LAUTECH Journal of Computing and Informatics (LAUJCI) – ISSN : 2714-4194
Volume 4. No.2, May 2024 – www.laujci.lautech.edu.ng

Multi-Objective Feature Selection Using Non-dominated Sorting Mechanisms and Bi-Directional Elimination for Heart Disease Classification

Ali Muhammad Usman
Computer Science Department,
Federal College of Education (Technical) Gombe, Nigeria
aliakko@yahoo.com

Kabiru Abubakar
Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi, Nigeria
muhammadkabircs@gmail.com

Haruna Chiroma
College of Computing and Engineering Computer Science and Engineering, University of Hafr Al Batin, Qaisumah, KSA
freedonchi@gmail.com

ABSTRACT

Heart Disease (HD) is the primary cause of death worldwide, and its early detection poses challenges for physicians due to a high probability of false positives resulting from the large number of features. To address this issue, Feature Selection (FS) is being employed to identify the most significant feature subsets for HD classification. FS aims to accomplish two main objectives: (1) reducing the number of features and (2) enhancing the classification performance of the selected features. However, most current approaches treat FS as a single objective, combining these aims into one. Consequently, this research introduces a multi-objective FS approach that treats the number of selected features and classification performance as distinct objectives to avoid conflicts arising from a single objective. The study utilizes the Non-dominated Sorting Genetic Algorithm (NSGA-III), a bi-directional elimination-based method, as the optimization algorithm during the search process. The proposed method is evaluated on nine HD datasets of varying complexity, employing five Machine Learning (ML) classifiers as evaluation measures. The results demonstrate that the proposed method performs favorably on most datasets in terms of both the number of selected features and classification performance. The accuracy of each ML classifier surpasses its respective Sensitivity, Specificity, and F-Measure. Support Vector Machine (SVM) excels in terms of the number of selected features, while Decision Tree Classifier (DTC) exhibits superior classification accuracy. K-Nearest Neighbors (KNN) yields promising results for high-dimensional datasets. Comparative analysis with existing studies establishes the superiority of the proposed method.

Key words: Feature Selection, Multi-objective Optimization, Bi-directional elimination, NSGAIII, Heart Disease

1. INTRODUCTION

It is reported that heart disease (HD) is one of the major causes of death in the world according to the World Health Organization, (World Health Organization, 2020). The projection shows that 17.9 million deaths worldwide from HD occurred in 2019 making it 32% mortality rate accounting for 85% of the mortality caused by both HD and stroke (Pathan, Nag, Pathan, & Dev, 2022). Most of the mortality occurs in low- and middle-income nations (Kolukisa & Bakir-Gungor, 2023; Reddy et al., 2022; Abdollahi, & Nouri-Moghaddam, 2022). Early detection of HD is crucial to start treatment with counseling and medication. All efforts made by the expert physicians to detect it went in vain considering the death rate has been incrementing year in and year out (Ramesh et al., 2022). Therefore, computer scientists are currently employing the idea of data mining to forecast the early onset of HD among patients (Pathan, Nag, Pathan, & Dev, 2022; Kolukisa & Bakir-Gungor, 2023; Reddy et al., 2022;

Ramesh et al., 2022). To know the exact features or attributes that contribute significantly to the disease, the term feature selection is used (Abdollahi, & Nouri-Moghaddam, 2022). The idea of FS is thus used in numerous sorts of studies for the classification of HD (Pathan, Nag, Pathan, & Dev, 2022; Kolukisa & Bakir-Gungor, 2023; Reddy et al., 2022; Abdollahi, & Nouri-Moghaddam, 2022; Ramesh et al., 2022).

The goal of feature selection (FS) is to identify the best group of features that significantly contribute to the disease. Using FS will significantly cut down on the number of tests, which saves money (Reddy et al., 2022; Abdollahi, & Nouri-Moghaddam, 2022). FS may function as a filter or a wrapper. Filter measures the usefulness of a subset of features by training a model on it, whereas wrapper techniques test the relevance of a subset of features by their association with the dependent variable. Additionally, more features make filters faster and more useful, whereas a wrapper performs better (Reddy et al., 2022; Liu & Yu, 2005). Choosing the best subset of features with a high classification performance is the goal (Liu & Yu, 2005). Based on that, we suggest implementing a wrapper-based FS.

The wrapper approach to feature selection relied on the machine learning algorithm proposing to fit into a dataset (Liu & Yu, 2005). The wrapper method of feature selection is mainly classified into three, namely, forward selection, Bi-directional elimination also referred to as step-wise selection, and Backward elimination (Liu & Yu, 2005; Jović., Brkić & Bogunović, 2015).

Wrappers employed a greedy search strategy for feature selection by the combination of features as the criterion for the evaluation (Jović., Brkić & Bogunović, 2015). The evaluation criteria for classification can include things like accuracy, precision, recall, f1-score, error rate, sensitivity, and specificity, among others. In the last stage, the combination of features that have the optimal output from the machine learning algorithm is selected. By doing so it becomes computationally expensive, especially on datasets with many features (Jović., Brkić & Bogunović, 2015). On the other hand, there is the possibility of overfitting because it involves modeling the algorithm with different combinations of features.

To avoid the overfitting and computationally expensive issues explained in the preceding section, the paper proposed a multi-objective optimization algorithm; a non-dominated sorting genetic algorithm (NSGA-III) developed by (Liu & Yu, 2005). Instead of the greedy search. The NSGA has the capability of avoiding the problem of overfitting on large dimensional datasets when tuned along with the stepwise refinement selection. The NSGA-III was proposed to deviate from the shortcomings of greedy algorithms and first-generation evolutionary algorithms facing the challenge of lack of elitism and the sustaining of diverse Pareto sets by using shared parameters. It uses a fast non-dominated sorting algorithm, niche preservation, tournament selection, and reference point population abilities (Liu & Yu, 2005); Deb & Jain, 2013).

