

Detection of Heart Related Disease using Machine Learning

Bindu ¹, Manjunath Swamy C²

¹² Mtech Computer Science and engineering,

¹²New Horizon college of engineering.

Abstract: Every human is much worried about his health. Health first all other are rest. The Heart is a muscular organ which pumps blood to entire body. Heart pumps the oxygenated blood to all other parts of the body using arteries and collects deoxygenated blood from all the parts of the body using veins. Heart Disease is one of the disorder that is related to arteries and blood veins of the heart. It is not that easy to know the presence of the disease in a human if known also its difficult to say which type of category it belongs to and which type of treatment must be given to that patient. Data science is one of the more important things in today's current world for early prediction of disease and solves large data problems. Large amount of data is produced across a variety of fields. The Big Data is used to collect and organize data and valuable data need to be extracted. We need a large amount of data to analyze the useful information. In order to get that we use Data Mining and Machine Learning Techniques to build a model which can analyze the patterns and give the required prediction or the results. In this we are going to deal with the prediction of heart related disease using Support Vector Machine technique.

Keyword: Machine Learning, Data Mining, Health disease, Classification, Prediction, SVM.

I. Introduction

Every human is much worried about his health. Health first all other are rest. The Heart is a muscular organ which pumps blood to entire body. Heart pumps the oxygenated blood to all other parts of the body using arteries and collects deoxygenated blood from all the parts of the body using veins. Heart Disease is one of the disorder that is related to arteries and blood veins of the heart. It is not that easy to know the presence of the disease in a human if known also its difficult to say which type of category it belongs to and which type of treatment must be given to that patient. Data science is one of the more important things in today's current world for early prediction of disease and solves large data problems. Large amount of data is produced across a variety of fields. The Big Data is used to collect and organize data and valuable data need to be extracted. We need a large amount of data to analyse the useful information. In order to get that we use Data Mining and Machine Learning Techniques to build a model which can analyse the patterns and give the required prediction or the results.

Analysis and detection of the disease related to heart can be done by using Machine Learning and Data Mining techniques, we can develop a model which can help them to improve the prediction of presence of disease with in a person and when huge set of data is taken then including large number of population. The prediction models will help in detecting whether the patient is suffering from heart related disease or not. It also determines the health of the patient. The key difference between human and computer is the ability to think and learn from experience. Humans learn from the experiences at other hand Computers execute human-made algorithms to get the knowledge, get trained and make the predictions. Machine learning algorithms are found to be very effective to particular type of learning.

II. PROBLEM DEFINITION

Heart Disease is a disorder related to arteries and blood veins. It been difficult for doctors to find the type of disease accurately and exactly about heart disease diagnosis. Machine Learning is branch of data science which helps in early prediction of things and even solve the problems when huge dataset is given. World health organization[WHO] says that by using the data mining techniques we can predict the early stage of the disease related to heart and also give concerned medication for it.

Data mining is having two models which areas predictive model and the descriptive model. Predictive model is the model which is been brought to predict the output of the problem by using modeling techniques for predictive analysis. But the descriptive model is the model which is been brought up to clear understanding of the data which out pointing out at any variable by using descriptive model analysis methods. Our project uses various different machine learning techniques and brings with the accurate predictive and descriptive models

III. PROJECT PURPOSE

In order to predict whether the given set of population has heart related disease. To get the prior knowledge about the heart related disease we are using machine learning models to predict if the patient is having disease or not. The data set has been collected first and then, later the analysis is made. Here the data is images of the patient's heart taken in different angles. The dataset is collected in the form of images and fed to the algorithm, finally get the output data which helps in determining if that person is suffering from heart related disease. This helps in identification of the problem of a person by just running one algorithm.

IV. PROJECT FEATURES

Our scheme was developed to reduce the time that is taken for determining the disease and the part infected. The accuracy of the system is very high which is calculated based on training accuracy and testing accuracy. Training accuracy is calculated by considering all the input samples that is every sample is trained and the correctness of the system is calculated. Testing accuracy is calculated by verifying input data to undergo the process of classifying the images to be healthy or infected. This is being determined by using bits. In which '0' indicates healthy heart of a person and '1' indicates that the person has an infected heart. Totally we are considering forty-eight images as our input data and performing data pre-process as the initial step to clean up the data.

IV. MACHINE LEARNING

In software development process literature survey plays a very important role. Before developing any model it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then next steps is to determine which operating system and language can be used for developing the model. Programmer need lot of external support for developing the models. External supports includes guidelines from senior developers, from the internet sources or from the books. Before starting the project all the external factors are taken into consideration. We have to analyze the Machine Learning Outline Survey.

