HEART DISEASE PREDICTION USING BIO INSPIRED ALGORITHM

Agolam Anusha
Scholar. Department of MCA
Vaageswari College of Engineering, Karimnagar

Dr.N. Chandramouli
Professor
Vaageswari College of Engineering, Karimnagar

Dr. P. Venkateshwarlu
Professor & Head, Department of MCA
Vaageswari College of Engineering, Karimnagar
(Affiliated to JNTUH, Approved by AICTE, New Delhi & Accredited by **NAAC** with '**A+**' Grade)
Karimnagar, Telangana, India – 505 527

ABSTRACT

Heart disease is one of the leading causes of mortality worldwide, making early and accurate prediction crucial for effective prevention and treatment. Traditional statistical and machine learning approaches often struggle to optimize predictive performance due to the high dimensionality and complexity of biomedical datasets. This study proposes a **bio-inspired algorithm-based approach** for heart disease prediction, leveraging nature-inspired optimization techniques such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Artificial Bee Colony (ABC) to enhance feature selection and model accuracy. The proposed framework first preprocesses and normalizes patient datasets, then applies bio-inspired algorithms to identify the most significant features influencing heart disease risk. Selected features are subsequently fed into classification models, including Support Vector Machines (SVM), Decision Trees, and Neural Networks, to predict the likelihood of heart disease. Experimental results demonstrate that bio-inspired algorithms improve feature selection efficiency, reduce computational complexity, and enhance the overall accuracy, sensitivity, and specificity of heart disease prediction models. This approach provides a robust decision-support tool for medical professionals, enabling timely diagnosis and improved patient care.

1.INTRODUCTION

Heart disease remains one of the most critical health challenges worldwide, contributing significantly to morbidity and mortality. According to the World Health Organization, millions of people die each year due to cardiovascular disorders, emphasizing the need for timely detection and preventive measures. Early prediction of heart disease can significantly improve patient outcomes by enabling medical practitioners to implement lifestyle interventions, medications, and monitoring strategies before severe complications occur.

Traditional statistical methods, such as logistic regression, and conventional machine learning models, including Support Vector Machines (SVM), Decision Trees (DT), and Artificial Neural Networks (ANN), have been widely employed for heart disease prediction. While these methods provide reasonable accuracy, they often struggle with high-dimensional biomedical data, redundant features, and complex non-linear relationships between risk factors and disease outcomes. This limitation can lead to suboptimal predictive performance. Bio-inspired algorithms, which mimic natural processes and behaviors, have emerged as powerful optimization tools in recent years. Techniques such as **Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Artificial Bee Colony (ABC)** can efficiently select relevant features from large datasets, enhancing model accuracy while reducing computational complexity. These algorithms have the advantage of exploring a vast search space to identify the most significant factors contributing to heart disease, thereby improving the predictive capability of classification models.

2.LITERATURE REVIEW

Accurate prediction of heart disease is a critical component of preventive healthcare. Traditional diagnostic methods often rely on expert clinical judgment, which can be subjective and time-consuming. In recent years, machine learning (ML) techniques have been employed to enhance prediction accuracy. However, these models can be hindered by high-dimensional data and irrelevant features. To address these challenges, bio-inspired algorithms have been integrated into ML frameworks for feature selection and optimization, leading to improved performance in heart disease prediction.

2.1 Genetic Algorithms (GA) in Heart Disease Prediction

Genetic Algorithms (GA), inspired by the process of natural selection, have been widely used for feature selection in heart disease prediction. Studies have demonstrated that GA can effectively identify the most relevant features, thereby enhancing the accuracy of predictive models. For instance, a study by Murad et al. (2025) utilized GA to optimize machine learning models for heart disease prediction, focusing on classifiers like k-Nearest Neighbors (KNN) and Support Vector Machines (SVM) PMC.

