

HEART DISEASE PREDICTION USING BIO INSPIRED ALGORITHMS**MUSA OM PRAKASH¹, Dr J REDDEPPA REDDY², K RAJU³**

1 & 2, Associate Professor, CSE department, Brilliant Institute of Engineering & Technology, Hyderabad, TS.

3 Assistant Professor, CSE department, Brilliant Institute of Engineering & Technology, Hyderabad, TS.

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 for a 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 healthcare industry and the professionals in the diagnosis of heart related diseases. This project presents a review of various models based on algorithms and approaches and analyzes 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 the most effective one among all. Also, we forecast the output by taking some random data.

1. INTRODUCTION

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 to 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. 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 forecasts from the viewpoint of huge information investigation. Different explorers 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.

2. LITERATURE SURVEY

In 2010, A. Rajkumar and G.S. Reena applied machine learning algorithms similar to 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 (DHDPS) 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 a web-based questionnaire application. Historical data set of heart patients from the 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 executing associative rule mining using a new approach that combines the conception of sequence numbers and clustering for heart attack prediction. By using this approach the first dataset of heart disease patients has been converted into binary format and also applied to the proposed system on binary transitional data.

The Data set of heart disease patients has been taken from the 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 an introductory parameter in associative rule mining. To become an element of a frequent

item is set to an item that should satisfy the support threshold. In this exploration, the transactional data table is divided into clusters based on skipping fragments (disjoint subsets of the actual transitional table) also Sequence Number and Sequence ID of each item has been calculated. On the basis of Sequence ID frequent item sets have 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 the 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 consists of 303 records and Statlog Heart Disease dataset consists 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 the dataset were linked and replaced with most applicable values using Replace Missing Values (RMV) filter from Weka3.6.6. The table below summarizes the relative results of our research. With the results it has been observed that neural networks provide accurate results as compared to decision trees and Naive Bayes.

3. EXISTING SYSTEM

Heart diseases have emerged as one of the most prominent cause of death all around the world. According to World Health Organisation, heart related diseases are responsible for the taking 17.7 million lives every year, 31% of all global deaths. In India too, heart related diseases have become the leading cause of mortality [1]. Heart diseases have killed 1.7 million Indians in 2016, according to the 2016 Global Burden of Disease Report, released on September 15, 2017. Heart related diseases increase the spending on health care and also reduce the

productivity of an individual. Estimates made by the World Health Organisation (WHO), suggest that India have lost up to \$237 billion, from 2005-2015, due to heart related or Cardiovascular diseases . Thus, feasible and accurate prediction of heart related diseases is very important. Generally a huge amount of data is generated during the clinical tests. These results need to be analyzed by the specialized doctor. However, the manual classification of these test results is time consuming and there may be a chance of misdiagnosis

PROPOSED SYSTEM

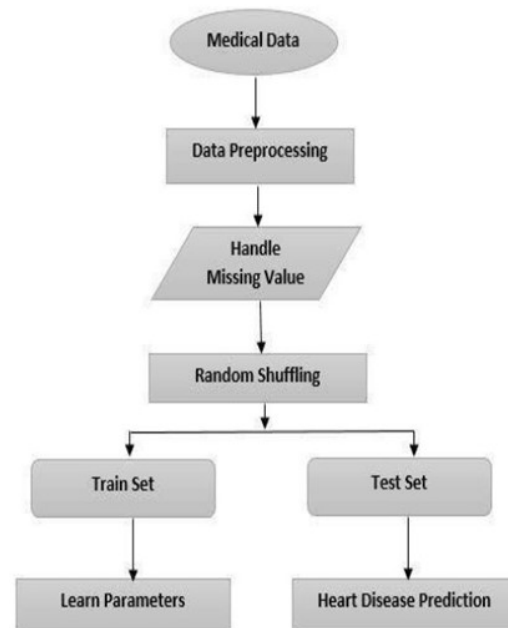
After evaluating the results from the existing methodologies, we've used python operations to perform heart disease classification for the data attained from the UCI repository. It provides an easy to- use visual representation of the dataset, working environment and building the predictive analytics. 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.

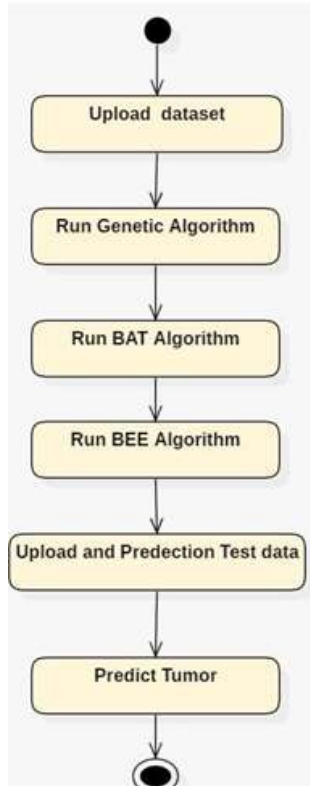
4.SYSTEM ARCHITECTURE .

