

Heart Disease Prediction By Voting Classifier based Hybrid Model

Asst.Prof.Lanke RaviKumar
Computer Science
Jain University

City-Bangalore, Country-India
ravikumar.l@jainuniversity.ac.in

Bavirisetty Venkata Manideep
Computer Science
Jain University

City-Bangalore, Country-India
19btcrs013@jainuniversity.ac.in

Gollakaram Sasidhar
Computer Science
Jain University

City-Bangalore, Country-India
19btcrs022@jainuniversity.ac.in

Athish Sagar Gowda
Computer Science
Jain University

City-Bangalore, Country-India
19btcrs010@jainuniversity.ac.in

Meka Koteswara Sidhartha
Computer Science
Jain University

City-Bangalore, Country-India
19btcrs112@jainuniversity.ac.in

Abstract— This research compares and contrasts the efficacy of many popular categorization and prediction algorithms for making accurate diagnoses of heart disease. Although these algorithms excel at making predictions, their accuracy falls short. The author is creating a new technique, Hybrid Machine Learning, by merging two classification algorithms, Linear Model and Random Forest, to improve the accuracy of predictions made using the heart dataset. The Voting classifier will be constructed from a Linear Model and a Random Forest inside the hybrid algorithm, and the classification voting algorithm will compare the two algorithms' prediction accuracy and choose the one with the highest score. Hence, a more accurate method for predicting cardiovascular disease may be developed by applying hybrid models consistently.

Keywords— Heart disease, Hybrid Model, Genetic algorithm, Prediction, Machine Learning

I. INTRODUCTION

Artificial neural networks (ANNs) outperform Data Mining on every measures, including performance, computational speed, and complexity level [1]. Networks of basic processing components (called "neurons") that use local data and communicate with one another make up artificial neural networks (ANNs)[2]. Several difficult problems in the real world have been handled with the help of ANN, including the prediction of future trends using a company's massive data archive. Successful applications of ANN have been made in every area of engineering, from health and medical to engineering and manufacturing to marketing and ocean research[3]. In order to aid in the prediction of cardiovascular illness, this research introduces a swarm intelligence approach called Particle swarm optimisation (PSO) combined with a feed forward neural network.

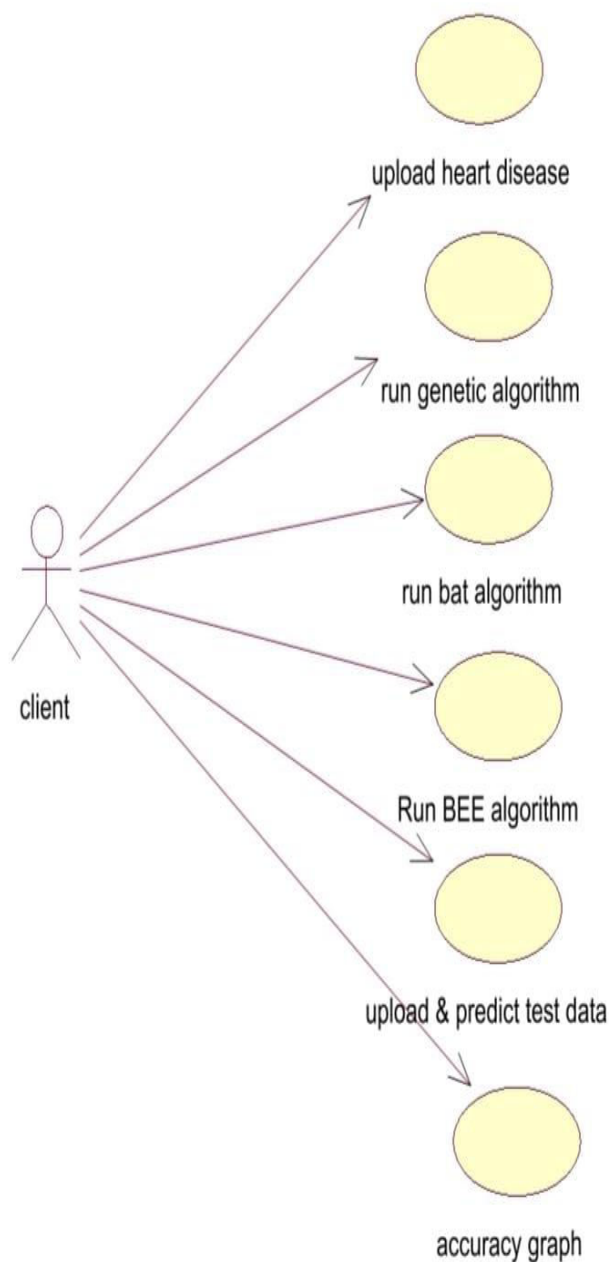
The "emergent collective intelligence of groupings of simple agents," as defined by Bonabeau [4].

