Received: October 18, 2023. Revised: December 7, 2023. 134

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Heart Disease Prediction Using Random Forest Based Hybrid Optimization Algorithms

Ravichandra Torthi[1]* Ajay Dilip Kumar Marapatla[2] Soumya Mande[1]
Harish Kumar Varma Gadiraju[3] Chalapathiraju Kanumuri[4]

[1]Department of Electronics and Communication Engineering,
Ellenki College of Engineering and Technology, Hyderabad, India
[2]Department of AIML and IoT, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India
[3]Department of Electrical and Electronics Engineering,
Sagi Rama Krishnam Raju Engineering College, Bhimavaram, India
[4]Department of Electronics and Communication Engineering, S.R.K.R Engineering College, Bhimavaram, India
* Corresponding author's Email: Chandra.torthi@gmail.com

Abstract: Nowadays, heart diseases have become a leading cause of mortality worldwide and it affects a huge number of individuals. The early and accurate prediction of heart disease risk factors plays a crucial role in preventing opposing results. Additionally, it is necessary to recognize heart disease quickly and accurately by analyzing patient's data. This paper proposed a novel approach for predicting heart disease through machine learning techniques. The proposed Bat Algorithm (BA) and Particle Swarm Optimization (PSO) based Random Forest (RF), named BAPSO-RF is utilized for selecting optimum features that can enhance the heart-disease prediction accuracy. The proposed BAPSO-RF is evaluated on UCI heart disease dataset which contains 14 attributes and 270 records. The proposed BAPSO-RF model attains better results by utilizing metrics like accuracy, precision, recall, and f1-score values of about 98.71%, 98.67%, 98.23%, and 98.45% correspondingly which ensures early and accurate prediction of heart disease compared to existing techniques like hybrid of Genetic Algorithm and Particle Swarm Optimization (PSO) with Random Forest (GAPSO-RF), stacked Genetic Algorithm (GA) and Genetic Algorithm with Radial Basis Function (GA-RBF).

Keywords: Bat algorithm, Grid search, Heart disease prediction, Particle swarm optimization, Random forest.

1. Introduction

In recent years, the healthcare field has perceived an innovative transformation with the incorporation of machine learning techniques into its practices [1]. By analyzing an enormous quantity of medical data, machine learning techniques offer the probability to transform patient care, diagnostics, and treatment approaches [2]. The major promising study parts in the healthcare domain and the technical group are concentrated on medical applications like creating of Computer-Aided Diagnosis (CAD) system for heart disease prediction [3]. By utilizing machine learning models, healthcare experts can leverage knowledge and proficiency within these pre-trained models and apply it to various healthcare tasks like disease diagnosis, prediction, and clinical image analysis. This technique saves time, and improves the accuracy and efficiency of healthcare systems [4, 5]. The early and precise prediction of heart disease has a critical goal for ensuring better patient results and reducing the burden on healthcare systems [6]. In this instance, healthcare distributor has more prominence on heart disease prediction. By estimating patient health records, developing innovative approaches for data analysis might enable early diagnosis of heart disease [7, 8].

Heart disease is the most serious and way of the deadliest human illness worldwide, accounting for above 70% of all mortalities and a yearly death rate of more than 17.7 million [9]. It is difficult to recognize high-risk patients with heart disease

because of the contribution of various other risk factors like high blood pressure and diabetes [10]. Additionally, some other factors like unhealthy breathing situations and high stress level contributes larger spread of risk for heart disease [11]. The heart fails to pump the essential quantity of blood to various organs to instigate the usual functionality of the human body. The symptoms of heart sickness include swollen feet, physical body fatigue, irregular heartbeats, exhaustion with signs, chest pain, and dizziness [12, 13]. To alleviate these errors, it is desirable to have a computer technique that precisely predicts the probability of diseases and minimizes the imprecisions through the prediction process [14]. The powerful and high-scale machine learning algorithms are utilized for predicting heart disease and reducing high-cost diagnosis and treatment issues [15]. The major contribution of this research as follows:

- During feature selection, the proposed BAPSO-RF utilized BA to generate fast converging for specified objective functions and it shifts between the exploration to the exploitation stage rapidly at the primary stage. The PSO is involved in enhancing the optimization performance of the disease prediction model.
- The fitness function of BA and PSO is enhanced through an RF classifier for enhancing accuracy in classification.
- The proposed BAPSO-RF approach was evaluated by utilizing metrics like accuracy, precision, recall, and f1-score.

The rest part of the research is described as follows, relative research is given in Section 2. The proposed method is explained in Section 3. The results and comparative analysis of the proposed method are given in Section 4 and Section 5 is a conclusion of the paper.

2. Literature Review

Al-Ssulami [16] introduced machine learning and data augmentation techniques for improved coronary heart disease prediction. This developed model provides an augmented dataset through duplicate misclassified occurrences selectively in a leave-one-out cross-validation procedure for model overfitting. This developed model utilized the UCI heart disease dataset which corresponds with a certain classifier for providing distinct augmented datasets. This experiment was repetitive through various smaller base datasets and every produced dataset constantly yields higher accuracy. The developed model enhances the robustness and generalization in a large dataset. However, this model enhances the computation time at each iteration.