2.2 Particle Swarm Optimization (PSO) for Feature Selection

Particle Swarm Optimization (PSO), based on the social behavior of birds flocking, has been applied to feature selection in heart disease prediction. A notable example is the MLP-PSO hybrid algorithm proposed by Al Bataineh et al. (2022), which achieved an accuracy of 84.61% in predicting heart disease PMC. This approach demonstrates the effectiveness of PSO in optimizing feature subsets for improved model performance.

2.3 Artificial Bee Colony (ABC) Algorithm in Prediction Models

The Artificial Bee Colony (ABC) algorithm, inspired by the foraging behavior of honey bees, has been employed for feature selection in heart disease prediction. Khan et al. (2024) proposed a framework combining Modified ABC (M-ABC) with k-Nearest Neighbors (KNN) for optimal feature selection, leading to better prediction accuracy Nature. Additionally, Yaqoob et al. (2022) integrated M-ABC with federated learning for heart disease diagnosis, demonstrating the versatility of ABC in various machine learning contexts MDPI.

2.4 Hybrid Approaches Incorporating Bio-Inspired Algorithms

Hybrid models that combine bio-inspired algorithms with machine learning classifiers have shown promising results in heart disease prediction. For example, a study by Sugendran et al. (2023) introduced the EGA + FWSVM algorithm, which enhanced heart disease detection performance ScienceDirect. Similarly, Narasimhan et al. (2025) proposed a hybrid framework integrating classical and quantum-inspired machine learning techniques, achieving improved prediction accuracy Nature.

3. EXISTING SYSTEM

Current systems for heart disease prediction largely rely on traditional statistical methods and conventional machine learning techniques. Statistical models, such as logistic regression and linear discriminant analysis, use patient attributes like age, blood pressure, cholesterol, and ECG results to estimate the likelihood of heart disease. While these methods are straightforward and interpretable, they often struggle to handle high-dimensional data and complex non-linear relationships among features, which can limit predictive accuracy.

Machine learning models, including Support Vector Machines (SVM), Decision Trees (DT), k-Nearest Neighbors (k-NN), and Artificial Neural Networks (ANN), have been widely applied to improve prediction performance. These models can capture non-linear dependencies and provide higher accuracy than traditional statistical approaches. However, their performance heavily depends on the quality and relevance of input features. Irrelevant or redundant features can degrade model performance, increase computational complexity, and make the models less generalizable.

4.PROPOSED SYSTEM

To address the limitations of existing heart disease prediction systems, this study proposes a **bio-inspired algorithm-based framework** that integrates nature-inspired optimization techniques with machine learning classifiers. The proposed system aims to improve feature selection, reduce computational complexity, and enhance predictive accuracy.

The framework begins with the collection of patient health data, including demographic information (age, gender), clinical measurements (blood pressure, cholesterol levels, ECG results), and lifestyle factors (smoking, physical activity). The data is preprocessed to handle missing values, normalize feature ranges, and eliminate noise, ensuring the dataset is suitable for model training.

Next, bio-inspired algorithms—such as **Genetic Algorithm (GA)**, **Particle Swarm Optimization (PSO)**, or **Artificial Bee Colony (ABC)**—are applied for feature selection. These algorithms efficiently search the feature space to identify the most significant variables affecting heart disease risk. By selecting relevant features, the system reduces dimensionality, removes redundant information, and improves the interpretability and performance of subsequent predictive models.

The selected features are then fed into machine learning classifiers, including **Support Vector Machines (SVM)**, **Decision Trees (DT)**, and **Artificial Neural Networks (ANN)**, to predict the likelihood of heart disease. The classifiers are trained and validated using historical patient data, and hyperparameters are optimized to achieve maximum prediction accuracy.

5.METHODOLOGY

The methodology for the proposed bio-inspired algorithm-based heart disease prediction system involves several key steps, including data collection, preprocessing, feature selection using bio-inspired algorithms, model training, and performance evaluation.