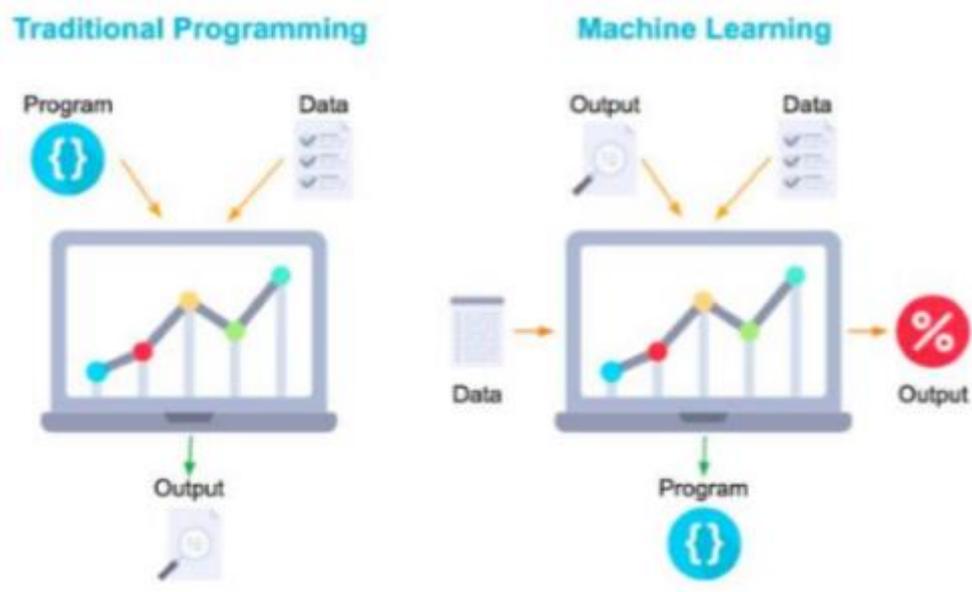


Figure : Machine Learning

Machine learning algorithms have the ability to learn from data and make predictions based on that data. It is an application of AI that provides ability to the system to learn automatically and improve from the experiences without being explicitly programmed. The primary aim is to allow the computers to learn automatically without human intervention or assistance and adjust accordingly. In Machine Learning, computers get programmed by themselves by learning from the given input data-set. Machine Learning helps to automate the machine to process automatically. Machine learning is the part of Artificial Intelligence, it makes easy to predict the futures.

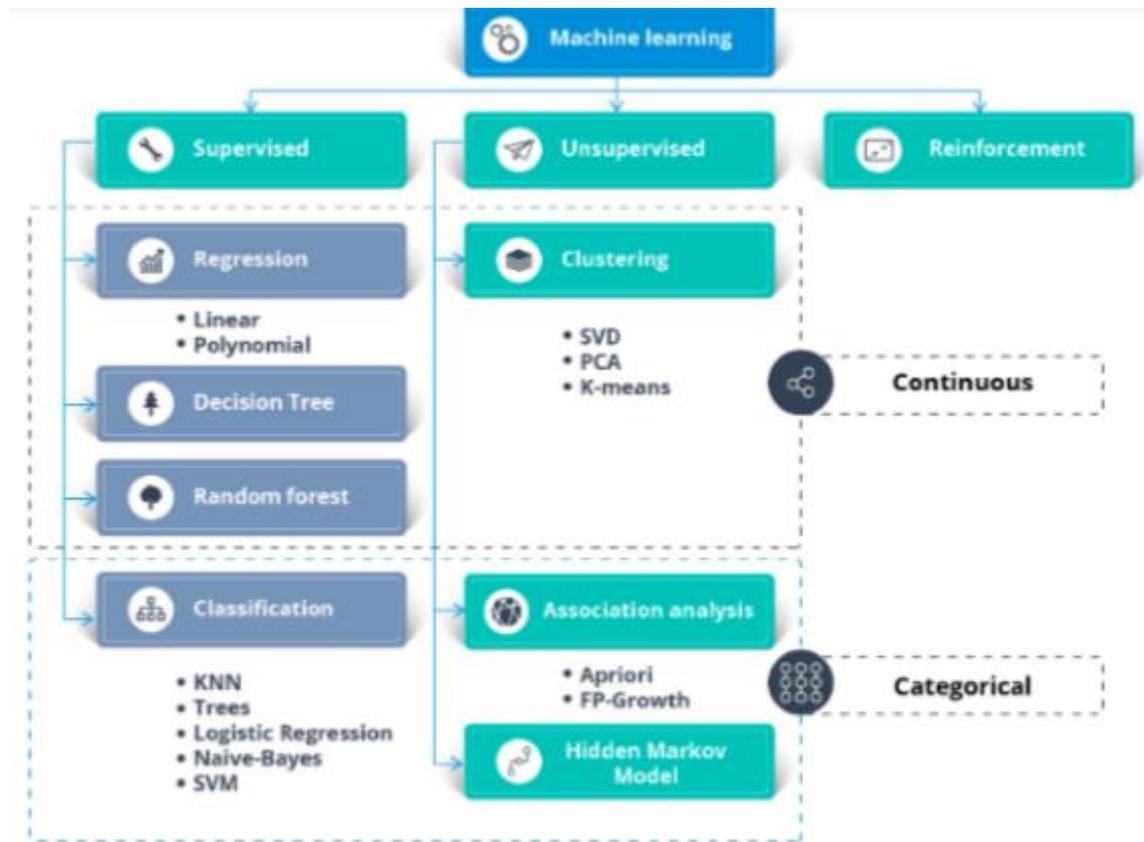


Figure : Machine Learning Algorithm

Machine Learning Methods are Supervised Learning, Unsupervised Learning, Semi supervised Learning, Reinforcement Learning. Machine learning easily identifies the patterns and predicting trends. No human intervention is required. It can handle multi-dimensional and multi-variety data at the same time. It can train the data and provide accuracy. It is used over wide range of applications. It reduces cost and time. In Machine Learning computers gets programed by itself by learning from the given input data-set. Machine Learning helps to automate the machine to process automatically. Machine learning is the part of Artificial Intelligences, it makes easy to predict the futures

Supervised Learning:

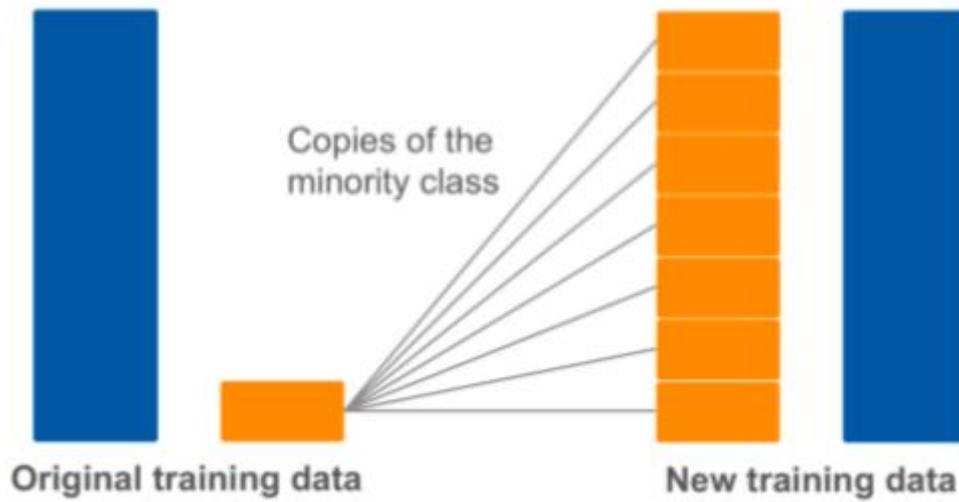
Supervised machine learning algorithms can be applied to the labeled data for an accurate prediction from the dataset. The analysis starts with training dataset, the learning algorithm. Produces a inferred function to make prediction about the output values. The system is about to provide targets for any new input after sufficient training. The algorithms compare its output with the correct, intend output and find errors in order to modify model accordingly.

