"Heart Disease Prediction and Severity Level Classification": A Machine Learning approach with Feature Selection technique

2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) | 979-8-3503-3509-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCCNT56998.2023.10308175

Sobia Mir
School of Computer Science and Engineering
Shri Mata Vaishno Devi University
Katra, Jammu and Kashmir, India
sobiamir24@gmail.com

Sunanda
School of Computer Science and Engineering
Shri Mata Vaishno Devi University
Katra, Jammu and Kashmir, India
sunanda.gupta@smvdu.ac.in

***Abstract*— Heart disease is a major public health concern that affects millions of people worldwide. Machine learning algorithms have shown great potential in predicting heart disease, but selecting the most relevant features is crucial for accurate prediction. In this study, we explore two feature selection techniques- DFO and GWO and their impact on heart disease prediction and severity level of heart disease using machine learning algorithms. Our model is evaluated on two datasets Heart.csv for binary classification and processed-Cleveland for multiclass. We assess the effectiveness of various algorithms, including Random Forest, Decision tree, XGBoost, KNN, and Voting classifier, by examining their accuracy, precision, recall, and F1-score. Using Dragonfly Optimization as feature selection technique resulted in an accuracy of 100% for XGBoost and 99.69% for Decision Tree for predicting the presence and absence of heart disease (Binary), and accuracy of 100% for decision tree and 99.72% for XGBoost for predicting the severity level of the disease (Multiclass). Our results demonstrate the effectiveness of feature selection in improving heart disease prediction and severity level accuracy and highlight the potential of machine learning in assisting clinical decision making.**

Keywords*— *heart disease, machine learning, feature selection, severity classification.

I. INTRODUCTION

Heart disease is a leading cause of morbidity and mortality worldwide. According to the World Health Organization (WHO), cardiovascular diseases are responsible for approximately 17.9 million deaths annually, accounting for approximately 31% of all global fatalities. Early detection and accurate prediction of heart disease can greatly improve patient outcomes, but traditional diagnostic methods can be time-consuming and costly. Machine learning algorithms have emerged as a promising approach for predicting heart disease by analyzing large datasets and identifying patterns and relationships among various risk factors. However, the accuracy of these algorithms can be influenced by the selection of features used for prediction. Feature selection techniques aim to identify the most relevant and informative features from a dataset, thereby improving the accuracy and interpretability of the model. In this study, we aim to explore the effectiveness of different feature selection techniques in predicting heart disease using machine learning algorithms. We will evaluate the performance of various algorithms and compare their accuracy, precision, recall, and F1-score. Our study can contribute to the development of effective and efficient prediction models for heart disease and improve patient outcomes. Recent literature has shown a growing interest in using machine learning algorithms for heart disease prediction. Feature selection techniques have also been explored in previous studies to improve heart disease prediction accuracy.

II. LITERATURE SURVEY

A variety of experiments and research projects have been conducted in recent years, with the relevant major papers being published, with rising development in the fields of medical science and machine learning.

[1] Rahma Atallah et al. in 2019 utilized the well-known UCI repository called the CLEVELAND DATASET to develop an enhanced and dependable model for identifying the likelihood of having a heart condition. Through the implementation of an ensemble learning approach, their model achieved an accuracy rate of 90%, surpassing that of other classifiers.

[2] Amin Al Haq et al. conducted a study in 2019 to evaluate the effectiveness of various algorithms. They employed ensemble learning methods such as bagging, boosting, and stacking to improve the accuracy of their models. The classification capabilities of the SVM were compared to those of ML algorithms, and relief feature selection was used to enable the SVM to classify features with 88% accuracy. The use of ensemble learning approaches significantly increased the accuracy of the support vector machine classification to 92.30%. The BPNN demonstrated a 93% accuracy rate in their experiments. Based on all of the experimental results and outcomes, the deep neural network method (BPNN) was found to be the most effective in detecting heart disease.

[3] In 2019, Liaqat Ali et al. put forth an automated diagnostic system for diagnosing heart disease. The proposed classification utilized a deep neural network for classification and 2 statistical models for feature refinement. The system's effectiveness was evaluated using six different evaluation criteria. Through the implementation of a chi-square DNN model, the proposed method increased the accuracy of heart disease detection to 93.33%. Although the study did not explore the predicted time complexity of a hybrid diagnostic system.

