

Machine Learning Algorithms for Disease Prediction using IoT Environment

Shital Markad, Dr. S.D.Shirbahadurkar

Department of Electronics and Telecommunication, Zeal College of Engineering and Research , Pune, India

ABSTRACT: Prediction system is very essential in today's era for different domains like health care, cloud computing, green computing etc. Such system basically uses different machine learning as well as supervised learning algorithms. In this work effectively implemented the smart device concept with the help of Internet of Things under the sensor network. We propose Heart Disease Prediction (HDP) base approach which uniquely calculates the different vital parameters of human body. This all data will be available on cloud at any moment of time. Furthermore, our work describes about the mobile app which uses sensor data and here after machine learning algorithms have been adopted to find the various ranges of vitals. It will work on three different zones like green, yellow and red zone. Green indicates normal, yellow is indicating moderate values whereas red is nothing but danger zone i.e. emergency. Finally with these vital parameters, cardiac index has been measured which is nothing but the proposed work. Theoretical analysis and implemented results shows that Various Machine Learning (ML) algorithm gives us higher efficiency in achieving our goals. Thus the proposed work advances the hypothetical establishment and makes benefaction in remotely monitoring patient and arranges traveling to Internet of Things in this research culture.

KEYWORDS: Supervised learning, Machine Learning, Heart disease, Data mining techniques, Clustering, and Classification, Prediction system.

I. INTRODUCTION

Machine learning approach will sensitize and analyzed the trained data for inculcating heart efficiency. Two major interfaces for doctor and patient will enable the transfer the information to each other. This framework detects underlying heart conditions in real time for sake of patient's life and generates alarm in sense of SMS, Email etc. based on high and low threshold values set a right for doctor as well as for responsible registered any family member. This will enrich the system with feeling of attention, care which add value to patient's health. In over last few years, healthcare systems have evolved rapidly purpose of providing wireless communication among doctor and patient via wearable devices referred to as "telemedicine." The technology should be to provide real-time monitoring of diseases such as heart disease failure, asthma, hypotension, hypertension, etc. located far from medical facilities such as remote areas or a people out of healthcare for enhancements Due to the behavioral change common to all age categories, heart disease is the leading cause of death in all of these circumstances. Literature narrates that about 2.8 billion people die from heart problems due to abdominal obesity that ultimately affects cholesterol levels, ups and downs in blood pressure, and more importantly, stress hormones cause ultimate heart disease. Many wearable devices have studied popular parameters of heart function such as BP, blood glucose level, blood oxygen saturation, ECG etc. Often discussed in detail is the need for hormonal imbalance due to stress factor, i.e. the individual's mood (behavioral state of health) and the impact of good / bad cholesterol.

II. LITERATURE SURVEY

Junior Carvalho, et al. [1] The authors submitted an integrated system for evaluating viability and enhancing the classification of cardiac diseases. This produces a time decrease in the electrocardiogram-ECG signal that can be achieved without any significant loss by reducing the amount of data samples. The aim of the urbanized system is to look from the heart for signals. The ECG signals are subjected to the preliminary filtering method and then using a clustering algorithm Gustafson – Kessel Fuzzy to exercise for the organization and correlation of signals. The term describes severe heart disorders including angina, myocardial infarction and coronary artery disorders. Also, the device could immediately be used by "on duty" physicians from any field of practice, which could require the first or initial diagnosis of cardiopathy. If any system senses a heart condition, this program endows other testimonials with PPV with a better diagnosis of disease and thus provides high confidence compared to other approaches. Another major factor is

the fact that this system was identical to many other systems by accessing the entire data set and this system used a fuzzy clustering algorithm to minimize the data set, thus reducing its use.

According to de Mirmozaffari Junior et.al Data Mining[2], in this work a resolution is being projected into a database to extract secret pattern from the scientific dataset. The database is composed of 209 occurrences and 8 attributes. The method was used in WEKA and MATLAB applications, and the accuracy of prophecy within the Apriori algorithm is compared in just three stages. MATLAB as the basis for better performance is ground-breaking. Wide ranges of Apriori algorithms sturdy system have been evaluated to predict heart disease in data mining. A single model is emerging which consists of one filter and one assessment system. Three strong rules are applied, along with different estimation methods, to find the superior algorithm. Apriori rules are calculated in terms of their real number of supports, greater precision and taking into account strong laws. We brought the high-performance technologies into effect. The experiment will serve as a practical method for physicians to predict, and then prescribe in turn unpredictable cases.