In this study, we propose NSGA-III along with bi-directional elimination for FS on HD datasets. To the best of our knowledge, this is the pioneer study applying NSGA-III specifically for HD classification. However, the closest work to our proposal in recent times is the study conducted by Gutowski, Schang, Camp, & Abraham (2022). that applied hybrid bi-directional NSGA-III with the gain ratio on the collusion detection features in the Internet of Vehicles, which differs from our proposal.

Thus, the summary of the contributions in this study is as follows:

1. Explored the concept of NSGA-III instead of the greedy search approach for FS in HD classification with improved performance.
2. Used both the feature size and classification performance as a separate objective in the

NSGA-III for better FS.

3. Used the concept of bi-directional stepwise refinement with NSGA-III for FS on HD datasets.
4. The performance of the proposed NSGA-III-based bi-directional FS along with other state-of-the-art approaches indicates improvement.

The entire manuscript is organized into five Sections. Apart from the Section 1, Introduction. Sections 2 and 3 are Review of Related Works and Methodology, respectively. While Section 4 is the Results and Discussion. Finally, Section 5 is Conclusions and Recommendations for Future Studies.

2. FEATURE SELECTION ON HEART DISEASE

Evolutionary algorithms (EA) have been applied to FS in different studies. This sub-section presents the basic background of the EA. EAs are also referred to as evolutionary computation, however, this word is typically used as a general phrase to refer to optimization algorithms inspired by natural intelligent agents such as particle swarm optimization (PSO), cuckoo search, and artificial immune systems, among others Kanwal, Rashid, Nisar, Kim & Hussain (2021).

Babaoglu, Findik, & Ülker (2010) presented an article that compares FS models using binary particle swarm optimization along with GA to determine HD using a support vector machine as the classifier. The experiments were run on HD datasets. The results obtained indicated that the binary PSO algorithm outperforms the GA in terms of accuracy, the number of selected features, and computational complexity. On the other hand, to identify the best subsets of features and aid in the early prediction of heart disease. Jinny & Mate (2021) presented a GA-based FS on heart disease datasets. Modern gradient boosting approaches along with hyperparameter optimization are used to improve the prediction. The results show that more than 20 percent of the features were reduced while maintaining accuracy.

Balamurugan, Ratheesh & Venila (2022) demonstrated a successful HD prediction system that combined stochastic gradient boosting and FS with a recursive feature elimination method. The characteristics are grouped using deep GA and the adaptive Harris Hawk optimization clustering technique. The technique enhances the deep neural network's performance by supplementing its initial weights with an improved GA that suggests better weights for the neural network. In comparison to state-of-the-art methodologies, the testing findings demonstrate that the new FS and classification methodology achieves an accuracy of 98.36% and 97.3%, respectively.

Khourdifi & Baha (2019). proposed a heart disease prediction system using a combined PSO and ant colony optimization algorithm as the search technique. A fast correlation-based FS was used to get the best subsets of features. Besides, five different ML classifiers were used to record the accuracy, f-measure, and sensitivity. The results showed a better performance than other state-of-the-art approaches.

Usman, Yusof, & Naim (2018) developed a heart disease prediction system using cuckoo search algorithms as well as cuckoo optimization algorithms. Both algorithms use the general filter generation to find the best subset of features. The experiment was conducted on five heart disease datasets of various degrees of complexity. The results obtained on the four different classifies showed that cuckoo search performs better on datasets with fewer features whereas cuckoo optimization algorithms perform well on a high dimensional dataset.

Acharjya (2020) introduced a hybrid FS using the concepts of rough set theory along with a cuckoo search algorithm on some heart disease datasets. The proposed method performs better than many others presented in the manuscript both in terms of accuracy and some selected features.

Dubey, Inoue, Birmann & Silva (2022) suggests a novel FS technique that makes use of a chaotic map function to generate fresh random subsets of features and the idea of evolution to direct the algorithm to the best path. Although the experiments were run on 10 datasets including a heart disease dataset using various ML algorithms, they significantly outperformed other approaches in the literature on each dataset.

Vivekanandan & Iyengar (2017), proposed a modified differential evolution (DE) algorithm to perform FS for heart disease and optimization of selected features. Various performance measures of integrating the modified DE with fuzzy AHP and a feed-forward neural network in the prediction of heart disease are evaluated. The accuracy of the proposed hybrid model is 83%, which is higher than that of some other existing approaches. Despite all efforts made by various researchers to solve FS problems and consequently improve the classification performance of the selected features. The alarming rate of heart disease is still worrisome. This is likely because all the articles reviewed in this subsection are based on a single objective FS, whereby mostly classification accuracy is considered at the detriment of the number of selected features or vice versa.

To resolve this conflict in selecting either accuracy or some selected features in the single-objective FS. The term multi-objective FS is introduced, which considers both classification accuracy and several selected features as a two-objective optimization problem.

Gaspar-Cunha (2010) proposed a multi-objective evolutionary Algorithm (MOEA) for FS on the SPECT heart disease dataset. A support vector machine was used to get the classification accuracy. The experimental results showed the suitability of the method compared to the existing works in the literature.

For the prediction and analysis of heart disease, Bashir, Qamar & Khan (2015) offers an ensemble framework based on an improved bagging approach and a multi-objective weighted voting scheme. Five different datasets related to heart disease were used in the experiment. The experimental results demonstrate that the suggested framework works with all types of qualities and achieves high diagnosis accuracy, with 84.16 percent accuracies in sensitivity, specificity, and f-measure, as well as 93.29 percent accuracies in sensitivity and 96.70 percent in specificity.