[4] In 2020, Norma Latif Fitriyani and colleagues presented an effective model for detecting heart disease. The model incorporated machine learning algorithms such as SMOTE_ENN, DBSCAN, and XGBoost. DBSCAN was utilized to identify and eliminate outlier data, while SMOTE_ENN balanced the uneven training dataset, and XGBoost was employed to train and generate the prediction model. Experimental results indicated that the proposed model outperformed existing models and prior research,

achieving accuracy levels of up to 98.40% and 95.90% for the Cleveland and Statlog datasets, respectively.

[5] In 2020, Jian Ping Li et al. introduced a machine learning-based diagnostic system for predicting heart disease. This system was developed using various algorithms, with traditional feature selection algorithms utilized to address the feature selection problem. The Cleveland heart disease dataset was employed for this study, and the effectiveness of the diagnostic system was evaluated using performance evaluation indicators. Compared to previously proposed methods, the support vector machine's accuracy utilizing the feature selection algorithm (FCMIM) was 92.3%, which was deemed favorable

[6] In 2020, Jikuo Wang et al.presented a 2-level model based on stacking, comprised of a base-level and a meta-level. The base-level prediction classifiers served as input for the meta-level classifiers. The classifier with the lowest correlation was initially determined using the greatest information and Pearson-correlation coefficients, and an enumeration approach was employed to identify the top classifiers to combine for achieving the best final result. Based on experimental results, the proposed model was successful in detecting CHD, achieving sensitivity, accuracy, and specificity levels of 95.84%, 95.43%, and 94.44%, respectively.

[7] Bayu Adhi Tama et al. proposed a CHD prediction model using ensemble in 2020. The suggested methodology involved stacking three separate ensemble learners. To evaluate the proposed approach, four publicly accessible datasets were utilized: Z-Alizadeh Sani, Cleveland, Statlog, and Hungarian. Based on the experimental findings, the model was able to surpass state-of-the-art CHD recognition approaches in terms of accuracy, F1, and AUC value. The accuracy achieved for the Z-Alizadeh Sani, Cleveland, Statlog, and Hungarian datasets was 98.13%, 93.55%, 86.49%, and 91.18%, respectively.

[8] ABDALLAH ABDELLATIF et al.proposed a model in 2022 that integrates Inf-FSs, IWRF, and BO. By utilizing these three techniques, the most important features are selected, the classification problem of imbalanced data in medical datasets is addressed, and the weighting factor is adjusted. To evaluate the generated model and compare it to earlier research, two open datasets were used. Based on the experimental findings, the proposed model was more successful in producing better outcomes without altering the data distribution. Additionally, the suggested improved weighted random forest outperforms the standard RF in terms of CVD detection by 3.62% and 6.3%.

[9] Abdallah Abdellatif et al. in 2022 proposed a decision support system for detecting the presence of heart disease and determining its severity level. The model incorporated ET, SMOTE, and HB algorithms. The model was evaluated using various performance metrics, and the results demonstrated its superiority over other models, achieving up to 99.2%, 99.33%, 99%, and 0.983 for the Cleveland dataset, and 98.52%, 98.8%, 98.080%, and 0.960% for the Statlog dataset. For the multiclass classification of severity levels, the model achieved 0.939%, 95.73%, 96.35%, 95.73%, 95.78%, and MCC, accuracy, precision, recall, and f1-score, respectively.

[10] In 2022, Ghulab Nabi Ahmad and colleagues proposed a machine learning-based medical diagnosis system for predicting heart disease using GridSearchCV. They employed several machine learning models such as KNN, GBC, SVM, and LR in combination with GridSearchCV, and verified their performance using 5-fold cross-validation. Multiple datasets, including Cleveland, Switzerland, Hungary, and Long BeachV, were considered in the experiments. The results indicated that the XGBoost classifier with GridSearchCV achieved the highest testing accuracy of 100% and training accuracy of 99% for all datasets. Moreover, among all the other classifiers, XGBoost also exhibited the best accuracy even without GridSearchCV.