Swarm behaviour occurs when a large number of similar, basic agents interact locally with one another and their surroundings without any overarching leadership. Swarm-based algorithms are a new class of population-based algorithms inspired by nature that can solve many complicated problems efficiently, quickly, and cheaply [5, 6].

So, we might describe Swarm Intelligence (SI) as a relatively young branch of AI that attempts to mimic the coordinated actions of social swarms seen in the natural world. Individually, these agents (insects or swarm members) are not very clever or capable, but they interact in predictable ways to accomplish tasks that are essential to their survival. Individuals in a swarm may have direct or indirect social interactions with one another.

In response to pupil displeasure, a typical swarm intelligence system will generate and show individualised encouragement messages. Our strategy for dealing with students' anger was tested in a commercial maths ITS with a big user base called Mindspark. We used the theory-driven approach in Mindspark to track student displeasure with the ITS in real-time. Based on the identified causes of dissatisfaction and data from Mindspark log files, we created inspirational messages to address students' feelings of helplessness. We developed and presented the inspirational messages using attribution theory, complimenting students' efforts, and demonstrating empathy. We have created an algorithm that triggers the display of the inspirational messages whenever it is determined that the learner is experiencing frustration. After applying the algorithm, we measured the average number of frustration occurrences every session to gauge the success of our method. Frustration levels dropped when people were exposed to the inspirational messages, and the difference was statistically significant (p 0.05).

USER CASE DIAGRAM



II. RELATED WORKS

An effective method for summarising, analysing, presenting, and maintaining massive datasets through data mining is necessary in today's competitive world. This calls for familiarity with all data mining techniques so that the best may be selected for desired datasets; such methods can provide answers to previously intractable puzzles. Data, according to studies, doubles in size every three years.

In 2010, A. Rajkumar and G.S. Reena used Naive Bayes, KNN (K- closest neighbours), and a diagnostic list—all machine learning algorithms—to predict cardiac disease. Table 4 compares the results of utilising the Tanagra tool for data classification with the results of using 10-fold cross confirmation to estimate the data. There are 3000 examples in all, and 14 types of data are included. Seventy percent of the data is utilised for training, while the remaining thirty percent is used for testing. Ten-fold cross-verification was used to arrive at these comparative findings.

The Naive Bayes method is deemed to have the best overall performance when compared to the other algorithms used for classification. Compared to KNN and Decision Lists, it provides the highest accuracy with the least amount of effort required to construct a model.[7]

Decision Support in Heart Disease Prediction System (DSHDPS) was created by G. Subbalakshmi, K. Ramesh, and M. Chinna Rao in 2011 using the data mining modelling approach of Naive Bayes. Patients' risk of developing heart disease may be estimated using factors such as chest discomfort, age, sex, cholesterol, blood pressure, and blood sugar. The application is mandatory as a web-based questionnaire. The Decision Support System was trained and tested using a historical data set of cardiac patients from the Cleveland database of the UCI repository (DSS). Naive Bayes is a machine learning approach that is preferred when data is abundant, the qualities are distinct from one another, and we want to outperform other models in terms of accuracy. The Naive Bayes classification method shines when dealing with high dimensional data. Naive Bayes, despite its apparent lack of sophistication, often outperforms more complex techniques of categorization.[8]