Below diagram depicts the whole system architecture of project.



Activity Diagram

A graphical representation of the work process of stepwise exercises and activities with support for decision, emphasis and simultaneousness, used to depict the business and operational well-ordered stream of parts in a framework furthermore demonstrates the general stream of control.



5. SYSTEM IMPLEMENTATION

1. Upload Dataset
2. Training data
3. Testing data
4. Accuracy Graph

Upload Dataset

The Heart disease data is uploaded by admin without any particular scenario but with the details of articles. The most importantly large amount of can be handled in order to do practically. The data that are handling throughout the project can be done in this module. Users have permission to view data but not edit the data in online they can request the user to get the data.

Training data

The data is trained in three algorithms, they are Genetic algorithm, BAT algorithm, and BEE algorithm. The algorithm which gives more accuracy is considered as a best algorithm for testing the data. Besides of these three algorithms we can use CNN algorithm, Decision Tree algorithm etc.

Testing data

The test data is uploaded to the algorithm which gives more accuracy in the training of data. In the

testing of data we get to know whether the patient is having heart disease or not and we also predict the stage of disease like First stage, Second stage, Third stage.

Accuracy Graph

Data can be analyzed with the help of graphs like bar chart. This will bring the accuracy of the algorithms like Genetic Algorithm have 100% accuracy, BAT algorithm have 51% accuracy and BEE algorithm have 54% accuracy.

6.1 TYPES OF TESTING

■ Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application. It is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

■ Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfactory, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

■ Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

7.RESULTS

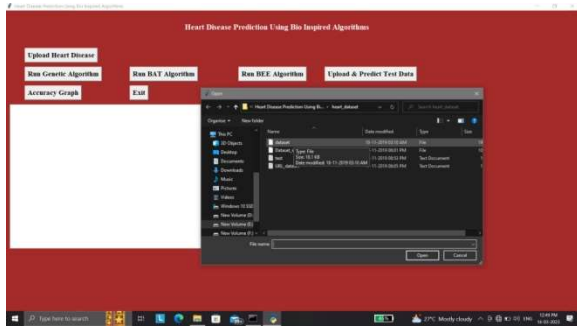


fig.7. 1 uploading of dataset



fig.7.2 dataset loaded

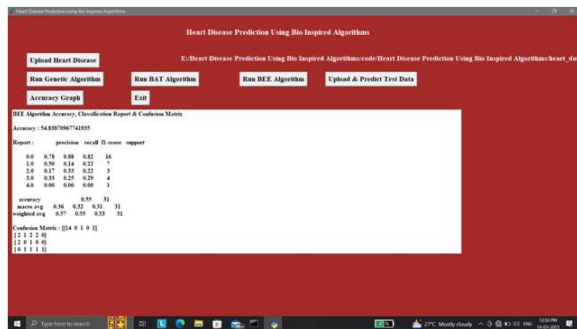


fig.7.5 accuracy of BEE algorithm

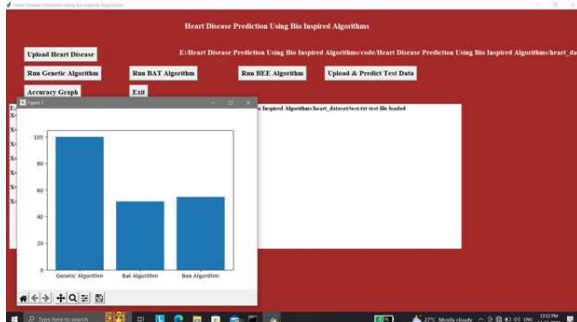


fig.7.6 accuracy Graph

8. CONCLUSION & FUTURE WORK

Heart Disease Prediction has been developed using three bio-inspired algorithm modeling techniques. This project predicts a person with heart disease by extracting the patient's 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 being diagnosed with a previous heart disease. The algorithms used in this building of 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 algorithms. Also, it is concluded that the accuracy of the BAT algorithm is 45% accuracy and the BEE algorithm is 38% accuracy. Finally Genetic algorithms will give us 100% accuracy. From this we conclude that the Genetic algorithm will give higher accuracy than BAT and BEE algorithms. This also provided a view accuracy of all algorithms in graph format. In the graph x axis represents Algorithm Name and y-axis represents Accuracy of those algorithms.

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.

[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.