El-Shafiey [17] developed a hybrid of Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) optimization technique for predicting heart disease according to Random Forest (RF). The developed GAPSO-RF model executes multivariate numerical analysis in the first step for selecting the most substantial features through the initial population which enhances prediction accuracy. Then, the discriminative mutation approach was executed in GA and the GAPSO-RF model integrates an improved GA and PSO for global and local search respectively. The developed model was computationally affordable because of small system requirements. However, the developed model has some limitations such as temporal complexity, small-data issues, and high computational cost.

Abdollahi and Nouri-Moghaddam [18] implemented an ensemble technique to predict heart disease using feature selection based on machine learning techniques. The feature selection method designates the important features for enhancing accuracy and minimizing the computational time of classification. To estimate the hyperparameter tuning and machine learning model, cross-validation is utilized. The classifiers' performance was examined on designated features as chosen by the performance metrics. The developed stacked GA model employed UCI dataset which comprised 270 records and 13 features. This ensemble technique has some benefits like smaller computational time and better generalization ability than conventional algorithms. However, this developed model has huge computational overhead and a probability of overfitting issues.

El-Hasnony [19] presented multi-label active learning based on five various selection approaches for heart disease prediction. The developed model is utilized to minimize labeling costs by designating the relevant query information iteratively. The selected method by label ranking classifier consumes enhanced hyperparameters through grid search for executing predictive modeling in every stage of the heart disease dataset. The prevalence of employing active learning for analyzing heart disease improves generality over memorization of the generated model. The developed model was utilized to solve the memorizing learning model problems. However, this model has slow convergence and high computational cost.

Doppala [20] suggested a hybrid machine-learning technique for predicting coronary disease through feature selection over the heart disease dataset. The developed method utilized a GA-RBF

for coronary prediction through enhanced accuracy of feature selection mechanism. The developed GA-RBF technique goal was attribute reduction which is the solution for obtaining better effectiveness and the feature selection mechanism affects the target predicted values. The GA-RBF model has an excellent learning capability, a simple layout, and the ability to manage datasets through numerous features. However, the computational cost and time required for each generation has increased.

Budholiya [21] introduced an optimized Extreme Gradient Boosting (XGBoost) classifier for efficient heart disease prediction. The developed model utilized Bayesian optimization which is an effective technique for optimizing the hyperparameter of the XGBoost classifier. Additionally, the One-Hot (OH) encoding technique for encoding the absolute features is utilized in data preprocessing in heart disease datasets to enhance prediction accuracy. The developed model performance was estimated on the heart disease dataset and compared with Extra Tree (ET) and Random Forest (RF) classifiers. The developed model has a built-in ability to handle mission data during prediction. However, this model consumes significant amount of data while working with large datasets.

Bhatt [22] developed a machine-learning technique based on efficient heart disease prediction to minimize the fatality produced by heart diseases. The developed method with k-mode clustering through Huang starting can enhance the classification accuracy. Machine learning methods like random forest, XGBoost, decision tree, and Multilayer perceptron are utilized in this research. The GridSearchCV was utilized to tune the hyperparameters of this model for optimizing the result. The developed model dataset was divided based on gender which can be beneficial for prediction. However, this developed model considers only a restricted set of clinical and demographic variables and does not consider other potential risk factors.

Rani [23] presented a machine learning-based hybrid decision support system to predict heart disease. The developed model utilized a multivariate attribution through a chained algorithm for handling missing values. The hybrid feature selection algorithm integrates the elimination of recursive features and a Genetic algorithm for selecting features from accessible datasets. The naïve Bayes, support vector machine, logistic regression, adaboost, and random forest are utilized for classification. The developed model iterates over numerous generations which helps to produce better solutions. However, the disease severity was not diagnosed which minimizes the model's performance.

