International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

Human Heart Condition Prediction using Machine Learning

Akshay Pathare[1], Abhijit Sonawane[2], Akash Shete[3], Santosh Punde[4], Santosh Waghmode[5]

[1,2,3,4]*Student, Dept. of Computer Engineering, Imperial College of Engineering and Research, Pune, Maharashtra, India*
[5]*Professor, Dept. of Computer Engineering, Imperial College of Engineering and Research, Pune, Maharashtra, India*

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 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.*

Key Words: **Heart Classification, Support Vector Machine, Random Forest Classifier, Machine Learning, etc**

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 diagonistics)and unsure(too noisy to make decision ;retake the recordings).

2. MACHINE LEARNING

Machine learning is the science of getting computers to act without being explicitly programmed. In the past decade, machine learning has given us self-driving cars, practical speech recognition, effective web search, and a vastly improved understanding of the human genome. Machine learning is so pervasive today that you probably use it dozens of times a day without knowing it. Many researchers also think it is the best way to make progress towards human-level AI. 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. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly. But, using the classic algorithms of machine learning, text is considered as a sequence of keywords; instead, **an** approach based on semantic analysis mimics the human ability to understand the meaning of a text.

2.1 METHODS OF MACHINE LEARNING

Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can

also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

In contrast, **unsupervised machine learning algorithms** are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

Semi-supervised machine learning algorithms fall somewhere in between supervised and unsupervised learning, since they use both labeled and unlabeled data for training – typically a small amount of labeled data and a large amount of unlabeled data. The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labeled data requires skilled and relevant resources in order to train it / learn from it. Otherwise, acquiring unlabeled data generally doesn't require additional resources.

Reinforcement machine learning algorithms is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behaviour within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

3. CLASSIFICATION OF ALGORITHM

3.1. LOGISTIC REGRESSION CLASSIFIER

Logistic regression is a classification algorithm, used when the value of the target variable is categorical in nature. Logistic regression is most commonly used when the data in question has binary output, so when it belongs to one class or another, or is either a 0 or 1. In Logistic Regression, Logistic Function is used. It also called Sigmoid Function. The sigmoid function/logistic function is a function that resembles an "S" shaped curve when plotted on a graph. It takes values between 0 and 1 and "squishes" them towards the margins at the top and bottom, labeling them as 0 or 1. The equation for the Sigmoid function is this:

$$y=1/(1+e^x)$$

3.2.K NEAREST NEIGHBOURS(KNN)

KNN is a non-parametric and lazy learning algorithm. Non-parametric means there is no assumption for underlying data distribution. In other words, the model structure determined from the dataset. This will be very helpful in practice where most of the real world datasets do not follow mathematical theoretical assumptions. Lazy algorithm means it does not need any training data points for model generation. All training data used in the testing phase. This makes training faster and testing phase slower and costlier. Costly testing phase means time and memory. In the worst case, KNN needs more time to scan all data points and scanning all data points will require more memory for storing training data.

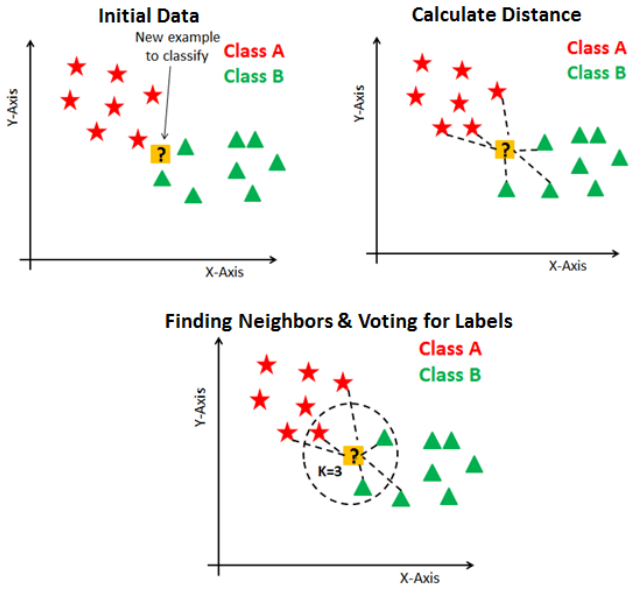


Fig. K Nearest Neighbour Algorithm

3.3. DECISION TREE CLASSIFIER

Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, the decision tree algorithm can be used for solving regression and classification problems too. The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data (training data). In Decision Trees, for predicting a class label for a record we start from the root of the tree. We compare the values of the root attribute with the record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.

