machine, Random Forest Classifier, Machine Learning*

1. INTRODUCTION

Cardiovascular disease (CVD) continues to be the leading cause of morbidity and mortality worldwide with an estimated 17.5 million people having died from CVD related conditions in 2012, representing 31% of all global deaths . However, with patient to doctor ratios as high as 50,000:1 in some regions of the world, access to expert diagnosis is often impeded. A potential solution to this is to provide automated diagnosis on the mobile phone or in the cloud.

Typical methods for heart sound classification can be grouped into: artificial neural network-based classification, support vector machine-based classification, hidden Markov model-based classification and clustering based classification.
In this project, we will be doing classification on prerecorded audio files of heartbeat sounds into the three level: Normal , Abnormal(refer for further diagnostics) and unsure(too noisy to make decision ;retake the recordings).

2. EXISTING SYSTEMS

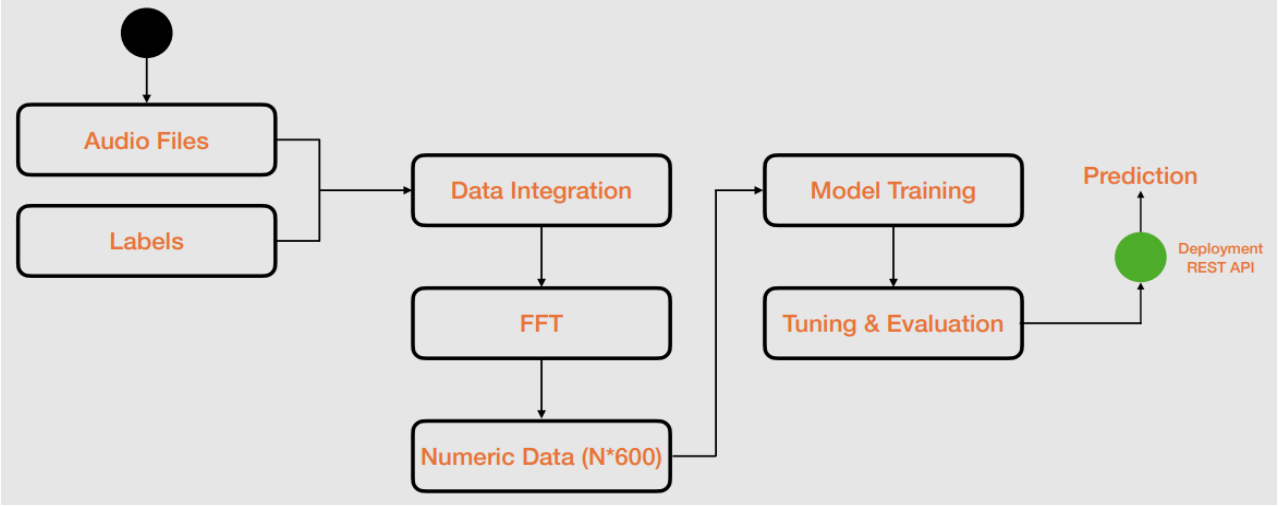
Title: Heart Disease Classification Using Nearest Neighbor Classifier With Feature Subset Selection
Author: M. Akhil Jabbar, B. L. Deekshatulu, Priti Chandra
Description: In this project we investigate by applying KNN with feature subset selection to predict heart disease of a patient for Andhra Pradesh population. .K nearest neighbor is one of the widely used data mining technique in classification. It is a straight forward classifier where samples are classified based on the class of their nearest neighbor .Medical data bases are high volume in

nature. If medical data contains redundant, irrelevant attributes classification will produce less accurate results. Feature subset selection is a dimensionality reduction technique used to remove redundant features and to increase accuracy. By applying feature subset selection on medical data we can determine the attributes which contributes more towards the disease which indirectly reduces no. of clinical tests to be taken by a patient .This project investigates to apply K nearest neighbor with feature subset selection in the diagnosis of heart disease. The experimental results show that applying feature subset selection to KNN will enhance the accuracy in the diagnosis of heart disease for Andhra Pradesh population.

Title: Human Heart Sounds Classification using Ensemble Methods
Author: S. Ali , S. M. Adnan
Description: In this project, sounds classification on well-known ensemble techniques [xx] was done. To our information, for such classification task, no such type of ensemble methods has been used. In classifying human heart sounds, Ensemble methods have proved effective results as compared with the results of individual classifiers. Our study also compares the results of suggested setup of ensemble methods with the existing results in [ix,xxi] and shows the substantial enhancement over them. This article is structured as follows. The proposed methodology and experimental setup are given in Section II. Section III presents the results of our experimental setup, analysis of results and comparison with the existing work. Conclusion is presented in the last section.