According to [3] the prediction of a heart attack received a professional advance from the heart disease database. Originally, the heart disease database is huddled using the K-means clustering algorithm, which will extort heart attack-appropriate data from the database. This method allows to expertise the number of fragments via its k parameter. Accordingly, the regular trends are excavated using the extracted data algorithm MAFIA (Maximum Regular Element Set Algorithm), related to heart disease. The machine learning algorithm is modeled on using the selected major patterns to effectively predict heart attack. They have engaged the ID3 algorithm as the training algorithm to prove level of heart attack with the decision tree. The results indicated that the prediction system designed is capable of effectively predicting a heart attack.

According to [4] authors, a prototype was described using data mining techniques mostly from Naïve Bayes and WAC (Weighted Associated Classifier). The dataset contains important factors such as age, sex, diabetes, height, weight, blood pressure, cholesterol, fasting of blood sugar, hypertension, disease. The program shows whether patient was at risk of heart failure or not.

Authors have proposed a patented heart disease prediction scheme using two different approaches, Naive Bayes and Logistic Regression as per[5]. A combinational approach is required to improve predictive accuracy for heart disease, as defined by survey.

According to the authors[6], heart disease is one of the major causes of death in the world region and precipitate prediction of disease is significant. The computer-aided devices are helping the doctor predict and manage heart disease as a gizmo. The goal of this paper is to expand heart-related cardiovascular disease and to brief about open decision support systems for computing and researching heart disease through data mining and smart hybrid techniques. A lot of DSS remains to be predicted on heart disease with several methodologies. Worldwide life expectancy estimates show a rise in number of heart disease. Therefore, it is important to build an powerful, competent, trustworthy automated system that predicts heart disease precisely based on the symptoms of gender / age and provincial expertise experts in the field at the lowest cost.

Explain that figures reveal a heart disease is one of the most significant factors behind deaths worldwide, according to [7]. Data mining techniques are very common in difficult scientific support systems and have the ability to detect hidden patterns and associations in medical data. Until now, data mining classification approaches are used to evaluate the various types of heart-based problems. The goal of this paper is to build a heart disease prediction framework using clustering methods for data mining. This paper races the various methods of clustering, k-mean, EM and the most remote first algorithm for prophecy of heart disease. The end result shows the first clustering algorithm is the best algorithm for testing other algorithms. Since the ratio of properly reported events to the cluster is the highest and the time required to construct the model is negligible. The System can be further extended. You can use more input attributes, and it can be further expanded by through cluster no. The same experiment can also be conducted on other data mining methods, such as R. And the classifier ensemble can also be made with the unique classifiers in order to estimate their effectiveness. The above algorithms can be applied to certain datasets to test whether or not the same algorithm provides the highest precision.

In this paper[8] the authors suggested incorporating access to a clustering approach and the regression methodology. DBSCAN is the clustering method used, and regression is subjected to multiclass logistic regression. Through running the DBSCAN clustering algorithm the entire dataset is divided into disjoint clusters. Clusters resulting were found to have fewer occurrences. Such clusters are focused on spatial regression of multiple groups. This finding stems from the clustering approach learned through an unsupervised process. When regression is complete we have reached the same variety of cardiac arrhythmia it is at a conclusion. The predicted system achieves an average accuracy of 80 per cent when measured with various other current approaches. By combining the use of DBSCAN clustering and multi-class logistical regression algorithms, a prophecy tool for cardiac arrhythmia type is designed. Through combining PCACRA with other approaches this approach is found to be 80 per cent effective.



Banu et al. The aim of Proposed[9] is to scrutinize broad medical diagnostic data using data mining tools and distribute useful information known as knowledge. Mining is a process by which vast data sets are investigated to gain secret and previously unknown connections and patterns of information detection to promote deeper understanding of medical data to thwart heart disease. Classification techniques for Naïve bayes(NB), Decision tree (DT), Neural Network (NN), Genetic Algorithm (GA), Artificial Intelligence (AI) and Clustering algorithms such as KNN, and Support Vector Machine (SVM) are available in many DM techniques. Through accessing entity technique, and also by combining two or more techniques, several studies for mounting prophecy model have been given. This paper, via data mining from 2004 to 2016, provides for a fast and easy evaluation and analysis of obtainable prophecy model. A survey conducted in this paper from 2004 to 2015 provides the scheme of various models that can be obtained and the different methodologies used for data mining. It also determines the accuracy of those versions obtained. This is true that not all transparent solutions have used the analytics of big data. The use of big data analytics along with data mining will deliver talented results to get the finest precision in manipulating the prophecy model.

