

Heart Disease Prediction Using Bio Inspired Algorithms

Dr. Dinesh K¹, Indu R², Nalani Kanth L N³, *Niharika Priya R⁴,
Pavan Kumar Reddy Y⁵

¹Associate Professor, ²Student, ³Student, ⁴Student, ⁵Student

Department of CST, Madanapalle Institute of Technology and Science,
Madanapalle, Andhra Pradesh , India .

drkdinesh@mits.ac.in, 18691A2528@mits.ac.in, 18691A2549@mits.ac.in,

*18691A2551@mits.ac.in, 18691A2554@mits.ac.in

Abstract: *Heart related diseases or Cardiovascular Diseases (CVDs) are the main reason for a huge number of deaths in the world over the last many decades and has emerged as the most life-threatening disease, not only in India but also in the whole world. Prediction of cardiovascular disease is a critical challenge in the area of clinical data analysis. So, there's a need of dependable, accurate and possible system to diagnose similar diseases in time for proper treatment. Machine Learning algorithms and approaches have been applied to various medical datasets to automate the analysis of large and complex data. Multiple experimenters, in recent times, have been using several machine learning approaches to help the health care industry and the professionals in the diagnosis of heart related diseases. This project presents a review of various models based on like algorithms and approaches and analyses their performance. The main aim of this design is to give an effective algorithm to predict heart disease. So, at the end we compare our algorithm (Genetic algorithm) with BAT and BEE algorithms and we prove that the produced algorithm is effective one among all. Also, we forecast the output by taking some random data.*

Keywords: *Cardiovascular Diseases, Machine Learning Algorithms, Genetic Algorithm,*

1.INTROCUCTION

As indicated in a report by McKinsey, half of Americans have at least one persistent sickness, and 80% of American clinical consideration expense is spent on persistent illness treatment. With the improvement of expectations for everyday comforts, the rate of persistent sickness is expanding. The US has spent a normal of 2.7 trillion used every year on persistent infection treatment. This sum includes 18% of the whole yearly gross domestic product of the US. The medical services issue of ongoing illnesses is likewise vital in numerous different nations. In China, persistent sicknesses are the primary driver of death, as indicated by a Chinese report on sustenance and persistent illnesses in 2015, 86.6% of passing are brought about by constant infections. Thus, performing risk evaluations for constant diseases is fundamental. With the development in clinical information, gathering electronic wellbeing records (EWR) is progressively advantageous. Furthermore, first introduced a bio-enlivened elite execution heterogeneous vehicular telematics worldview, with the end goal that the assortment of versatile clients' wellbeing related continuous large information can be accomplished with the organization of cutting edge heterogeneous vehicular organizations. Patients' factual data, test results furthermore, illness history are kept in the EWR, empowering us to distinguish potential information driven answers for decrease the expenses of clinical contextual investigations. Proposed an ideal large information sharing calculation to deal with to entangle informational collection in telehealth with cloud strategies. One of the applications is to identify high-risk patients which can be used to diminish clinical expense since high-risk patients frequently require costly medical services. Also, in the principal paper proposing medical services digital actual framework, it creatively presented the idea of expectation based medical care applications, counting wellbeing risk evaluation. Expectation utilizing customary sickness risk models ordinarily implies an ai calculation (e.g., strategic relapse and relapse investigation, and so forth), and particularly a directed learning calculation by the utilization of preparing information with names to prepare the model. In the test set, patients can be ordered into gatherings of either high-chance or generally safe. These models are important in clinical circumstances and are generally considered. Not with standing, these plans have the accompanying qualities and deformities.

The informational collection is ordinarily little, for patients and infections with explicit circumstances, the attributes are chosen through experience. In any case, these prechosen qualities perhaps not fulfil the progressions in the illness and its affecting elements. With the improvement of enormous information examination innovation, more consideration has been paid to illness forecast from the viewpoint of huge information investigation, different explores have been led by choosing the qualities naturally from countless information to work on the exactness of hazard order instead of the recently chosen qualities. Nonetheless, that current work for the most part thought of organized information. For unstructured information, for instance, utilizing convolutional brain organization (CNN) to separate text attributes consequently has previously drawn in wide consideration and additionally accomplished awesome outcomes. Be that as it may, to the best of our insight, none of past work handle Chinese clinical text information by CNN. High-risk patients which can be used to diminish clinical expense since high-risk patients frequently require costly medical services. Also, in the principal paper proposing medical services digital actual framework, it creatively presented the idea of expectation based medical care applications, counting wellbeing risk evaluation. Expectation utilizing customary sickness risk models ordinarily implies an AI calculation (e.g., strategic relapse and relapse investigation, and so forth), and particularly a directed learning calculation by the utilization of preparing information with names to prepare the model.