Title: Heart Disease Identification Method Using Machine Learning classification in E-Healthcare
Author: Jianping Li1 , Amin Ul Haq
Description: In this Project, we proposed an efficient and accurate system to diagnosis heart disease and the system is based on machine learning techniques. The system is developed based on classification algorithms includes Support vector machine, Logistic regression, Artificial neural network, K-nearest neighbor, Naïve bays, and Decision tree while standard features selection algorithms have been used such as Relief, Minimal redundancy maximal relevance, Least absolute shrinkage selection operator and Local learning for removing irrelevant and redundant features. We also proposed novel fast conditional mutual information feature selection algorithm to solve feature selection problem. The features selection algorithms are used for features selection to increase the classification accuracy and reduce the execution time of classification system. Furthermore, the leave one subject out cross-validation method has been used for learning the best practices of model assessment and for hyperparameter tuning. The performance measuring metrics are used for assessment of the performances of the classifiers. The performances of the classifiers have been checked on the selected features as selected by features selection algorithms. The experimental results show that the proposed feature selection algorithm (FCMIM) is feasible with classifier support vector machine for designing a high-level intelligent system to identify heart disease. The suggested diagnosis system (FCMIM-SVM) achieved good accuracy as compared to previously proposed methods. Additionally, the proposed system can easily be implemented in healthcare for the identification of heart disease.

3. PROPOSED SYSTEM



3.1 Extracting the training dataset

First, we have to collect recorded audio files of Heartbeat Sound of peoples who are suffering from cardiovascular or heart diseases and audio files of heartbeat sound of Normal People also . We had use ECG (Electrocardiogram) machine to record heartbeat sound. We had also taken audio files available on internet also. After that we stored all files in our system folders.

Approximately, We have taken more than 3000 audio files of normal or abnormal peoples for analysing data.We had set '-1' for normal and '1' for abnormal. For reading/accessing the files present in different folders we use Pandas module of machine learning. We had converted audio files into the numeric form by using FFT (Fast Fourier Transform) function of scipy module of machine learning.

3.2 Feature Extraction and Labelling

As soon as the audio files is converted into the numerical, we used list() to store all files which was concatenated .All files in different folders were stored in list named as Features. After that we did labelling to all audio files by giving label as 'HeartStatus'. We used list names as label to store label values. At the end, we also find out the length of Label and Feature list.

3.3 Train Test Split

From sklearn.model_selection we import train_test_split in order to split the data inside the dataset. We trained our 75%-80% of the data and remaining 25%-20% of data is used to test the model.

3.4 Train model using Random Forest Classifier

After splitting the data inside the dataset ,75%-80% of the split data is used to train the Random Forests classifier, Random Forest Classifier algorithm will get processed on the data and will generate some trained data and remaining 20%-25% data will be used for testing the data at later stage which will be used to generate the accuracy, f1-score, confusion matrix and other performance metrics.

Result of training code of finding accuracy, score, confusion matrix:

	precision	recall	f1-score	support
-1	1.00	1.00	1.00	645
1	1.00	0.99	1.00	165
accuracy			1.00	810
macro avg	1.00	1.00	1.00	810
weighted avg	1.00	1.00	1.00	810

```
Confusion Matrix:

True Negative: 645
False Positive: 0
False Negative: 1
True Positive: 164
[[645   0]
 [  1 164]]
```

3.5 Test on holdout set

After the model is trained and validated with our algorithm, we get the final estimate of the machine learning model performance. After this we test our unseen data using test set to check the performance of our model.

3.6 Plotting Wave Of Audio File

We also plotted the wave by using Matplotlib module of machine learning. In that, we used Matplotlib module to plot wave of heartbeat audio files by using or with respect to factors Time and Amplitude. Because of this waves ,Easy to understand the speed/condition of heartbeats.