Mane, suggested by TejaswiniU[10] indicates that heart disease is one of the diseases due to the regular incidence of fatality, and according to the world health organization the rate is high for that. Therefore, heart disease is estimated for the big data process, and as Big Data is evaluated so using Hadoop Map's diminished model. K-Means improved on the classification theory i.e., for clustering and the decision tree algorithm. ID3 can be accessed in the hybrid method. Because the second estimate is too large, the program is very useful for promoting forecasting methods, based on other constraints such as chest pain, cholesterol, age, Bp resting, Thalac and many more. This system will improve and be swift in taking medical decisions. It will also have an effect on the diagnosis humanisation process. This way the forecast for heart failure would be very helpful. In this way writers had cultivated with their conflicts and concerns about the big data and its resources. Individuals within the medical field are influenced by different parameters to the heart. Better K-Means is the algorithm which takes the precision of the centroid assortment more into account than the simple K-Means.

III. PROPOSED METHODOLOGY

First system collect the current input states from each sensors, then convert it from Analog to digital using ADC, once conversion has done, it will received by microcontroller, and at the same time it has stored into the database. The runtime monitoring system parallels real all events from database and show it to Graphical User Interface (GUI) then proposed machine Learning algorithm has works in the middle ware of system, It will always check all input values from desired threshold, if any time values shows below minimum support as well as maximum resistance, then it will automatically executes the output appliances. At the same time system measure the activity state of dangerous level, system also measure the time count of specific state, and whenever it cross the time scenario, it will execute the buzzer as well as GPS messaging system.

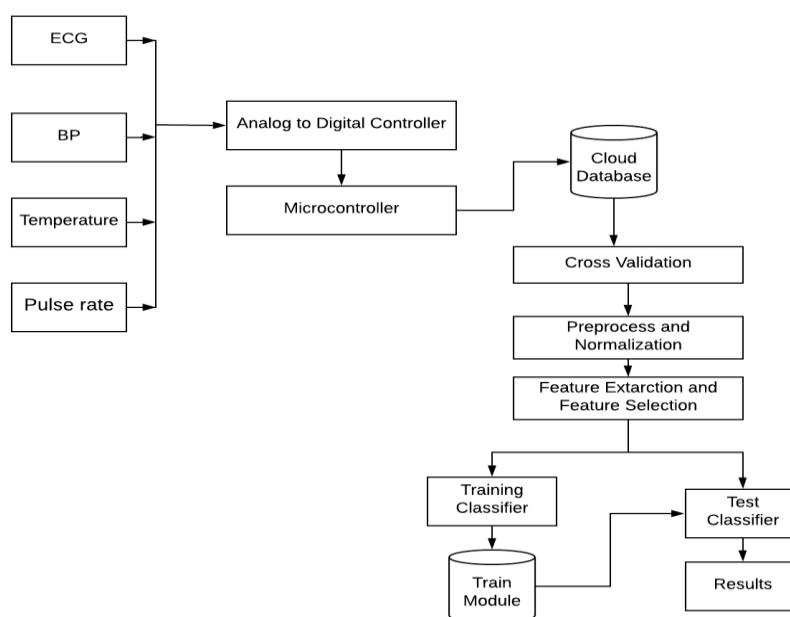


Figure 1 : Proposed System Architecture



Mainly proposed method has separated into two different phases, training and testing.

Training

- Gather information from internet similar to artificial information also actual time tolerant audit information.
- Concern information withdrawal approach like information pre-processing, information clean-up, information acquisition, outlier discovery also information alteration.
- Some time ago total these phase information has keep into the record called as backdrop information, which is use at the time of time testing.

Testing

- Initially we deploy anIoT-based healthcare scheme surroundings wherever we use some sensors as wearable devices.
- After that we have associated every one sensors to Raspberry Pi, also gather information as of sensor suing lot allowance approach.
- Every one collected has build up into worldwide record by association oriented design.
- In testing we study every testing also preparation information at the same time.
- Apply dissimilar classifiers also forecast the potential by choice creation system
- If any patient sufferings with dangerous condition about heart then using SMS API system automatically send SMS to user mobile number including heart severity parameters.
- Finally it gives the examination correctness with True useful also fake unenthusiastic of system.

