HEART DISEASE PREDICTION USING BIO-INSPIRED ALGORITHMS

[1]**V. Sundara Ratnam,**   [2]**Bathini Rukmini,**   [3]**G. Meghana Reddy,**   [4]**Sukasi Dhatrija,**   [5]**T. Jahnavi Latha**

[1]Assitsant Professor, [2,3,4,5]UG Students, Dept. Computer Science and Engineering-Data Science, Mallareddy Engineering college for Women, Hyderabad, India

ABSTRACT

Heart related diseases or Cardio Vascular Diseases (CVDs) are the main reason for a huge number of death in the world over the last few decades and has emerged as the most life-threatening disease, not only in India but in the whole world. So, there is a need of reliable, accurate and feasible system to diagnose such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart related diseases. This paper presents a survey of various models based on such algorithms and techniques and analyze their performance. Models based on supervised learning algorithms such as Random Forest (RF) and Deep Learning based Bio-Inspired Algorithms such as Genetic Selection CV, BAT, ABC(BEE) and ensemble models are found very popular among the researchers.

INTRODUCTION

Heart disease is a significant global health concern, and early detection plays a crucial role in improving patient outcomes. Traditional methods of heart disease prediction often rely on clinical data and medical examinations, which may be time-consuming and costly. As a result, there is a need for a more efficient and accurate approach to predict heart disease risk using bio-inspired algorithms.

The goal of this project is to develop a heart disease prediction system that utilizes bio-inspired algorithms to analyze relevant patient data and provide accurate predictions of heart disease risk. The system will leverage the power of bio-inspired algorithms, such as genetic algorithms, particle swarm optimization, or ant colony optimization, to optimize the prediction model and improve its performance.

PROBLEM OVERVIEW

Heart disease is a leading cause of morbidity and mortality worldwide, making early detection and accurate prediction crucial for effective intervention and treatment. Traditional methods of heart disease prediction heavily rely on clinical data and medical examinations, which can be time-consuming, expensive, and limited in their ability to handle complex patterns and high-dimensional data.

HEART DISEASE STATISTICS

According to a news article by **Interactive Atlas of Heart Disease and Stroke[1],** heart disease proves to be the leading cause of death for both women and men. The article states the following:

- About 610,000 people die of heart disease in the United States every year– that's 1 in every 4 deaths.
- Heart disease is the leading cause of death for both men and women. More than half of the deaths due to heart disease in 2009 were in men.
- Coronary Heart Disease (CHD) is the most common type of heart disease, killing over 370,000 people annually.

- Every year about 735,000 Americans have a heart attack. Of these, 525,000 are a first heart attack and 210,000 happen in people who have already had a heart attack.

This makes heart disease a major concern to be dealt with. But it is difficult to identify heart disease because of several contributory risk factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate, and many other factors. Due to such constraints, scientists have turned towards modern approaches like Data Mining and Machine Learning for predicting the disease.

Overall, this project addresses the need for an efficient and accurate approach to heart disease prediction using bio-inspired algorithms, ultimately striving to enhance patient outcomes, reduce healthcare costs, and contribute to the advancement of cardiovascular health management.

EXISTING SYSTEM

Before the advent of machine learning techniques, heart disease prediction relied on traditional statistical methods and clinical risk assessment models. These approaches incorporated various risk factors, medical history, and diagnostic test results to estimate the likelihood of developing heart disease.

The existing heart disease prediction methods, which are based on traditional statistical models and risk assessment tools, have been widely used in clinical practice to estimate the risk of developing heart disease. However, they may have limitations in handling complex and nonlinear relationships between risk factors and disease outcomes, as well as in accommodating large-scale and diverse datasets. Machine learning techniques have emerged as a powerful approach to address these limitations and improve the accuracy of heart disease prediction models.

The base paper, "Disease Prediction by ML over Big Data from Healthcare Communities" by Min Chen, provides valuable insights into disease prediction using machine learning techniques over large-scale healthcare data. While the base paper does not specifically focus on heart disease prediction using bio-inspired algorithms, it lays a foundation for understanding the broader landscape of disease prediction in the context of healthcare. The literature survey presented here encompasses relevant research articles that discuss heart disease prediction, bio-inspired algorithms, and their application in the healthcare domain.

