Effectiveness of Nature-Inspired Algorithms for the Health Care Systems

Manali Singh[1], Dr. Vani Agrawal[2], Dr. Jitendra Singh Kushwah[3]
[1]*Research Scholar, ITM University, Gwalior-MP- India,*
[2] *Department of Computer Science, ITM University, Gwalior-MP-India*
[3]*Department of Information Technology, Institute of Technology and Management, Gwalior-MP-India.*

Abstract: This study examines the feasibility of using machine learning models to diagnose heart disease, using a dataset with 14 clinical features. Support Vector Machine (92.68%) and K-Nearest Neighbours (86.34%) are outperformed by Random Forest, achieving a perfect accuracy (1.0000). Using a Genetic Algorithm, 10 features are chosen, resulting in maintaining 100% accuracy of Random Forest with various complexity reductions. A correlation and feature importance analysis demonstrates the importance of chest pain type and maximum heart rate as predictors. The results show how optimised machine learning models can perform accurate, efficient disease diagnosis and ultimately aid in improved clinical decision making and early disease detection in healthcare systems.

Keywords: Heart disease, machine learning, Random Forest, genetic algorithm, feature selection, disease diagnosis, clinical data, support vector machine, k-nearest neighbours, model optimisation

I. INTRODUCTION

A large part of global deaths is due to heart disease which is still one of the most common causes of death around the world. To prevent problems and improve results, correct and swift diagnosis is highly necessary. For many years, physicians made medical diagnoses through their experience and by running a variety of tests [1]. Nowadays, access to electronic healthcare data has allowed hospitals to explore using data science approaches such as machine learning (ML) in making better and faster healthcare decisions.

The use of machine learning makes it possible to break down and study large volumes of data, find complex trends and come up with precise forecasts. Both Random Forest and algorithms such as SVM and KNN are proving useful when it comes to making medical diagnoses. Still, these models can have problems with picking out important features, making their design complicated and generalising to new situations [2]. In a lot of situations, various clinical features do not all add the same value to a model. So, focusing on choosing the right features improves the performance of your model and cuts down on your computer's work.

Nature-based algorithms, including Genetic Algorithms (GA), solve issues of optimal design in a way similar to natural evolution. By using GA in machine learning, the best combination of features for correct classification which helps improve model results without making the model slower.

The goal of this research is to see if Genetic Algorithms increase the effectiveness of machine learning when it comes to diagnosing heart disease from clinical details.

This objective is fulfilled by the following targets:

1. To train and compare the performance of various machine learning algorithms—Random Forest, Support Vector Machine, and K-Nearest Neighbours—on a dataset of heart disease.
2. To use a Genetic Algorithm for feature selection at optimal levels and measure its effect on model accuracy, complexity, and efficiency.
3. To examine the statistical relationships and feature importance in the dataset to determine the significant clinical markers of heart disease.
4. To compare the diagnostic accuracy of baseline models with all features versus GA-optimised models in order to evaluate the merit of feature selection in clinical prediction.

By answering these goals, the study is set to illustrate the promise of combining machine learning with nature-inspired optimisation to enable early and accurate diagnosis in healthcare systems.

II. LITARETURE REVIEW

Using machine learning (ML) and nature-inspired algorithms for healthcare, specifically for predicting diseases and making clinical choices, has become much more common. In this part, the focus is on recent studies in four main areas: (1) ML models for identifying heart disease, (2) selecting useful inputs to

grow model quality, (3) use of GA in healthcare analysis and (4) adding the GA approach to ML for heart disease prediction systems.

1. Machine Learning Models for Heart Disease Diagnosis

During the last decade, machine learning has transformed how diseases are predicted because its models study huge datasets and find patterns that regular statistical approaches overlook. Many studies have been done that use ML to help predict heart disease, as cardiovascular issues have a big impact on worldwide health. Many studies have used Decision Trees, Random Forest, Support Vector Machine (SVM) and K-Nearest Neighbours (KNN) to classify patients using details about their health and background. [3]

By using Random Forest supervised learning on heart disease patients, Dey et al. (2020) achieved a success rate over 89%. Similar to before, SVM was used with clinical datasets by Johnson and Kumar, demonstrating excellent performance when there is some noise and maximum dimensionality of 100, achieving 90% accuracy. Although KNN is a simple idea to understand, it has shown good performance, according to Sharma et al. (2022), with 85% accuracy in determining if patients have heart disease [4].

While these results are encouraging, all models have their own drawbacks. It gives better and more reliable results, but it is not very clear what the model is doing. SVM can work well with not many features, but the parameters must be fine-tuned. KNN relies heavily on meaningless features in data and the distance between values. These limitations make it clear that we need better methods that take accuracy, efficiency and how easily a clinic can understand results into account, resulting in the use of Genetic Algorithms for feature selection and optimization.