Important Terminology related to Decision Trees:
Root Node: It represents the entire population or sample and this further gets divided into two or more homogeneous sets.
Splitting: It is a process of dividing a node into two or more sub-nodes.
Decision Node: When a sub-node splits into further sub-nodes, then it is called the decision node.
Leaf / Terminal Node: Nodes do not split is called Leaf or Terminal node.

Pruning: When we remove sub-nodes of a decision node, this process is called pruning. You can say the opposite process of splitting.
Branch / Sub-Tree: A subsection of the entire tree is called branch or sub-tree.
Parent and Child Node: A node, which is divided into sub-nodes is called a parent node of sub-nodes whereas sub-nodes are the child of a parent node.

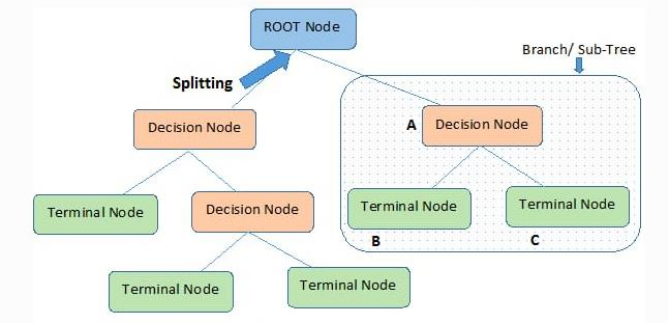


Fig. Decision Tree Classifier

3.4. RANDOM FOREST CLASSIFIER

Random Forest is a learning method that operates by constructing multiple decision trees. The final decision is made based on the majority of the trees and is chosen by the random forest. Random forest is a supervised 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.

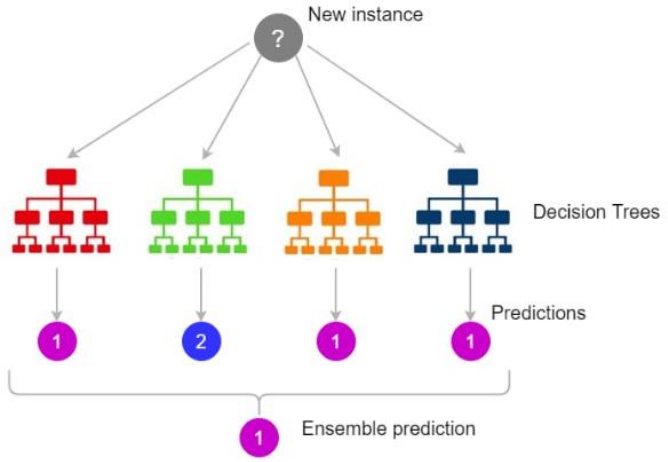


Fig. Random Forest Classifier

Working of Random Forest Algorithm

We can understand the working of Random Forest algorithm with the help of following steps –

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.

3.5. SUPPORT VECTOR MACHINE

Support Vector Machine, abbreviated as SVM can be used for both regression and classification tasks. But, it is widely used in classification objectives. To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, that is the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.

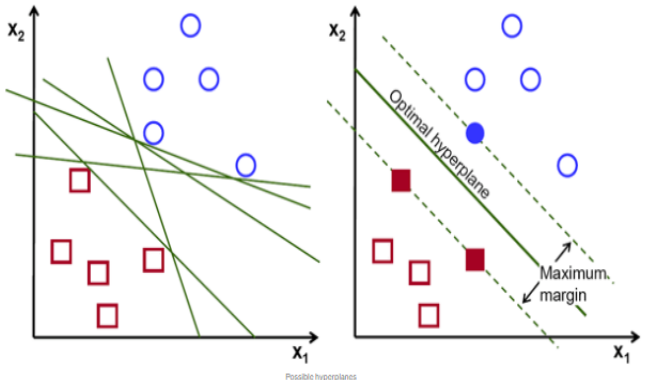


Fig. SVM in Machine Learning

Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features. **Support vectors** are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane. These are the points that help us build our SVM.