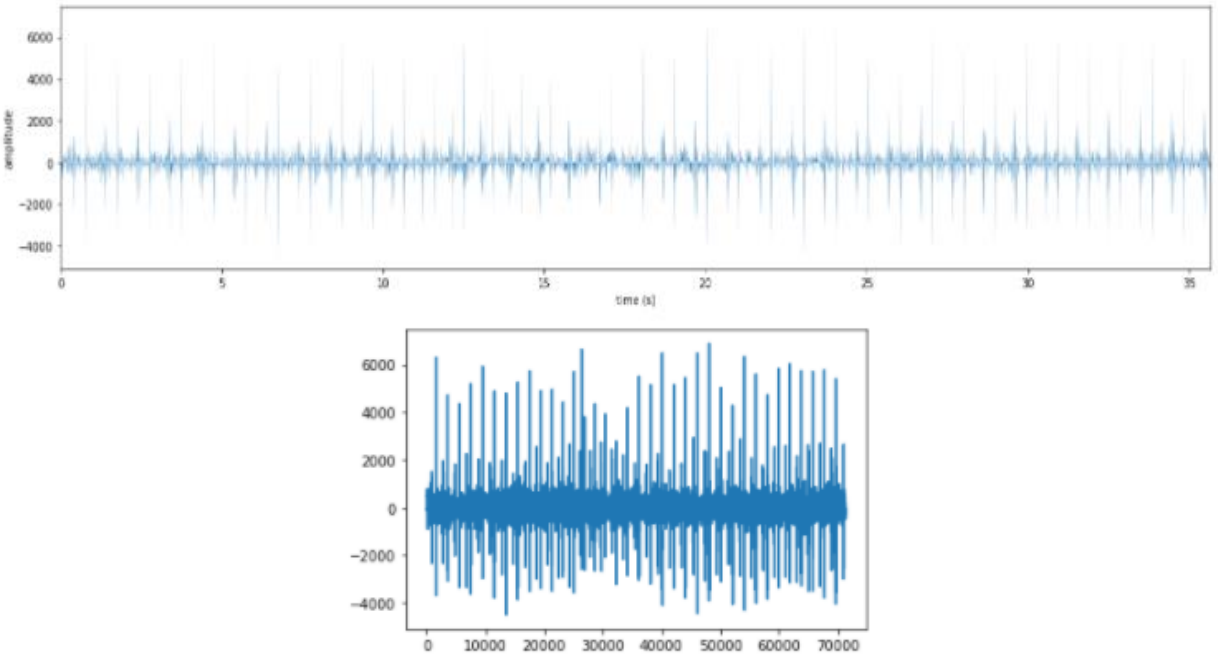


Figure: Sample Plotted Wave for Audio Files

3.7 Flask connection

Flask is a web framework for Python, meaning that it provides functionality for building web applications, including managing HTTP requests and rendering templates. In this section, we will create a basic Flask application. We used Flask Server to deployment of API. Because of Flask Server we are able to fetching training and testing data .

3.8 Result

After selecting the audio file from the system, we got output as Normal Condition or Abnormal Condition. Graph of wave to understand more efficiently the condition of heartbeats.

4. ALGORITHM AND TERMS USED

4.1 Machine Learning

Machine learning having ability to learn without explicitly program and perform operation in efficient manner, three types of machine learning algorithm perform on various algorithm. Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. In this Project we have used Random Forest Classification Algorithm of Supervised Learning .

4.2 Supervised Learning

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers afunction from labeled training data consisting of a set of training examples. Supervised learning is that the type of machine learning during which machines are trained using well "labeled" training data, and on basis of that data, machines predict the output. The labeled data means some input file is already tagged with the right output. A supervised learning algorithm aims to seek out a mapping function to map the input variable(x) with the output variable(y).

4.3 Random Forest Classification:

Random Forest Classification is a supervised Machine Learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. As we know that a forest is made up of trees and more trees means more robust forest. Similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.

4.4 How does the algorithm work?

It works in four steps:

1. Select random samples from a given dataset.
2. Construct a decision tree for each sample and get a prediction result from each decision tree.
3. Perform a vote for each predicted result.
4. Select the prediction result with the most votes as the final prediction.