Classification:

In Classification incoming data is labeled based on former data sets and algorithm is manually trained to recognize type of object and based on the object data is categorized. Machine should learn how to distinguish data, binary recognition or image.

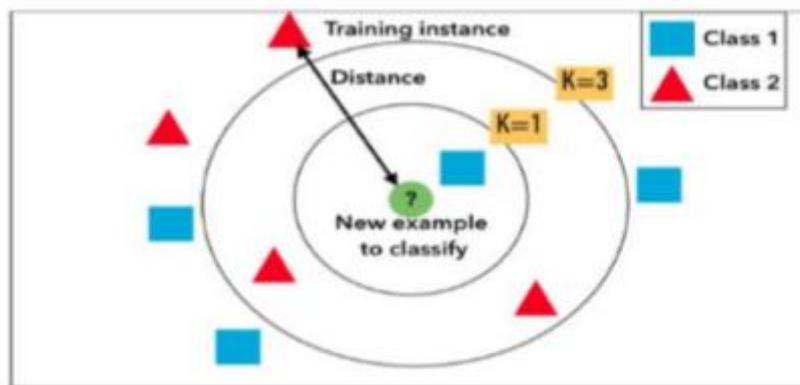
Naïve Bayes:

Naïve Bayes comes under the classification method. It is based on the Bayesian theorem. It is a simplest and most efficient algorithm. The efficiency depends on the following -a. Independent values :We assume that the attribute values is independent of values of other attribute b. Hidden values Here we assume that prediction process will not affect by hidden attributes



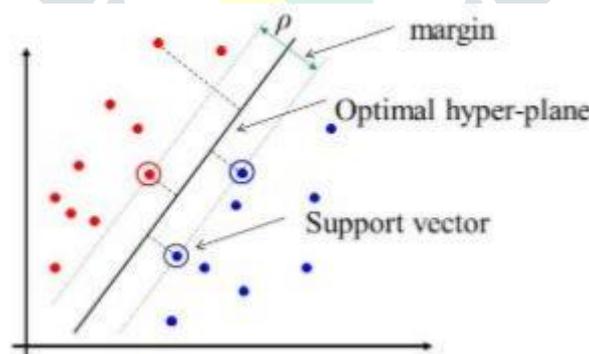
Nearest Neighbor:

Nearest Neighbor comes under the classification method. It takes a group of labelled points and with the labelled points learns to label other points. To label a new point, it looks at the labelled points closest to that new point and has those neighbors point are nearby it is consider as next point.



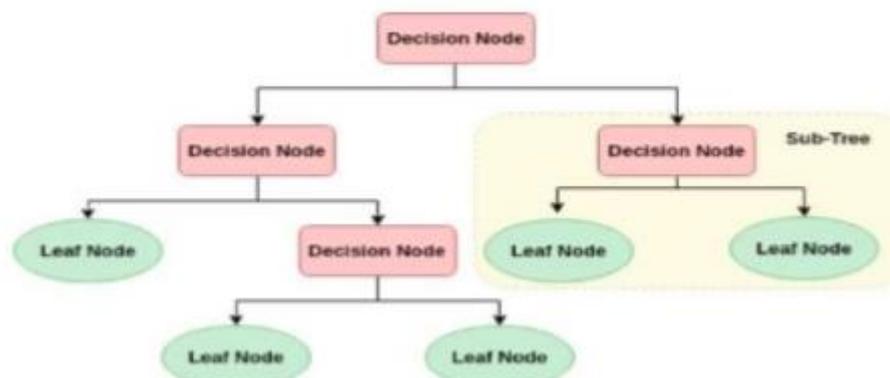
Support Vector Machines:

Support Vector Machine (SVM), comes under the classification method or regression method to achieve the prediction decision it derive features from the variables, and then manage them in linear combination.



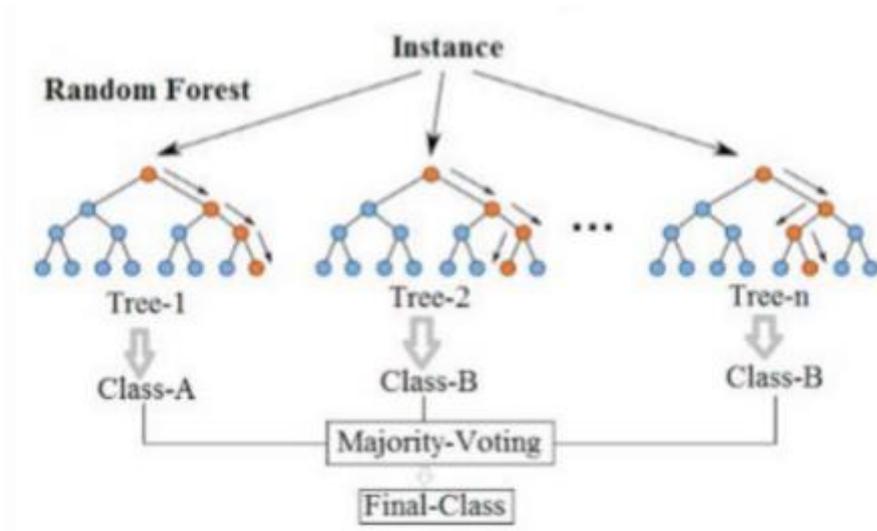
Decision Trees:

It builds the model in the form of tree structure by breaking down the data-set into subsets. Finally tree has decision nodes and leaf nodes. Top most node is called as root node and bottom nodes are called as leaf nodes which represents decision or classification.