TABLE I. COMPARISION OF LITERATURE SURVEY

Authors	Year	Dataset	Classifiers	Feature selection used	Accuracy
Atallah et al. [1]	2019	Cleveland	Machine learning ensemble techniques: SGD, KNN, Random Forest, Logistic Regression, Hard Voting Ensemble Method	No	Accuracy of Hard Voting Ensemble Method: 90%
Haq et al [2]	2019	Cleveland	SVM, LR, KNN, NB, ANN, DT and Ensemble learning techniques such as bagging, boosting, and stacking, BPNN algorithm	Relief	DT-78.54%, NB- 86%, SVM- 92.30% BPNN-93%
Ali et al. [3]	2019	Cleveland	DNN	Chi square (for feature elimination)	prediction accuracy of 93.33%
Fitriyani et al. [4]	2020	Statlog and Cleveland	SMOTE-ENN, and XGBoostbased ML algorithm.	N/A	Accuracy: 95.90% for statlog dataset. 98.40% for Cleveland dataset.
Li et al. [5]	2020	Cleveland	KNN, LR, ANN, SVM, NB, and DT	Relief, MRMR, LASSO, LLBFS, FCMIM	Accuracy of SVM with FCMIM: 92.37%

Wang et al. [6]	2020	Z-Alizadeh Sani, Statlog, SPECTF, big dataset	ML: Extra tree, RF, ADB, SVC, multi-MLP, XGB, GPC, GNB, LR, GB	Chi square, RFECV, XGB	Accuracy: 95.43% on Z-Alizadeh Sani Accuracy: 90.7% on statlog 92.2% on SPECTF dataset 73.2% on big dataset
Tama et al. [7]	2020	Z-Alizadeh Sani, statlog, Cleveland, and Hungarian	RF, GB machine, extreme gradient boosting machine	Correlation-based FS + PSO	Accuracy: Two-tier ensemble PSO-based feature selection-98.13% (Z-Alizadeh Sani), 93.55(statlog), 86.49(Cleveland), 91.18% (Hungarian dataset).
Abdellatif et al. [8]	2022	Statlog and Cleveland	Improved Weighted Random Forest, and Bayesian optimization	INF-FSs	Accuracy: IWRF (with FS)-97.7%(statlog) 95.9%(Cleveland) 98.3% with optimization (statlog), 97.2% with optimization (Cleveland).
Abdellatif et al. [9]	2022	Statlog and Cleveland	SVC, KNN, LR, SGD, TREEBASED ENSEMBLE SMOTE, ET.	No	Accuracy: (SMOTE+ ET) 99.2%(Cleveland), 98.52%(Statlog), 95.73% (severity level).
Ahmad et al. [10]	2022	Cleveland, Hungary, Switzerland, longbeachCV	LR, SVM, XGBoost, RF. KNN GridSearchCV	No	Accuracy: XGBoost with GridSearchCV: -100%
Proposed Method	2023	Cleveland-statlog-hunagary-Final, processed-cleveland	Random forest, XGBoost, Decision tree, KNN, Voting	Dragonfly optimizer, grey wolf optimizer	Accuracy:((binary) XGBoost =100%, Decision tree=99.69% (using DFO) Accuracy: (multiclass) XGBoost=99.72%, decision tree=100% (using DFO)

III. PROPOSED METHADOLOGY

A. Architecture of the proposed methadology

The proposed approach aims to achieve accurate prediction for heart disease prediction in two scenarios: when cardiovascular disease (CVD) is absent or present, and CVD severity levels which are Absent, low, medium, high, severe. Our evaluation of this proposed approach involves using two datasets obtained from the UCI machine learning repository: the Heart-Cleveland-Statlog-Hungary-Final dataset for binary classification which contains 14 attributes and the processed-Cleveland dataset for multiclass evaluation which also contains 14 attributes. Fig. 1 shows the flowchart of our proposed approach. This approach involves the following steps:

Step 1: *Data Collection:* The first step is to collect a dataset of patient information,
including their medical history, symptoms, and test results.

Step 2: *Data Pre-processing:* Before feeding the data to the machine learning models, it is important to pre-process the data to improve the accuracy of the prediction. This includes normalisation to scale the numerical data between 0 and 1, and label encoding to convert categorical data to numerical data.

Step 3: *Feature Selection:* The next step is to select the most relevant features that are useful in predicting the presence and absence of heart disease and severity of heart disease. We have used two feature selection technique Dragonfly Optimiser and Grey Wolf Optimiser.

Step 4: *Model Training:* After selecting the relevant features, the next step is to train the machine learning models using the pre-processed data. We trained 5 different machine learning models XGBoost, KNN, DT, RF, and Voting.