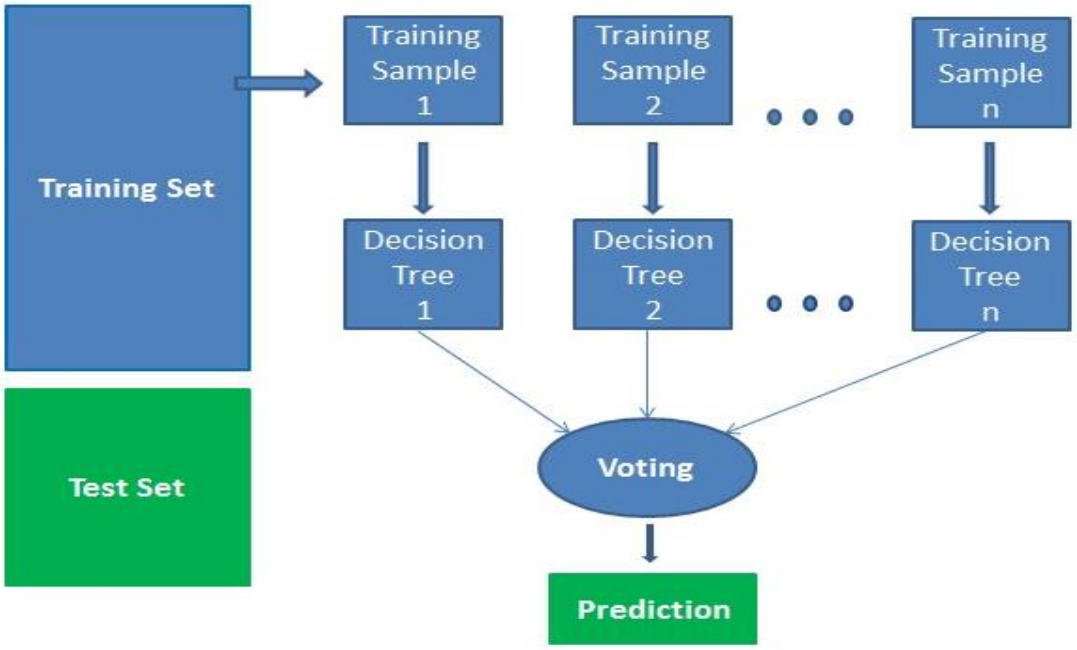


Figure: Working Of Random Forest Classification Algorithm

4.5 Fast Fourier Transform/Fourier Analysis

Fourier analysis is a method for expressing a function as a sum of periodic components, and for recovering the signal from those components. When both the function and its Fourier transform are replaced with discretized counterparts, it is called the discrete Fourier transform (DFT). The DFT has become a mainstay of numerical computing in part because of a very fast algorithm for computing it, called the Fast Fourier Transform (FFT).

4.6 Classification Report

A Classification report is employed to live the standard of predictions from a classification algorithm.

4.6.1 Terms Of Classification Report

1. Accuracy:

It defines how often the model predicts the right output.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision:

It is often defined because the number of correct outputs provided by the model or out of all positive classes that have predicted correctly by the model, what percentage of them were true. algorithm.

$$\text{Precision} = \frac{TP}{TP + FP}$$

3. Recall:

It is defined as the out of total positive classes, how our model predicted correctly. The recall must be as high as possible.

$$Recall = \frac{TP}{TP + FN}$$

5. RESULT

This is the User Interface (UI) of our page where user will be prompted to check whether the heart condition is Normal or Abnormal.