In 2011, M.A.Jabbar, Priti Chandra, and B.L.Deekshatulu used a novel method that integrates the idea of sequence numbers with clustering to develop a prediction system by executing associative rule mining for heart attract prediction. In addition to using the suggested technique to the binary transitional data, the first dataset of patients with heart disease has been converted to binary format using this method. Heart disease patient data with 14 key characteristics was obtained from the Cleveland repository at the University of California, Irvine. Cluster-Based Association Rule Mining with Sequential Numbers is a common name for this approach (CBARBSN). Associative rule mining begins with the introduction of the support parameter. In order to be included in the often occurring category, an item must have the required level of support.[9] In this investigation, we have generated the Sequence Number and Sequence ID for each item, and we have also clustered the transactional data table based on skipping fragments (disjoint sub sets of the real transitional table). Sequence IDs have been used to identify similar patterns among various clusters, and from these patterns, a Global Item Set has been generated. Age > 45, Blood pressure > 120, Maximum Heart Rate > 100, Old Peak > 0 and Thal > 3 = > Heart Attack, as determined by experiment (Common frequent item set start in both clusters in this experiment). As compared to the previously created system, the execution time of our suggested method to mine association rules is shorter (i.e., ms when support = 3), and the prosecution time varies dramatically as support grows. In 2012, Chaitrali S. Dangare and Sulabha S. Apte created an Intelligent Heart

Disease Prediction System by using data mining and machine learning approach algorithms such as Decision Trees (J48), Naive Bayes, and Neural Networks to datasets related to heart disease. This investigation made use of two different data sets. There are a total of 303 entries in the Cleveland Heart Disease dataset, and 270 records in the Statlog Heart Disease dataset. In addition to the standard 13 indicators for diagnosing cardiac disease, the dataset also includes two additional features, namely, body mass index and tobacco usage. Both the 13-attribute and the 15-attribute datasets were used for the comparative analysis. The whole collection of 573 records was split in half, with 303 records utilised for training and 270 records used for testing.[10]

Experiments are conducted using the data mining and machine learning programme Weka3.6.6. Using the Replace Missing Values (RMV) filter in Weka3.6.6, we were able to connect and replace missing values in the dataset with the most relevant ones. The following table provides a quick overview of our findings. The findings show that, compared to decision trees and Naive Bayes, neural networks provide more reliable predictions. A. Taneja used many machine learning techniques, including the Decision Tree (J48 algorithm), ANN, Naive Bayes, and data mining, to predict cardiac events in 2013. PGI Chandigarh has generously provided a dataset consisting of 7339 cases and 15 characteristics. For this study, we used WEKA 3.6.4. The random 10-Fold Cross Validation method is used for training and testing models.[11]

From the initial pool of 15, only the top 8 were considered useful enough to be included in the final set through the First Search technique. Two sets of data were collected for each experiment: one with all 15 characteristics and one with just 8. It has been determined that the J48 trimmed in the case of chosen attributes has done well in accuracy with 95.56%, while the Naive Bayes with all attributes case delivers less accuracy (91.96%), but takes the least time to create a model in the whole trial.[12]

III. PROPOSED METHODOLOGY

In this assignment, we will use Bio Inspired 4 features optimisation algorithms such the Genetic Algorithm, Bat Algorithm, Bee Algorithm, and ACO to identify cases of heart disease in a dataset. In this case, I am implementing the Genetic, Bat, and Bee algorithms since the ACO method, which was designed in Python to solve the Traveling Salesman Problem and discover the shortest route, is incompatible with the heart disease dataset. After a thorough analysis of the findings in terms of accuracy, sensitivity, and specificity, as well as comparison to other well-known data sets, it was determined that these results are among the best results produced in relevant past research.

- (1) This module is used to upload a database of prior patients with heart disease.
- (2) Pre-process Module: Using this module, we may filter out records that are lacking required information. Each classifier will use its share of the dataset, labelled training data, to construct a train model; this model will then be put through its paces in a test data setting.
- (3) To improve our classification accuracy, we will use the test data on the model we trained using the SVM module.