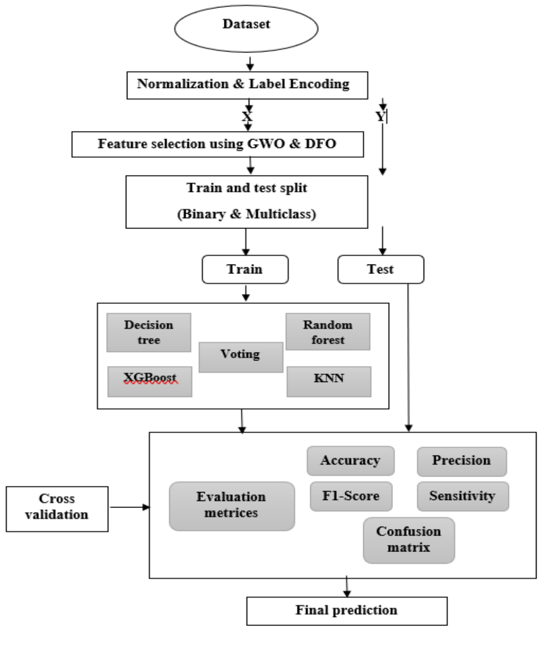

Fig. 1. Flowchart of the proposed mode

Step 5: *Model Evaluation:* Once the models are trained, their training performance is evaluated using appropriate evaluation metrics such as accuracy, precision, recall, and F1 score.

Step 6: *Testing:* The final step is to test the performance of the best performing model on a test data. This helps to verify the model's effectiveness in predicting heart disease accurately.

Step 7: *Cross validation:* After that 10-fold cross validation is used for evaluating the performance of the proposed model. It provides a more accurate estimate of model performance.

Step 8: In the last step, the model is evaluated and benchmarked with the existing models and the best model is chosen.

B. Feature selection techniques

1) Grey Wolf Optimization:

A metaheuristic optimization algorithm Grey Wolf Optimization (GWO) is inspired by the hunting behaviour of grey wolves. The algorithm starts by initializing a population of grey wolves, and then evaluates their fitness based on an objective function. The algorithm then selects the alpha wolf, beta wolf, and delta wolves by considering their superior fitness values, and proceeds to adjust their positions using a series of search equations. The remaining wolves undergo updates to their positions, which are determined by referencing the positions of the alpha, beta, and delta wolves. This process is repeated until a stopping criterion is met, such as a maximum number of iterations or a desired level of convergence. The GWO algorithm has been shown to be effective at solving a variety of optimization problems.

In the study [13], the authors present a new binary version of the GWO algorithm for feature selection in classification tasks. The proposed binary version uses two different approaches. The two approaches are utilized to search for the feature subset that maximizes classification accuracy and minimizes the number of selected features. The performance of the proposed binary GWO is compared to two commonly used optimizers, Particle Swarm Optimization and Genetic Algorithms, using 18 datasets from the UCI repository. The results demonstrated that the suggested binary GWO displays effectiveness in identifying optimal combinations of features, irrespective of the initialization process or the stochastic operators employed. A study [14] proposed a modified GWO algorithm for feature selection in datasets with a high dimensionality. During the initialization phase, the modified GWO algorithm integrated the Relief algorithm and Coupla entropy, while also introducing two novel search strategies. The experiments conducted on 10 gene expression datasets, characterized by high dimensionality and small sample sizes, revealed that the proposed algorithm selected less than 0.67% of the features. This selection process resulted in improved classification accuracy and demonstrated competitive performance when compared to advanced feature selection methods. The authors highlighted the algorithm's ability to balance exploration and exploitation and the significant improvements made to the search capability of the Gray Wolf Optimization algorithm.

The flowchart shown in Fig. 2 represents the iterative process of the Grey Wolf Optimization algorithm, where the wolves move towards the best solution found so far in each iteration. The algorithm aims to find the optimal subset of features that maximizes the performance metric used for evaluation.

The GWO algorithm is based on the leadership hierarchy of the grey wolf pack, which consists of an alpha, beta, delta, and omega wolf. The position of each wolf in the pack represents a potential solution to the optimization problem being solved. The algorithm starts with a randomly generated initial population of wolves, and then iteratively updates the positions of the wolves based on their fitness values.