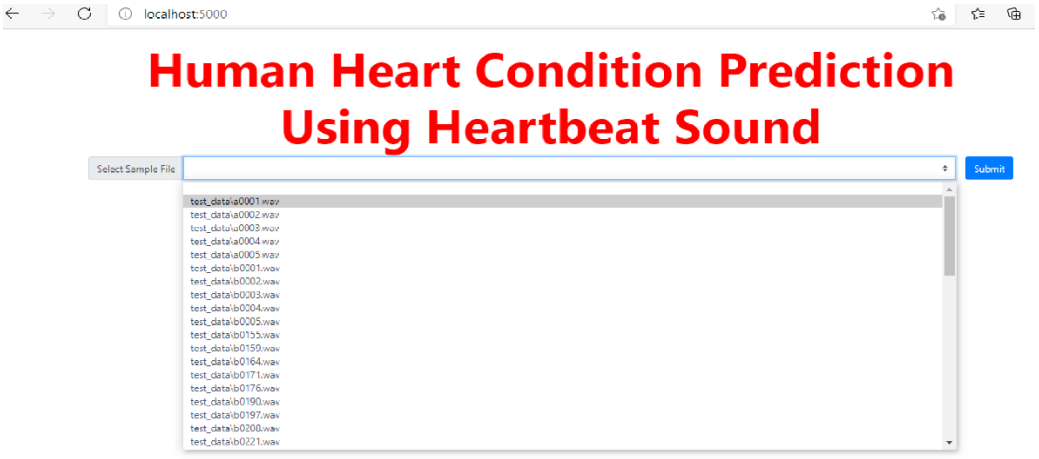


Figure: UI to select audio file from system

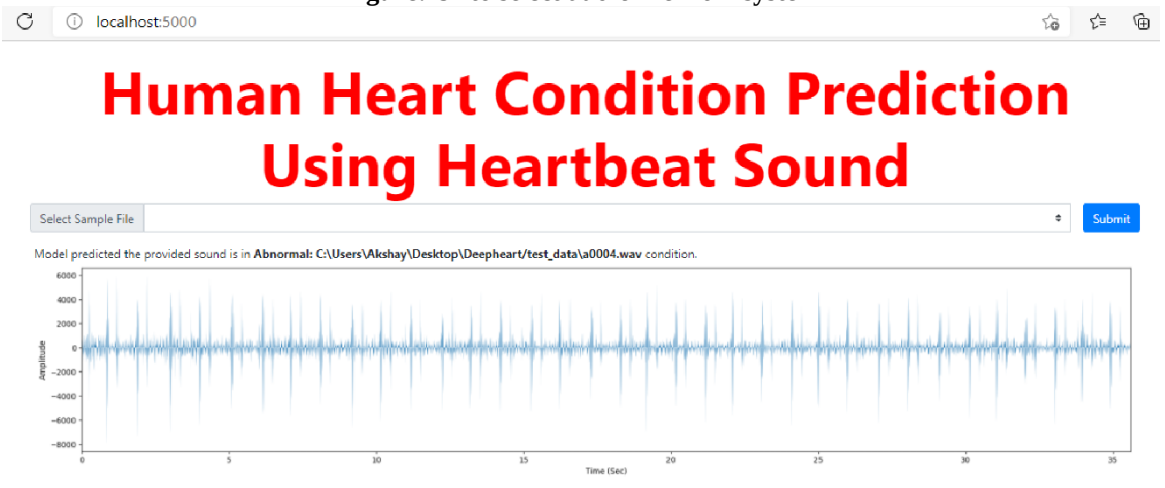


Figure: Output after Selecting File From System

6. FUTURE SCOPE

- User Friendly System for routine check-up.
- Reduction in Response Time.
- Android Application.

7. CONCLUSION

We concluded that, It is a very efficient and real time project. It is very useful for people who are facing problems with the existing system. Project shows that automated heart sound segmentation and classification techniques have the potential to screen for pathologies in a variety of clinical applications at a relatively low cost. This Project presents a machine learning approach for classification of heart sound recordings. We have extracted several features in both the time- and frequency domains. Future research and development could concentrate on the creation of an algorithm that is able to distinguish between the different types of diseases. Noise immunity of the algorithm and its tolerance towards dissimilarities in recording circumstances should also be improved in the future.

8. ACKNOWLEDGEMENT

We are pleased to present "Human Heart Condition Prediction Using Machine Learning" project Implementation Paper. We would like to express special thanks of gratitude towards the our project coordinator and guide prof. Santosh Waghmode sir for guiding us in every phase of project. Thank you, sir, for support, patience and faith in our capabilities and for giving us flexibility in terms of working and reporting schedules. Also thanks to authors, researchers of various books and articles I used to make this project eminent.

9. REFERENCES

[1] Liu C, Springer D, et al. An open access database for the evaluation of heart sound algorithms. Physiological Measurement 2016;37(9).

[2] Springer DB, Tarassenko L, Clifford GD. Logistic regression-hsmm-based heart sound segmentation. IEEE Transactions on Biomedical Engineering April 2016; 63(4):822–832. ISSN 0018-9294.

[3] Schmidt S E; Holst-Hansen CGCTESJJ. Segmentation of heart sound recordings by a duration-dependent hidden markov model. Physiological Measurement 04 2010;31.

[4] Maglogiannis I, Loukis E, Zafiropoulos E, Stasis A. Support vectors machine-based identification of heart valve diseases using heart sounds. Computer Methods and Programs in Biomedicine 2009;95.

[5] Uguz H. Adaptive neuro-fuzzy inference system for diag- ˇ nosis of the heart valve diseases using wavelet transform with entropy. Neural Computing and Applications 2012; 21(7):1617–1628. ISSN 1433-3058.

BIOGRAPHIES

Akshay Baban Pathare
Author (Team Leader)
B.E.(Computer Engineering)
Imperial College of Engineering and Research Pune, Maharashtra

Abhijit Yuvraj Sonawane
Co-Author (Team Member)
B.E.(Computer Engineering)
Imperial College of Engineering and Research Pune, Maharashtra

Santosh Machhindra Punde
Co-Author (Team Member)
B.E.(Computer Engineering)
Imperial College of Engineering and Research Pune, Maharashtra

Akash Rajendra Shete
Co-Author (Team Member)
B.E.(Computer Engineering)
Imperial College of Engineering and Research Pune, Maharashtra