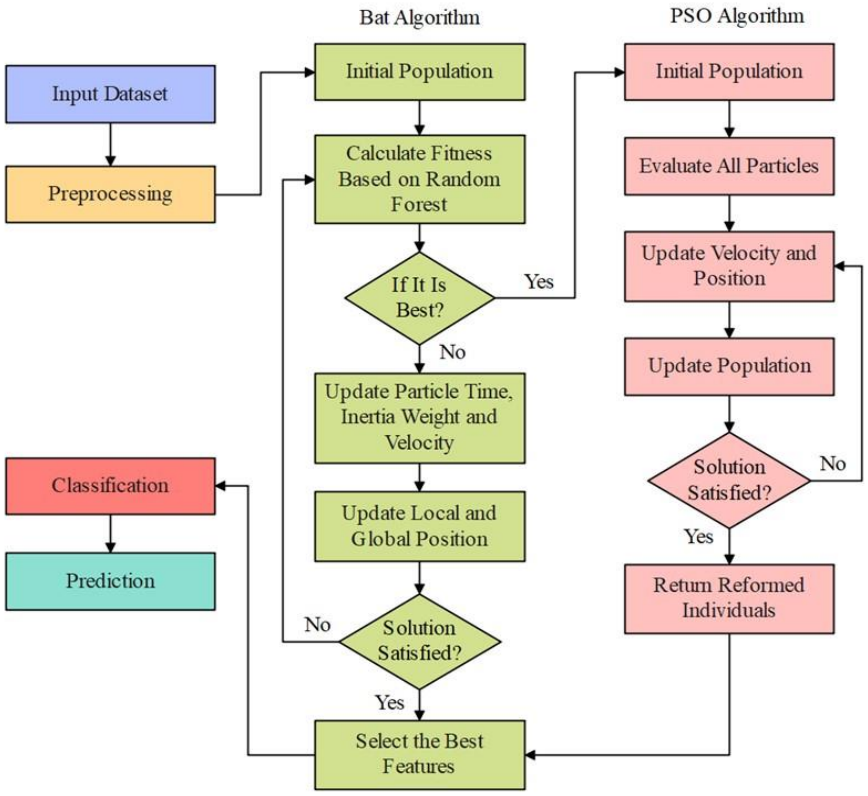

Figure. 1 Block Diagram of Proposed Methodology

Kavitha Chandrashekar and Anitha Tuluvanooru Narayanreddy [22] suggested an Ensemble Feature Optimized (EFO) learning model which utilized an improved XGBoost and feature level cross-validation approach for efficient heart disease prediction. To obtain high accuracy, this model utilized a cross-validation approach that employed efficient feature rank model for enhance the prediction accuracy through improving the prediction error. This model produced the better accuracy when the data was imbalance. However, this model has high training error rate for multiclass classification.

From the overall analysis, the GAPSO-RF model has temporal complexity, small-data issues, and high computational cost. The stacked GA model has huge computational overhead and a probability of overfitting issues. The GA-RBF model has enhanced the computational cost and time required for each generation. The optimized XGBoost with OH encoded model consumes a significant amount of data while working with large datasets. The EFO model has high error rate for multiclass classification. Hence, these limitations are considered and overcomes by proposed BAPSO in this manuscript.

3. Proposed Methodology

The UCI heart disease dataset is utilized in this paper which contains 14 attributes and 270 records. This dataset is standardized using the min-max normalization method which improves the model performance. The preprocessed features are selected by using the Bat Algorithm and Particle Swarm Optimization (BAPSO). During feature selection, the proposed BAPSO generates fast converging for the specified objective functions and it shifts between exploration to the exploitation stage rapidly at the primary stage. The PSO is involved in enhancing the optimization performance of the disease prediction model. The fitness function of BA and PSO is enhanced through an RF classifier for enhancing accuracy in classification. The block diagram of the proposed methodology is presented in Fig. 1.

3.1 Dataset

The proposed methodology was evaluated by utilizing the publicly accessible heart disease dataset from University of California, Irvine (UCI) repository [25] which contains 14 attributes and 270 records. The dataset was divided into 75% for training and 25% for testing. The target is to classify the absence or presence of heart sickness from the medical data given to a patient. This is a benchmark dataset due to the influences of real patient information and which is widely utilized to test numerous data processing methods. The description of this dataset is illustrated in Table 1.

Table 1. Dataset features and data types

Features	Type
Age	Integer
Blood pressure	Integer
Blood sugar	Numerical two values
Ca	Numerical two values
Chest pain	Numeric values
C	Two-digit number
Cholesterol	Integer
Depression	Real number
Electrocardiographic	Three-digit number
Exercise-included	Numerical two values
Heart rate	Integer
Sex	Numerical two values
Slope	Numeric values
Thal	Numeric values

3.2 Preprocessing

Data preprocessing is the process of converting the raw data into a desired format, the dataset from several resources may consist of incomplete data. So, for further analysis, this data is required to be filtered and normalized. Data normalization is the process of preprocessing the input data. The UCI heart disease dataset is standardized using min-max normalization method which improves the model performance [26]. The highest score of the feature is transformed into 1, the smallest score of the feature is transformed into 0 and other values of the feature are transformed into an integer between 0 and 1. The mathematical representation of min-max normalization is shown in Eq. (1),

$$f(x,y) = \frac{f(x,y) - Z_{min}}{Z_{max} - Z_{min}} \quad (1)$$

Where, f is the input image, x and y are pixel location in image, Z_{max} and Z_{min} are the maximum and minimum pixel values. After the preprocessing step, the features are selected by using an optimization algorithm.

3.3 Feature Selection

The Bat Algorithm and Particle Swarm Optimization (BAPSO) are used for selecting features from preprocessed images. Bat Algorithm (BA) generates fast converging for the specified objective functions and it shifts between exploration to the exploitation stage rapidly at the primary stage. The PSO is computationally cheap because of its few system requirements. The BAPSO algorithm is

designated for feature selection due to its discrete features in attaining optimal solutions for optimization issues. These features are utilized to recognize the optimal resources from the disease prediction. The BA is a nature-inspired optimization technique according to the echolocation features of bats. The bats naturally employed a sonar category called echolocation for detecting prey [27], tracing problems in the track, and staying in crevices. The echoes formed by bats are loud and the difference in pulses was relevant for hunting approaches. Usually, frequency-moderated signals are formed which come back to bat in the form of the octave. The frequency ranges from 25kHz to 150kHz and the distinctive pulse count is 10-20 per sec which can rise to 200 pulses per sec. Using echolocation, the hunting behavior of bats is correlated with the optimal resource recognition from the disease prediction. There are three common rules for expressing the bat algorithm such as,

- Bats already know the echo difference among prey and background obstacles
- Based on the targets, bats alter the emanated pulse frequency and pulse emission rates automatically.
- The loudness differs between huge positive into minimum persistent.