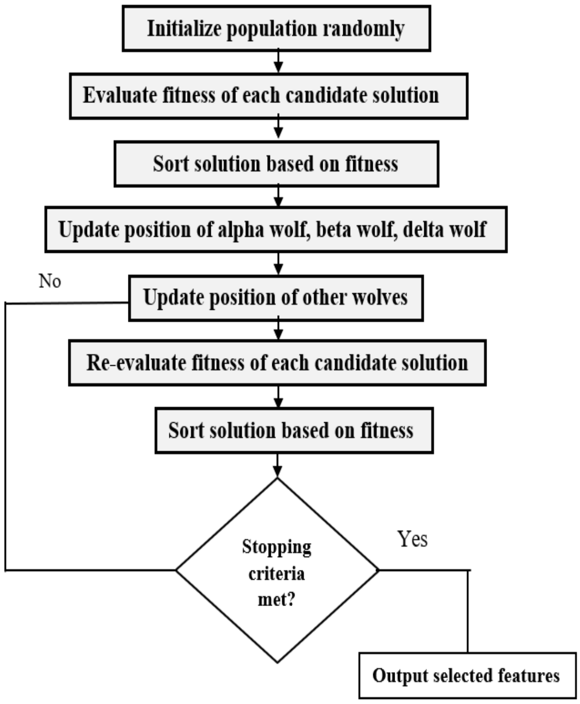

Fig. 2. Working of Grey Wolf Optimizer.

The fitness function in GWO defines how well a particular wolf's position solves the optimization problem. The fitness function is problem-specific and is usually defined based on the objective function of the optimization problem. For example, if we are optimizing a mathematical function, the fitness function could be defined as the negative value of the objective function, as we are trying to minimize the function. The GWO algorithm iteratively updates the positions of the wolves based on the positions of the alpha, beta, delta, and omega wolves. The update equations for each wolf are as follows:

Alpha wolf:

$$X\_alpha = X\_alpha + A * D\_alpha \tag{1}$$

Beta wolf:

$$X\_beta = X\_beta + A * D\_beta \tag{2}$$

Delta wolf:

$$X\_delta = X\_delta + A * D\_delta \tag{3}$$

Omega wolf:

$$X\_omega = X\_omega + A * D\_omega \tag{4}$$

Where $X\_alpha,\ X\_beta,\ X\_delta,\ and\ X\_omega$ are the positions of the alpha, beta, delta, and omega wolves, respectively. A is the search step size, which is a function of the iteration number and the maximum number of iterations. $D\_alpha,\ D\_beta,\ D\_delta,\ and\ D\_omega$ are the distance vectors of each wolf to the alpha, beta, delta, and omega wolves, respectively, and are calculated by following equations:

$$D\_alpha = abs\big(C1 * X\_alpha - (C2 * X1 + C3 * X2)\big) \tag{5}$$

$$D\_beta = abs\big(C1 * X\_beta - (C2 * X1 + C3 * X3)\big) \tag{6}$$

$$D\_delta = abs\big(C1 * X\_delta - (C2 * X1 + C3 * X3)\big) \quad (7)$$

$$D\_omega = abs\big(C1 * X\_omega - (C2 * X1 + C3 * X4)\big) \quad (8)$$

Where $C1$, $C2$, and $C3$ are random coefficients $X1$, $X2$, $X3$, and $X4$ are random positions of four different wolves.

2) Dragonfly Optimization:

Dragonfly Optimization (DFO) is a relatively new nature-inspired optimization algorithm that is based on the social behaviour of dragonflies. DFO has shown promising results in solving various optimization problems, including feature selection in machine learning. Feature selection is an important pre-processing step in machine learning that aims to identify the most informative subset of features from a high-dimensional dataset.

In recent years, DFO has been applied to feature selection in several machine learning applications, demonstrating its effectiveness in improving the classification accuracy and reducing the feature subset size.

The paper [11] proposes a Spark-based distributed dragonfly algorithm for feature selection, which outperforms existing methods on accuracy and efficiency. The proposed approach is evaluated on various datasets and outperforms state-of-the-art techniques in terms of accuracy and efficiency, demonstrating its suitability for big data feature selection. Similarly, authors in [12] proposed a system to categorize breast cancer tumours into benign and malignant using wrapper feature selection based on the Dragonfly optimization technique. The study proposes a strategy to create a subset of features that can accurately classify breast cancer. The study uses a hard vote classifier to evaluate the chosen feature subsets by the Dragonfly optimization algorithm. The experiment's findings demonstrate that the accuracy of the selected features using the voting classifier is higher than using all features.