With the use of an Active Contour Model-based region-growing technology, Priyatharshini & Chitrakala (2019) introduced a Dual-Phase Multi-objective Optimization strategy. Then, a very accurate embedded FS approach is built with a skilled classifier to find calcified objects in the segmented artery. The degree of coronary artery calcium plaque is then measured using the Agatston scoring system. Clinical practice was used to obtain coronary CT images from the AS+CT scanner with a 3 mm slice thickness. Experimental findings show that the suggested strategy increases the precision of lesion detection for improved treatment planning.

Three steps make up the PSO-based multi-objective FS approach that Rostami, Forouzandeh, Berahmand & Soltani (2020) created. In stage 1, a graph representation model of the original features is displayed. The next stage involves calculating the feature centralities for each node in the graph, and the third stage involves the final FS using an enhanced PSO-based search procedure. The suggested method outperforms earlier similar methods in terms of efficiency and efficacy, according to the results on five medical datasets.

Habib, Aljarah, Faris & Mirjalili (2020) presented a multi-objective PSO for FS on various medical datasets. The experimental results indicated the proposed MOPSO performed better than MOEA/D and NSGA-II in terms of classification accuracy as well as some selected features. However, PSO tunes to work based on the MOEA, and hence it was not a surprise to have a better result. The major deficit of the MOPSO is the computational complexity which will take longer time on datasets with higher features.

Li, He, Chen & Pan (2021) suggested a three-objective FS multi-objective large-scale cooperative coevolutionary algorithm. To find the ideal feature subset quickly and effectively, a cooperative searching framework is created. Second, the framework establishes three goals—feature number, classification accuracy, and overall information gain—to direct the evolution of feature combinations. Thirdly, dual indicator-based representations are developed for balancing the convergence and diversity of the representative solution in the framework's coevolution process, while cluster-based decomposition strategies are developed for the framework's decomposition process to reduce computation. Finally, a system for diagnosing heart disease based on the framework is created to demonstrate its viability. Numerical tests show that the suggested strategy exceeds its rivals in terms of both classification accuracy and metrics of the number of features.

Finally, recently on FS to medical binary classification, Gutowski, Schang, Camp, & Abraham (2022) put forth a multi-objective optimization approach. It is based on a Genetic Algorithm and a 3-dimensional Compass, which are intended to direct the search toward a desired trade-off between the number of features, accuracy, and area under the ROCCurve (AUC). On several real-world medical datasets, the proposed Genetic Algorithm with Multi-Objective Compass (GAwC) performs better than any existing competitive GA-based MOFS techniques.

Most of the works conducted as well as the review of the related works on Multi-objective FS are on medical datasets. Besides, the little ones that perform the FS on heart disease are just one single dataset: to be precise the Cleveland and SPECT datasets. Even those that used the heart disease datasets are filter-based FS, where the accuracy obtained from them might not likely be as significant as that of the wrapper-based approaches. Therefore, this study explores the most recent NSGA-III along with wrapper-based stepwise refinement to select the best subsets of features that contribute significantly to heart disease classification.

3. THE PROPOSED METHOD

3.1 Wrapper-Based Feature Selection

The general wrapper method is shown in Algorithm1. It assesses the goodness of each created subset B by using the ML algorithm on the data with the subset of features B and assessing the effectiveness of the outcomes. Consequently, the outcomes of the FS will vary according to the ML used. Different wrapper algorithms can be produced by varying the search strategies using the function generate (DTS) and mining algorithms ML. The selection of feature subsets is governed by ML algorithms; hence the wrapper model typically performs better since the discovered feature subsets are better matched to the preset mining algorithm. As a result, it requires more computing than the filter model.

Algorithm 1 General Wrapper-Based Algorithm

INSERT: $DTS(A_0, A_1, \cdots, A_{n-1})$ // Dataset with N attributes/features for training
B_0 // Begin finding of subset
δ // End criteria
RESULTS: F_{best} // Optimal subset of features

BEGIN
SET: $B_{best} = B_0$;
$\gamma_{best} = eval(B_0, DTS, ML)$; // evaluate B_0 using a ML algorithm
DO BEGIN
$B = generate(DTS)$; // generate subsets for evaluation
$\gamma = eval(B, DTS, ML)$; // evaluate the current subset F by the ML
if (γ out perform γ_{best})
$\gamma_{best} = \gamma$;
$B_{best} = B$;
END DO (δ is Found);
RETURN B_{best};
END;

3.2 Data Selection Approach

The two basic FS techniques used in machine learning are forward selection and backward elimination. Whichever strategy you choose will ultimately depend on the precise facts and objectives you have. Both have advantages and disadvantages.

Forward selection involves starting with a null model or an empty set, fitting each feature into the model one at a time, and choosing the feature with the lowest p-value. By attempting combinations of the previously chosen feature with all other existing features, fit a model with two features at this point. Once more, pick the feature with the lowest p-value. The next step is to fit a model with three features by combining the two features you chose earlier with the other two features. Repeat this process until the set of features that were chosen has a p-value for each feature that is below the significance level. Although this approach is straightforward to use, it may not yield the best selection of features and may be computationally expensive.

Up until the time when we get the whole set of significant features, this process is repeated repeatedly. One such approach that is frequently employed in machine learning is backward elimination. This method is beneficial because it can lessen the likelihood that the data will be overfitted and improve the interpretability of the linear regression model. Although this approach is more computationally efficient, it might not also find the best combination of features.