These general rules are correlated to the optimum feature selection procedure wherein differences among required and other resources are identified for the distribution method. Then, the source recognition chooses optimum features through parameter altering, and lastly highest optimum score was taken as more classification procedure. The bat motion is expressed for determining the optimization procedure. Primarily, consider the bat frequency (f_x), velocity (v_x^t), location (M_x^t) and solution space (d) for tth iteration [28]. The better solution display between the bat is specified as M_* according to the defined function. The above rules are mathematically expressed in Eq. (2), (3) and (4),

$$f_x = f_{min} + (f_{max} - f_{min})\varphi \tag{2}$$

$$v_x^t = v_x^{t-1} + (M_x^{t-1} - M_*)f_x \tag{3}$$

$$M_x^t = M_x^{t-1} + v_x^t \tag{4}$$

Here, φ is the random function vector attained from a uniform distribution within the range of [0,1]. Primarily, entire bats are allocated through random frequency within the range of $[f_{max}, f_{min}]$.

For this purpose, the BA is specified as a frequency-tuning algorithm which achieves the best features in exploitation and exploration. Moreover, pulse rates and loud variations are expressed for altering the system among exploration into the exploitation phase. In this phase, loudness variation (l_x) and pulse emission rate (r_x) are differed in iteration process. Generally, loudness was reduced whether the bat recognized prey though (r_x) was enhanced. The loudness designates among l_{max} and l_{min} and smallest loudness is considered as $l_{min} = 0$. In this case the bat recognizes prey and quits pulse emanation which is denoted in Eq. (5) and (6),

Table 2. Parameters of optimization model

Parameters	Values
Bat size	**15**
Acceleration constants	1.04
Highest loudness, pulse rate and frequency	1, 2, 3
Smallest loudness, pulse rate and frequency	0
Number of maximum iterations	150
Pulse rate constant	0.9
Loudness constant	0.96

$$l_x^{t+1} = \rho l_x^t \tag{5}$$

$$r_x^{t+1} = r_x^0(1 - e^{-\sigma t}) \tag{6}$$

Where ρ and σ are constant, r_x^0 is the primary pulse rate. For $0 < \rho < 1$ and $\sigma > 0$ the functions are altered into $l_x^t = 0, r_x^t = r_x^0$. The ρ is designated in the range of $0.9 - 0.98$. Table 2 illustrates the parameters of the optimization model.

The PSO is involved in enhancing the optimization performance of the disease prediction model. The PSO is a stochastic technique that is expressed according to the characteristics of birds flocking in the food searching procedure. The random population originated as particles transmit data around search space. The additional particles accomplished similar thus the data was swapped through other particles. The best solution of PSO was labeled as the global best and overall enduring particles were required to transfer from the present position into the optimum position [29]. The curve movement will be defined and this procedure was frequently utilized to attain the best solution. Present velocity values are interned through particles in the population thus the following best solution can be attained [30]. The velocity of all particle position vectors is denoted in eq. (7),

$$M_{xy}^{t+1} = M_{xy}^t + v_{xy}^{t+1} \tag{7}$$

Where, M_{xy}^{t} is the position vector of ith iteration in xth particle and v_{xy}^{t+1} is the velocity vector of $t+1$th iteration in ith particle. The velocity is expressed in eq. (8),

$$v_{xy}^{t+1} = wv_{xy}^{t} + c_1 r_1\left(pbest_{xy}^{t} - M_{xy}^{t}\right) + \left(c_2 r_2 gbest_{xy}^{t} - M_{xy}^{t}\right) \quad (8)$$

Where, v_{xy}^{t+1} is the velocity vector of $t+1$ th iteration in ith particle, c_1 and c_2 are the uniform distribution system. The particle fitness rate is specified as $pbest$ and $gbest$ that is according to every particle in the fitness function. According to the fitness function, every particle position is estimated which is shown in eq. (9),

$$f = (M_{xy}^{t} - v_{max})^2 + (M_{xy}^{t} - x_{max})^2 \quad (9)$$

The fitness function of every particle is equated to update the position of $pbest$ and $gbest$, whether the position is relatively better than the earlier position than the current position is taken as the best and the all-fitness function is modernized which is presented in eq. (10) and (11),

$$pbest_{xy}^{t} \begin{cases} M_{xy}^{t} & if\ f\left(M_{xy}^{t}\right) < pbest_{xy}^{t} \\ pbest_{xy}^{t} & Otherwise \end{cases} \quad (10)$$

$$gbest_{xy}^{t} = \min\left(pbest_{x}^{t}, pbest_{x+1}^{t}, \ldots, pbest_{s}^{t}\right. \quad (11)$$

To update each particle velocity and position eq. (10) and (11) were utilized and a similar process was executed for tuning the BA fitness function. The PSO last velocity is presented in eq. (12),

$$v_{xy}^{t+1} = wv_{xy}^{t} + \left(c_1 r_1\left(pbest_{xy}^{t} - M_{xy}^{t}\right) + \left(c_2 r_2 gbest_{xy}^{t} - M_{xy}^{t}\right)\right) f_x \quad (12)$$

Where, v_{xy}^{t+1} is the velocity vector of $t+1$ th iteration in ith particle, c_1 and c_2 are the uniform distribution system and M_{xy}^{t} is the position vector of ith iteration in xth particle.