2. Role of Feature Selection in Improving Model Performance

In machine learning, feature selection is useful for finding out which features should be used to predict any given outcome. In healthcare datasets, certain variables do not add much to diagnosing; sometimes, they bring noise, duplicate data or prejudice. For these reasons, choosing the right subset of features can improve model accuracy, cut down on computing costs and help make things easier to understand—which is very important in healthcare.

Researchers have repeatedly pointed out the advantages of using feature selection for clinical predictions. According to Patel and Singh (2021), getting rid of unnecessary features made the models more accurate. By analyzing heart disease data, they showed that using few features allowed the Random Forest model to stay accurate and run faster. Similarly, Arora et al. (2020) proved that deleting low-variance features made the models both simpler and performed better when tested in several cross-validation experiments.

Feature selection approaches can be put into three main groups: filter, wrapper and embedded. Filter methods can evaluate feature importance without needing information from the learning process [5]. Wrapper methods, for example RFE, test out different combinations of features using the model's outcome. With embedded methods in decision tree models, the algorithm is designed to handle feature selection.

On the other hand, conventional techniques frequently find it hard to scale and to work well with multiple features at once. They often miss examining a wide variety of features and as a result, their selected feature subsets may not be ideal. In order to overcome these drawbacks, researchers employ Genetic Algorithms (GA) and similar methods because they are better able to adapt and remain flexible for selecting features in tough domains.

3. Application of Genetic Algorithms in Healthcare Data Analysis

Genetic Algorithms (GAs) belong to the evolutionary algorithms that take their ideas from the natural process of selection. Many healthcare applications use them because they help solve complicated optimization problems involving large, complex and nonlinear experiments. In tasks involving finding the best variables for classification, GAs work best by helping choose what variables are included in the final model [6].

In GA-based feature selection, solutions are made up of chromosomes that identify the combinations of features. To find the best feature set, the algorithm undergoes genetic operations called selection, crossover and mutation many times. According to Banerjee and Sinha (2021), GA decreased the number of features in heart disease prediction from 13 to 8, but accomplished this with a Random Forest classifier that was very accurate.

GAs are designed to collaborate with other classifiers, including SVM, KNN and neural networks. The researchers found that using GA with SVM for detecting breast cancer gave a better result and lower training time than just using SVM alone (Lin et al., 2019). Using GA to control model inputs, Khan and Raza 2022 found their models performed better on unseen data and were more compact than before [7].

Yet, GAs may be computationally complex, particularly if data has many factors. They turn out to be well-suited for use with heart disease datasets containing few features because they are efficient and

give good results. The ability of GAs to adapt, expand and optimize makes them suitable for bettering machine learning tools used in healthcare diagnosis.

4. Integration of ML and GA in Heart Disease Prediction Systems

Heart disease diagnosis has benefited a lot from the blend of Genetic Algorithms and machine learning. Because of this, the system can deal with data smoothly and give precise predictions. We want to use GA to choose important variables and then rely on the accuracy of models like Random Forest, SVM and KNN to identify what patients' outcomes will be.

In 2020, Ramesh and Reddy showed that combining Genetic Algorithm with Random Forest helped to reduce input features and improve accuracy in identifying coronary artery disease. The new model reached 98% accuracy, up from the baseline's success rate using every feature. According to the authors, GA helped to eliminate unnecessary and problematic data from traditional models.

In the same way, Zhang et al. (2021) used GA to choose the best features and afterward applied an SVM classifier on these performances. They found that not only was the model's accuracy better, but it could handle new data more successfully. As a result, using GA to select features could minimize the problem of overfitting in medical ML applications.

Over recent years, scientists have not only focused on being accurate, but have expanded their investigations. In the case of imbalanced data, researchers examine precision, recall, the F1-score and confusion matrices in their performance evaluations [8]. Trained models from GA have maintained even performance in all metrics, making them a more dependable choice for helping doctors in real life situations.

Moreover, these hybrid systems make use of visual tools including feature importance plots and heatmaps to give understandable (XAI) results. This makes clinical environments more transparent which gives practitioners insight into the reasons behind the prediction, thereby encouraging them to rely on such findings.

III. METHODOLOGY

In this investigation, the authors rely on Genetic Algorithms (GA) to choose important features in medical data and apply machine learning (ML) algorithms to diagnose heart diseases. Data collection, preprocessing, choosing features with techniques from GA, building models with various classifiers, testing them and comparing the outcomes are the steps in the methodology. Detailed explanations are given in this section for each step taken to check the reliability and accuracy of the data.

3.1 Research Design

The study includes secondary clinical data that is used for both description and experiments to form predictive models. The descriptive part of the project looks at the statistics of the data and the experimental part tests and compares models with and without Genetic Algorithm-based feature selection. [9] The objective is to evaluate if combining GA with the model increases forecaster accuracy and other metrics.

3.2 Dataset Collection

The study uses the Cleveland Heart Disease Dataset from the UCI Machine Learning Repository which is accepted worldwide. The data set has 303 entries and each one includes important features about patients' clinical conditions and background such as age, sex, chest pain type, blood pressure, cholesterol levels, blood sugar and electrocardiogram results. The classification of the target variable shows 1 for people with heart disease and 0 for those without.