1. "Prediction of Heart Disease Using Machine Learning Techniques: A Survey" by A. S. Mohiuddin, R. A. Awan, and S. H. Mirza (IEEE Access, 2019):

This survey paper provides a comprehensive overview of various machine learning techniques applied to heart disease prediction. It discusses the use of different algorithms, such as decision trees, support vector machines, and neural networks, and highlights the strengths and limitations of each approach. Understanding the existing ML techniques for heart disease prediction will provide a foundation for exploring the potential of bio-inspired algorithms in this domain.

2. "Bio-Inspired Optimization Algorithms for Healthcare Data Analytics: A Review" by A. Ahmed, M. M. Al-Shabi, and K. Elleithee (Journal of Healthcare Engineering, 2019):

This review article explores the application of bio-inspired optimization algorithms in healthcare data analytics. It discusses the use of genetic algorithms, particle swarm optimization, and ant colony optimization in various healthcare applications, including disease prediction. The article provides insights into the strengths, challenges, and potential benefits of utilizing these algorithms in healthcare contexts.

3. "A Hybrid Model for Predicting Heart Disease using Random Forest and Genetic Algorithm" by R. Jindal, K. Kaur, and H. S. Bhatti (International Journal of Biomedical Engineering and Technology, 2017):

This research paper presents a hybrid model that combines random forest and genetic algorithms for heart disease prediction. The study demonstrates the effectiveness of using bio-inspired algorithms to optimize the performance of machine learning models. It highlights the potential of genetic algorithms in feature selection and model optimization to improve prediction accuracy.

4. "Heart Disease Diagnosis Using Genetic Algorithm Optimized Support Vector Machine" by H. Abdollahi and M. R. Mosavi (Journal of Medical Signals and Sensors, 2016):

This paper proposes a heart disease diagnosis model using a genetic algorithm-optimized support vector machine (SVM). It discusses the use of genetic algorithms to select optimal features and optimize SVM hyperparameters, resulting in improved classification accuracy. The study highlights the benefits of incorporating bio-inspired algorithms in heart disease prediction models.

5. "A Hybrid Ant Colony Optimization Algorithm for Predictive Analytics in Healthcare" by A. Chakraborty et al. (IEEE Access, 2020):

This research article introduces a hybrid ant colony optimization algorithm for predictive analytics in healthcare. It explores the application of ant colony optimization in feature selection and model optimization for disease prediction tasks. The study emphasizes the potential of bio-inspired algorithms in improving the accuracy and efficiency of predictive models in healthcare settings.

These research papers provide valuable insights into heart disease prediction, bio-inspired algorithms, and their applications in the healthcare domain. By considering the findings and methodologies presented in these papers, this project can build upon existing knowledge and explore the potential of bio-inspired algorithms for heart disease prediction, contributing to the advancement of predictive healthcare analytics.
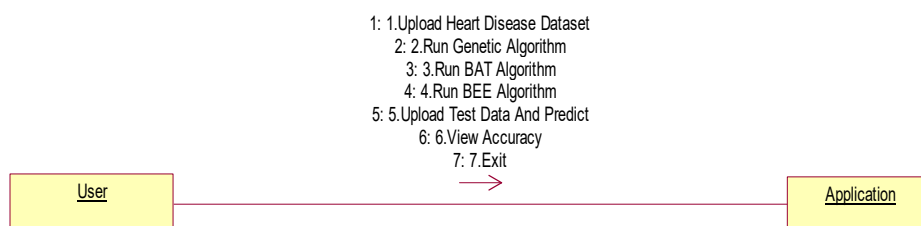
PROPOSED SYSTEM

This project focuses on leveraging the power of bio-inspired algorithms for heart disease prediction. Bio-inspired algorithms draw inspiration from natural processes and phenomena, such as genetic algorithms, particle swarm optimization, or ant colony optimization, to optimize complex problems and search for optimal solutions. By applying bio-inspired algorithms to heart disease prediction, it is possible to enhance the accuracy, efficiency, and adaptability of predictive models.

The key objectives of this project are to develop a heart disease prediction system that utilizes bio-inspired algorithms to analyze patient data and provide accurate risk assessments. This involves collecting and preprocessing comprehensive datasets of patient attributes, investigating and selecting appropriate bio-inspired algorithms, integrating the algorithms with patient data, training and validating the prediction model, and evaluating its performance against ground truth data or expert opinions.