Bi-directional elimination (Stepwise Selection) Combining forward selection and backward elimination, bidirectional elimination checks each step's inclusion or exclusion of features. It is like the forward selection, but when a new feature is added, it also evaluates the significance of previously added features. If any previously picked features are determined to be inconsequential, that feature is then simply eliminated by backward selection. Bi-directional elimination involves the following steps, in brief:

i. To enter and exit the model, select a significance level (SL in = 0.05 and SL out = 0.05 with a 95% confidence level, for example).
ii. Continue with the forward selection process (the newly added feature must have a p-value less than SL to enter.
iii.Complete all phases of backward elimination (any previously included feature with a p-value > SL out is prepared to exit the model).
iv. To obtain the ultimate optimal set of traits, repeat steps 2 and 3 as necessary.

3.3 Proposed Multi-Objective Algorithm

A multi-objective optimization problem is an optimization problem that involves multiple objective functions. In mathematical terms, a multi-objective optimization problem can be formulated as

$$\text{Min/Max} \quad f_m(x) \qquad m = 1,2, \dots , M$$

$$\text{subject to } g_j(x) \geq 0 \qquad j=1.2. \dots. J$$

where the integer $m \geq 2$. $f_m(x)$ is the m^{th} objective function to be minimize or maximize. In this case the number of selected features are minimize while the classification accuracy is improved in a bi-directional approach.

The concepts of bi-directional; whereby both forward and backward elimination strategies are ensured to avoid the problem of over fitting. The DTS (datasets) are divided into training and testing tests. Then the reference points of each population solution are generated. Bi-directional elimination is applied to evaluate B0 on each dataset by taking into cognizance the feature size and classification performance as two separate objectives.

Five different ML classifiers were employed to determine and record the best classification performance against the chosen subsets of features. These concepts are done using the tournament selection. Similarly, to escape from the local optima and improve searchability niche preservation was used and stored for the next generation. By doing so the complexity of the proposed wrapper-based approach is drastically reduced.

Algorithm 2 Proposed NSGA-III and Bi-directional Wrapper-Based Feature Selection

INSERT: $DTS(A_0, A_1, \cdots, A_{n-1})$ // Dataset with N attributes/features for training
B_0 // Begin finding of subset using bi-directional elimination
δ // End criteria
RESULTS: F_{best} // Optimal subset of features

```
 1: BEGIN
        Divide HD datasets into training and testing set
        Calculate the reference point of each population
 2:   SET: B_best = B_0;
 3:     γ_best = eval(B_0, DTS, ML); // evaluate B_0 using a ML algorithm
        Apply bi-directional elimination
        Use non-dominated sorting mechanism
 4:     DO BEGIN
 5:         B = generate(DTS); // generate subsets for evaluation
 6:         γ = eval(B, DTS, ML); // evaluate the current subset F by the ML
 7:         if (γ outperform γ_best)
 8:             γ_best = γ;
 9:             Use tournament selection
10:             Apply and store niche preservation for next generations
11:         Calculate the accuracy, and others along with the feature size
12:         B_best = B;
13:     END DO (δ is Found);
14:     RETURN B_best;
15: END;
```

3.4 Description of Heart Disease Datasets

Table 1 presents a comprehensive overview of the datasets employed in this study. The research focuses on seven datasets pertaining to heart disease. The "Dataset" column in the table denotes the specific datasets utilized for the study. The "#Features" column represents the total number of features or attributes associated with each dataset. Conversely, the "#Instances" column indicates the number of records contained within each dataset. It is important to note that the table does not include a class label, as all the datasets consist of binary class labels.

Furthermore, it is worth mentioning that all the datasets used in this study are accessible through the renowned machine learning repository known as the University of California Irvine (UCI) machine learning repository. These datasets can be obtained from the following web address: https://archive.ics.uci.edu/ml. Prior to analysis, the datasets have undergone a thorough cleaning process, ensuring the removal of any incomplete instances. Additionally, a few instances that do not contribute significantly to the analysis have been imputed using the reverse and forward fill functionality of the Python Anaconda Navigator

3.5 Performance Evaluation

A dependent criterion used in the proposed wrapper-based approach requires a predetermined ML algorithm in FS and uses the performance of the ML algorithm applied to the selected subset to determine which features are selected. As such, five ML classifiers used include Support Vector Machine (SVM), Gaussian Naıve Bayes (GNB), k-nearest Neighbor (KNN), Random Forest Classifier (RFC), and Decision Tree Classifier (DTC). In each of these classifiers, we recorded the number of selected features, and classification accuracy, along with F-measure, Sensitivity, and Specificity.

Table 1. Heart Disease Datasets

Name of Datasets	#-of-Features	#-of-Instances	Classes
Cleveland	13	303	2
Echocardiogram	12	131	2
Eric	7	209	2
Hungarian	13	294	2
LongBeach VA	13	200	5
SPECT	22	267	4
SPECTF	44	267	4
Z-Alizadeh Sani	55	303	2

4. RESULTS AND DISCUSSION

This section describes the results of the proposed b-directional-based NSGA-III wrapper method. The results are presented in Table 2 and Table 3, respectively. The first subsection describes the results of the selected features. Followed by the evaluation criteria used. The result of each dataset is presented differently. Finally, the complete results are presented, discussed, and analyzed.

4.1 Number of Selected Features

The number of chosen features as well as the accuracy recorded by each dataset using the proposed method is depicted in Table 2. The table describes the results obtained by the proposed wrapper-based method that used both NSGA-IIII and a Bi-directional elimination approach on related HD datasets. From the table, ”#Features” represented the exact number of features on the dataset before applying the proposed FS method. Each of the five classifiers displayed the number of features selected ”Sel. Fts” and Classification accuracy ”Acc.” respectively.