- (4) The fourth method is called Naive Bayes, and it entails creating a train model with the help of the Naive Bayes algorithm and then utilising test data to determine the accuracy of the model's classification.
- (5) At this step, the accuracy of the train model is evaluated using the Logistic Regression technique.
- (6) ANN The produced model of the deep learning artificial neural network may be evaluated for correctness using test data.
- (7) High-Risk Fuzzy Linear Modeling Offer a Hybrid Algorithm that Combines Linear Modeling with Random Forest. Both algorithms will be utilised to produce a hybrid model, and the best performing method will be selected using a voting classifier.
- (8) One such module is the supplementary Extreme Machine Learning Extension Module, which boasts improved prediction accuracy above other available methods thanks to its use of a cutting-edge Extreme Machine Learning algorithm. An innovative approach to pattern categorization and function approximation, Extreme Learning Machine (ELM) has been developed recently. This strategy relies on a single hidden layer of nodes and is conceptually identical to a single feed forward neural network, with weights between inputs and hidden nodes being randomly assigned and remaining constant during training and prediction. In contrast, it doesn't take long to train the weights that link the hidden nodes to the outputs. In the experimental literature, ELMs have been shown to have adequate prediction accuracy at a substantially lower computing cost than networks trained using the back-propagation approach.
- (9) The ninth section, "Graph," provides a comparative visual representation of the accuracy of each method.

IV. RESULTS AND DISCUSSION

The 'heart dataset' folder contains the Cleveland Heart disease dataset used in the aforementioned study. The home page of the project is shown in the fig.1.

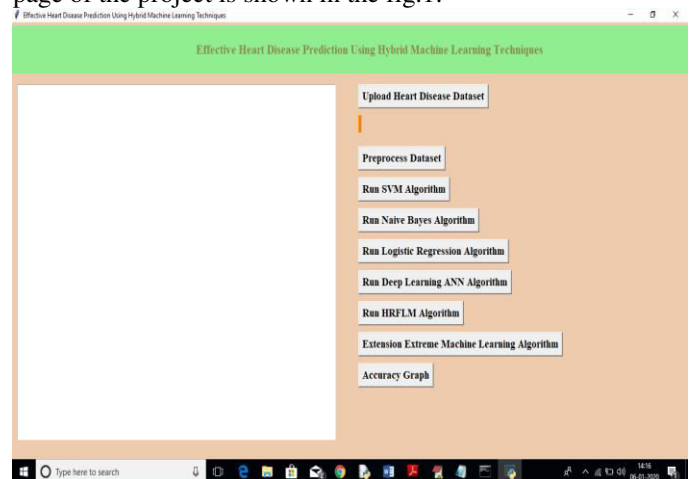


Fig. 1. Home Page

To submit a heart dataset, use the top page and its "Upload Heart Disease Dataset" button. Clicking the "Pre-process Dataset" button will apply a pre-processing procedure to the dataset, removing any non-numerical information. After some pre-processing, the dataset was reduced in size to only

297 records, and it was clear that the programme had arbitrarily divided the remaining data into a train dataset and a test dataset. We used 237 records for the training set, and 60 for the test set. Since the application randomly selects 60 records for testing, the accuracy of the same method will constantly vary. The results of applying test data to the SVM model generated from the train dataset and clicking the "Run SVM Algorithm" button are shown in Fig. 2. The SVM model achieved an accuracy of 62%. Fig. 3 displays the results of running the Naive Bayes algorithm, which showed an accuracy of 72 percent. Logistic regression, when its accuracy was checked by selecting "Run Logistic Regression Algorithm," achieved 69%, as seen in Fig. 4. Accuracy for ANN was calculated by pressing the "Run Deep Learning ANN Algorithm" button, and the results are shown in Fig. 5. Using the "Run HRFLM Algorithm" button yielded 84% accuracy for the planned task, as seen in fig.6. Using the "Extreme Machine Learning Extension" option yielded a 93% accuracy for the EML extension method, outperforming every other technique tested (fig.7). The accuracy graph shows algorithm names along the x-axis and algorithm performance along the y-axis. Fig. 8 shows that among the proposed algorithms, HRFLM and the extension algorithm provide the highest levels of accuracy.