In the test set, patients can be ordered into gatherings of either high-chance or generally safe. These models are important in clinical circumstances and are generally considered. Notwithstanding, these plans have the accompanying qualities and deformities. The informational collection is ordinarily little, for patients and infections with explicit circumstances, the attributes are chosen through experience. In any case, these pre-chosen qualities perhaps not fulfil the progressions in the illness and its affecting elements. With the improvement of enormous information examination innovation, more consideration has been paid to illness forecast from the viewpoint of huge information investigation, different explores have been led by choosing the qualities naturally from countless information to work on the exactness of hazard order, instead of the recently chosen qualities. Nonetheless, that current work for the most part thought of organized information. For unstructured information, for instance, utilizing convolutional brain organization (CNN) to separate text attributes consequently has previously drawn in wide consideration and additionally accomplished awesome outcomes. Be that as it may, to the best of our insight, none of past work handles Chinese clinical text information by CNN.

2.LITERETURE SURVEY

In 2010, A. Rajkumar and G.S. Reena applied machine learning algorithms similar as Naive Bayes, KNN (K- nearest neighbors) and diagnosis list for heart disease prediction. Tanagra tool is used to classify the data and the data estimated using 10-fold cross confirmation and the results are compared in table 4. The data set consists of 3000 cases with 14 different attributes. The dataset is divided into two parts, 70 of the data are used for training and 30 are used for testing. The results of comparison are based on 10-fold cross confirmation. Comparison is formed among these classification algorithms out of which the Naive Bayes algorithm is called as the better performance algorithm. Because it takes minor time to build a model and also it gives us best accuracy as compared to KNN and Decision Lists.

In 2011, G. Subbalakshmi, K. Ramesh and M. Chinna Rao developed a Decision Support in the Heart Disease Prediction System (DSHDPS) with data mining modeling technique, namely, Naive Bayes. By using heart disease attributes like chest pain, age, sex, cholesterol, blood pressure and blood sugar can predict the liability of patients getting a heart disease. It is enforced as web-based questionnaire application. Historical data set of heart patients from Cleveland database of UCI depository was used to train and test the Decision Support System (DSS). The cases to prefer Naive Bayes machine learning algorithm for predicting heart disease are as follows when data is high, when the attributes are independent of each other and when we want to achieve high accuracy as compared to different models. When the dimensionality of the inputs is high then Naive Bayes classifier technique is specifically suited. Despite its simplicity, Naive Bayes can frequently outperform more sophisticated classification methods.

In 2011, M.A.Jabbar, Priti Chandra and B.L.Deekshatulu in this study elaborate a prediction system by execute associative rule mining using a new approach that combines the conception of sequence numbers and clustering for heart attract prediction. By using this approach first dataset of heart disease patients has been converted into binary format also apply proposed system on binary transitional data.

The Data set of heart disease patients has been taken from Cleveland database of UCI depository with 14 essential attributes. The algorithm is well comprehended as Cluster Based Association Rule Mining Based on Sequence Number (CBARBSN). Support is a introductory parameter in associative rule mining. To become element of a frequent item is set to an item should satisfy support threshold. In this exploration transactional data table is divided into clusters based on skipping fragments (disjoint sub sets of actual transitional table) also Sequence Number and Sequence ID of each item has been calculated. On the base of Sequence ID frequent item sets has been discovered in different clusters and a common frequent item set has been taken as Global Item set. It has been observed from the experiment that Age > 45 and Blood pressure > 120 and Max Heart rate > 100 and old Peak > 0 and Thal > 3 = > Heart attack (Common frequent item set start in both clusters in this experiment). In our proposed algorithm execution time to mine association rules is lower (i.e., ms when support = 3) and as support increases prosecution time changes drastically as compared to preliminarily developed system.