By looking at the table one can see clearly that the selected features for the Cleveland dataset are 4 on SVM, and RFC, while 5 on GNB and KNN respectively. DTC has the least number of selected features on the Cleveland dataset. Similarly, on the other datasets like Echocardiogram, Eric, Hungarian, Long Beach VA and Statlog have 3 as the number of selected features as the least on DTC and SVM. RFC mostly has 4 as the number of selected features except for Statlog which recorded 5 Similar to GNB on the same datasets.

KNN performs well on large dimensional datasets such as SPECT, SPECTF, and Z-Alizadeh Sani compared to most of the ML classifiers employed in this study. On the other hand, RFC performance is worse on most of the high-dimensional datasets. DTC performance is commendable on the smaller datasets while on average GNB performed better in terms of selecting the most suitable subsets of features on all the datasets. SVM selects a fewer number of features compared to all other ML classifiers and hence performs much better in this regard. It can be noticed that from the table ”Average” number of features under the SVM is 4.2 which affirms the fact that SVM selects fewer subsets of features on all the datasets using the proposed method. Alternatively, GNB has the highest number of selected features 7.7 followed by RFC 7.4.

Table 2. Number of Selected Features Vs. Classification Accuracy of the Five Classifiers

Name of Dataset	# of Features	SVM		GNB		KNN		RFC		DTC	
		Sel Fts	Acc	Sel Fts	Acc	Sel Fts	Acc	Sel Fts	Acc	Sel Fts	Acc
Cleveland	13	4	97	5	89	5	91	4	96	3	96
Echocardiogram	12	3	98	4	95	5	90	4	98	3	98

Eric	7	3	94	4	91	5	89	4	91	3	93
Hungarian	13	3	96	4	90	4	96	4	95	3	96
Long Beach VA	13	3	91	4	89	5	89	4	90	3	91
SPECTF	44	6	91	15	88	7	92	18	90	15	90
SPECT	22	6	94	11	89	7	94	13	92	9	94
Statlog	13	3	93	5	90	5	89	5	91	3	93
Z-Alizadeh San	55	7	90	16	88	8	94	11	94	10	94
Total	192	**38**	844	69	809	51	824	67	837	52	**845**
Average	21.3	**4.2**	93.8	7.7	89.9	5.7	91.6	7.4	93.0	5.8	**93.9**

4.2 Performance Measure

In terms of classification Accuracy, Table 2 shows the accuracy of the proposed methods on the five ML classifiers. An accuracy of 98% was found on the Echochardiagram dataset for SVM, RFC, and DTC. Also, Cleveland attained 97% accuracy on SVM more than all other ML classifiers. Similarly, 96% and 94% accuracy was obtained on Hungarian and SPECTF datasets respectively for SVM, KNN, and DTC. Moreover, a 94% accuracy was achieved on Z-Alizadeh Sani datasets on KNN and DTC.

One can observe that SVM performs better mostly on all the binary class datasets. Whereas KNN outperformed most of them on the large dimensional datasets. The performance of DTC and GNB is regardless of the size and class of the dataset. Generally, DTC performs better in terms of classification accuracy on all the datasets. As can be seen from the table it has the highest average accuracy of 93.9% followed by SVM with 93.8% on all the nine datasets. Even though SVM performs much better in terms of several selected features it still performs well on classification accuracy with DTC being better with 0.01%.

Therefore, using the concept of a non dominated sorting mechanism helps in achieving better classification performance both in terms of number of selected features and classification accuracy for classification on HD datasets.

Table 3. Performance Measure of The Proposed Method On All The Datasets

Dataset		Cleve-land	Echocardiogram	Eric	Hungarian	Long Beach VA	SPECTF	SPECT	Statlog	Z-Alizadeh San	Avg
SVM	**Sens.**	95.1	96.0	92.1	94.1	89.2	89.2	92.1	91.1	88.2	91.9
	Spec.	95.5	96.5	92.6	94.6	89.6	89.6	92.6	91.6	88.7	92.4
	F-Mea	96.2	97.2	93.2	95.2	90.2	90.2	93.2	92.2	89.2	93.0
	Acc.	97.0	98.0	94.0	96.0	91.0	91.0	94.0	93.0	90.0	93.8
GNB	**Sens.**	87.2	93.1	89.2	88.2	87.2	86.2	87.2	88.2	86.2	88.1
	Spec.	87.7	93.6	89.6	88.7	87.7	86.7	87.7	88.7	86.7	88.5
	F-Mea	88.2	94.2	90.2	89.2	88.2	87.3	88.2	89.2	87.3	89.1
	Acc.	89.0	95.0	91.0	90.0	89.0	88.0	89.0	90.0	88.0	89.9
KNN	**Sens.**	89.2	88.2	87.2	94.1	87.2	90.2	92.1	87.2	92.1	89.7
	Spec.	89.6	88.7	87.7	94.6	87.7	90.6	92.6	87.7	92.6	90.2
	F-Mea	90.2	89.2	88.2	95.2	88.2	91.2	93.2	88.2	93.2	90.8
	Acc.	91.0	90.0	89.0	96.0	89.0	92.0	94.0	89.0	94.0	91.6
RFC	**Sens.**	94.1	96.0	89.2	93.1	88.2	88.2	90.2	89.2	92.1	91.1
	Spec.	94.6	96.5	89.6	93.6	88.7	88.7	90.6	89.6	92.6	91.6
	F-Mea	95.2	97.2	90.2	94.2	89.2	89.2	91.2	90.2	93.2	92.2