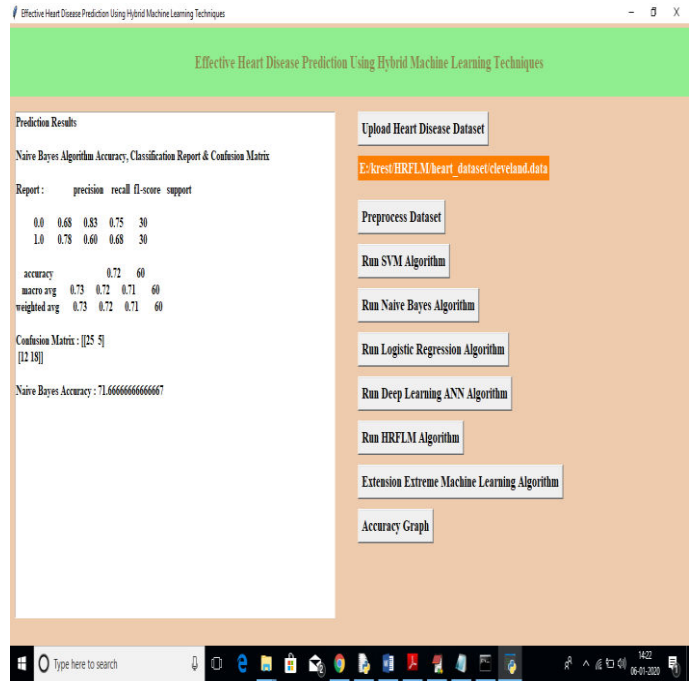


Fig. 3. NB Accuracy

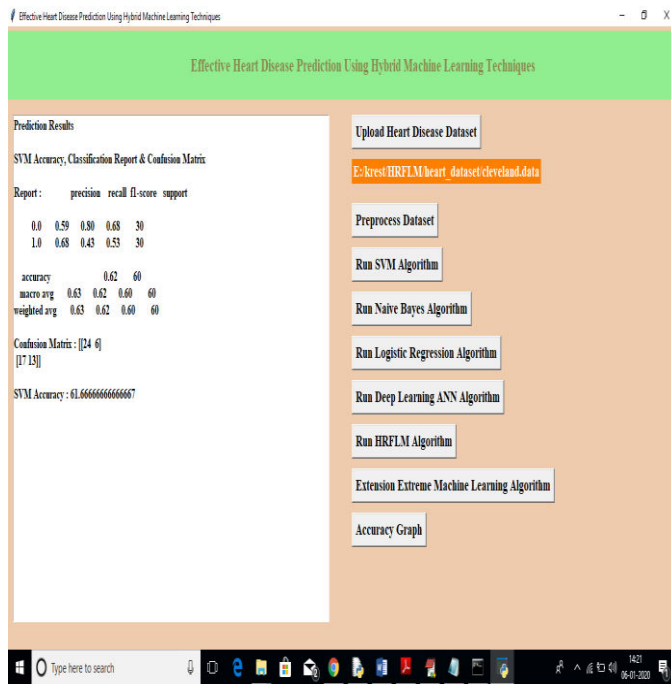


Fig. 2. SVM Accuracy

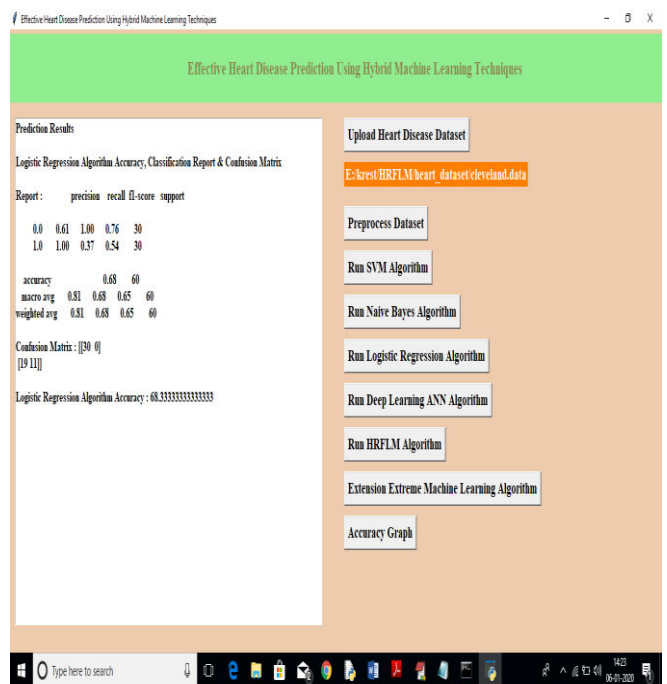


Fig. 4. LR Accuracy

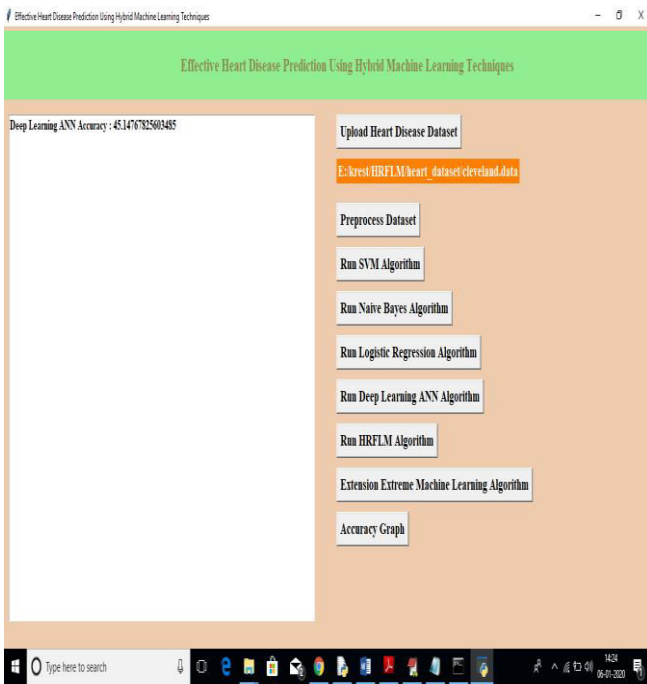


Fig. 5. ANN Accuracy

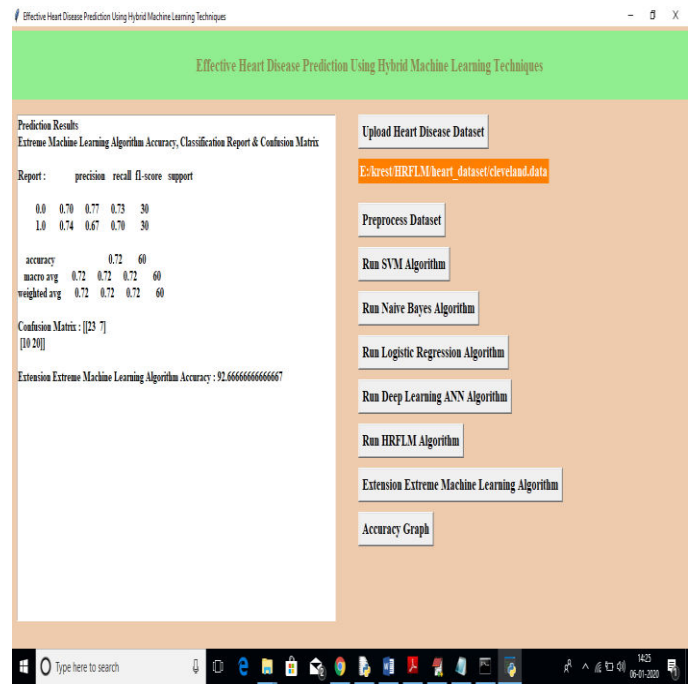


Fig. 7. EML Accuracy

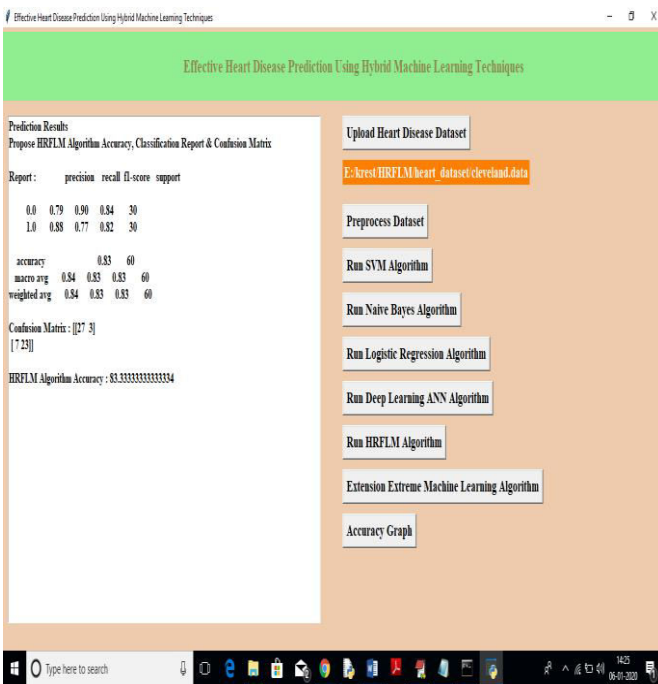


Fig. 6. HRFLM Accuracy

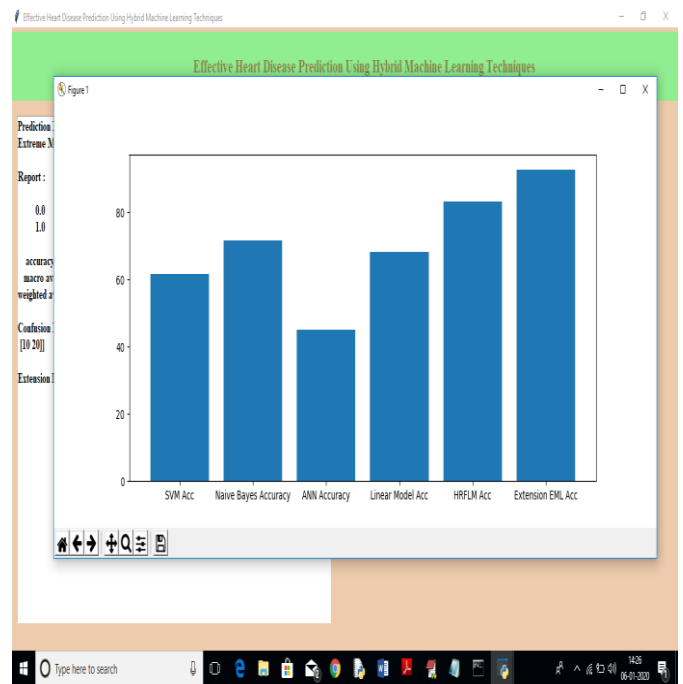


Fig. 8. Accuracy Graph

V. CONCLUSION

Three bio-inspired algorithm modelling methodologies were used to create a method for predicting heart disease. Using a dataset that contains patients' medical history such chest discomfort, sugar level, blood pressure, serum cholesterol, maximal heart rate attained etc., this study predicts person with heart disease. If a patient has already been diagnosed with heart disease, the Heart Disease Detection System may help. In order to construct the specified model, we use the Genetic, BAT, and BEE algorithms. Our model's accuracy is

one hundred percent. In order to improve the model's ability to determine if a person has heart disease, more data must be used during training. We can make more accurate, timely, and affordable patient predictions with the help of these computer-aided methods. As these bio-inspired methods are superior and can make more accurate predictions than humans, they are useful not just for patients but also for their physicians. Because of the work done here to clean the information and apply genetic algorithm, we can more accurately forecast which patients will be diagnosed with heart disease. The BAT algorithm is shown to have a 45% success rate, whereas the BEE algorithm only has a 38% success rate. One day, a genetic algorithm will allow us to achieve perfect precision. This leads us to believe that the Genetic Algorithm is superior to the BAT and BEE algorithms in terms of accuracy. As a bonus, this allowed us to see how precise each method was in graphical form. The accuracy of several algorithms is shown in a graph, with their names along the x-axis and the numbers along the y-axis.