In 2012, Chaitrali S. Dangare and Sulabha S. Apte implemented data mining and machine learning technique algorithms namely Decision Trees (J48), Naive Bayes, and Neural Networks on Heart disease datasets to make Intelligent Heart Disease Prediction System. In this exploration two datasets were used. The Cleveland Heart Disease dataset consisting of 303 records and Statlog Heart Disease dataset consisting of 270 records. Generally used 13 attributes two further attributes, i.e., obesity and smoking are included in the dataset for effective diagnosis of heart disease. Comparative results were examined on the both 13 attribute dataset and 15 attribute dataset independently. The Whole 573 records were divided into two data sets one is used for training consisting of 303 records and another for testing consisting of 270 records. Weka3.6.6 data mining and machine learning tool is used for experiment. Missing values in dataset were linked and replaced with most applicable values using Replace Missing Values (RMV) filter from Weka3.6.6. Give below table summarizes the relative results of our research. With the results it has been observed that neural network provides accurate result as compare to decision trees and Naive Bayes.

In 2013, A. Taneja, applied machine learning algorithms namely Decision Tree (J48 algorithm), Artificial Neural Networks (ANN), Naive Bayes and data mining for heart disease prediction. A dataset of 7339 instances with 15 attributes has been taken from PGI Chandigarh. The tool WEKA 3.6.4 was used for the experiment. For model training and testing the 10-Fold Cross Validation techniques is used randomly. The First Search method was used to select the best attributes from the already available 15 attributes and among them only 8 attributes has been selected. Each experiment was done on two different scenarios, first one containing all 15 attributes and second case only 8 selected attributes. From all these experiments the comparative results has been obtained and from these comparative results it has been found that the J48 pruned in selected attributes case has performed well in accuracy with 95.56% and Naive Bayes with all attributes case gives less accuracy 91.96% but takes least time to build a model in whole experiment.

2. PROPOSED SYSTEM

After evaluating the results from the existing methodologies, we've used python operations to perform heart disease classification for the data attained from the UCI repository. It provides an easy-to-use visual representation of the dataset, working environment and building the predictive analytics. Genetic Algorithm process starts from a preprocessing data phase followed by feature selection based on data cleaning, classification of modelling performance evaluation. also, BAT and BEE algorithms are used to improve the accuracy of the result.

Dimensionality reduction involves opting a mathematical representation such that one can relate the majority of, but not all, the variance within the given data, thereby including only most significant information. The data considered for a task or a problem, may consists of a lot of attributes or dimensions, but not all of these attributes may equally impact the output. A large number of attributes, or features, may affect the computational complexity and may indeed lead to overfitting which leads to poor results. therefore, dimensionality reduction is a very important step considered while building any model. Dimensionality reduction is generally achieved by two methods- feature extraction and feature selection.

The genetic algorithm is a method for solving both constrained and unconstrained optimization problems that is based on natural selection, the process that drives biological evolution. The genetic algorithm repeatedly modifies a population of individual solutions. At each step, the genetic algorithm selects individuals from the current population to be parents and uses them to produce the children for the next generation. Over successive generations, the population "evolves" toward an optimal solution. You can apply the genetic algorithm to solve a variety of optimization problems that are not well suited for standard optimization algorithms, including problems in which the objective function is discontinuous, nondifferentiable, stochastic, or highly nonlinear. The genetic algorithm can address problems of mixed integer programming, where some components are restricted to be integer-valued.

Bio-Inspired Genetic Algorithms with Formalized Crossover Operators for Robotic Applications. Genetic algorithms are widely adopted to solve optimization problems in robotic applications. In such safety-critical systems, it is vitally important to formally prove the correctness when genetic algorithms are applied.