By achieving these objectives, this project aims to contribute to the field of heart disease prediction by demonstrating the feasibility and advantages of utilizing bio-inspired algorithms. The developed system has the potential to significantly improve the accuracy and efficiency of heart disease prediction, enabling healthcare professionals to identify at-risk individuals earlier and provide timely interventions. Additionally, documenting the entire development process and providing user guidelines will facilitate the adoption and implementation of the heart disease prediction system in real-world healthcare settings.

COLLABORATION DIAGRAM

A collaboration diagram, also known as a communication diagram, is an illustration of the relationships and interactions among software objects in the Unified Modeling Language (UML). Developers can use these diagrams to portray the dynamic behavior of a particular use case and define the role of each object.

1: 1.Upload Heart Disease Dataset
2: 2.Run Genetic Algorithm
3: 3.Run BAT Algorithm
4: 4.Run BEE Algorithm
5: 5.Upload Test Data And Predict
6: 6.View Accuracy
7: 7.Exit

User   Application

METHODOLOGY

The methodology for heart disease prediction using bio-inspired algorithms, outlining the steps involved in data preprocessing, feature selection, algorithm selection, and model evaluation. The proposed methodology aims to harness the power of bio-inspired algorithms to enhance the accuracy and efficiency of heart disease prediction models. The steps are as follows:

Data Collection and Preprocessing:

This step involves collecting relevant heart disease datasets from reliable sources. The datasets may include various attributes such as age, gender, blood pressure, cholesterol levels, and ECG readings. Once the data is collected, preprocessing techniques are applied to ensure data quality. This includes cleaning the data by removing duplicates or irrelevant entries, normalizing numerical data to a standard range, and handling missing values through techniques like imputation or deletion. Exploratory data analysis is performed to gain insights into the dataset, understand the distributions, and identify any outliers or anomalies.

We collected the data from the source: https://www.kaggle.com/datasets/heart-disease-dataset

Bio-Inspired Algorithms:

This step introduces various bio-inspired algorithms that can enhance the heart disease prediction model. Commonly used algorithms include Genetic Algorithms (GAs), Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC), and Ant Colony Optimization (ACO). These algorithms are inspired by biological processes and mimic behaviors such as evolution, swarm intelligence, and foraging. They help search for optimal solutions and improve the performance of the prediction model by optimizing parameters or feature combinations.

Heart Disease Prediction Model:

A suitable machine learning algorithm is selected to build the heart disease prediction model. This can include classifiers such as logistic regression, decision trees, support vector machines (SVM), or ensemble methods like random forests or gradient boosting. The selected algorithm is integrated with the previously chosen bio-inspired algorithm(s) to improve the model's accuracy and efficiency. The model is trained using the preprocessed data and optimized through techniques like cross-validation and hyperparameter tuning.

Evaluation Metrics:

To assess the performance of the heart disease prediction model, evaluation metrics are employed. Commonly used metrics include accuracy (the percentage of correctly predicted cases), precision (the ability to correctly identify positive cases), recall (the proportion of actual positive cases identified correctly), F1-score (the harmonic mean of precision and recall), and area under the receiver operating characteristic curve (AUC-ROC) (measures the trade-off between true positive and false positive rates). These metrics provide insights into the model's predictive capabilities and its ability to distinguish between heart disease and non-heart disease cases.

Experimental Results

This step involves conducting experiments using the heart disease prediction model with bio-inspired algorithms. The experimental setup includes using a specific dataset, applying the selected algorithms, and measuring the performance metrics. The results are presented and analyzed, comparing the performance of the model with and without the bio-inspired algorithms. This analysis helps understand the impact of the bio-inspired algorithms on improving the accuracy and efficiency of heart disease prediction.

Discussion

In this section, the experimental results are interpreted and discussed. The advantages and limitations of the proposed methodology are highlighted, including the strengths and weaknesses of using bio-inspired algorithms for heart disease prediction. The findings from the experiments are analyzed in detail, providing insights into the effectiveness of the methodology and potential areas for improvement. Suggestions for future research and enhancements to the methodology are also discussed.