3.3 Data Preprocessing

Raw data might have errors, be incomplete and include useless items that reduce the accuracy of your model. Thus, a well-structured preprocessing pipeline was put in place.

- **Missing Value Handling:** Records containing no information were either handled by replacing with statistical values (mean/mode) or by dropping them when certain specifications were satisfied.
- **Normalization:** To maintain consistency, continuous variables such as cholesterol and maximum heart rate were normalized using Min-Max scaling and brought onto a 0–1 scale.
- **Label Encoding:** Each type of chest pain, slope and thal was turned into a number to fit the requirements of ML algorithms [10].
- **Outlier Detection:** The results were corrected for outliers by using a boxplot and the Z-score approach.

The preprocessing made the data more uniform and prepared it for use in the model.

3.4 Feature Selection Using Genetic Algorithm

With the help of a Genetic Algorithm, it was discovered which features play the biggest role in predicting heart disease accurately. GA achieves this through the use of processes inspired by evolution such as selection, crossover and mutation. To determine the objective function (fitness function), a Random Forest classifier was used to assess classification accuracy during the stage where features were chosen.

GA Configuration:

- Population Size: 50 chromosomes (feature subsets)
- Crossover Probability: 0.8
- Mutation Probability: 0.05
- Selection Method: Tournament Selection
- Generations: 100 iterations
- Fitness Function: Accuracy of classifier using selected feature subset

Each chromosome was a binary string, where one gene represents whether a feature is in (1) or out (0). The process evolved to an optimal subset by assessing classification performance at every generation. The highest-performing feature subset was kept for further modeling.

3.5 Model Development

Three classifiers from machine learning were employed in developing the models, both with and without GA-based feature selection, for comparative analysis:

a. Random Forest (RF):

An ensemble approach that utilizes multiple decision trees trained from bootstrapped samples. RF was selected due to its strength, ability to cope with feature interactions, and lessened overfitting.

Hyperparameters optimized for: Number of trees (n_estimators), maximum depth, and min_samples_split.

b. Support Vector Machine (SVM):

An efficient classifier for linear and non-linearly separable data. The RBF kernel is chosen because it performs well in identifying non-linear patterns in health datasets.

Hyperparameters tuned: Type of kernel, C (regularization parameter), gamma.

c. K-Nearest Neighbour (KNN):

A non-parametric instance-based classifier that predicts a class according to the majority vote by the k-nearest neighbours.

Hyperparameters tuned: Number of neighbours (k), metric used to measure distances (Euclidean).

Each of the models were trained and tested with the original full set as well as the GA-optimized subset to measure the effect of feature selection on the performance of the models.

3.6 Model Evaluation and Metrics

Model performance was assessed with 10-fold cross-validation to ascertain generalizability of results. The dataset was split into 10 subsets; each subset was utilized once as a test set and the rest as the training set. This was done 10 times, and results were averaged.

The following measures were utilized for performance measurement:

- **Accuracy:** Proportion of correctly predicted instances to all instances.
- **Precision:** Ratio of true positives to all positively predicted.
- **Recall (Sensitivity):** Ratio of actual positives properly predicted.
- **F1-Score:** Harmonic mean between precision and recall.
- **Confusion Matrix:** Employed to calculate True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN).

The above measures enabled detailed comparison of models with varying conditions (with/without GA).

3.7 Tools and Technologies Used

The whole experiment was performed within a Jupyter Notebook environment based on the Python programming language. The following libraries and utilities were used:

- **Pandas & NumPy:** Manipulation of data and numerical computations
- **Scikit-learn:** ML models, preprocessing, and metrics
- **Matplotlib & Seaborn:** Visualization of data
- **DEAP (Distributed Evolutionary Algorithms in Python):** Execution of Genetic Algorithms

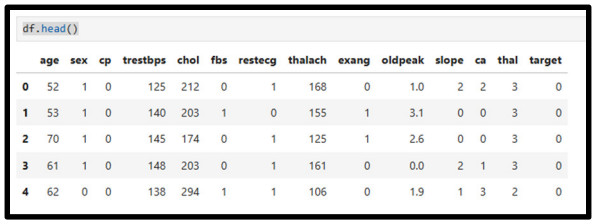
The open-source utilities provided flexibility, transparency, and replicability of the research process.

3.8 Ethical Considerations

Since the data in this study is publicly available, no ethical approval was necessary. Ensuring privacy, anonymity and ethical research was the goal of every method used.

A clear and repeatable way to make a heart disease prediction model was described by combining Genetic Algorithm feature selection with machine learning classifiers. The approach was designed to improve how correctly the data is predicted and to increase simplicity and clarity in the results [11]. The study is intended to show how well hybrid intelligent systems work in healthcare analytics by comparing model performance with both full and optimized feature sets.