	Acc.	96.0	98.0	91.0	95.0	90.0	90.0	92.0	91.0	94.0	93.0
DTC	**Sens.**	94.1	96.0	91.1	94.1	89.2	88.2	92.1	91.1	92.1	92.0
	Spec.	94.6	96.5	91.6	94.6	89.6	88.7	92.6	91.6	92.6	92.5
	F-Mea	95.2	97.2	92.2	95.2	90.2	89.2	93.2	92.2	93.2	93.1
	Acc.	96.0	98.0	93.0	96.0	91.0	90.0	94.0	93.0	94.0	93.9
Total		1855.6	1895.2	1812.1	1871.4	1780.4	1784.4	1831.9	1804.2	1820.0	1828.3
Average		92.8	94.8	90.6	93.6	89.0	89.2	91.6	90.2	91.0	91.4

4.3 From Table 3, the accuracy is mostly better than others on all the ML classifiers employed in the study. The average accuracy is better across all the ML evaluation measures. The difference between the Specificity and Sensitivity is insignificant, merely less than or equal to 0.5% on most of the datasets. On the other hand, the difference between the accuracy and F-Measure is less than one percent on most of the datasets. Specificity is slightly better than Sensitivity, while F-measure is slightly better than Specificity.

Table 3 below shows the results of all the proposed NSGA-III and Bi-directional wrapper-based methods using various ML classifiers based on the performance measure presented earlier as the wrapper evaluation criteria. From the table ”Sens.”, ”Spec.”, ”F-Mea.”, and ”Acc.” stands for Sensitivity, Specificity, F-Measure, and Accuracy, respectively. Therefore, using the concepts of nondominated sorting mechanism along with bi-directional elimi- nation as a wrapper-based FS one can obtain a better classification performance on most of the heart disease datasets.

4.4 The Overall Results

The overall performance measure on all the nine heart disease datasets along with the five ML classifiers is depicted as shown in Fig 1. From the table, the y-axis represents the percentage from 80-100 attained by each of the datasets on various ML classifiers. Similarly, the x-axis represents the performance measures employed in the study for the respective ML classifier used.

So far none of the ML classifiers recorded less than eighty percent on all the datasets. From the figure, one easily identifies which ML performs better on what dataset. On the other hand, one can also easily identify which ML performs less on which dataset and so on. Therefore, by merely looking at the figure one can see that mostly SVM and DTC dominate all the datasets in terms of Accuracy, F-measure, Specificity, and Sensitivity. Moreover, the classification accuracy outpaced others on all the datasets. Thus, the ML classifiers can easily be used along with the concept of a nondominated sorting mechanism along with bi-directional elimination to explore the best subsets of features and produce promising wrapper evaluation results.

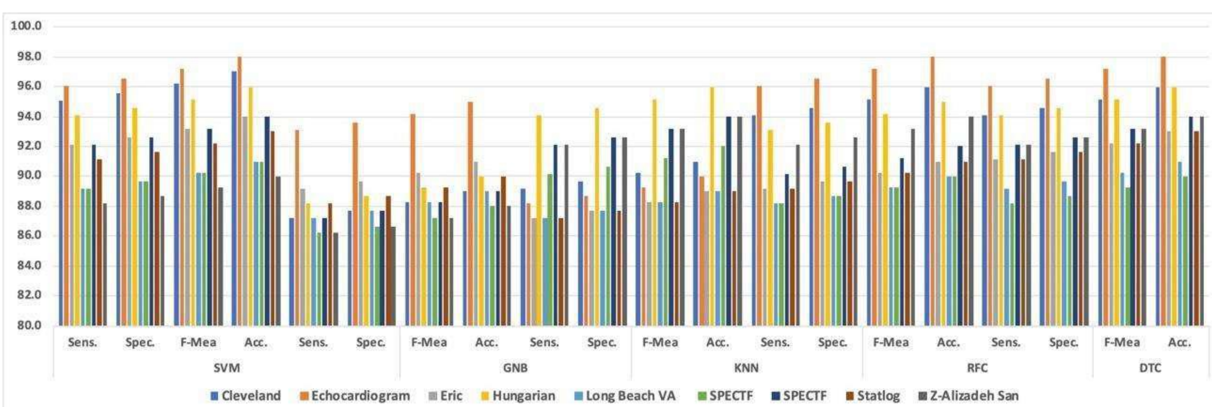

Fig 1: Accuracy, Sensitivity, Specificity and F-Measure of the Results Obtain

Fig 2. shows the highest average performance measures on all the ML classifiers employed in the study. The description of the figure is like others described earlier.

Upon observing the figure, it becomes apparent that the bars on the SVM graph exhibit a predominantly larger size in comparison to the other graphs. Whereas GNB has the least bars its performance is said to be the worst among them. Despite a good performance exhibited by the SVM, Table 3 indicated that the average performance of DTC is 93.9 while that of SVM is 93.8. DTC outperformed SVM with 0.1 percent and is hence considered the best in this study. Therefore, all the wrapper evaluation measures used in this study can evolve the set of features using the nondominated sorting mechanism along with the bi-directional elimination and obtain a better classification performance.

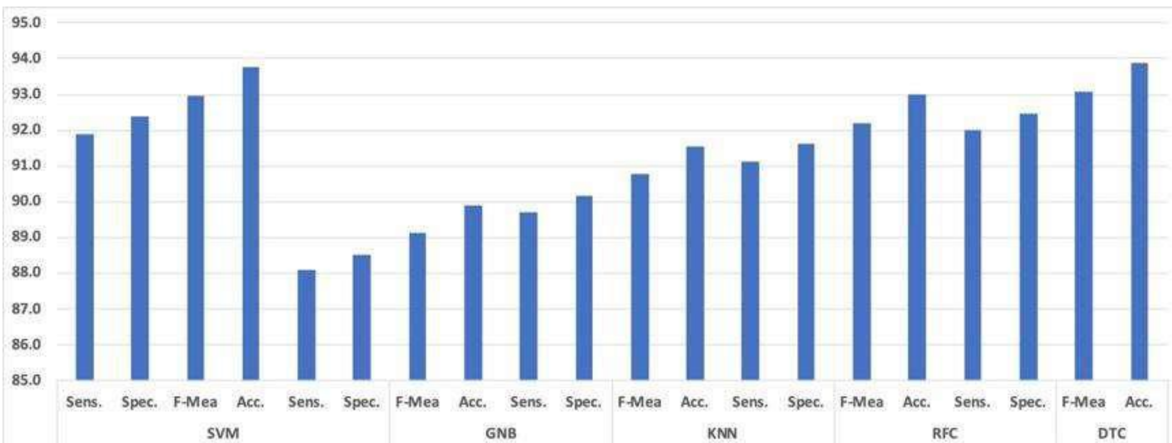

Fig 2: Highest Average Performance Possess By Each Classifier On All the Datasets.