4. HOW ALGORITHM IS USED IN ML?

The use of Machine Learning and its prowess had grown exponentially over the last few years. Machine mastering elements is a stepwise system to construct an environment friendly computer learning project. The major purpose of the ml is to discover a answer to the hassle or project. Machine

mastering lifestyles involves seven foremost steps, which are given below:

1. Gathering the required data to perform operations.
2. Data Analyzing
3. Data guidance
4. Data Wrangling
5. Train Data
6. Test the model
7. Deployment
8. Review or Feedback

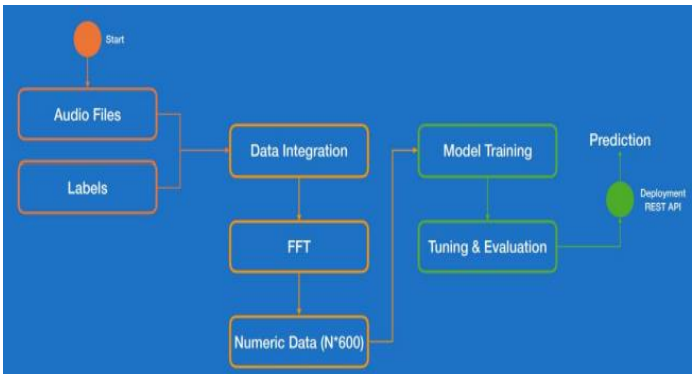
5. EXISTING SYSTEM

Various human's and healthcare companies developed different heart condition classification systems but they used deep learning or other methods which gives less accuracy results. Right now, We are working on that to give more accuracy results. False Positive rate is high in the existing systems. In the present system required computational time is massive.

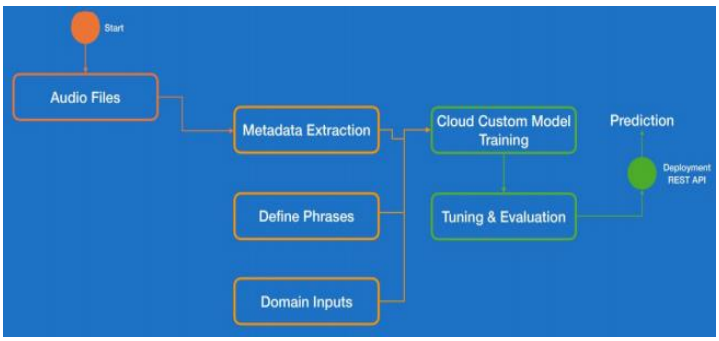
6. PROPOSED SYSTEM

Following Steps we are performed in proposed system.

1. Gathering Audio Files recorded from ECG machine.
2. Labeling the Audio Files.
3. Integration of audio files.
4. Converting audio files into numeric data by using Fast Fourier Transform(FFT).
5. Creating Model and Perform train and test on Model.
6. Model Tunning & Evolution.
7. Predicting output by using prediction code.



System Architecture



FUTURE SCOPE

It will give high accuracy results of human heart codition in less time. It will help to patients have cardiovascular diseases to do routine check ups at low cost.

CONCLUSIONS

Research 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.

ACKNOWLEDGEMENT

We are pleased to present "**Human Heart Condition Prediction Using Machine Learning**" project survey 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.

REFERENCES

[1] Z. Jiang, and S. Choi. "A cardiac sound characteristic waveform method for in-home heart disorder monitoring with electric stethoscope." Expert Systems with Applications 31.2 (2006): 286-298

[2] Z. Sharif, M. S. Zainal, A. Z. Sha'ameri, S. H. S. Salleh. "Analysis and classification of heart sounds and murmurs based on the instantaneous energy and

frequency estimations." TENCON 2000. Proceedings. Vol. 2. IEEE, 2000.

[3] E. Çomak, A. Arslan, and I. Türkoğlu. "A decision support system based on support vector machines for diagnosis of the heart valve diseases." Computers in Biology and Medicine 37.1 (2007): 21-27.

[4] G. E. Guraksin, U. Ergun, and O. Deperlioglu. "Classification of the heart sounds via artificial neural network." International Journal of Reasoning-based Intelligent Systems 2.3-4 (2010): 272-278.

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