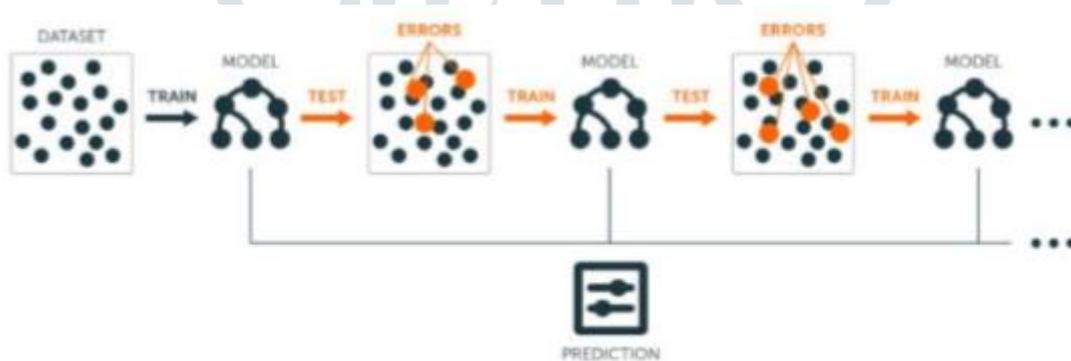
Random forest:

Random forest which is also known as random decision forests method are ensemble for classification, regression and also for other tasks. It works by creating multitude decision trees, trees are created during training period and outputs are classes or mean prediction.



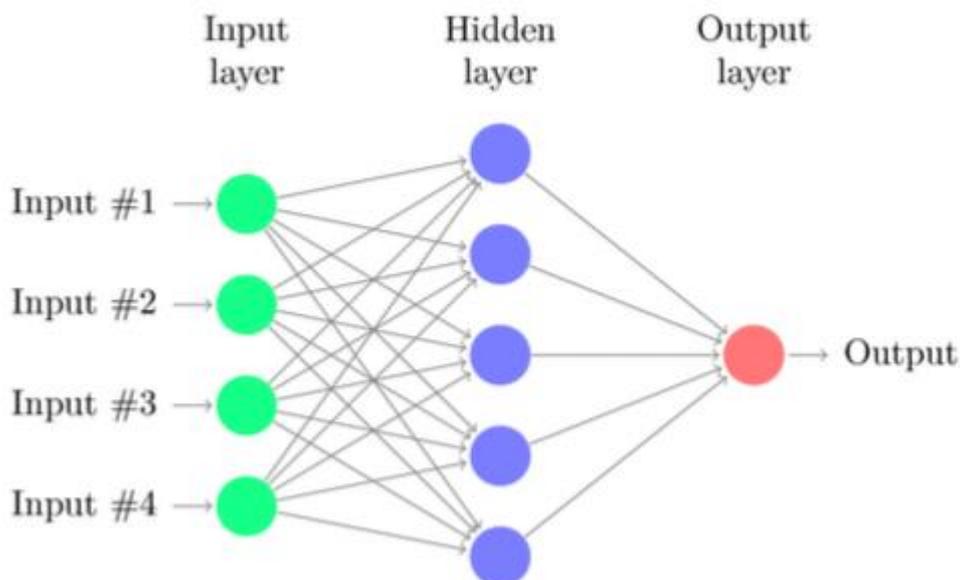
Boosted Trees:

Boosted trees technique can be used for classification or regression model. In the form of decision tree prediction model is developed.



Neural Networks:

They are the set of algorithms modeled similar to human brain, which designed to pattern recognition. Through machine perception it interpret data. It can process all real world data such as text, voice, image, videos, and time series. They group the unlabeled data based on the similarities among the input data and classifies data into labeled data.



Neural network mainly contain 3 layers: Input, Hidden, Output

a. Input: In the input layer raw information is feed into the network

b. Hidden: Based on the input provided it checks the weight and relation between raw data and hidden values. There may be more than one hidden layer

c. Output: Detection of Heart Related Disease using Machine Learning

Output depends on the activities carried out by the hidden layers

V. PYTHON

Python is a high level general purpose, dynamic, multi-platform, object oriented, interpreted language. Python was developed by Guido Rossum in 1991. It is open source software. There are two main branches of Python those are Python 2.x and Python 3.x Python 3.x new improvement is that it removes backward compatibility with its past releases of Python. It supports both procedure oriented and object oriented programming. This language has wide range of applications. Its mainly used for coding, scripting, used even with selenium. It's a developer friendly language and use for developing GUI applications. Its famous for its code readability and simple syntax. It has no concept of using semicolon. It strictly follows indentation.

Python Converts python code to byte code with the help of Syntax checker and translator. Python virtual machine (PVM) is an interpreter which converts byte code to readable output. The python Framework is a language-neutral platform for writing programs that can easily and securely interoperate. The Python framework provides the foundation for components to interact seamlessly, whether locally or remotely on different platforms. It standardizes common data types and communications protocols so that components created in different languages can easily interoperate.

Step 1: The interpreter reads a python code or instruction. Then it verifies that the instruction is well formatted, i.e. it checks the syntax of each line. If it encounters any error, it immediately halts the translation and shows an error message.

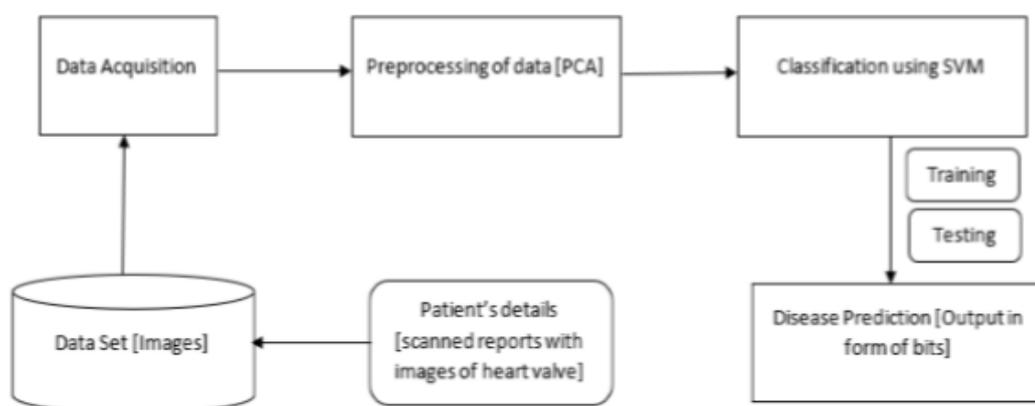
Step 2 : If there is no error, i.e. if the python instruction or code is well formatted then the interpreter translates it into its equivalent form in intermediate language called "Byte code".Thus, after successful execution of Python script or code, it is completely translated into Byte code.