4.5 Comparison With Existing Studies

To be fair in our comparison and to determine the performance of the proposed method we compared it with another multi-objective FS approach that used similar HD datasets. Despite long searching, only one work was reported on multi-objective FS that used an HD dataset. The work of Li, He, Chen, & Pan (2021) uses only the Cleveland HD dataset. The detailed comparison is depicted in Table 4.

Table 4. Performance Measure of The Proposed Method On All The Datasets

Name	Accuracy	Classifier	Selected Features
Proposed Method	97	SVM	4
MLFS-CCDE	86	ELM	6

From the table the name of the column name stands for the name of the method, Accuracy is the highest accuracy achieved by the method. The classifier column shows the best classifier with the highest accuracy and number of selected features by the method. The work of Li, He, Chen, & Pan (2021) selects only six features out of the 13 features of the Cleveland HD dataset and attained an accuracy of 86 percent. There is no doubt that the proposed method in this study outperformed it both in terms of selected features where only four features were selected and yet attained an accuracy of 97 percent.

5. CONCLUSIONS

In the quest to avoid the use of a single objective optimization problem to solve FS the two aims of FS are combined into a single objective at the detriment of one and hence create a conflict. This study presented a multi-objective FS using a bi-directional wrapper-based FS along with a nondominated sorting algorithm (NSGA-III) for HD classification. Another major issue of heart disease is how to get the best subsets of features that contribute significantly to the disease classification without overfitting. Therefore, the concept of bi-directional that utilizes the advantages of both forward and backward elimination was used to curb the issues of overfitting while searching for the best subsets of features. Moreover, a multi-objective optimization algorithm precisely; NSGA-III was introduced as the search technique within the wrapper-based FS rather than using the existing single-objective algorithms.

Nine HD datasets were used in the experiment. The subsets of features were evaluated using five different ML classifiers. The results indicated that the proposed method can significantly select the best subsets of features using the bi-directional non dominated sorting mechanism and yield better classification performance with fewer subsets of features. Comparison amongst the classifiers shows that SVM performed better in terms of the number of selected features compared to all other ML classifiers.

KNN performs better on high dimensional datasets both in terms of selected features and other performance metrics compared to others. On the other hand, DTC performs slightly better than SVM in terms of classification performance. SVM performs better mostly on binary class datasets. Moreover, the proposed method performs better than the existing state-of-the-art approach.

Future research endeavors will encompass the exploration of multi-objective feature selection (FS) techniques on the heart disease (HD) dataset, including scenarios involving more than two classes and imbalanced class labels. It is imperative to investigate alternative multi-objective optimization algorithms such as MOEA/D, MOGA, and others to ascertain their potential for FS on HD datasets, which may yield more promising outcomes. Moreover, an underrepresented aspect in current literature pertains to the treatment of FS as a hybrid multi-objective process that integrates both filter and wrapper methodologies. Therefore, adopting the perspective of filter-wrapper multi-objective FS for HD datasets has the potential to enhance heart disease classification by identifying appropriate subsets of features and achieving superior classification performance.

6. ACKNOWLEDGMENTS

The authors wish to thank the Directorate of Research and Development of Federal College of Education (Technical) Gombe, Nigeria for finding this study worthy to be sponsored by TetFund through its Institutional Based Research (IBR) Grant.

7. REFERENCES

[1] Pathan, M. S., Nag, A., Pathan, M. M., & Dev, S. (2022). Analyzing the impact of feature selection on the accuracy of heart disease prediction. Healthcare Analytics, 2, 100060.

[2] Kolukisa, B., & Bakir-Gungor, B. (2023). Ensemble feature selection and classification methods for machine learning-based coronary artery disease diagnosis. Computer Standards & Interfaces, 84, 103706.

[3] Reddy, K. V. V., Elamvazuthi, I., Aziz, A. A., Paramasivam, S., Chua, H. N., & Pranavanand, S. (2022). An efficient prediction system for coronary heart disease risk using selected principal components and hyperparameter optimization. Applied Sciences, 13(1), 118.

[4] Abdollahi, J., & Nouri-Moghaddam, B. (2022). A hybrid method for heart disease diagnosis utilizing feature selection based ensemble classifier model generation. Iran Journal of Computer Science, 5(3), 229-246.

[5] Ramesh, T. R., Lilhore, U. K., Poongodi, M., Simaiya, S., Kaur, A., & Hamdi, M. (2022). Predictive analysis of heart diseases with machine learning approaches. Malaysian Journal of Computer Science, 132-148.

[6] Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. IEEE Transactions on knowledge and data engineering, 17(4), 491-502.

[7] Jović, A., Brkić, K., & Bogunović, N. (2015, May). A review of feature selection methods with applications. In 2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO) (pp. 1200-1205). Ieee.

[8] Deb, K., & Jain, H. (2013). An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints. IEEE transactions on evolutionary computation, 18(4), 577-601.