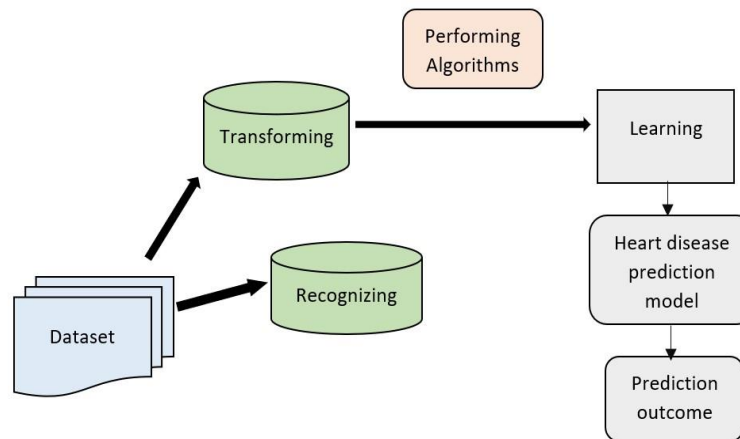


Fig 2.1 Infrastructure Setup of the Proposed System

3. EXPERIMENTATION

In this project student want to detect heart disease from dataset using Bio Inspired 4 features optimizing algorithms such as Genetic Algorithm, Bat, Bee and ACO. Here ACO algorithm is design in python to solve Travelling Salesman Problem to find shortest path and it cannot be implemented with heart disease dataset, so we are implementing 3 algorithms called Genetic, Bat and Bee.

Bio inspired algorithms design to optimized features used in dataset for training classification algorithms to increase prediction accuracy, sometime some datasets may have irrelevant values inside dataset and those irrelevant attributes or values may degrade classification accuracy so using optimize algorithms we can reduce features (attribute values) from dataset. This optimize algorithms will be applied on dataset to check whether all values are related to dataset or not, if any attribute found unrelated then it will removed from dataset.

To implement this algorithms we are using Heart disease dataset which contains 14 attributes and 4 class labels where 0 refers to No heart Disease and 1 refers to stage1 disease and 2 and 3 refers stage 3 and 4 disease.

Below are some values from dataset to train algorithms

age,sex,cp,trestbps,chol,fbs,restecg,thalach,exang,oldpeak,slope,ca,thal,class

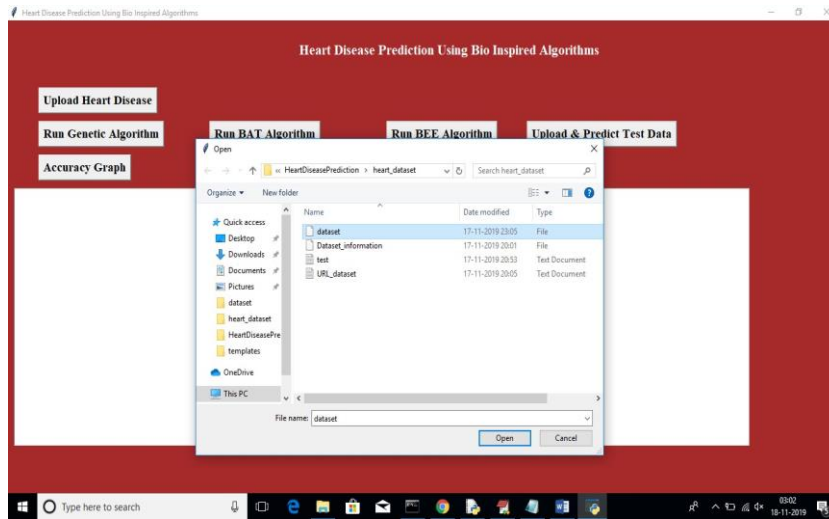
63.0,1.0,1.0,145.0,233.0,1.0,2.0,150.0,0.0,2.3,3.0,0.0,6.0,0

67.0,1.0,4.0,160.0,286.0,0.0,2.0,108.0,1.0,1.5,2.0,3.0,3.0,2

67.0,1.0,4.0,120.0,229.0,0.0,2.0,129.0,1.0,2.6,2.0,2.0,7.0,1

37.0,1.0,3.0,130.0,250.0,0.0,0.0,187.0,0.0,3.5,3.0,0.0,3.0,0

First records contain dataset column names and remaining records are the values of dataset. In last column we have class values as 0, 2, 1 and 3 as disease stage.



Test dataset also contains record values but it will not have class labels and application will apply that test values on train dataset to predict it class labels. Some values from test dataset.