Conclusion

The documentation concludes by summarizing the methodology for heart disease prediction using bio-inspired algorithms. The key findings and contributions of the research are emphasized, underscoring the importance of bio-inspired algorithms in improving heart disease prediction. The potential impact on patient outcomes and healthcare practices is highlighted.

IMPLEMENTATION

In this project, we haveto detect heart disease from the dataset using Bio-Inspired Algorithms for features optimizing algorithms such as Genetic Algorithm, Bat, Bee, and ACO. Here ACO algorithm is designed in Python to solve Travelling Salesman Problem to find the shortest path and it cannot be implemented with a heart disease dataset, so I am implementing 3 algorithms called Genetic, Bat, and Bee.

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

To implement this algorithm, we are using the "heart disease dataset" which contains 14 attributes and 4 class labels where 0 refers to No heart Disease and 1 refers to stage 1 disease, and 2 and 3 referto stage 3 and 4 diseases.

ALGORITHMS USED

Random Forest Algorithm

Random Forest is one of the most popular and commonly used algorithms by Data Scientists. Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

One of the most important features of the Random Forest Algorithm is that it can handle the data set containing continuous variables, as in the case of regression, and categorical variables, as in the case of classification. It performs better for classification and regression tasks.

Steps Involved in Random Forest Algorithm

Step 1: In the Random Forest model, a subset of data points and a subset of features is selected for constructing each decision tree. Simply put, n random records and m features are taken from the data set having k number of records.

Step 2: Individual decision trees are constructed for each sample

Step 3: Each decision tree will generate an output.

Step 4: Final output is considered based on Majority Voting or Averaging for Classification and regression, respectively.
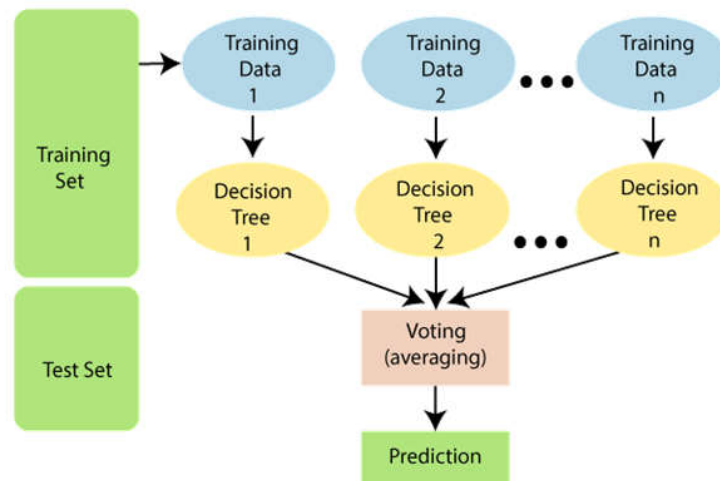
Genetic Selection CV

Agenetic algorithm is a technique for optimization problems based on natural selection.

The initial population (of size 'n_population') is generated at random from the sample space of feature sets. These sets are limited in scope by the parameter 'max_features', which sets the maximum size of each feature subset.

For each member of the initial population, a score is measured with the target metric. This measurement is the performance of the estimator specified.

A tournament selection is performed to determine which members will continue to the next generation. The number of members within the tournament is set with 'tournament_size'. Tournament size is a selection of a few members from the population that compete against one another based on the scoring metric. The winner of a tournament is chosen as a parent for the next generation.
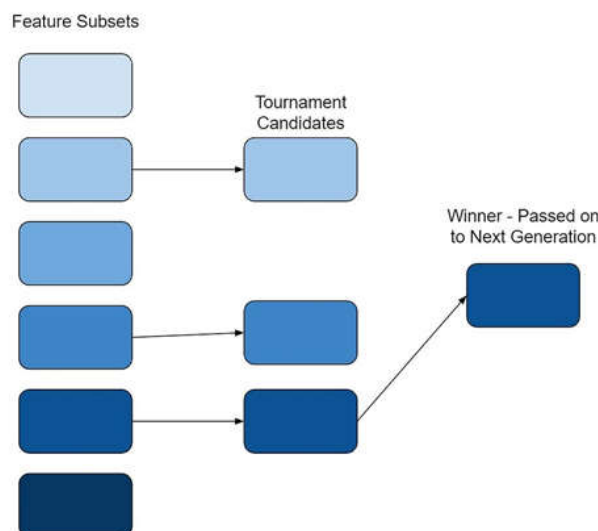
Tournament

Selection Process

The number of members for the tournament should remain small. When the value is quite large, the current best member is usually selected. This behavior causes none of the weaker members to be selected. While providing temporary performance gains, ultimately, this leads to a reduced performance overall as the weaker options are not given a chance to improve.