REFERENCES

- [1] K. P. Kumar, V. Rohini, J. Yadla and J. VNRaju, "A Comparison of Supervised Learning Algorithms to Prediction Heart Disease," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-5, doi: 10.1109/ICECONF57129.2023.10084035.
- [2] V. K. H M and S. D S, "Performance Analysis of Base and Meta Classifiers and the Prediction of Cardiovascular Disease using Ensemble Stacking," 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2023, pp. 584-589, doi: 10.1109/IDCIoT56793.2023.10053464.
- [3] R. Jahed, O. Aseer and A. Al-Mousa, "Using Personal Key Indicators and Machine Learning-based Classifiers for the Prediction of Heart Disease," 2023 International Conference on Smart Computing and Application (ICSCA), Hail, Saudi Arabia, 2023, pp. 1-6, doi: 10.1109/ICSCA57840.2023.10087430.
- [4] A. Pant, A. Rasool, R. Wadhvani and A. Jadhav, "Heart disease prediction using image segmentation Through the CNN model," 2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2023, pp. 374-380, doi: 10.1109/Confluence56041.2023.10048835.
- [5] M. S. Gangadhar, K. V. S. Sai, S. H. S. Kumar, K. A. Kumar, M. Kavitha and S. S. Aravinth, "Machine Learning and Deep Learning Techniques on Accurate Risk Prediction of Coronary Heart Disease," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 227-232, doi: 10.1109/ICCMC56507.2023.10083756.
- [6] P. Mohana Priya, "Disease Prediction and Classification using Intelligent Algorithms," 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2023, pp. 494-498, doi: 10.1109/ICAIS56108.2023.10073698.
- [7] L. G and P. Sujatha, "Assessment of Classification Techniques for Heart Disease Prediction," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-5, doi: 10.1109/ICECONF57129.2023.10083836.
- [8] L. D. Gopiseti, S. K. L. Kummera, S. R. Pattamsetti, S. Kuna, N. Parsi and H. P. Kodali, "Multiple Disease Prediction System using Machine Learning and Streamlit," 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2023, pp. 923-931, doi: 10.1109/ICSSIT55814.2023.10060903.
- [9] G. N. Ahmad, H. Fatima, Shafiullah and M. Haris, "Hybrid Machine Learning Algorithms for Optimal Diagnosis of Heart Disease with Feature Analysis," 2023 International Conference on Power, Instrumentation, Energy and Control (PIECON), Aligarh, India, 2023, pp. 1-6, doi: 10.1109/PIECON56912.2023.10085781.
- [10] S. T. Himi, N. T. Monalisa, M. Whaiduzzaman, A. Barros and M. S. Uddin, "MedAi: A Smartwatch-Based Application Framework for the Prediction of Common Diseases Using Machine Learning," in IEEE Access, vol. 11, pp. 12342-12359, 2023, doi: 10.1109/ACCESS.2023.3236002.
- [11] K. K. Baseer, S. B. A. Nas, S. Dharani, S. Sravani, P. Yashwanth and P. Jyothirmai, "Medical Diagnosis of Human Heart Diseases with and without Hyperparameter tuning through Machine Learning," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 1-8, doi: 10.1109/ICCMC56507.2023.10084156.
- [12] W. A. W. A. Bakar, N. L. N. B. Josdi, M. B. Man and M. A. B. Zuhairi, "A Review: Heart Disease Prediction in Machine Learning & Deep Learning," 2023 19th IEEE International Colloquium on Signal Processing & Its Applications (CSPA), Kedah, Malaysia, 2023, pp. 150-155, doi: 10.1109/CSPA57446.2023.10087837.