[9] Li, H., He, F., Chen, Y., & Pan, Y. (2021). MLFS-CCDE: multi-objective large-scale feature selection by co-operative coevolutionary differential evolution. Memetic Computing, 13, 1-18.

[10] Babaoglu, İ., Findik, O., & Ülker, E. (2010). A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine. Expert Systems with Applications, 37(4), 3177-3183.

[11] Jinny, S. V., & Mate, Y. V. (2021). Early prediction model for coronary heart disease using genetic algorithms, hyper-parameter optimization and machine learning techniques. Health and Technology, 11(1), 63-73.

[12] Kanwal, S., Rashid, J., Nisar, M. W., Kim, J., & Hussain, A. (2021, July). An effective classification algorithm for heart disease prediction with genetic algorithm for feature selection. In 2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC) (pp. 1-6). IEEE.

[13] Balamurugan, R., Ratheesh, S., & Venila, Y. M. (2022). Classification of heart disease using adaptive Harris hawk optimization-based clustering algorithm and enhanced deep genetic algorithm. Soft Computing, 26(5), 2357-2373.

[14] Khourdifi, Y., & Baha, M. (2019). Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. International journal of Intelligent engineering & systems, 12(1).

[15] Usman, A. M., Yusof, U. K., Naim, S., & Naim, S. (2018). Cuckoo inspired algorithms for feature selection in heart disease prediction. International Journal of Advances in Intelligent Informatics, 4(2), 95-106.

[16] Acharjya, D. P. (2020). A hybrid scheme for heart disease diagnosis using rough set and cuckoo search technique. Journal of Medical Systems, 44(1), 1-16.

[17] Dubey, A., Inoue, A. H., Birmann, P. T. F., & Silva, S. R. D. (2022, July). Evolutionary feature selection: a novel wrapper feature selection architecture based on evolutionary strategies. In Proceedings of the Genetic and Evolutionary Computation Conference (pp. 359-366).

[18] Vivekanandan, T., & Iyengar, N. C. S. N. (2017). Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease. Computers in biology and medicine, 90, 125-136.

[19] Gaspar-Cunha, A. (2010, May). Feature selection using multi-objective evolutionary algorithms: application to cardiac SPECT diagnosis. In Advances in Bioinformatics: 4th International Workshop on Practical Applications of Computational Biology and Bioinformatics 2010 (IWPACBB 2010) (pp. 85-92). Berlin, Heidelberg: Springer Berlin Heidelberg.

[20] Bashir, S., Qamar, U., & Khan, F. H. (2015). BagMOOV: A novel ensemble for heart disease prediction bootstrap aggregation with multi-objective optimized voting. Australasian physical & engineering sciences in medicine, 38, 305-323.

[21] Priyatharshini, R., & Chitrakala, S. (2019). An efficient coronary disease diagnosis system using dual-phase multi-objective optimization and embedded feature selection. In Coronary and Cardiothoracic Critical Care: Breakthroughs in Research and Practice (pp. 20-43). IGI Global.

[22] Rostami, M., Forouzandeh, S., Berahmand, K., & Soltani, M. (2020). Integration of multi-objective PSO based feature selection and node centrality for medical datasets. *Genomics*, *112*(6), 4370-4384.

[23] Habib, M., Aljarah, I., Faris, H., & Mirjalili, S. (2020). Multi-objective particle swarm optimization: theory, literature review, and application in feature selection for medical diagnosis. Evolutionary Machine Learning Techniques: Algorithms and Applications, 175-201.

[24] Gutowski, N., Schang, D., Camp, O., & Abraham, P. (2022). A novel multi-objective medical feature selection compass method for binary classification. Artificial Intelligence in Medicine, 127, 102277.

[25] Almutairi, M. S., Almutairi, K., & Chiroma, H. (2023). Selecting Features That Influence Vehicle Collisions in the Internet of Vehicles Based on a Multi-Objective Hybrid Bi-Directional NSGA-III. Applied Sciences, 13(4), 2064.

[26] World Health Organization, (2020). Hearts: technical package for cardiovascular disease management in primary health care: risk-based CVD management. World Health Organization.

Author's Brief Profile

Ali Muhammad Usman holds a PhD, Msc and B. Tech (Hons) degrees in Computer Science from Universiti Sains Malaysia, Adamwawa State University Mubi and Abubakar Tafawa Balewa University Bauchi. Ali is currently a lecturer at the Computer Science Department, Federal College of Education (Technical), Gombe. And a visiting lecturer at Abubakar Tafawa Balewa University Bauchi. He has published several international and journal conferences. His research interests include data mining, computational intelligence, Web/software engineering, optimization, multi-objective optimization, and data security. He is a member of the Computational Intelligence Society of the IEEE. He can be reached by phone on +2348032878120 and through E-mail aliakko@yahoo.com

Kabir Abubakar is currently a masters student at the Department of Mathematical Science, Abubakar Tafawa Balewa University Bauchi, Nigeria. He obtained B Tech (Hons) in Computer Sciece from Modibbo Adama University Yola. His research areas include Software Development, Multi-Objective Optimization Problems, Data Mining, and Artificial Intelligence. He is currently with Briatek Computers Nig. Ltd as a Director and CTO.

Haruna Chiroma is currenly a Lecturer at University of Hafr Al Batin · Computer Science and Engineering Doctor of Philosophy and a visiting lecturer at Abubakar Tafwa Balewa University, Bauchi. He holds a Ph.D. Artificial Intelligence. He was ranked in the top 2% most influential scientists in the world. Expertise on neural networks, currently working on large language models. To understand teaching, he obtained Post-Graduate Diploma in Education. Authored book on computing research, published by IoP Publishers.. Publish in many journals, conferences, books, book chapters and serve as an editor for many journals.