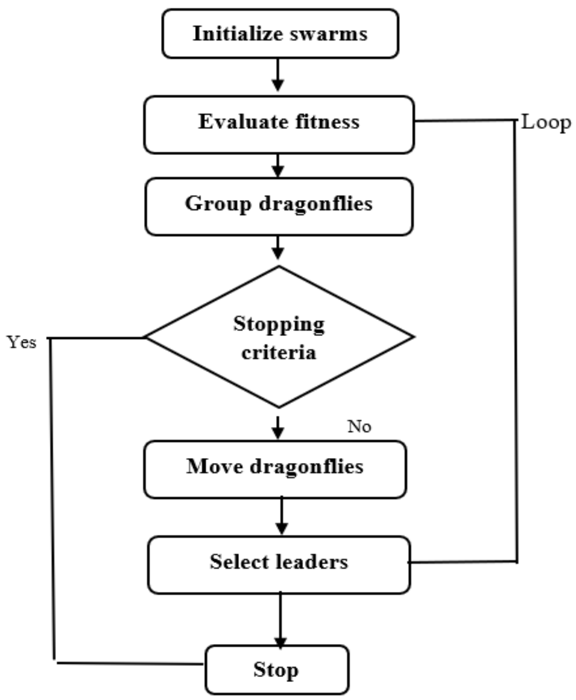

Fig. 3. working of dragonfly optimizer.

The flowchart shown in Fig. 3 is a diagrammatic representation of the Dragonfly Algorithm for feature selection. The flowchart depicts the steps involved in the algorithm and how they are interconnected.

The mathematical equations for the DFO algorithm is described as follows:

a) Initialization: Initialize the population of dragonflies

$$X = \{x_1, x_2, \dots, x_n\} \quad (9)$$

where n is the population size, and xi is a d-dimensional vector representing a candidate solution.

b) Searching phase: Each dragonfly randomly explores the search space by adjusting its position xi according to the following equation:

$$x_i(t+1) = x_i(t) + \alpha\ rand * \left(x_j(t) - x_k(t)\right) \quad (10)$$

where t is the current iteration, αrand is a random parameter between -1 and 1, j and k are two random indices in the population X.

c) Chasing phase: Each dragonfly follows the leader dragonfly with the highest fitness value. The position of each dragonfly is updated according to the following equation:

$$x_i(t+1) = x_i(t) + \beta * exp\big(-\gamma * d(x_i, x_{best})\big) * \big(x_{best} - x_i(t)\big) \quad (11)$$

where β , γ are the scaling factors, $d(x_i, x_{best})$ is the Euclidean distance between x_i and x_{best} , and x_{best} is the position of the best dragonfly in the population.

d) Swarming phase: All dragonflies converge towards the best solution found so far by adjusting their positions according to the following equation:

$$x_i(t+1) = x_i(t) + \varepsilon * \big(x_{best} - x_c(t)\big) \quad (12)$$

where ε is the swarm factor, and $x_c(t)$ is the centroid of the current population.

e) Termination: The algorithm terminates when a stopping criterion is met, such as a maximum number of iterations or a minimum fitness value.

In the context of feature selection, the fitness function is typically defined as the classification accuracy of a machine learning model trained on the selected subset of features. The DFO algorithm can be used to search for the optimal subset of features that maximizes the fitness function.

IV. EXPERIMENTS AND RESULTS

A. Heart disease prediction (Binary):

1) Without feature selection:

Table 1 presents the results (in percentage) of heart disease prediction using five different classifiers without feature selection. The Decision Tree Classifier (DT) achieved the highest accuracy and F1-score, making it a good model for heart disease prediction. The KNeighbors Classifier(KNN) had the lowest accuracy, while the RandomForestClassifier(RF) had a high accuracy but not as high as the Decision Tree Classifier. The Voting Classifier(VC) achieved lower accuracy. Finally, the XGBClassifier achieved(XGB) the almost same performance as the Decision Tree Classifier, indicating that it is also a good model for predicting heart disease. Overall, these results suggest that ML models have the low accuracy and other parameters as compared to performance of ML models with feature selection.