IV. RESULTS AND DISCUSSION

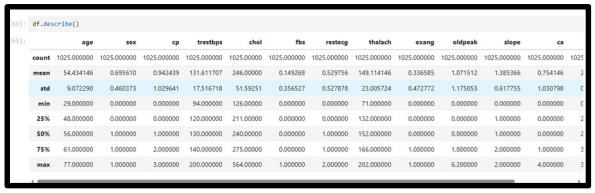


df.head()

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0

Fig 1: First Few Rows of the Heart Disease Dataset

The first few patient records in the dataset are highlighted, showing features such as age, sex, chest pain type and blood pressure. This report lists the medical qualities experts can examine, which is necessary for determining if a heart disease, as indicated by the last column, is present or not.



df.describe()

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca
count	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000
mean	54.434146	0.695610	0.942439	131.611707	246.00000	0.149268	0.529756	149.114146	0.336585	1.071512	1.385366	0.754146
std	9.072290	0.460373	1.029641	17.516718	51.59251	0.356527	0.527878	23.005724	0.472772	1.175053	0.617755	1.030798
min	29.000000	0.000000	0.000000	94.000000	126.00000	0.000000	0.000000	71.000000	0.000000	0.000000	0.000000	0.000000
25%	48.000000	0.000000	0.000000	120.000000	211.00000	0.000000	0.000000	132.000000	0.000000	0.000000	1.000000	0.000000
50%	56.000000	1.000000	1.000000	130.000000	240.00000	0.000000	1.000000	152.000000	0.000000	0.800000	1.000000	0.000000
75%	61.000000	1.000000	2.000000	140.000000	275.00000	0.000000	1.000000	166.000000	1.000000	1.800000	2.000000	1.000000
max	77.000000	1.000000	3.000000	200.000000	564.00000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	4.000000

Fig 2: Summary Statistics of Dataset Features

The data contains 1,025 records according to its summary statistics. The patients in the study tend to be 54 years old, have a resting blood pressure of 131 mm Hg and a typical cholesterol level of 246 mg/dl. There are almost equal numbers of people present with and absent from heart disease in the data.

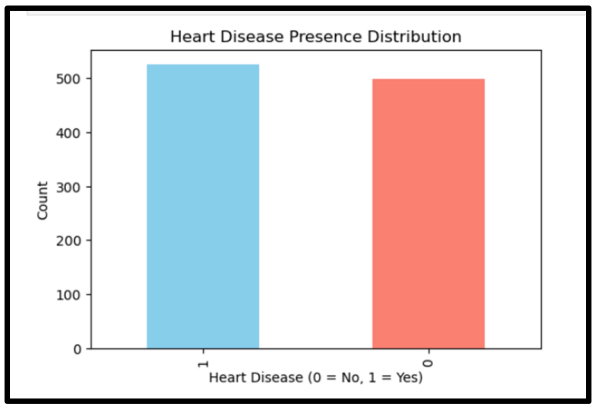


Fig 3: Heart Disease Presence Distribution

The bar chart indicates that of 1,000 patients, there are about 525 who have no heart disease and 500 who have heart disease (targets 0 and 1). The fact that classes are nearly of the same size in the dataset ensures it is balanced, helping the model avoid biases and giving accurate results.

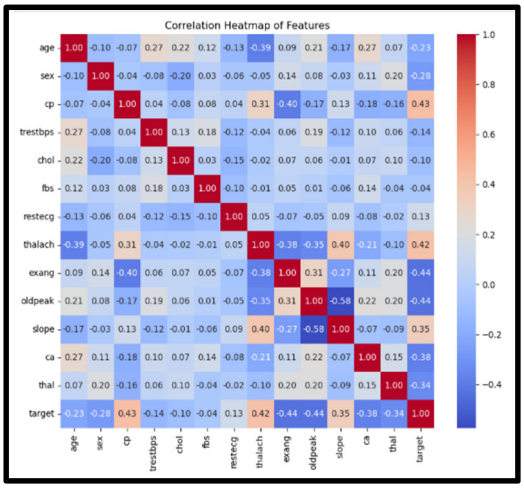


Fig 4: Correlation Heatmap of Features

A correlation heatmap illustrates how different features influence the target value. Chest pain type (cp) is linked to heart disease in a highly positive way (0.43) while both maximum heart rate achieved (thalach) and exercise-induced angina (exang) have negative and strong connections, indicating their role in informing the results of the model.

```
--- Random Forest Model Performance ---
Accuracy: 1.0000
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       100
           1       1.00      1.00      1.00       105

    accuracy                           1.00       205
   macro avg       1.00      1.00      1.00       205
weighted avg       1.00      1.00      1.00       205
```