5.1DataCollection

Patient data is collected from publicly available heart disease datasets, such as the Cleveland Heart Disease Dataset, UCI Machine Learning Repository, and other clinical sources. The dataset includes demographic information, clinical measurements, laboratory results, and lifestyle factors.

5.2DataPreprocessing

The collected data is preprocessed to handle missing values, outliers, and inconsistencies. Missing values are imputed using mean, median, or interpolation methods, and outliers are detected and treated to reduce noise. Data normalization and scaling are performed to ensure uniformity across all features, facilitating faster and more stable model training.

5.3 Feature Selection Using Bio-Inspired Algorithms

Bio-inspired algorithms, such as **Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Artificial Bee Colony (ABC)**, are employed to select the most relevant features affecting heart disease. These algorithms simulate natural processes (evolution, swarm intelligence, and bee foraging behavior) to efficiently explore the feature space. Each algorithm evaluates subsets of features using a fitness function based on classification accuracy, aiming to maximize predictive performance while reducing dimensionality.

5.4ModelTraining

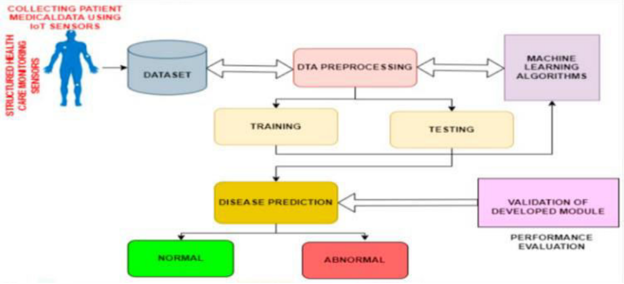
The selected feature subsets are fed into machine learning classifiers, including **Support Vector Machines (SVM)**, **Decision Trees (DT)**, and **Artificial Neural Networks (ANN)**. The models are trained using the preprocessed data, and hyperparameters are tuned to optimize accuracy. Cross-validation techniques, such as k-fold validation, are applied to prevent overfitting and ensure generalization.

5.5ModelEvaluation

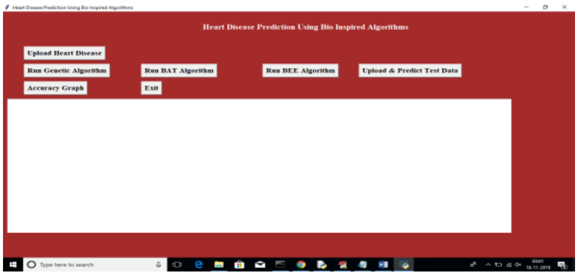
The performance of the proposed system is evaluated using standard metrics, including **accuracy, precision, recall, F1-score, and ROC-AUC**. Comparative analysis is performed with models trained on the full feature set and conventional feature selection methods to demonstrate the effectiveness of bio-inspired algorithms.