TABLE II. PERFORMANCE OF ML MODELS WITHOUT FEATURE SELECTION

Model	Accuracy	Precision	Recall	F1-Score	Specificity
DT	95.53	95.26	94.08	95.22	94.78
KNN	73.17	73.07	73.78	73.43	72.58
RF	92.07	92.08	92.08	92.08	92.05
VC	83.41	78.99	91.26	84.68	75.49
XGB	94.53	94.43	94.08	94.25	94.36

2) Using Dragonfly Optimization (DFO):

In our experiment DFO selected 10 features out of 14 features. Table 2 shows the performance of different machine learning models in terms of accuracy, precision, recall, F1-score, and specificity. Random forest has the accuracy of 99.57% with 100% recall and an F1-score of 99.59%. The decision tree has perfect precision and F1-score, with a recall of 99.19%. XGBoost achieved highest scores across all metrics. KNN has lower accuracy (83.36%) but relatively high recall (87.90%). The voting model has an accuracy of 84.87%, with precision and specificity (87.28% and 86.84%, respectively) being better than recall and F1-score (83.06% and 85.12%, respectively).

TABLE III. PERFORMANCE OF MACHINE LEARNING MODELS USING DFO.

Model	Accuracy	Precision	Recall	F1-score	Specificity
DT	99.57	100	99.19	99.59	100
KNN	83.36	81.95	87.90	84.82	78.94
RF	99.57	99.20	100	99.59	99.12
VC	84.87	87.28	83.06	85.12	86.84
XGB	100	100	100	100	100

3) Using grey wolf optimization(GWO):

In our experiment GWO selected 12 features out of 14. Table 3 shows the per-formance of five classification algorithms on a given dataset, measured using accuracy, precision, recall, F1-score, and specificity. The Random Forest and XGBoost algorithms had the highest accuracy, precision, recall, F1-score, and specificity. The K-Nearest Neighbours (KNN) algorithm had the lowest perfor-mance, while the voting ensemble had moderate performance. Overall, the XGBoost algorithm had the highest performance, achieving an accuracy of 95.37%, precision of 94.11%, recall of 97.70%, F1-score of 95.88%, and specific-ity of 92.52%.

TABLE IV. PERFORMANCE OF MACHINE LEARNING MODELS USING GWO

Model	Accuracy	Precision	Recall	F1-score	Specificity
DT	90.33	90.29	92.36	91.32	87.85
KNN	79.84	75.39	72.51	73.93	71.02
RF	94.97	94.44	94.89	93.13	90.85
VC	85.29	85.82	87.78	86.79	82.24
XGB	95.37	94.11	97.70	95.88	92.52

The above results shows that DFO performed better than GWO and selected the best features for heart disease prediction. Among 5 machine learning models XGBoost and Random Forest performed best

B. Severity level classification (Multiclass):

1) Using dragonfly optimization

Table 4 shows the performance of 5 machine learning models for the severity level prediction of heart disease. The models were assessed based on their accuracy, precision, recall, F1-score, and specificity. The Decision Tree and XGBoost models achieved the highest accuracy scores of 100% and 99.72%, respectively. The Decision Tree model had a perfect score in all metrics, indicating it correctly classified all the data points. XGBoost also had high precision, recall, and F1-score values. The Random Forest model achieved an accuracy of 98.03%, which is lower than the Decision Tree and XGBoost models. However, it had a high specificity score of 98.59%, suggesting that it correctly predicted negative cases. KNN had the lowest accuracy of 81.63%, indicating its poor performance in this task. The model's precision, recall, and F1-score values were also relatively low compared to other models. The Voting ensemble model achieved an accuracy of 89.50%, which is lower than all other models except for KNN. However, it had a relatively high F1-score, indicating its ability to balance precision and recall. Overall, the results suggest that Decision Tree and XGBoost are the best models for predicting the severity level of heart disease, while KNN and Voting are less effective for this task.

TABLE V. PERFORMANCE OF MACHINE LEARNING MODELS ON SEVERITY LEVEL USING DFO.