Hardware Specification

- Raspberry pi. or ADC (Analog to Digital Converter) as microcontroller
- Blood pressure sensor
- ECG for QR, RR, PR intervals
- Temperature sensor for Body Tempreture
- FGA10 photodiode (900nm to 1700nm range)

IV. SYSTEM ANALYSIS

Algorithms 1 : Proposed HDP Algorithm

Input: Input values for all parameters HashMap<Double Value, String class> which contains the all attributes values like {BP, Heart_Rate, Cholestrol, Stress, Sugar, ECG, Oxygen_saturation, Hemoglobin, CI} etc. Policy patterns {P1,P2,Pn}

Output: Generate sample report for individual patient.

Step 1 :for each read Hashmap

$$\text{Extracted_Attribute}[i][j] \sum_{i=0,j=0}^n (a_{[i]}, a_{[j]}.....a_{[n]}, a_{[n].})$$

Step 2: if Extracted_Attribute[j] similar to P[1]

NormalPos = +1

MasterLits1. Add ← (NormalPos)

Step 3 : if Extracted_Attribute[j] similar to P[2]

AbnPos = +1

MasterLits2. Add ← (AbnPos)

Step 4 : if Extracted_Attribute [j] similar to P[n]

DenPos = +1

MasterLits3. Add ← (DenPos)

Step 5: end for

Step 6 :calculate the fitness factor for all classes using below formula for all class list

$$f = \sum_{k=0}^n \frac{F(x)}{\text{Sum}F(x)}$$



Step 7: Weight $_CurrentList[w] = \frac{MasterLits[i]}{TotalTest} * 100$

Step 8 : Sort $_CurrentList[w]$ using desc order

Step 9 : Recommend $_CurrentList[0]$ for final class for patient profile.

Step 10 :end procedure

Algorithm 2 : Naïve Bayes

Input: User input file data record which contains all body parameters sensor values, Patient id Pid, Timestamp T.

Output: Classified label

Step 1: Read R {All attributes } from current parameters.

Step 2: Map with train features with each sample.

Step 3 :calculate average weight of train DB with same evidences

$$AvgTScore = \sum_{k=0}^n (Sc)$$

Step 4 :evaluate AvgTScore> threshold

Step 5: ReturnAvgTScore

Algorithm 3 :Q- Learning Algorithm

Input: inp[1...n] all input parameters which is generated by sensors, Threshold group TMin[1...n] and TMax[1...n] for all sensor, Desired Threshold Th.

Output: Trigger executed for output device as lable.

Step 1 :Read all records from database (R into DB)

Step 2: Parts [] \leftarrow Split(R)

$$CVal = \sum_{k=0}^n Parts[k]$$

Step 3:

Step 4: check (Cval with Respective threshold of TMin[1...n] and TMax[1...n])

Step 5: T \leftarrow get current state with timestamp

Step 6 : if(T.time> Defined Time)

Read all measure of for penalty TP and reward FN

Else continue. Tot++

Step 7: calculate penalty score = (TP *100 / Tot)

Step 8 : if (score >= Th)

Generate event

end for

VI. RESULTS AND DISCUSSION

The result section is the final stage of research that includes experimentation, the data confirmed, and the evaluation and discussions to be concluded. The research is conducted by conducting various experimentation to check the efficiency of the proposed algorithm in terms of various parameters such as computation time, dataset type, and distinct algorithm input. A consternation matrix is a simple production analysis tool used in classification problems. Used to depict a prediction model's test result. Each matrix column represents instances in a projected class, while each row represents instances in an actual class. The four different experiments conducted to evaluate the confusion matrix with different dataset types. In the first experiment system carried out around 253 instances, the second 410, the third 750 and the final 1020 were used to find the consternation matrices.



Table 5.1: Confusion Matrix Analysis (Exp. 1)

Confusion matrix		Predicted	
		Negative	Positive
Actual	Negative	124	2
	Positive	5	122

The entries in the confusion matrix have the following meaning in the context of a data mining problem. The multiclass classification matrix defines some basic terms. Accuracy is the percentage of accurate predictions overall. Calculated by using the equation:

$$\text{Accuracy} = \frac{a+d}{a+b+c+d}$$

The proposed accuracy has calculate using below formula and it achieve around 97.23% correctly prediction is better than all the existing approaches.

$$97.23 = \frac{124+122}{124+113+9+7}$$

The figure 5.2 illustrate the multiple experiment analysis with various data size or records, the average accuracy is around acceptance level, even data size has changed as well as some algorithm parameters has tunes.