Step 3: Byte code is sent to the Python Virtual Machine(PVM).Here again the byte code is executed on PVM.If an error occurs during this execution then the execution is halted with an error message.

IV. SYSTEM ARCHITECTURE

When heart patient visits the doctor for the treatment, and doctor suggests him to take up scanning. During scanning the images of that patient are collected. Data set is collected from the hospital and it's the images of the patient's heart. This data is collected and preprocessed by some of the algorithms like PCA, the data is trained and tested by using Classification algorithm and finally gives out the result saying the patient is suffering from heart related disease or not. This helps the doctors judge easily if the patient has the disease or not by making use of this machine learning approach.

In this approach the data is been feed into the system and the system is trained based on some inputs given to it. Later by analyzing those images the system analysis the condition of the patient and results it as output by informing the doctor about the patients condition.



Data Processing:

Before applying the algorithm on the dataset we need to preprocess the data. Once we collected the raw data we need to process it, before we do any analysis on that data. Usually data from such sources will be very huge, messy, duplicate data can be present, there can be some missing values. In order to get the accurate analysis we need to cross

check the data. The common errors which we came across: 1. Missing values 2. Unrevent data for our analysis such as address and the names of the students and staff etc.

Data Cleaning:

Data cleaning includes mainly two methods where either of them can be followed in

order the clean the present data those are 1 Remove Missing Values 2 Remove Noisy Data 3 Removing Missing values can be done by ignoring the tuples or filling the values manually but this is time consuming or by filling missed values with the global constant. Or by using central tendency for each attribute. Or by calculating mean value of the class. Even by the use of Most probable value.

Data Transformation:

Data transformation is also termed as data distrezation. The following methods can be

done to perform data transformation. 1 Smoothing – this is same as cleaning 2 Attribute Construction – Construction of new attributes for easy identification 3 Normalization – this includes various techniques like Min Max Normalization, Score normalization and Decimal Scaling. 4 Concept Hierarchy Generation – this include Specification of partial ordering of attributes at schema level. Specification of portion of hierarchy by explicit data grouping. Specification of set of attributes but not their Partial Ordering.

Data Reduction:

Data reduction is also termed as data selection. The following methods are been used to perform the operation of reducing the data. Data can be reduced in three ways 1 Dimensionality Reduction 2 Numerosity Reduction 3 Dimensionality Reduction – Huge data is been converted into small data by

applying techniques like Lossy Reduction and Lossless Reduction. 4 Numerosity Reduction – includes sampling, reduction based on regression, histogram reduction and others

Binning:

This is one of the technique that is been followed to remove the noisy data in the given data set. There are three types of binning those are 1 Bin Mean 2 Bin Boundaries 3 Bin Median. Bin Mean – Splits the data into different parts by grouping three values in each calculating the mean of every part splited. Bin Boundaries - Splits the data into different parts by grouping three values in each finding out the nearest element off and replacing it with the same Bin Median - Splits the data into different parts by grouping three values in each finding the middle value among the present values and replacing the middle value.

Selection of best attribute and feature Extraction:

Feature extraction is the process by which we reduce the redundant data when set of data is given and analysis is to be made on it. By this we can also reduce the number of resources which needs to be processed. Selection of the best attribute helps the ML algorithms to fasten and speed up the training of the dataset by reducing overfitting

Classifier Modelling:

Different classifiers are been used which is been discussed above in the survey and the best among them is been predicted based on the given input dataset

Results:

Finally determining whether the in the given set of data if the patient is suffering from heart related disease and same in case of huge set of data.