Model	Accuracy	Precision	Recall	F1-score	Specificity
DT	100	100	100	100	100
KNN	81.63	80.30	79.98	77.36	84.76
RF	98.03	94.93	89.92	91.97	98.59
VC	89.50	73.67	70.22	82.25	91.87
XGB	99.72	99.21	99.16	99.57	99.49

V. CONCLUSION

The application of machine learning techniques such as XGBoost, KNN, DT, RF, and Voting in combination with feature selection techniques Dragonfly Optimiser and Grey Wolf Optimiser have proven to be effective in predicting heart disease with high accuracy. Specifically, using Dragonfly Optimisation as feature selection technique resulted in an accuracy of 100% for XGBoost and 99.69% for Decision for heart disease prediction (Binary), and accuracy of 100% for decision tree and 99.72% for predicting the severity level of the disease using the same algorithm. This is a remarkable achievement as accurate prediction of heart disease can lead to early detection, timely treatment, and improved patient outcomes. Therefore, the use of machine learning techniques, particularly with the incorporation of feature selection techniques, can greatly enhance the accuracy and effectiveness of heart disease prediction models. Further research and development in this field can potentially lead to the creation of more advanced and efficient heart disease prediction models that can contribute to better patient care and outcomes.

REFERENCES

[1] R. Atallah and A. Al-Mousa, “Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method,” 2019 2nd International Conference on New Trends in Computing Sciences, ICTCS 2019 - Proceedings, Oct. 2019, doi: 10.1109/ICTCS.2019.8923053.

[2] A. U. L. Haq et al., “Identifying the Predictive Capability of Machine Learning Classifiers for Designing Heart Disease Detection System,” 2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2019, pp. 130–138, Dec. 2019, doi: 10.1109/ICCWAMTIP47768.2019.9067519.

[3] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, “An Automated Diagnostic System for Heart Disease Prediction Based on $\chi 2$ Statistical Model and Optimally Configured Deep Neural Network,” IEEE Access, vol. 7, pp. 34938–34945, 2019, doi: 10.1109/ACCESS.2019.2904800.

[4] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, “HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System,” IEEE Access, vol. 8, pp. 133034–133050, 2020, doi: 10.1109/ACCESS.2020.3010511.

[5] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, “Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare,” IEEE Access, vol. 8, pp. 107562–107582, 2020, doi: 10.1109/ACCESS.2020.3001149.

[6] J. Wang et al., “A Stacking-Based Model for Non-Invasive Detection of Coronary Heart Disease,” IEEE Access, vol. 8, pp. 37124–37133, 2020, doi: 10.1109/ACCESS.2020.2975377.

[7] B. A. Tama, S. Im, and S. Lee, “Improving an Intelligent Detection System for Coronary Heart Disease Using a Two-Tier Classifier Ensemble,” Biomed Res Int, vol. 2020, 2020, doi: 10.1155/2020/9816142.

[8] A. Abdellatif, H. Abdellatef, J. Kanesan, C. O. Chow, J. H. Chuah, and H. M. Gheni, “Improving the Heart Disease Detection and Patients’ Survival Using Supervised Infinite Feature Selection and Improved Weighted Random Forest,” IEEE Access, vol. 10, pp. 67363–67372, 2022, doi: 10.1109/ACCESS.2022.3185129.

[9] 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, doi: 10.1109/ACCESS.2022.3191669.

[10] G. N. Ahmad, H. Fatima, Shafiullah, A. Salah 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, doi: 10.1109/ACCESS.2022.3165792.

[11] H. Chen, D. Liu, L. Han, S. Yao, C. Jin, and X. Hu, “A Spark-based Distributed Dragonfly Algorithm for Feature Selection,” in 2020 15th International Conference on Computer Science & Education (ICCSE), IEEE, Aug. 2020, pp. 419–423. doi: 10.1109/ICCSE49874.2020.9201896.

[12] N. Mohammed Majeed and F. Mahmood Ramo, “Implementation of Features Selection Based on Dragonfly Optimization Algorithm,” Technium: Romanian Journal of Applied Sciences and Technology, vol. 4, no. 10, pp. 44–52, Nov. 2022, doi: 10.47577/TECHNIUM.V4I10.7203.

[13] E. Emary, H. M. Zawbaa, and A. E. Hassanien, “Binary grey wolf optimization approaches for feature selection,” Neurocomputing, vol. 172, pp. 371–381, Jan. 2016, doi: 10.1016/J.NEUCOM.2015.06.083.

[14] H. Pan, S. Chen, and H. Xiong, “A high-dimensional feature selection method based on modified Gray Wolf Optimization,” Appl Soft Comput, vol. 135, p. 110031, Mar. 2023, doi: 10.1016/J.ASOC.2023.110031.