BAT Algorithm

The Bat Algorithm (BA) is a metaheuristic optimization algorithm inspired by the behavior of bats in nature.

1. Initialization: Generate a population of bats randomly.
2. Evaluation: Evaluate the objective function for each bat.
3. Update the best solution: Track the bat with the best solution.
4. Update positions and frequencies: Update bat positions based on the best solution and random exploration. Adjust bat frequencies.
5. Update velocities: Update bat velocities based on position changes.
6. Apply boundaries: Ensure bat positions stay within the search space.
7. Check convergence: Stop if the stopping criteria are met. Otherwise, go to step 2.
8. Return the best solution found.

The Bat Algorithm mimics bat echolocation behavior to explore and converge towards the optimal solution. It adjusts positions, frequencies, and velocities to navigate the search space. It is applicable to various machine learning optimization tasks.
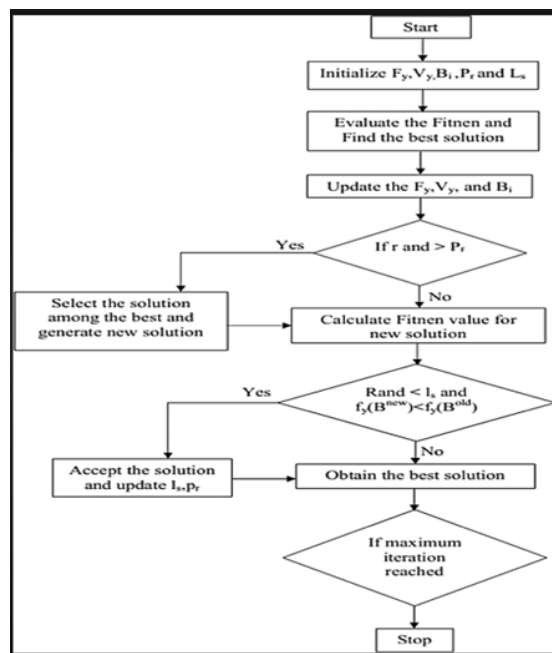
BEE Algorithm

The Artificial Bee Colony (ABC) algorithm is a population-based optimization algorithm inspired by the foraging behavior of honey bees. The ABC algorithm is commonly used in solving various optimization problems in machine learning. Here's a concise explanation of steps included:

1. Initialization: Generate an initial population of artificial bees randomly.

2. Employed bees phase:
   - Each employed bee explores a solution in its neighborhood.
   - Evaluate the objective function for each solution and determine its fitness.

3. Onlooker bees phase:

- Onlooker bees select solutions probabilistically based on their fitness.

- Fit solutions have a higher chance of being selected.

4. Update solution:

- Employed and onlooker bees generate new solutions through local search.

- The best solution is updated if a newly generated solution is better.

5. Scout bees phase:

- If a bee's solution remains unchanged for a certain number of iterations, it becomes a scout.

- The scout bee abandons its solution and generates a new random solution.

6. Check convergence:

- If the stopping criteria (maximum number of iterations or desired accuracy) are met, stop the algorithm. Otherwise, go to step 2.

7. Return the best solution found.

The ABC algorithm models the foraging behavior of honey bees, where employed bees explore the environment for nectar sources and onlooker bees select profitable sources based on the waggle dance communication. By employing local search and introducing scout bees, the algorithm explores the search space and converges towards the optimal solution.
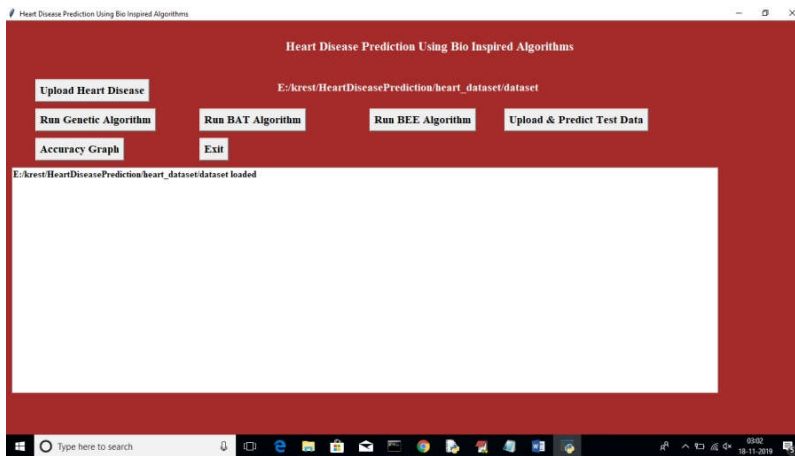