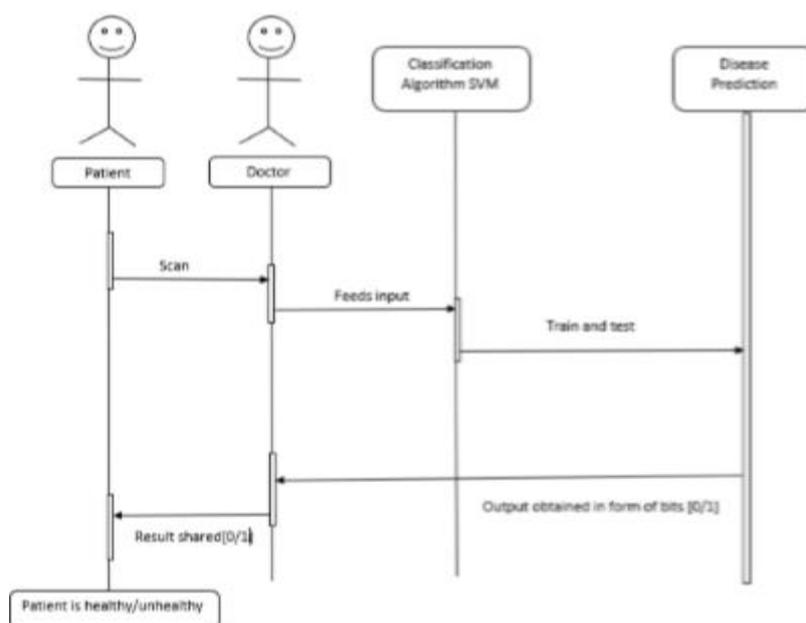


Figure : Sequence Diagram

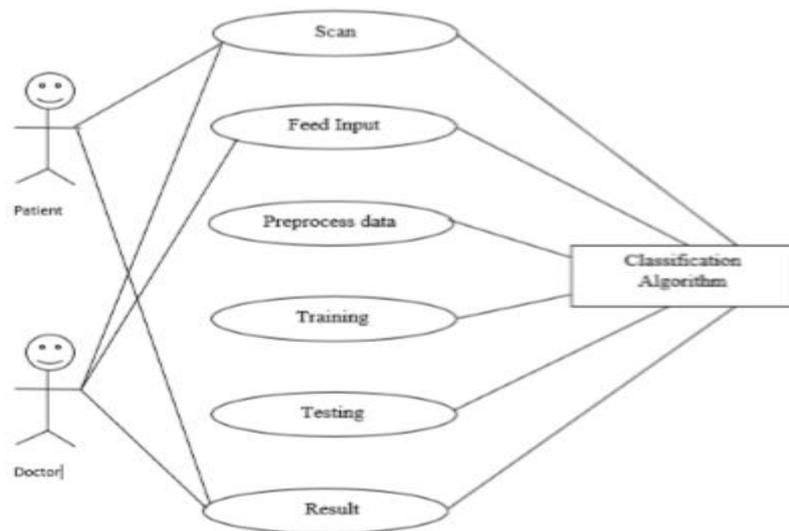


Figure : Use Case Diagram

IIV. IMPLEMENTATION

To develop the prediction model following steps need to be carried out

1. Problem Statement
2. Collection of data
3. Preprocessing
4. Data Mining
5. Machine Learning Algorithm
6. Result Evaluation
7. Knowledge flow representation

Problem Statement: Here we are looking forward to predict the Patient heart disease based on the scanning done to him. Which helps the medical field to determine the result accurately, faster and correctly.

Data Set: Once we get the problem statement, the next process is to preprocess the data. We have collected the data from the internet sources, where hospitals provided the information of the patient with the scanned images.

Data Pre-processing : Once we collected the raw data we need to process it, before we do any analysis on that data. Usually data from such sources will be very huge, messy, duplicate data can be present, there can be some missing values. In order to get the accurate analysis we had to cross check the data. The common errors which we came across: 1Missing values 2Unrevent data for our analysis 3 Invalid entries

Data Mining:

Once we get the data cleaned and transform it to required format. We had to closely examine the factors which will directly affect the result and which indirectly affect so keeping both in mind we begin the further process to analyze the patterns more deeply. We have divided the whole data into 60:40 where 60% of data is taken as training set and the remaining 40% is used for testing.

Machine Learning Algorithm:

numpy - This package is mainly used for systematic computations of data. This is an alternative to list. It also contains huge mathematical procedures to perform operations on the arrays. Hence they are termed as ndarrays. This is an alternative to list.

Scipy - Scipy is the extension of numpy. This is one of the opensource library of python. Scipy is built on numpy. It provides various special functions like elliptic, beta, gamma functions to coordinate with numpy as produce the result.

PCA - It is a statistical routine which helps in converting various observational variables into linear variables. Its mainly used to get the relation between two variables.

SKlearn is one of the most useful library in python. This is one of the open source machine learning library. It contains many efficient tools related to machine learning. This helps in resolving the data based on its functions like clustering, classification, regression etc. this mainly helps in prediction operations.

```

-----+-----1-----+-----2-----+-----3-----+-----4-----+-----5-----
1  import numpy as np
2  import scipy
3  import pickle
4  from sklearn.decomposition import PCA
5  from scipy import misc
6  import glob
7  import sys
8  import cv2
9  import os
10 from sklearn import svm
11 import warnings

```

Figure : Python Libraries

Preprocessing the data:

When we receive an image we need to extract the features from it. `imread` – is used to read the image. `CvtColor` – is used to display the color of the image. `Imshow` – using this function we do display them. We need to put the image in the right format we can achieve this by using the threshold value. Median blur – it gives the type of the image. Hanning window is used to calculate the width of the image. We can use different methods to extract the features of the image like Spectrogram, Histogram, Convolution 2D, convolution 3D. But in our project we are using the method of spectrogram FFT. Finally all the features are extracted from the image and the step of cleaning taken such that only the images which follow the measurements calculated finally are taken as input dataset.

```

12 warnings.simplefilter(action='ignore', category=FutureWarning)
13 os.chdir('HEART')
14 image = []
15 labs = []
16 def preprocess(image_path):
17     img2 = cv2.imread(image_path)
18     gray = cv2.cvtColor(img2, cv2.COLOR_BGR2GRAY)
19     final = cv2.resize(gray, (640, 480))
20     cv2.imshow("grayscale_img2", final)
21     ret, thresh1 = cv2.threshold(final, 30, 255, cv2.THRESH_BINARY)
22     ## cv2.imshow("binarized_img2", thresh1)
23     finall = cv2.medianBlur(thresh1, 3)
24     cv2.imshow("filtered_img2", finall)
25     width, height = len(finall), len(finall[0])
26     N = 640
27     w = np.hanning(N)
28     np.shape(finall)
29     np.shape(w)
30     scale = 1/(width*height)
31     finall = np.fft.fft2(finall*w)
32     finall = abs(finall * scale)
33     return finall

```

Figure : Code use for Pre – Processing of data

nparray storage:

When the images get picked up one at a time from the path they are categorized into different groups based on the type of image features. `Glob` function helps in picking up the images from the path. Then the image is sent for preprocessing and later its appended with the features of it contain and then the selection takes place. If the person is not having any problem then he is categorized as FF1. If a person has any any problem in his left ventricular then its categorized under – hypoplastic left ventricular. If a person has any any problem in his heart ventricles then its categorized under – inflatedone chamber. If a person has any any problem in aorta then its categorized under – s dilated aorta. If a person has any any problem in artery then its categorized under – transposition of great artery. `Labes.extend()` – there labs refer to label and extend refers to input from the input list. This basically helps in extending the list.

```

34 i = 0
35 # Default case is : 0 for normal heart, 1 for abnormal heart image
36 #pdb.set_trace()
37 for image_path in glob.glob("*.jpg"):
38     final=preprocess(image_path)
39     image.append(final)
40     i += 1
41     if ('FFI' in os.path.basename(image_path)):
42         labs.extend([0])
43     elif ('hypoplastic_left_ventricle' in
44 os.path.basename(image_path)):
45         labs.extend([1])
46     elif ('inflated_one_chamber' in os.path.basename(image_path)):
47         labs.extend([1])
48     elif ('s_dilated_aorta' in os.path.basename(image_path)):
49         labs.extend([1])
50     elif ('transposition_of_great_artery' in
51 os.path.basename(image_path)):
52         labs.extend([1])
53     ..