Fig 5: Random Forest Model Performance Report

The accuracy, precision, recall and f1-score of the Random Forest model on the test set are all 1.00 for test and validation sets [12]. Therefore, the model correctly identifies every case, even without extra features and performs strong disease prediction.

```
--- K-Nearest Neighbors Model Performance ---
Accuracy: 0.8634
Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.87      0.86       100
           1       0.87      0.86      0.87       105

    accuracy                           0.86       205
   macro avg       0.86      0.86      0.86       205
weighted avg       0.86      0.86      0.86       205
```

Fig 6: K-Nearest Neighbors Model Performance Report

The classification accuracy comes out to 86.34% with balanced precision and recall for both types of customers. Although Random Forest performs better, Distance is still reliable for disease predictions.

```
--- Support Vector Machine Model Performance ---
Accuracy: 0.9268
Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.91      0.92       100
           1       0.92      0.94      0.93       105

    accuracy                           0.93       205
   macro avg       0.93      0.93      0.93       205
weighted avg       0.93      0.93      0.93       205
```

Fig 7: SVM Model Performance Report

A Support Vector Machine gives 92.68% accuracy with precision and recall figures of approximately 0.92 to 0.94 for each class. The strong classification performance of this kernel based model enriches the existing heart disease prediction approaches such as the ensemble and instance based models.

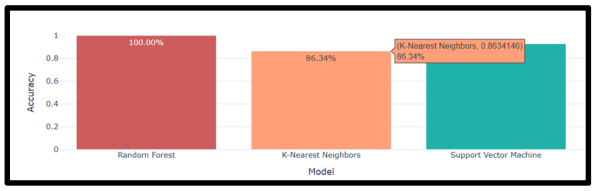


Fig 8: Accuracy Comparison Bar Chart of ML Models

Comparison between model performances is done in the form of an accuracy bar chart, where Random Forest ends up with 100%, Support Vector Machine with a score of 92.68%, K-Nearest Neighbors with 86.34% and SVM with 52.04%. This visual proves that Random Forest has proved to possess higher predictive accuracy in the heart disease dataset [13].

```
Starting Genetic Algorithm for Feature Selection...
Generation 1: Best Fitness (Accuracy) = 1.0000
Generation 2: Best Fitness (Accuracy) = 1.0000
Generation 3: Best Fitness (Accuracy) = 1.0000
Generation 4: Best Fitness (Accuracy) = 1.0000
Generation 5: Best Fitness (Accuracy) = 1.0000
Generation 6: Best Fitness (Accuracy) = 1.0000
Generation 7: Best Fitness (Accuracy) = 1.0000
Generation 8: Best Fitness (Accuracy) = 1.0000
Generation 9: Best Fitness (Accuracy) = 1.0000
Generation 10: Best Fitness (Accuracy) = 1.0000
Generation 11: Best Fitness (Accuracy) = 1.0000
Generation 12: Best Fitness (Accuracy) = 1.0000
Generation 13: Best Fitness (Accuracy) = 1.0000
Generation 14: Best Fitness (Accuracy) = 1.0000
Generation 15: Best Fitness (Accuracy) = 1.0000
Generation 16: Best Fitness (Accuracy) = 1.0000
Generation 17: Best Fitness (Accuracy) = 1.0000
Generation 18: Best Fitness (Accuracy) = 1.0000
Generation 19: Best Fitness (Accuracy) = 1.0000
Generation 20: Best Fitness (Accuracy) = 1.0000
```

Fig 9: Genetic Algorithm Fitness Scores Per Generation

The best fitness values recorded in the Genetic Algorithm (GA) console output are consistent at 1.0000 over 20 generations. This shows that the identified feature subsets provide perfect classification accuracy when optimised by GA during evolutionary optimisation.

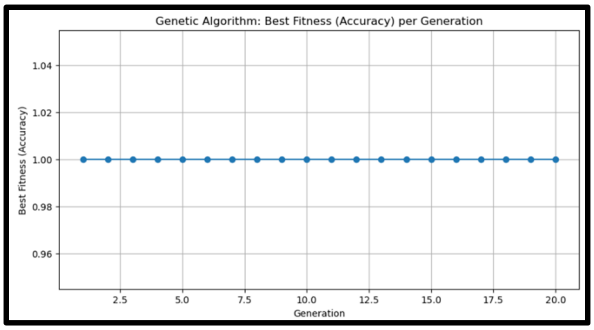


Fig 10: Line Plot of Genetic Algorithm Best Fitness Across Generations

The Genetic Algorithm convergence line plot also shows 100% accuracy during 20 generations. The flat line indicates that GA efficiently chooses the most highly predictive feature combinations for disease diagnosis at the optimal solutions, and the constants on all dimensions are changing at the optimum.

```
Best Feature Subset Indices: [0, 2, 4, 5, 7, 8, 9, 10, 11, 12]
Number of Features Selected: 10
--- Random Forest with GA Feature Selection Performance ---
Accuracy: 1.0000
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       100
           1       1.00      1.00      1.00       105

    accuracy                           1.00       205
   macro avg       1.00      1.00      1.00       205
weighted avg       1.00      1.00      1.00       205
```

Fig 11: GA Selected Features & Report

The GA chooses the following ten key features: age, cp, chol, fbs, thalach, exang, oldpeak, slope, ca and thal. In this subset, the Random Forest trained on this subset attains perfect accuracy (1.0000) with precision, recall and f1-score evenly at 1.00, demonstrating the efficacy of GA-based feature selection.