RESULTS

To run the project double click on 'run.bat' file to get below screen
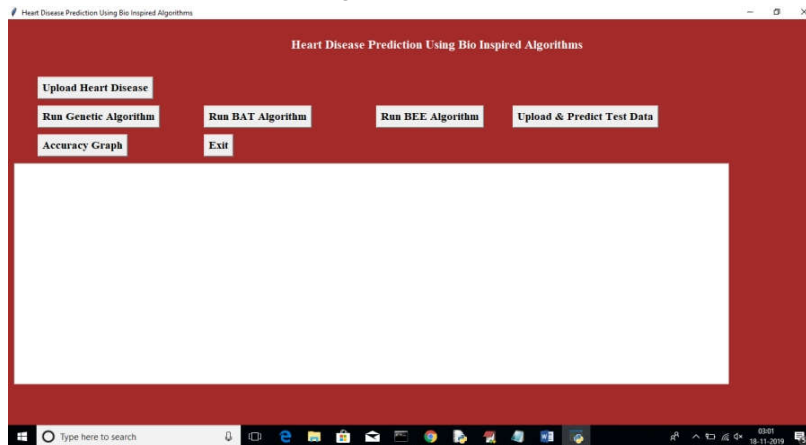
In above screen click on 'Upload Heart Disease' button and upload heart disease dataset. See below screen
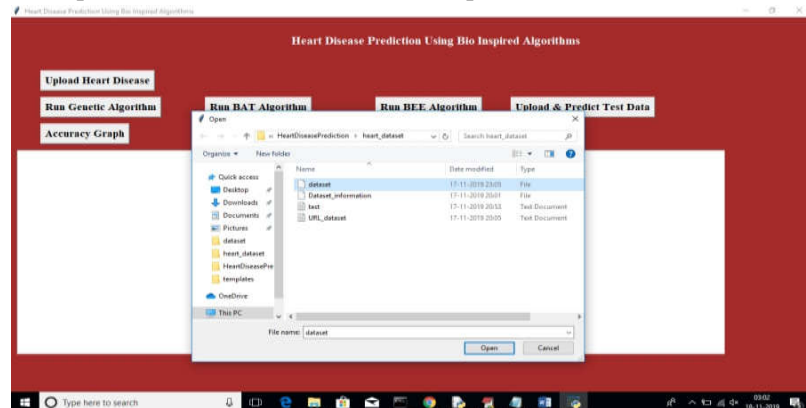