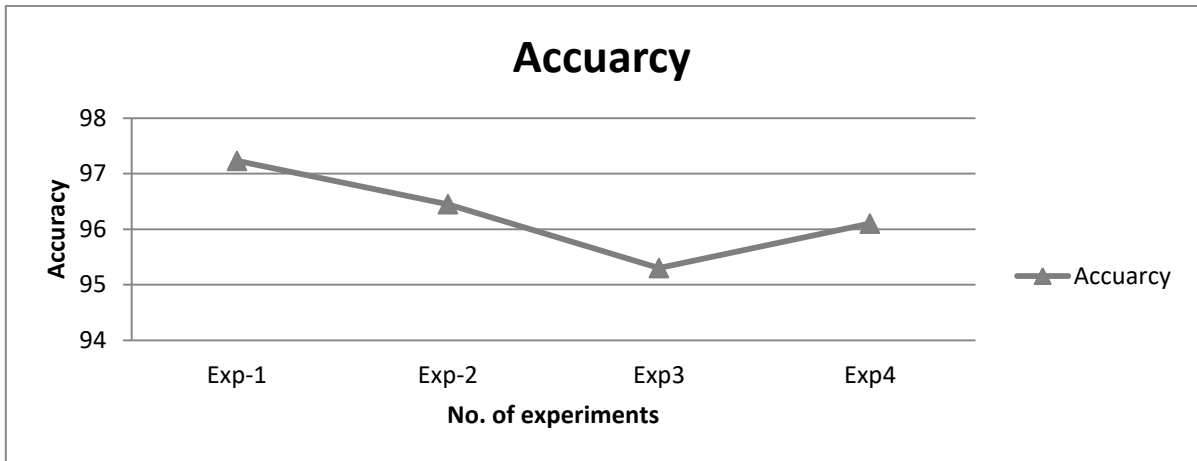


Figure 5.2 : Accuracy of proposed system with multiple experiment

The proposed FP has calculate using below formula and it achieve below 0.1 correctly prediction is better than all the existing approaches.

$$0.016 = \frac{2}{124}$$

The figure 5.13 illustrate False Positive ratio for proposed system. The experiment two has given maximum accuracy than first three experiments.

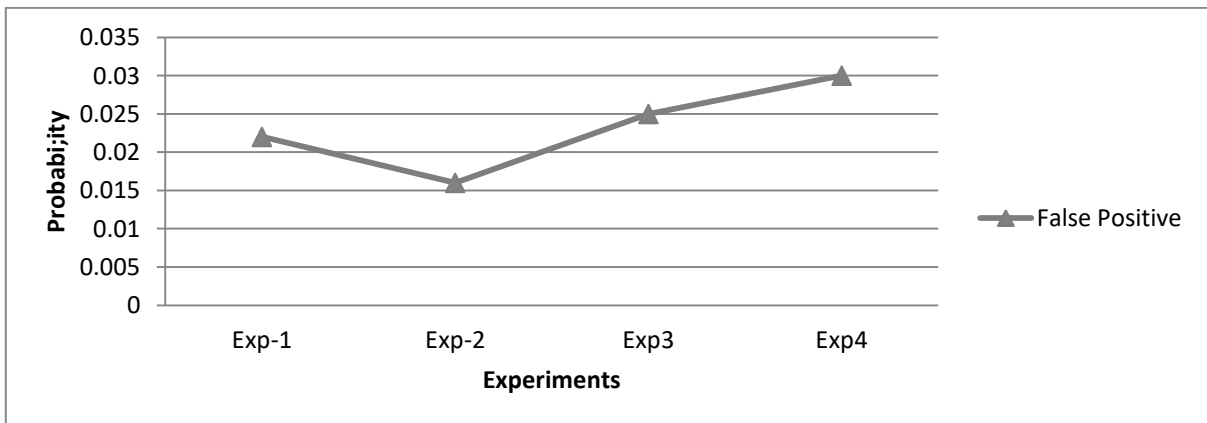


Figure 3: Error rate of proposed system with multiple experiments

The above figure 4 describes the false ratio of system with some algorithms; LR and HDP both consecutively generate the minimum false negative ratio