3.4 Classification

In this proposed technique, the Random Forest (RF) is utilized for binary classification. The RF builds numerous decision tree through the training time and produces class that has a mean estimation. The RF hyperparameter are finetuned through grid search. The best parameter set was extracted from grid search which is utilized to train RF to obtain the highest accuracy in classification. An extensive variety of parameters are executed in grid search which is presented in Table 3.

Table 3. Parameters and Grid search values

Parameters	Grid Search Values
No. of trees	50, 100, 200, 500, 1000, 2000
Min_samples_leaf	1, 5, 10, 12
Min_samples_split	2, 5, 10, 20
Max_depth	3, 5, 10, 15
Max_features	Auto, log2, sqrt

The RF utilized the bagging concept and integrated various decision trees to enhance prediction ability. The numerous data samples are randomly produced from the actual dataset through replacement and every decision tree is trained on various data samples. Features are randomly designated through tree construction. The prediction produced with numerous trees was integrated by a majority vote. The RF can be tuned with enhanced accuracy through parameter optimization like estimator number, number of features utilized for split node, minimum size of node, etc. Particularly, the decision tree designates optimum features from present feature sets. However, RF chooses the subset from the node set randomly that fits to decision trees and then selects optimum features from subsets [31]. The RF has the highest accuracy and is utilized in the big data set but have a good performance for some high-dimensional samples. The RF working process is presented in Fig. 2.

4. Experimental Result

In this paper, the proposed method is simulated by using a Python environment with the system configuration of RAM:16GB, Processor: Intel core i7 and Operating System: Windows 10. The parameters like accuracy, precision, recall and f1-score are utilized for estimate proposed technique performance which is shown in eq. (13), (14), (15) and (16),

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (13)$$

$$Precision = \frac{TP}{TP+FP} \quad (14)$$

$$Recall = \frac{TP}{TP+FN} \quad (15)$$

$$F1-Score = 2 \times \frac{Precision \times recall}{Precision+recall} \quad (16)$$

Where, TP, TN, FP and FN illustrate the True Positives, True Negatives, False Positives and False Negatives respectively.

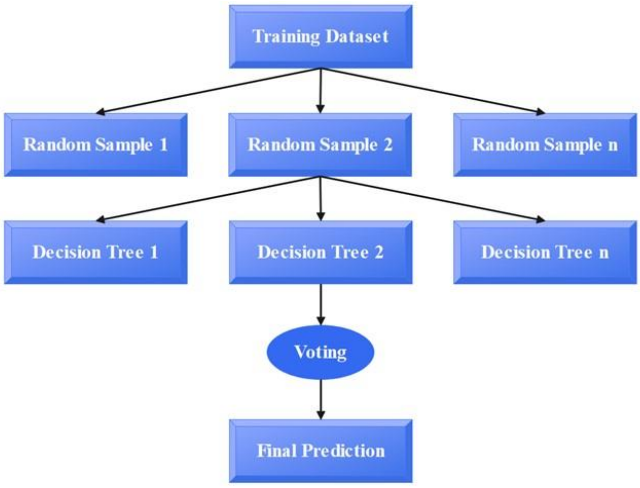

Figure. 2 Working process of Random Forest

Table 4. Quantitative analysis of the Optimization algorithm

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CSO	89.92	89.76	89.54	89.63
GWO	91.79	91.58	91.63	91.36
BA	93.43	93.39	93.28	93.11
PSO	94.82	94.71	94.57	94.41
BAPSO	96.57	96.48	96.31	96.23

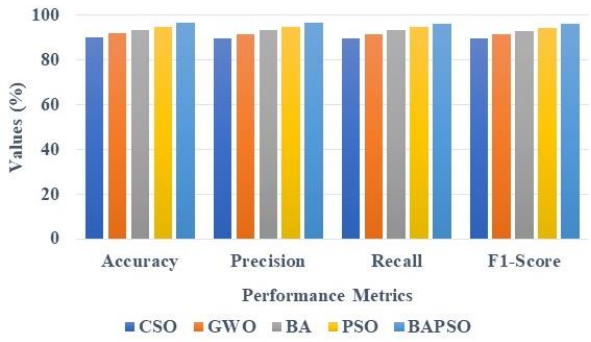

Figure. 3 Performance of optimization algorithm

Table 5. Quantitative analysis of classification with default features

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
NB	89.71	89.53	89.49	89.28
SVM	91.48	91.39	91.25	91.16
LR	92.87	92.61	92.58	92.37
DT	93.94	93.77	93.69	93.46
RF	94.79	94.61	94.48	94.57

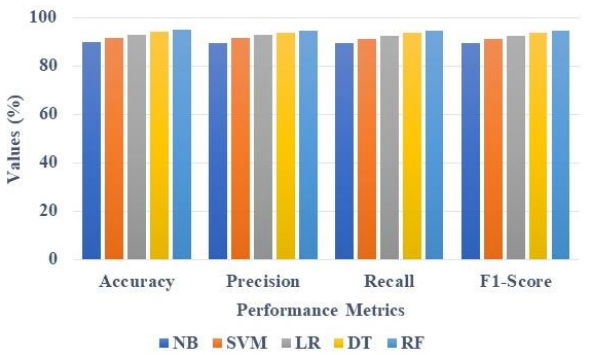

Figure. 4 Performance of classification with default features

4.1 Quantitative Analysis

This section shows the quantitative analysis of the proposed BAPSO-RF approach with regard to accuracy, precision, recall, and f1-score. Table 4, 5 and 6 illustrates the quantitative analysis of optimization algorithms, classification with default features and after feature selection by employing the UCI heart disease dataset respectively.