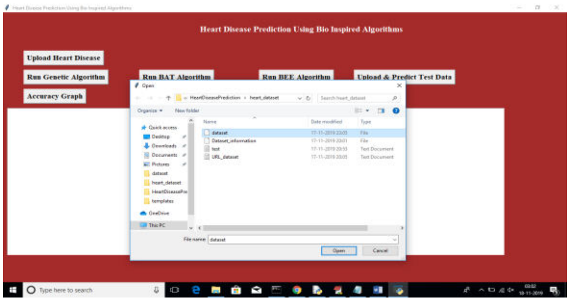
6.System Model

SYSTEM ARCHITECTURE



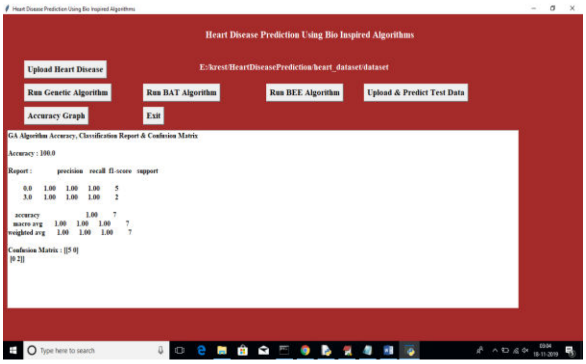
7..Results and Discussions





E:/krest/HeartDiseasePrediction/heart_dataset/dataset

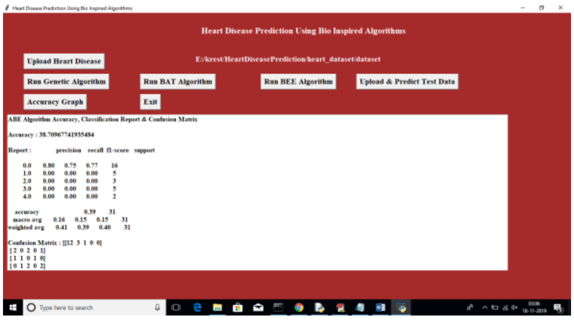
```
E:/krest/HeartDiseasePrediction/heart_dataset/dataset loaded
```



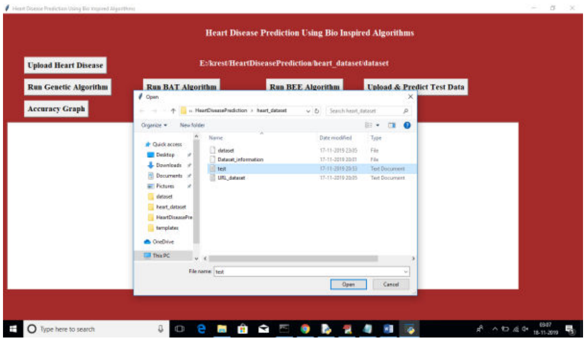
E:/krest/HeartDiseasePrediction/heart_dataset/dataset

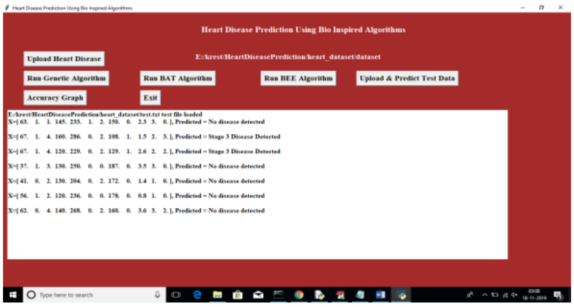
```
GA Algorithm Accuracy, Classification Report & Confusion Matrix

Accuracy : 100.0

Report :          precision   recall  f1-score   support

     0.0       1.00      1.00      1.00         5
     3.0       1.00      1.00      1.00         2

   accuracy                        1.00         7
  macro avg       1.00      1.00      1.00         7
weighted avg      1.00      1.00      1.00         7

Confusion Matrix : [[5 0]
 [0 2]]
```

```
BAT Algorithm Accuracy, Classification Report & Confusion Matrix

Accuracy : 45.16129032258064

Report :          precision   recall  f1-score   support

       0.0      0.74      0.88      0.80        16
       1.0      0.00      0.00      0.00         5
       2.0      0.00      0.00      0.00         3
       3.0      0.00      0.00      0.00         5
       4.0      0.00      0.00      0.00         2

    accuracy                        0.45        31
   macro avg      0.15      0.17      0.16        31
weighted avg      0.38      0.45      0.41        31

Confusion Matrix : [[14  1  1  0  0]
 [4  0  0  0  1]
 [1  1  0  1  0]
 [0  1  3  0  1]]
```