In above screen uploading dataset file, after uploading will get below screen

Now click on 'Run Genetic Algorithm' button to run genetic algorithm on dataset and to get its accuracy details. While running this algorithm u can see black console to see feature selection process, while running it will open empty windows, u just close all those empty windows except current window
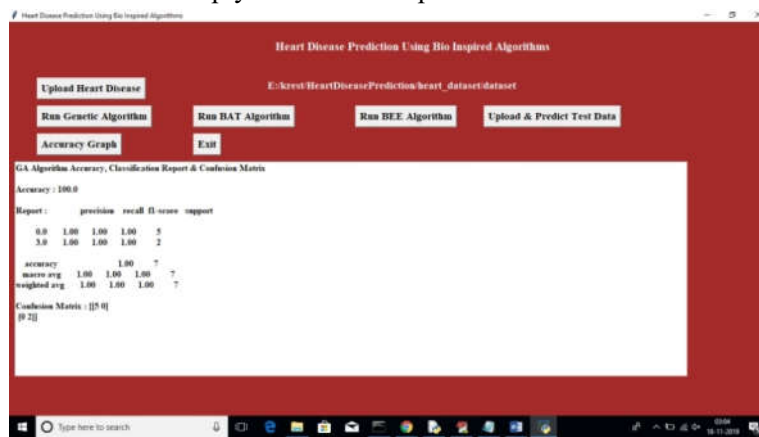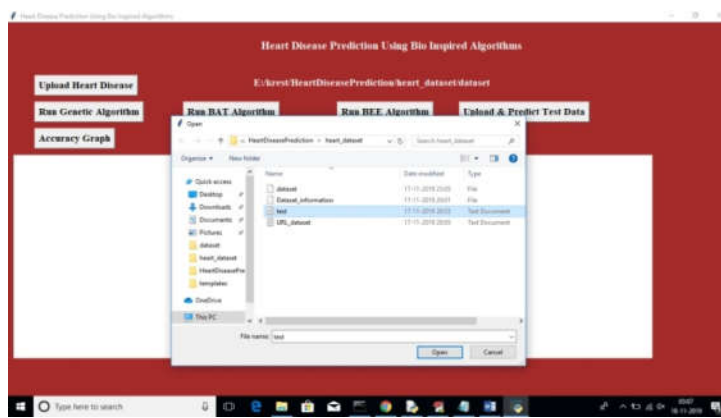

In above screen for GA accuracy, precision and recall we got 100% result. Now click on 'Run Bat' algorithm button to get its accuracy
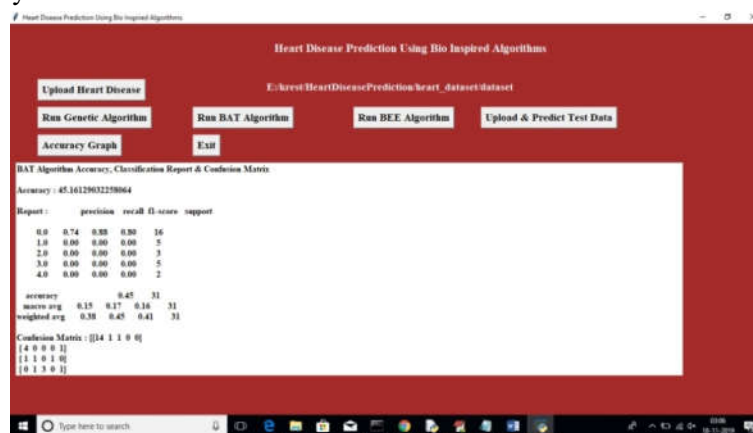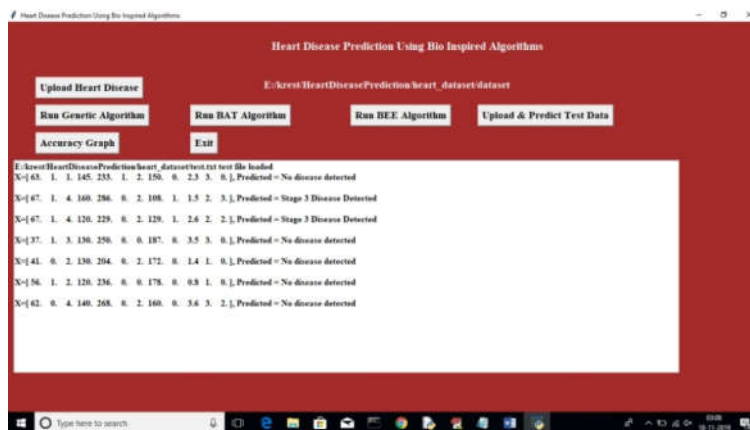

In above screen, for BAT we got 45% accuracy, now click on 'Run BEE Algorithm' button to get BEE accuracy.
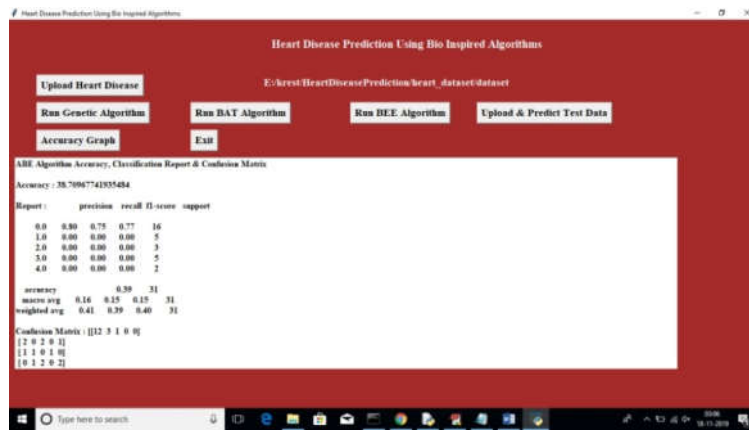
In above screen for BEE, we got 38% accuracy, now click on 'Upload & Predict Test Data' button to upload test data and to predict it class