Table 4 and Fig. 3 represent the performance of the optimization algorithm by utilizing metrics like accuracy, precision, recall and f1-score. The Cat Swarm Optimization (CSO), Grey Wolf Optimization (GWO), Bat Algorithm (BA), and Particle Swarm Optimization (PSO) are restrained and matched with the proposed BAPSO algorithm. The obtained result shows that the proposed BAPSO algorithm attains accuracy of 96.57%, precision of 96.48%, recall of 96.31% and f1-score of 96.23% which is comparatively higher than the existing optimization algorithms.

Table 5 and Fig. 4 represent the performance of classification with default features by utilizing metrics like accuracy, precision, recall and f1-score. The Naïve Bayes (NB), Support Vector Machine (SVM), Linear Regression (LR) and Decision Tree (DT) are measured and matched with RF model. The attained result shows that the RF model attains accuracy of 94.79%, precision of 94.61%, recall of

Table 6. Quantitative analysis of classification after feature selection

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
RF	94.79	94.61	94.48	94.57
GA-RF	95.42	95.39	95.27	95.22
BA-RF	96.88	96.73	96.62	96.71
PSO-RF	97.53	97.42	97.38	97.46
BAPSO-RF	98.71	98.67	98.23	98.45

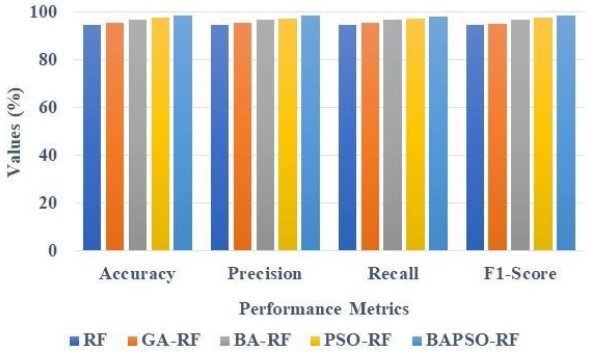

Figure. 5 Performance of classification after feature selection

94.48% and f1-score of 94.57% which is comparatively higher than the existing classifiers.

Table 6 and Fig. 5 represent the performance of classification after feature selection by utilizing metrics like accuracy, precision, recall and f1-score. The Random Forest (RF), Genetic Algorithm-RF (GA-RF), Bat Algorithm-RF (BA-RF), Particle Swarm Optimization-RF (PSO-RF) are restrained and matched with the proposed BAPSO-RF approach. The obtained result shows that the proposed BAPSO-RF approach attains accuracy of 98.71%, precision of 98.67%, recall of 98.23% and f1-score of 98.45% which is comparatively higher than the existing methods.

4.2 Comparative Analysis

This section illustrates the comparative analysis of proposed BAPSO-RF approach with evaluation metrics like accuracy, precision, recall and f1-score as shown in Table 7. The existing result such as [17], [18, 20, 21, 24] are utilized for estimating an ability of the classifier. The BAPSO-RF is trained, tested and validated by using UCI heart disease dataset. The result obtained from Table 7 shows that the proposed BAPSO-RF attains better performance when compared with the existing methods. The accuracy was improved to 98.71%, precision of 98.67%, recall of 98.23% and f1-score of 98.45%.

4.2.1. Discussion

In this section, the advantages of the proposed method and the limitations of existing methods are discussed. The existing method has some limitations such as the GAPSO-RF [17] model has temporal complexity, small-data issues, and high computational cost. The stacked GA [18] model has huge computational overhead and a probability of overfitting issues. The GA-RBF [20] model has enhanced the computational cost and time required for each generation. The optimized XGBoost with OH encoded [21] model consumes a significant amount of data while working with large datasets. The EFO [24] model has high error rate for multiclass classification. The proposed BAPSO-RF approach overcomes these existing model limitations. The proposed model is utilized to generate fast converging for the specified objective functions and it shifts between exploration into the exploitation stage rapidly at the primary stage. The PSO is in enhancing the performance of the disease prediction model. The fitness function of BA and PSO is enhanced through an RF classifier for enhancing accuracy in classification.