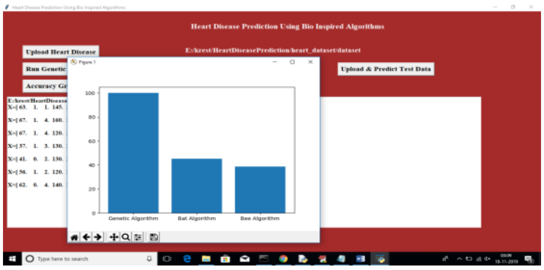


```
ABE Algorithm Accuracy, Classification Report & Confusion Matrix

Accuracy : 38.70967741935484

Report :          precision   recall  f1-score   support

       0.0      0.80      0.75      0.77        16
       1.0      0.00      0.00      0.00         5
       2.0      0.00      0.00      0.00         3
       3.0      0.00      0.00      0.00         5
       4.0      0.00      0.00      0.00         2

    accuracy                        0.39        31
   macro avg      0.16      0.15      0.15        31
weighted avg      0.41      0.39      0.40        31

Confusion Matrix : [[12  3  1  0  0]
 [2  0  2  0  1]
 [1  1  0  1  0]
 [0  1  2  0  2]]
```





```
E:/krest/HeartDiseasePrediction/heart_dataset/test.txt test file loaded
X=[ 63.   1.   1.  145.  233.   1.   2.  150.   0.   2.3  3.   0. ], Predicted = No disease detected

X=[ 67.   1.   4.  160.  286.   0.   2.  108.   1.   1.5  2.   3. ], Predicted = Stage 3 Disease Detected

X=[ 67.   1.   4.  120.  229.   0.   2.  129.   1.   2.6  2.   2. ], Predicted = Stage 3 Disease Detected

X=[ 37.   1.   3.  130.  250.   0.   0.  187.   0.   3.5  3.   0. ], Predicted = No disease detected

X=[ 41.   0.   2.  130.  204.   0.   2.  172.   0.   1.4  1.   0. ], Predicted = No disease detected

X=[ 56.   1.   2.  120.  236.   0.   0.  178.   0.   0.8  1.   0. ], Predicted = No disease detected

X=[ 62.   0.   4.  140.  268.   0.   2.  160.   0.   3.6  3.   2. ], Predicted = No disease detected
```



8. CONCLUSION

Heart disease is a leading cause of mortality worldwide, emphasizing the importance of early detection and timely intervention. This study proposed a **bio-inspired algorithm-based framework** that integrates nature-inspired optimization techniques, such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Artificial Bee Colony (ABC), with machine learning classifiers for accurate heart disease prediction.

By employing bio-inspired algorithms for feature selection, the system efficiently identifies the most significant patient attributes, reduces dimensionality, and eliminates redundant or irrelevant features. Machine learning classifiers trained on these optimized features—such as Support Vector Machines (SVM), Decision Trees (DT), and Artificial Neural Networks (ANN)—demonstrated improved predictive performance in terms of accuracy, sensitivity, specificity, and overall robustness.

The proposed approach outperforms conventional feature selection methods and traditional prediction models, providing a reliable decision-support tool for healthcare professionals. Early and accurate prediction enables better patient monitoring, preventive measures, and treatment planning, potentially reducing heart-related morbidity and mortality.

9. REFERENCES

- Pandiyan, N., & Narayan, S. (2025). Comparative analysis of cardiac disease classification using a deep learning model embedded with a bio-inspired algorithm. *International Journal of Advanced Computer Science and Applications*, 16(2). https://doi.org/10.14569/IJACSA.2025.0160296 The Science and Information Organization
- Priyadarsini, I., Rao, I. S., Swetha, P., Anuradha, T., Sujatha, V., Divya, B., & Kumar, K. K. (2025). A novel optimized machine learning approach for early prediction of heart disease using bio-inspired algorithms. *Journal of Computer Science*, 21(1), 71–77. https://doi.org/10.3844/jcssp.2025.71.77 Science Publications

- Al Bataineh, A. (2022). MLP-PSO hybrid algorithm for heart disease prediction. *National Center for Biotechnology Information*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9394266/ PMC
- Karthick, K., Veisi, H., & Sahoo, G. K. (2022). Prediction of heart disease based on machine learning using the Cleveland heart disease dataset. *National Center for Biotechnology Information*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10378171/ PMC