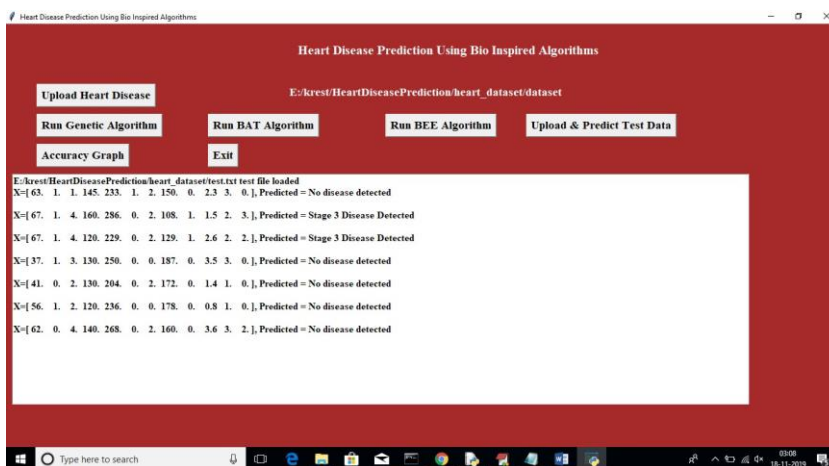
age,sex,cp,trestbps,chol,fbs,restecg,thalach,exang,oldpeak,slope,ca,thal

63.0,1.0,1.0,145.0,233.0,1.0,2.0,150.0,0.0,2.3,3.0,0.0,6.0

67.0,1.0,4.0,160.0,286.0,0.0,2.0,108.0,1.0,1.5,2.0,3.0,3.0

67.0,1.0,4.0,120.0,229.0,0.0,2.0,129.0,1.0,2.6,2.0,2.0,7.0

In above test dataset we can see there is no class name and application will predict it.



4. RESULTS

In above screen application has predicted disease stages. Now click on 'Accuracy Graph' button to view accuracy of all algorithms in graph format

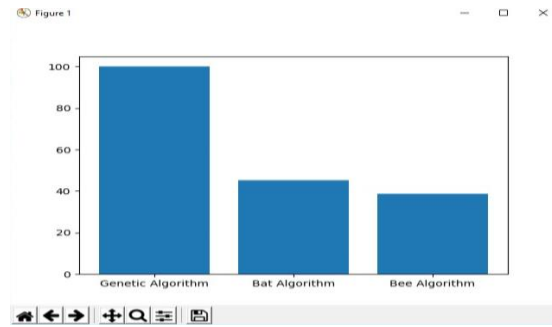


Fig 4.1 Accuracy Graph

5. CONCLUSION

Heart Disease Prediction has been developed using three bio-inspired algorithm modelling techniques. This project predicts person with heart disease by extracting the patient medical history that leads to a fatal heart disease from a dataset that includes patients' medical history such as chest pain, sugar level, blood pressure, serum cholesterol, maximum heart rate achieved etc. This Heart Disease detection system assists a patient based on his/her clinical information of them been diagnosed with a previous heart disease. The algorithms used in this building the given model are Genetic, BAT and BEE algorithms. The accuracy of our model is 100%. Use of more training data ensures the higher chances of the model to accurately predict whether the given person has a heart disease or not.

By using these, computer aided techniques we can predict the patient fast and better and the cost can be reduced very much. There are a number of medical databases that we can work on as these bio-inspired techniques are better and they can predict better than a human being which helps the patient as well as the doctors. Therefore, in conclusion this project helps us predict the patients who are diagnosed with heart diseases by cleaning the dataset and applying genetic algorithm. Also, it is concluded that accuracy of BAT algorithm is 45% accuracy and the BEE algorithm is of 38% accuracy. Finally Genetic algorithm will give us 100% accuracy. From this we conclude that Genetic algorithm will give high accuracy than BAT and BEE algorithms. This also provided a view accuracy of all algorithms in graph format. In graph xaxis represents Algorithm Name and y-axis represents Accuracy of those algorithms.

6 . REFERENCES

- [1] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Netw. Appl.*, vol. 19, no. 2, pp. 171–209, Apr. 2014.
- [2] J. Wang, M. Qiu, and B. Guo, "Enabling real-time information service on telehealth system over cloud-based big data platform," *J. Syst. Archit.*, vol. 72, pp. 69–79, Jan. 2017.
- [10] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, "Big data in health care: Using analytics to identify and manage high-risk and high-cost patients," *Health Affairs*, vol. 33, no. 7, pp. 1123–1131, 2014.
- [3] D. Tian, J. Zhou, Y. Wang, Y. Lu, H. Xia, and Z. Yi, "A dynamic and self-adaptive network selection method for multimode communications in heterogeneous vehicular telematics," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3033–3049, Dec. 2015.
- [4] K.Prasanna Lakshmi, Dr. C.R.K.Reddy, "Fast Rule-Based Heart Disease Prediction using Associative Classification Mining", *IEEE International Conference on Computer, Communication and Control (IC4-2015)*.
- [5] Boshra Bahrami, Mirsaeid Hosseini Shirvani, "Prediction and Diagnosis of Heart Disease by Data Mining Techniques", *Journal of Multidisciplinary Engineering Science and Technology (JMEST) ISSN: 3159-0040 Vol. 2 Issue 2, February–2015*.