Table 7. Comparative Analysis

Author	Method	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
El-Shafiey [17]	GAPSO-RF	UCI heart disease dataset	95.60	97.44	92.68	94.00
Abdollahi and Nouri-Moghaddam [18]	Stacked GA		97.57	N/A	96	N/A
Doppala [20]	GA-RBF		85.40	95	96	95
Budholiya [21]	Optimized XGBoost with OH encoded		91.80	N/A	85.71	90.56
Kavitha Chandrashekar and Anitha Tuluvanooru Narayanreddy [24]	EFO		98.61	N/A	N/A	N/A
Proposed method	BAPSO-RF		98.71	98.67	98.23	98.45

5. Conclusion

This paper proposed a Bat Algorithm (BA) and Particle Swarm Optimization (PSO) based Random Forest (RF), named as BAPSO-RF is utilized for selecting optimum features which can enhance the heart-disease prediction accuracy. The proposed BAPSO-RF is evaluated on UCI heart disease dataset which contains 14 attributes and 270 records. The UCI heart disease dataset is standardized using min-max normalization method which improves the model performance. The preprocessed features are selected by using Bat Algorithm and Particle Swarm Optimization (BAPSO). During feature selection, the proposed BAPSO generates fast converging for the specified objective functions and it shifts between exploration into exploitation stage rapidly at primary stage. The PSO is involved in enhancing the optimization performance of the disease prediction model. The fitness function of BA and PSO is enhanced through RF classifier for enhancing accuracy in classification. The proposed BAPSO-RF model attains better results by utilizing metrics like accuracy, precision, recall, and f1-score values of about 98.71%, 98.67%, 98.23% and 98.45% correspondingly which ensures accurate and early prediction of heart disease. In future, the hyper parameter tuning is applied in optimization algorithm for improving the model performance.

Notations:

Notations	Description
$f(x,y)$	Min-max normalization
f	Input image
x and y	Pixel location in image
Z_{max}	Maximum pixel values
Z_{min}	Minimum pixel values
(f_x)	Frequency of bat
(v_x^t)	Velocity of bat
(M_x^t)	Location of bat
φ	Random function vector
(l_x)	Loudness variation
(r_x)	Pulse emission rate
ρ and σ	Constant
r_x^0	Primary pulse rate
M_{xy}^t	Position vector of ith iteration in xth particle
v_{xy}^{t+1}	Velocity vector of $t+1$th iteration in ith particle
c_1 and c_2	Uniform distribution system
$pbest$ and $gbest$	Particle fitness rate
TP	True Positives
TN	True Negatives
FP	False Positives
FN	False Negatives

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

For this research work all authors' have equally contributed in Conceptualization, methodology, validation, resources, writing—original draft preparation, writing—review and editing.

References

[1] W. Li, M. Zuo, H. Zhao, Q. Xu, and D. Chen, "Prediction of coronary heart disease based on combined reinforcement multitask progressive time-series networks", *Methods*, Vol. 198, pp. 96-106, 2022.

[2] K.V.V. Reddy, I. Elamvazuthi, A.A. Aziz, S. Paramasivam, H.N. Chua, and S. Pranavanand, "Heart Disease Risk Prediction Using Machine Learning Classifiers with Attribute Evaluators", *Applied Sciences*, Vol. 11, No. 18, p. 8352, 2021.

[3] R.L. Priya, S.V. Jinny, and Y.V. Mate, "Early prediction model for coronary heart disease using genetic algorithms, hyper-parameter optimization and machine learning techniques", *Health and Technology*, Vol. 11, No. 1, pp. 63-73, 2021.

[4] A.K. Faieq, and M.M. Mijwil, "Prediction of heart diseases utilising support vector machine and artificial neural network", *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 26, No. 1, pp. 374-380, 2022.

[5] U. Nagavelli, D. Samanta, and P. Chakraborty, "Machine learning technology-based heart disease detection models", *Journal of Healthcare Engineering*, Vol. 2022, p. 7351061, 2022.

[6] M.S. Pathan, A. Nag, M.M. Pathan, and S. Dev, "Analyzing the impact of feature selection on the accuracy of heart disease prediction", *Healthcare Analytics*, Vol. 2, p. 100060, 2022.

[7] X.-Y. Gao, A.A. Ali, H.S. Hassan, and E.M. Anwar, "Improving the accuracy for analyzing heart diseases prediction based on the ensemble method", *Complexity*, Vol. 2021, p. 6663455, 2021.

[8] W. Sun, P. Zhang, Z. Wang, and D. Li, "Prediction of cardiovascular diseases based on machine learning", *ASP Transactions on*

Internet of Things, Vol. 1, No. 1, pp. 30-35, 2021.

[9] T.R. Mahesh, V.D. Kumar, V.V. Kumar, J. Asghar, O. Geman, G. Arulkumaran, and N. Arun, "AdaBoost ensemble methods using K-fold cross validation for survivability with the early detection of heart disease", *Computational Intelligence and Neuroscience*, Vol. 2022, p. 9005278, 2022.

[10] S. Nandy, M. Adhikari, V. Balasubramanian, V.G. Menon, X. Li, and M. Zakarya, "An intelligent heart disease prediction system based on swarm-artificial neural network", *Neural Computing and Applications*, Vol. 35, No. 20, pp. 14723-14737, 2023.

[11] G.N. Ahmad, H. Fatima, S. Ullah, A.S. Saidi, and Imdadullah, "Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques With and Without GridSearchCV", *IEEE Access*, Vol. 10, pp. 80151-80173, 2022.

[12] A. Saboor, M. Usman, S. Ali, A. Samad, M.F. Abrar, and N. Ullah, "A method for improving prediction of human heart disease using machine learning algorithms", *Mobile Information Systems*, Vol. 2022, p. 1410169, 2022.

[13] M.W. Nadeem, H.G. Goh, M.A. Khan, M. Hussain, M.F. Mushtaq, and V. Ponnusamy, "Fusion-Based Machine Learning Architecture for Heart Disease Prediction", *Computers, Materials & Continua*, Vol. 67, No. 2, 2481-2496, 2021.