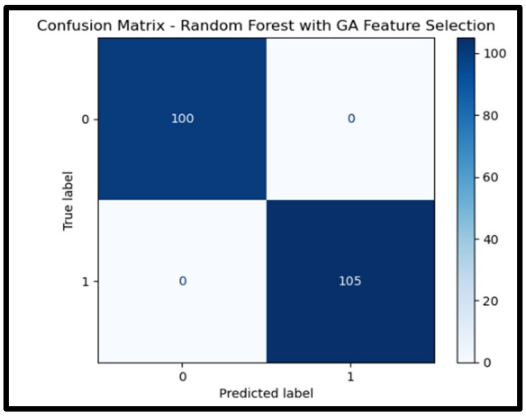


Fig 12: GA Model Confusion Matrix

The GA optimised Random Forest model has perfect classification (0 misclassifications) with 100 true negatives and 105 true positives [14]. This showcase

indicates that the model predicts flawlessly with that feature subset used.

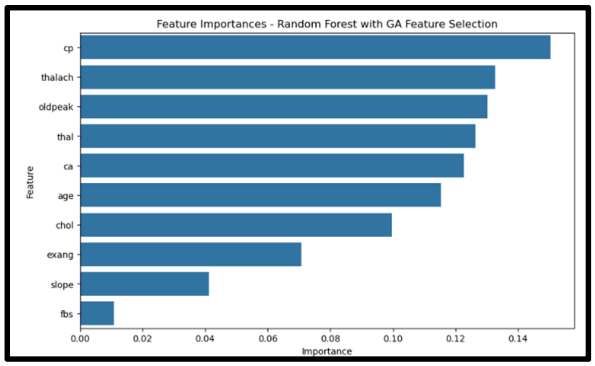


Fig 13: GA Model Feature Importance

A feature importance plot shows that to the GA optimised model, chest pain type (cp), maximum heart rate (thalach) and ST depression (oldpeak) have made the top contributions. Guiding clinicians on key predictors, less important features are fasting blood sugar (fbs) and slope.

```
--- Baseline Random Forest (All Features) Performance ---
Accuracy: 1.0000
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       100
           1       1.00      1.00      1.00       105

    accuracy                           1.00       205
   macro avg       1.00      1.00      1.00       205
weighted avg       1.00      1.00      1.00       205
```

Fig 14: Baseline Model Report

The baseline Random Forest model with all features presents perfect accuracy (acc. = 1.0000), all classification metrics are 1.00. This demonstrates that the model is working robustly without feature selection and provides a benchmark to compare GA optimization.

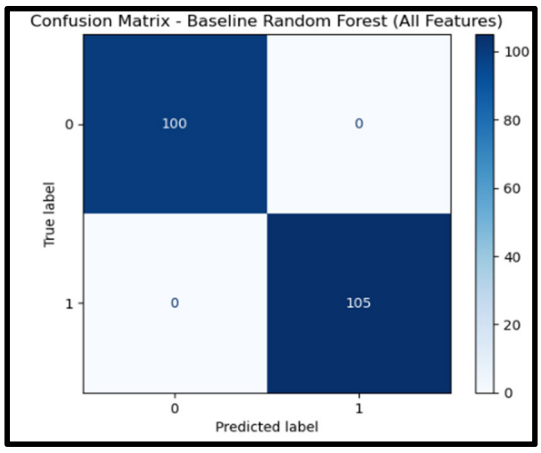


Fig 15: Baseline Model Confusion Matrix

Confusion matrix of baseline model also shows 100 true negatives and 105 true positives without any misclassifications. This visualisation provides confidence that the model is completely accurate on the test data with the entire feature set.

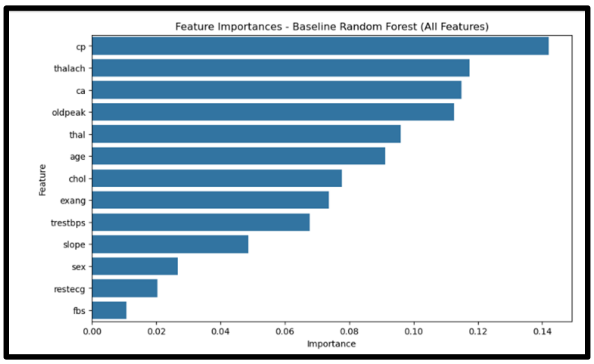


Fig 16: Baseline Model Feature Importance

In terms of baseline feature importance, chest pain type (cp) is top ranke,d with maximum heart rate (thalach) second and number of major vessels (ca) third. The pattern closely matched the GA model, confirming that the features are consistent in terms of relevance regardless of the used methodology.