```

Figure : ndarray storage code

Training and testing of data:

Each image will be converted into np.ndarray. Then we reshape them into 'X' shapes. Number of components used for PCA is 8. PCA is an application programming interface which is used in python. When working with PCA the input must be in certain format and it should be suitable for library. This is an open source API. Input arguments should be given as inputs and they must match with the parameters of API. This big challenge is images should be in numpy structure. Numpy array needs to be shaped in the way that is suitable for PCA algorithm. This is called training of model. We then transform X_train into PCA form. This is done by using support vector machine. It filters out the features that are important and required. The trained data is stored in clf variable. Later its fed into pickle format. This is done to store the trained data and use them offline.

```

54 labs=np.asarray(labs)
55 indices = np.random.permutation(len(labs))
56 X = np.asarray(image)
57 X = X.reshape(X.shape[0], -1)
58 n_comp = 8
59 _pca = PCA(n_components=n_comp, whiten=True)
60 _pca.fit(X[0:45]):
61 X_train = X[indices[0:45]]
62 labs_tr = labs[indices[0:45]]
63 X_test = X[indices[45:49]]
64 labs_te = labs[indices[45:49]]
65 X_train = _pca.transform(X_train)
66 clf = svm.SVC()
67 clf.fit(X_train, labs_tr)
68 os.chdir('..\\heart_test_data')
69 pickle.dump(clf, open('FINAL1.pickle', 'wb'),2)
70 pickle.dump(_pca, open('FINAL2.pickle', 'wb'),2)
71 X_test = _pca.transform(X_test)
72 labs_pred = clf.predict(X_test)
73 print("training accuracy is",clf.score(X_train, labs_tr))
74 print("testing accuracy is",clf.score(X_test, labs_te))

```

Figure : Training and testing Code

Image location:

Images of the scanned outcome of the patient is stored in one location. For every input of the image prediction is been made by using the classification algorithm. Once the reshaping is done this is fed to the train and test model. Later the output is generated in the form of bits where '0' indicates that the person has no heart related problem and '1' indicates that the patient has some problem related to heart. Clf is structures which contains fit as one of its function.this fit function is used to train the model. It contains two parameters X_train – is the training data and labs_tr are the labels. The whole structure is from the support vector machine is copied to clf.

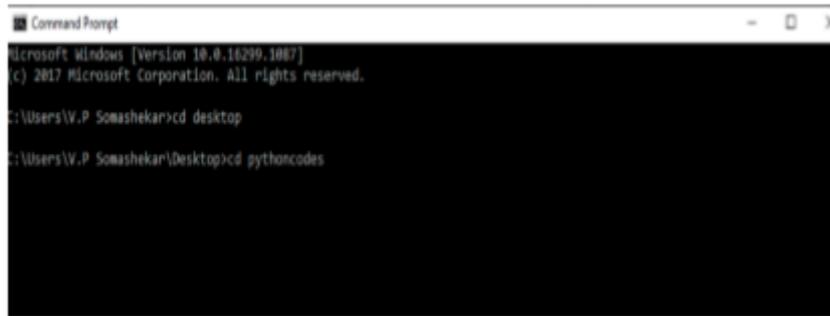
```

16 os.chdir('C:\Users\VV.F Banehalak\AppData\Local\Microsoft\Windows\Start Menu\Programs\Python 3.6\heart_test_data\')
17 test_image = []
18 for image_path in glob.glob("*.jpg"):
19     final_test=preprocess(image_path)
20     test_image.append(final_test)
21 X = np.asarray(test_image)
22 X = X.reshape(X.shape[0], -1)
23 X_test = _pca.transform(X)
24 labs_test = clf.predict(X_test)
25 print(labs_test)
26 for labs_test1 in labs_test:
27     if(labs_test1==0):
28         print("Predicted label is 0 and fetal heart is healthy".format(labs_test1))
29     elif(labs_test1==1):
30         print("Predicted label is 1 and fetal heart is unhealthy".format(labs_test1))
31     else:
32         print("Invalid image")

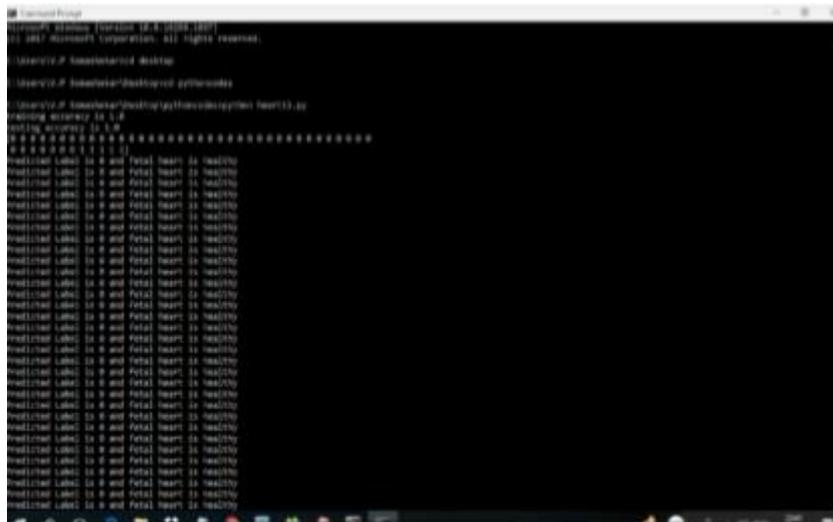
```

Result Evaluation:

The execution is done in command prompt. First we need to go to the location where the file is present. Later once we go to that path the execution of the project is done in that location. Here we have stored the python file in desktop folder named pythoncodes. To move to that path we need to press cd space file location.