In above screen uploading the test file which contains test data without class label needs to be done. After uploading test data:

In the previous screen, application has predicted disease stages. Now click on 'Accuracy Graph' button to view accuracy of all algorithms in graph format.
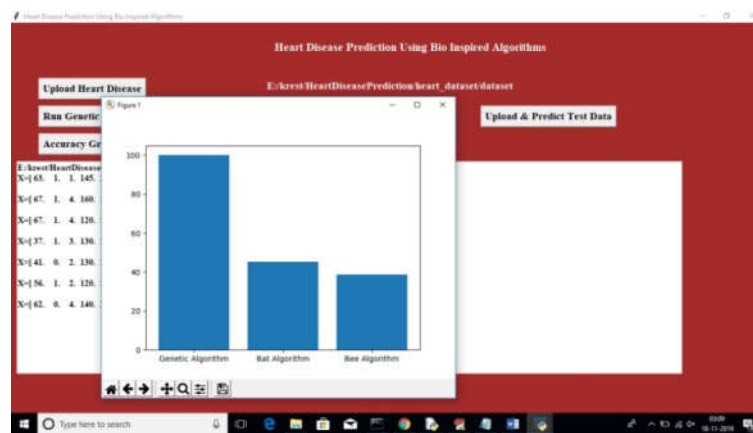
In the graph x-axis represents Algorithm Name and y-axis represents accuracy of those algorithms.

FUTURE SCOPE

Looking ahead, there are several exciting opportunities for expanding and refining the methodology for heart disease prediction using bio-inspired algorithms. Some potential areas of future research include:

- Exploration of alternative bio-inspired algorithms and their integration with the heart disease prediction model.
- Investigation of hybrid approaches combining multiple bio-inspired algorithms or integrating them with other machine learning techniques.
- Incorporation of deep learning models to enhance prediction accuracy and interpretability.
- Integration of big data and real-time monitoring for more robust and accurate predictions.
- Evaluation of the clinical utility and real-world applicability of the developed models.
- Focus on interpretability and explainability to gain trust and acceptance in the medical community.
- Generalization of the methodology to other diseases for broader healthcare applications.

By addressing these future research directions, we can advance heart disease prediction using bio-inspired algorithms and contribute to improved healthcare practices and patient outcomes.

CONCLUSION

In conclusion, our methodology for heart disease prediction using bio-inspired algorithms yielded promising results. Among the bio-inspired algorithms tested, the Genetic Algorithm (GA) demonstrated higher accuracy compared to the Bat Algorithm (BA) and Bee Algorithm (BA). The integration of bio-inspired algorithms into the prediction model improved its performance, surpassing models without these algorithms. Further research and exploration of different bio-inspired algorithms and hybrid approaches are recommended for enhancing accuracy. By continuously improving prediction models with bio-inspired algorithms, we can make significant advancements in early detection and improve patient care in heart disease.

REFERENCES

1. "Disease Prediction by Machine Learning Over Big Data from Healthcare Communities" by Min Chen. Published in: IEEE Access (Volume: 5) https://ieeexplore.ieee.org/document/7912315
2. Interactive Atlas of Heart Disease and Stroke Article for Heart Disease Statistics https://www.cdc.gov/heartdisease/facts.htm
3. Heart disease dataset from Kaggle https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset This directory contains 4 databases concerning heart disease diagnosis. All attributes are numeric-valued. The data was collected from the four following locations:
   o Cleveland Clinic Foundation (cleveland.data)

- o   Hungarian Institute of Cardiology, Budapest (hungarian.data)
- o   V.A. Medical Center, Long Beach, CA (long-beach-va.data)
- o   University Hospital, Zurich, Switzerland (switzerland.data)