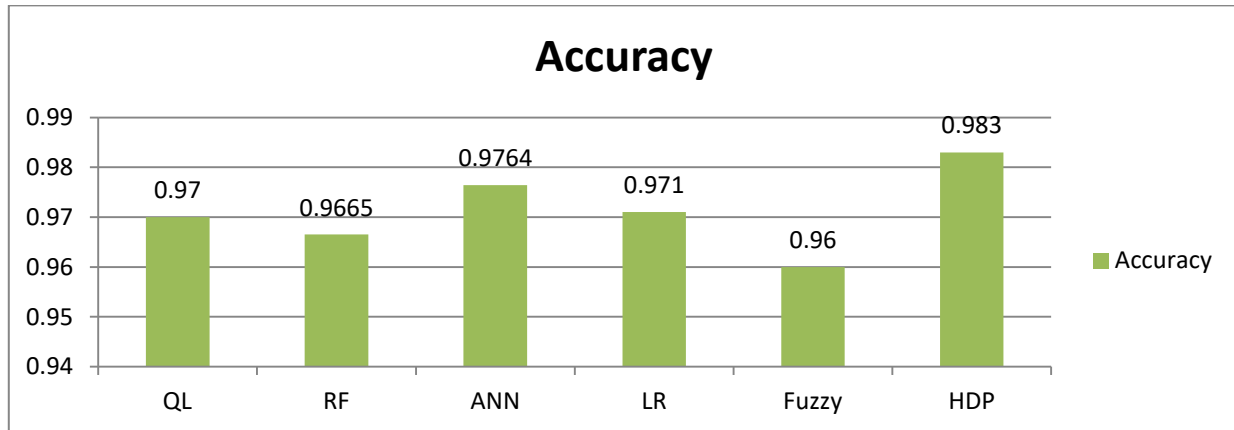


Figure 4 : Precision ratio of proposed system with multiple experiment

V. CONCLUSION

The typical approach we have found is that most of the tests are invasive that irritates patients, leaving them disappointed or indifferent about their health. Under these situations it is really difficult for them to cope. Therefore, the aim of this research is to provide them with a forum where each needy patient with a proposed non-invasive approach can obtain their vitality. In this scenario, patient can get in touch with the doctor 24X7 with internet technology and alert in case of emergency situations the implementation of the machine learning algorithm provides the system with efficiency, the system has implemented on different health care datasets. Using multiple experimental tests we conclude that the proposed method produces successful results compared with other disease prediction systems.

REFERENCES

- [1]. de Carvalho Junior, Helton Hugo, et al. "A heart disease recognition embedded system with fuzzy cluster algorithm." *Computer methods and programs in biomedicine* 110.3 (2013): 447-454.
- [2]. Mirmozaffari, Mirpouya, AlirezaAlinezhad, and AzadehGilanpour. "Data Mining Apriori Algorithm for Heart Disease Prediction." *Int'l Journal of Computing, Communications & Instrumentation Engg (IJCCIE)* 4.1 (2017).
- [3]. Khaing, HninWint. "Data mining based fragmentation and prediction of medical data." *Computer Research and Development (ICCRD)*, 2011 3rd International Conference on. Vol. 2. IEEE, 2011.
- [4]. Patel, Ajad, Sonali Gandhi, SwethaShetty, and BhanuTekwani. "Heart Disease Prediction Using Data Mining." (2017).
- [5]. Wghmode, MrAmol A., MrDarpanSawant, and Deven D. Ketkar. "Heart Disease Prediction Using Data mining Techniques." *Heart Disease* (2017).
- [6]. Vijayashree, J., and N. ChSrimanNarayanaIyengar. "Heart disease prediction system using data mining and hybrid intelligent techniques: A review." *Int. J. Bio-Sci. Biotechnol* 8 (2016): 139-148
- [7]. Singla, Meenu, and Kawaljeet Singh. "Heart Disease Prediction System using Data Mining Clustering Techniques."
- [8]. Cp, Prathibhamol, Anjana Suresh, and Gopika Suresh. "Prediction of cardiac arrhythmia type using clustering and regression approach (P-CA-CRA)." *Advances in Computing, Communications and Informatics (ICACCI)*, 2017 International Conference on. IEEE, 2017.
- [9]. Banu, NK Salma, and Suma Swamy. "Prediction of heart disease at early stage using data mining and big data analytics: A survey." *Electrical, Electronics, Communication, Computer and Optimization Techniques (ICECCOT)*, 016 International Conference on. IEEE, 2016.
- [10]. Mane, Tejaswini U. "Smart heart disease prediction system using Improved K-means and ID3 on big data." *Data Management, Analytics and Innovation (ICDMAI)*, 2017 International Conference on. IEEE, 2017.