[14] N. Absar, E.K. Das, S.N. Shoma, M.U. Khandaker, M.H. Miraz, M.R.I. Faruque, N. Tamam, A. Sulieman, and R.K. Pathan, "The Efficacy of Machine-Learning-Supported Smart System for Heart Disease Prediction", *Healthcare*, Vol. 10, No. 6, p. 1137, 2022.

[15] A. Abdellatif, H. Abdellatef, J. Kanesan, C.-O. Chow, J.H. Chuah, and H.M. Gheni, "An Effective Heart Disease Detection and Severity Level Classification Model Using Machine Learning and Hyperparameter Optimization Methods", *IEEE Access*, Vol. 10, pp. 79974-79985, 2022.

[16] A.M. Al-Ssulami, R.S. Alsorori, A.M. Azmi, and H. Aboalsamh, "Improving Coronary Heart Disease Prediction Through Machine Learning and an Innovative Data Augmentation Technique", *Cognitive Computation*, Vol. 15, No. 5, pp. 1687-1702, 2023.

[17] M.G. El-Shafiey, A. Hagag, E.-S.A. El-Dahshan, and M.A. Ismail, "A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest", *Multimedia Tools and Applications*, Vol. 81, No. 13, pp. 18155-18179, 2022.

[18] J. Abdollahi, and B. Nouri-Moghaddam, "A hybrid method for heart disease diagnosis utilizing feature selection based ensemble classifier model generation", *Iran Journal of Computer Science*, Vol. 5, No. 3, pp. 229-246, 2022.

[19] I.M. El-Hasnony, O.M. Elzeki, A. Alshehri, and H. Salem, "Multi-label active learning-based machine learning model for heart disease prediction", *Sensors*, Vol. 22, No. 3, p. 1184, 2022.

[20] B.P. Doppala, D. Bhattacharyya, M. Chakkravarthy, and T.-h. Kim, "A hybrid machine learning approach to identify coronary diseases using feature selection mechanism on heart disease dataset", *Distributed and Parallel Databases*, Vol. 41, No. 1-2, pp. 1-20, 2023.

[21] K. Budholiya, S.K. Shrivastava, and V. Sharma, "An optimized XGBoost based diagnostic system for effective prediction of heart disease", *Journal of King Saud University-Computer and Information Sciences*, Vol. 34, No. 7, pp. 4514-4523, 2022.

[22] C.M. Bhatt, P. Patel, T. Ghetia, and P.L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques", *Algorithms*, Vol. 16, No. 2, p. 88, 2023.

[23] P. Rani, R. Kumar, N.M.O.S. Ahmed, and A. Jain, "A decision support system for heart disease prediction based upon machine learning", *Journal of Reliable Intelligent Environments*, Vol. 7, No. 3, pp. 263-275, 2021.

[24] K. Chandrashekar, and A.T. Narayanreddy, "An Ensemble Feature Optimization for an Effective Heart Disease Prediction Model", *International Journal of Intelligent Engineering & Systems*, Vol. 16, No. 2, 2023, doi: 10.22266/ijies2023.0430.42.

[25] Dataset link: https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data.

[26] H. Kibriya, R. Amin, A.H. Alshehri, M. Masood, S.S. Alshamrani, and A. Alshehri, "A novel and effective brain tumor classification model using deep feature fusion and famous machine learning classifiers", *Computational Intelligence and Neuroscience*, Vol. 2022, p. 7897669, 2022.

[27] W.A.H.M. Ghanem, S.A.A. Ghaleb, A. Jantan, A.B. Nasser, S.A.M. Saleh, A. Ngah, A.C. Alhadi, H. Arshad, A.M.H.Y. Saad, A.E. Omolara, Y.A.B. El-Ebiary, and O.I. Abiodun,

"Cyber Intrusion Detection System Based on a Multiobjective Binary Bat Algorithm for Feature Selection and Enhanced Bat Algorithm for Parameter Optimization in Neural Networks", *IEEE Access*, Vol. 10, pp. 76318-76339, 2022.

[28] M. Zhang, Y. Yi, and W. Cheng, "Multistage condition monitoring of batch process based on multi-boundary hypersphere SVDD with modified bat algorithm", *Arabian Journal for Science and Engineering*, Vol. 46, No. 2, pp. 1647-1661, 2021.

[29] M. Mansouri, K. Dhibi, H. Nounou, and M. Nounou, "An Effective Fault Diagnosis Technique for Wind Energy Conversion Systems Based on an Improved Particle Swarm Optimization", *Sustainability*, Vol. 14, No. 18, p. 11195, 2022.

[30] S. Senthilkumar, V. Mohan, S.P. Mangaiyarkarasi, and M. Karthikeyan, "Analysis of single-diode PV model and optimized MPPT model for different environmental conditions", *International Transactions on Electrical Energy Systems*, Vol. 2022, p. 4980843, 2022.

[31] N. Tasnim, S.A. Al Mamun, M.S.S. Islam, M.S. Kaiser, and M. Mahmud, "Explainable Mortality Prediction Model for Congestive Heart Failure with Nature-Based Feature Selection Method", *Applied Sciences*, Vol. 13, No. 10, p. 6138, 2023.