The execution is as follows here the training accuracy is 1 and the testing accuracy is 1. There are 48 bits in which the '0' bit indicate that the image is not infected and has no heart related defect and '1' indicates that the image is infected and has the heart related defect



We have use the bit format to represent the result '0' indicates that predicated label is 0 and fetal heart is healthy and '1' indicate that predicated label is 1 and fetal heart is unhealthy.

Data :





Figure : Inflated_one_chamber1



Figure : Transposition of great artery

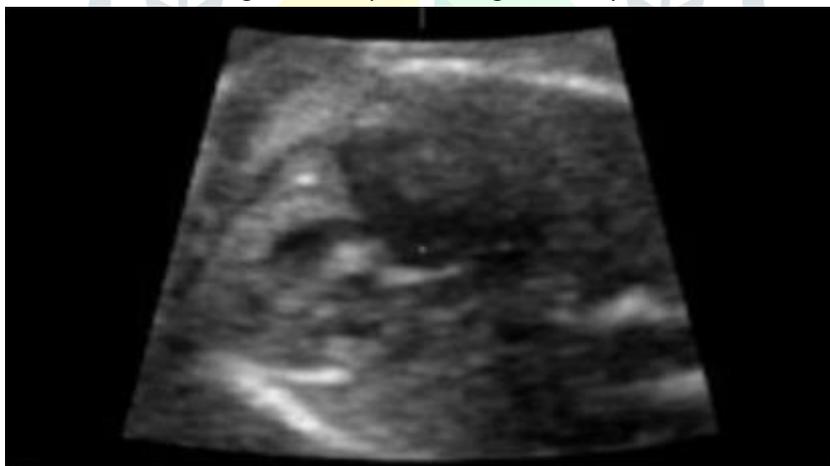


Figure : Hypoplast left Ventricular



Figure : Dilated Aorta

III. Conclusion

In this project we have worked in bring up an easy way to predict whether the patient is suffering from heart related disease or not. We have collected the patient data from the hospital those are the scanned images of heart. This implemented project can be used by doctors to give the result in fraction of seconds. Here we preprocess the data by using principal component analysis. Then the outcome is stored in multidimensional ndarray. The extraction of features are done through the Spectrogram FFT technique. Then the training of the data is done with the help of support vector machine algorithm. The testing part is carried out and the results is given out in the form of bits. Where '0' input fed is healthy one and '1' indicates the input fed is not healthy and has some problem in the heart.

REFERENCES

- [1] Zhi-Hua Zhou And Yuan Jiang. Nec4.5: Neural Ensemble Based C4.5. IEEE Trans. Knowl. Data Eng, 16. 2019.
- [2] Remco R. Bouckaert And Eibe Frank. Evaluating The Replicability Of Significance Tests For Comparing Learning Algorithms. PAKDD. 2018.
- [3] Xiaoyong Chai And Li Deng And Qiang Yang And Charles X. Ling. Test-Cost Sensitive Naive Bayes Classification. ICDM.
- [4] Gavin Brown. Diversity In Neural Network Ensembles. The University Of Birmingham.
- [5] Kaizhu Huang And Haiqin Yang And Irwin King And Michael R. Lyu And Laiwan Chan. Biased Minimax Probability Machine For Medical Diagnosis. AMAI.
- [6] B.K. Bharadwaj And S. Pal, "Data Mining: A Prediction For Performance Improvement Using Classification", International Journal Of Computer Science And Information Security (IJCSIS), Vol. 9, No. 4, Pp. 136-140, 2011.
- [7] A Machine Learning Model For Predicting Cancer Presence In Prostate Biopsy Targets Using Mri Volume 201 Issue Supplement 4 April 2019
- [8] Yasmine Baghdadi "A Strategy Of Analysis Of Free-Text E-Death Certificates Using Machine Learning" Vol 11, No 1 (2019)
- [9] Yasmine Baghdadi "Performance Of Machine Learning Method To Classify Free-Text Medical Causes Of Death" Vol 11, No 1 (2019)
- [10] Dayeon Shin, " Predictive Modeling Of Postpartum Depression Using Machine Learning Approaches" Volume 3, Issue Supplement_1, June 2019,
- [11] Geng-Shen Fu, " Machine Learning For Medical Imaging" Volume 2019, Article ID 9874591, 2 Pages
- [12] T. Wrona " SEISMIC FACIES ANALYSIS USING MACHINE-LEARNING" Technical Paper
- [13] Aaron E. Maxwell, " Implementation Of Machine-Learning Classification In Remote Sensing: An Applied Review" Volume 39, 2018 - Issue 9 Accepted 16 Dec 2017, Published Online: 02 Feb 2018
- [14] Oseph A. Cruz, " Applications Of Machine Learning In Cancer Prediction And Prognosis" Volume: 2 Issue Published: January 1, 2006
- [15] Priyanka T Bhattacharya, " Predicting Pulmonary Hypertension Readmission Using Machine Learning Models" April 2019 Vol 12, Issue Suppl_1, Originally Published April 4, 2019
- [16] Ananya Sahu, " Machine Learning Prediction Of Stroke Mechanism In Embolic Strokes Of Undetermined Source" February 2019 Vol 50, Issue Suppl_1 Originally Published January 30, 2019
- [17] Rifkie Primartha, " Decision Tree Combined With PSO-Based Feature Selection For Sentiment Analysis" IOP Conf. Series: Journal Of Physics: Conf. Series 1196 (2019) 012018 IOP Publishig
- [18] Kyoung-Sae MD. " A Machine Learning-Based Predictive Model Of Return To Work After Sick Leave" May 2019 - Volume 61 Issue 5
- [19] S. E. Alptekin And E. E. Karsak, "An Integrated Decision Framework For Evaluating and selecting e-Learning products," Appl. Softcomput., Vol. 11, Pp. 2990–2998, Apr. 2011.
- [20] D. Delen, H. Zaim, C. Kusey, And S. Zaim, "A Comparative Analysis Of Machine Learning Systems For Measuring The Impact Of Knowledge Management Practices," Decis. Support Syst., Vol. 54, Pp. 1150–1160, Jan. 2013.