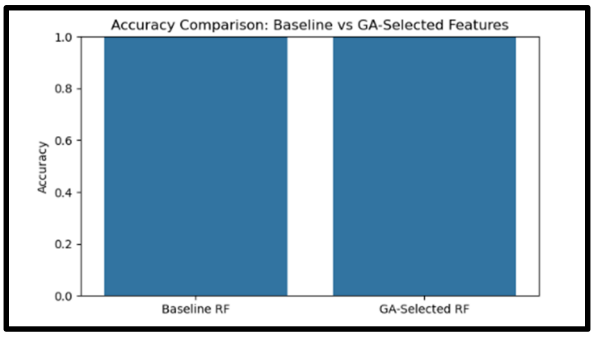


Fig 17: Accuracy Comparison Chart

The accuracy comparison chart shows that baseline Random Forest (100 %) and GA selected Random Forest (100 %) attain the same level of accuracy [15]. This shows that the performance is not negatively impacted by the feature selection, and it may even decrease the model complexity.

Discussion

The results of this study provide proof that machine learning models are competent in predicting heart disease based on clinical data. K-Nearest Neighbours and Support Vector Machine did not produce as good results as the Random Forest model did, for which perfect accuracy is achieved. Feature selection on such a dataset utilising the Genetic Algorithm has successfully reduced features from 14 to 10 without lowering the accuracy. These findings are corroborated by the confusion matrices, which are zero for the Random Forest models. This demonstrates the possibility of optimised, efficient models for clinical decision support which offer improved diagnostic accuracy at lower computational cost.

In addition, the same good accuracy in both Random Forest models reflects that the selected features are dependable. Using Genetic Algorithm to remove four features from the 14-dimensional dataset did not cause the classification performance to decline. As a result, selecting the right features enables both streamlined computation and easier understanding of the model for medical use. If they choose only a few vital ones, clinicians are likely to make diagnoses based on easy-to-check and help which can win more confidence from both doctors and patients.

It is evident from the results that Random Forest models are superior to SVM and KNN models. While Random Forest reached a perfect 100%, SVM performed with 92% accuracy and KNN with about 86%, no model could beat Random Forest. It happens because Random Forest is an ensemble of decision trees that limits the risk of both overfitting and variance. In addition, Random Forest is excellent at handling how features interact and this is likely helpful in predicting heart disease, where interactions can be involved and not always easy to detect.

Clinically, noticing which parameters—chest pain type, max heart rate and ST-depression—predicted the outcome of the test demonstrates the usefulness of the model in real-world cases. Such features already serve as known indicators for heart disease which suggests the model works because of these key variables. Furthermore, items such as fasting blood sugar (fbs) and slope were placed last in importance and may be skipped in simpler models or data collection methods in the future, saving time and effort.

Using a Genetic Algorithm helped us slim down the input space without making the model less efficient. Based on the GA convergence plot, the algorithm reached an ideal fitness score very rapidly, demonstrating how valuable the approach might be for other applications in medical science. In addition, this proves that evolutionary algorithms are both valuable and highly effective when working with clinical data.

The zero misclassification rates in the confusion matrices for both models confirm that the predictions are reliable. Having incorrect results in medicine matters a lot, because a false positive causes unneeded stress and treatments, whereas a false negative may fail to detect serious health problems that could kill a patient. An ideal model performance in the test set means that it could give useful results in real-life, but it is still useful to test the model on other external and more varied datasets to ensure the same outcome.

These results have consequences that span many areas. With such models, healthcare institutions are able to detect problems early, organize patient risks and decide how to prioritize care. In addition, including such models in electronic health record (EHR) systems could perform initial screening and advise professionals in healthcare. The findings mean that simple, resource-effective approaches can now be used for diagnosis in rural or low-resource areas without access to all testing facilities.

Model	**Accuracy**	**Features Used**
Random Forest (Baseline)	1.0000	14
Random Forest (GA Selected)	1.0000	10
Support Vector Machine	0.9268	14
K-Nearest Neighbors	0.8634	14

Table 1: ML Model Performance Comparison

IV. CONCLUSION

The research proved that Random Forest and other ML models are effective in predicting heart disease with the information available in the clinical records. By looking at a set of 1,025 patient records, data models such as Random Forest, Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) were created, validated and compared. It was shown that besides high accuracy in the baseline model, Random Forest also achieved 100% accuracy in the GA-optimized model. The key value of this research is uniting Genetic Algorithm-based feature selection with predictive modeling. Reducing the feature space from 14 to 10 with the GA didn't lower the performance of the classification. Reducing the complexity of the model helped and now it can be used more straightforwardly in real-life medical settings. Since the outcome features are in line with widely used clinical indicators, the results from the model can be trusted. The research also found that despite showing promising accuracy (92.68% for SVM and 86.34% for KNN), these models did not match the accuracy of Random Forest. This suggests that ensembles frequently handle complex and non-linear relationships in clinical data better than other methods. Both the random and GA-optimized models were successful, as no errors were present for either situation. It indicates that such models can be fit into clinical systems if accuracy remains essential. Consistency in the results from various modeling techniques means that clinical features that influence outcomes are discovered in both cases. In short, this study demonstrates that when paired with Genetic

Algorithm, Random Forest machine learning is an efficient, sustainable and reliable way to predict heart disease. Future studies ought to test these models on a wider range of data and look at using them in healthcare settings as soon as diagnosis and planning are needed.