- [6] N. Nori, H. Kashima, K. Yamashita, H. Ikai, and Y. Imanaka, "Simultaneous modeling of multiple diseases for mortality prediction in acute hospital care," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2015, pp. 855–864.
- [7] M. Satish, D. Sridhar, "Prediction of Heart Disease in Data Mining Technique", *International Journal of Computer Trends & Technology (IJCTT)*, 2015.
- [8] Lokanath Sarangi, Mihir Narayan Mohanty, Srikanta Pattnaik, "An Intelligent Decision Support System for Cardiac Disease Detection", *IJCTA, International Press* 2015.
- [9] S. Bandyopadhyay et al., "Data mining for censored time-to-event data: A Bayesian network model for predicting cardiovascular risk from electronic health record data," *Data Mining Knowl. Discovery*, vol. 29, no. 4, pp. 1033–1069, 2015.
- [10] B. Qian, X. Wang, N. Cao, H. Li, and Y.-G. Jiang, "A relative similarity based method for interactive patient risk prediction," *Data Mining Knowl. Discovery*, vol. 29, no. 4, pp. 1070–1093, 2015.
- [11] A. Singh, G. Nadkarni, O. Gottesman, S. B. Ellis, E. P. Bottinger, and J. V. Guttag, *Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration*, *J. Biomed. Inform.*, vol. 53, pp. 220–228, Feb. 2015.
- [12] P. Groves, B. Kayyali, D. Knott, and S. van Kuiken, *The 'Big Data' Revolution in Healthcare: Accelerating Value and Innovation. USA: Center for US Health System Reform Business Technology Office*, 2016.
- [13] M. Chen, Y. Ma, J. Song, C. Lai, and B. Hu, "Smart clothing: Connecting human with clouds and big data for sustainable health monitoring," *ACM/Springer Mobile Netw. Appl.*, vol. 21, no. 5, pp. 825–845, 2016.
- [14] K. Lin, M. Chen, J. Deng, M. M. Hassan, and G. Fortino, "Enhanced fingerprinting and trajectory prediction for iot localization in smart buildings," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 3, pp. 1294–1307, Jul. 2016.
- [15] M. Chen, P. Zhou, and G. Fortino, "Emotion communication system," *IEEE Access*, vol. 5, pp. 326–337, 2017, doi: 10.1109/ACCESS.2016.2641480.
- [16] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "HealthCPS: Healthcare cyber-physical system assisted by cloud and big data," *IEEE Syst. J.*, vol. 11, no. 1, pp. 88–95, Mar. 2017.
- [17] K. Hwang and M. Chen, *Big Data Analytics for Cloud/IoT and Cognitive Computing*. Hoboken, NJ, USA: Wiley, 2017.
- [18] Bo Jin, Chao Che, Zhen Liu, Shulong Zhang, Xiaomeng Yin, And Xiaopeng Wei, "Predicting the Risk of Heart Failure With EHR Sequential Data Modeling", *IEEE Access* 2018.
- [19] Aakash Chauhan, Aditya Jain, Purushottam Sharma, Vikas Deep, "Heart Disease Prediction using Evolutionary Rule Learning", "International Conference on "Computational Intelligence and Communication Technology" (CICT 2018). [3] Ashir Javeed, Shijie Zh
- [20] Aakash Chauhan, Aditya Jain, Purushottam Sharma, Vikas Deep, "Heart Disease Prediction using Evolutionary Rule Learning", "International Conference on "Computational Intelligence and Communication Technology" (CICT 2018). [3] Ashir Javeed, Shijie Zh
- [21] Senthilkumar Mohan, Chandrasegar Thirumalai, and Gautam Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", *IEEE Access* 2019.
- [22] Mamatha Alex P and Shaicy P Shaji, "Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique", *International Conference on Communication and Signal Processing* 2019.