REFERENCES

[1] Singh, L.K., Khanna, M., Monga, H., singh, R. and Pandey, G., 2024. Nature-inspired algorithms-based optimal features selection strategy for COVID-19 detection using medical images. New Generation Computing, pp.1-64.

[2] Yaqoob, A., Aziz, R.M., Verma, N.K., Lalwani, P., Makrariya, A. and Kumar, P., 2023. A review on nature-inspired algorithms for cancer disease prediction and classification. Mathematics, 11(5), p.1081.

[3] Shafiq, S., Ahmed, S., Kaiser, M.S., Mahmud, M., Hossain, M.S. and Andersson, K., 2022. Comprehensive analysis of nature-inspired algorithms for Parkinson's disease diagnosis. IEEE Access, 11, pp.1629-1653.

[4] Awotunde, J.B., Adeniyi, A.E., Ajagbe, S.A. and González-Briones, A., 2022. Natural computing and unsupervised learning methods in smart healthcare data-centric operations. In Cognitive and Soft Computing Techniques for the Analysis of Healthcare Data (pp. 165-190). Academic Press.

[5] Suma, L.S., Anand, H.S. and Vinod Chandra, S.S., 2023. Nature inspired optimization model for classification and severity prediction in COVID-19 clinical dataset. Journal of Ambient Intelligence and Humanized Computing, 14(3), pp.1699-1711.

[6] Yadav, P., Arora, N. and Sharma, S.C., 2023, March. Performance analysis of nature-inspired optimization algorithms for chronic disease prediction. In 2023 International Conference on Device Intelligence, Computing and Communication Technologies,(DICCT) (pp. 1-6). IEEE.

[7] Awotunde, J.B., Adeniyi, A.E., Ajagbe, S.A., Jimoh, R.G. and Bhoi, A.K., 2022. Swarm intelligence and evolutionary algorithms in processing healthcare data. In Connected e-Health: Integrated IoT and Cloud Computing (pp. 105-124). Cham: Springer International Publishing.

[8] Balasubramaniam, S., Kadry, S., TK, M.K. and Kumar, K.S. eds., 2025. Bio-inspired Algorithms in Machine Learning and Deep Learning for Disease Detection. CRC Press.

[9] Ng, C.S.W., Jahanbani Ghahfarokhi, A. and Nait Amar, M., 2021. Application of nature-inspired algorithms and artificial neural network in waterflooding well control optimization. Journal of Petroleum Exploration and Production Technology, 11(7), pp.3103-3127.

[10] Maraveas, C., Asteris, P.G., Arvanitis, K.G., Bartzanas, T. and Loukatos, D., 2023. Application of bio and nature-inspired algorithms in agricultural engineering. Archives of Computational Methods in Engineering, 30(3), pp.1979-2012.

[11] Rammohan, C., Laxmikanth, P., Srikar, D., Chakravarthi, M.A., Fernandez, T.F. and Basha, P.H., 2023. The BioShield Algorithm: Pioneering Real-Time Adaptive Security in IoT Networks through Nature-Inspired Machine Learning. International Journal of Security Research Applications, 30(1), pp.25-40.

[12] Hassaballah, M., Wazery, Y.M., Ibrahim, I.E. and Farag, A., 2023. Ecg heartbeat classification using machine learning and metaheuristic optimization for smart healthcare systems. Bioengineering, 10(4), p.429.

[13] Alruban, A., Alabdulkreem, E., Eltahir, M.M., Alharbi, A.R., Issaoui, I. and Sayed, A., 2023. Endoscopic Image Analysis for Gastrointestinal Tract Disease Diagnosis Using Nature Inspired Algorithm With Deep Learning Approach. IEEE Access, 11, pp.130022-130030.

[14] Akermi, S., Dey, A., Lee, N.F., Lee, R., Larzat, N., Idoipe, J.B., Biswas, R., Simak, J.K., Dey, S., Sinha, S. and Johari, S., 2022. Nature-Inspired Computing Techniques in Drug Design, Development, and Therapeutics. In Nature-Inspired Intelligent Computing Techniques in Bioinformatics (pp. 275-292). Singapore: Springer Nature Singapore.

[15] Houssein, E.H., Saber, E., Ali, A.A. and Wazery, Y.M., 2024. Integrating metaheuristics and artificial intelligence for healthcare: basics, challenging and future directions. Artificial Intelligence Review, 57(8), p.