Cardiovascular Diseases Classification Using the Flow Direction Optimization Algorithms

Fuad A. F. Yahia, Hossam M. J. Mustafa*

***Abstract*— Cardiovascular disease has gained significant attention from researchers in recent years because it is a leading cause of death worldwide. This paper introduces a classification method that employs an optimization algorithm to improve the accuracy of predicting cardiovascular disease development across various genders and age groups. Patient datasets often contain a substantial number of irrelevant, redundant, or noisy features, which can hinder the accuracy of the predictions. To address this issue, we propose the Flow Direction Algorithm (FDA), which selects the most relevant features of a disease to enhance the classification accuracy rate. In the prediction stage, we combined a Support Vector Machine (SVM) with the flow direction optimization algorithm (FDA) to identify the most relevant features. To enhance the classification results, this study investigated the FDA, OFDA, Genetic Algorithm (GA), and Particle Swarm Optimization (PSO) algorithms combination with K-Nearest Neighbors (KNN) and SVM classification algorithms. The performance of the proposed algorithms is evaluated using accuracy, recall, precision, and selected features ratio as measures. The proposed algorithms based on SVM and KNN are compared using three datasets: the Heart Failure Clinical Dataset (HFCD), the Heart Dataset (HD), and the Heart Disease Prediction Dataset (HDPD) obtained from the UCI repository. The experimental results demonstrated that the SVM and KNN algorithms performed better when combined with the FDA or OFDA optimization algorithms.**

***Index Terms*— Flow Direction Algorithm (FDA), Cardiovascular Diseases, Classification, feature selection**

I. Introduction

THE detection of diseases at an early stage is critical for successful treatment and patient recovery. Predicting the onset of disease several years before symptoms appear is essential for preventing factors that contribute to disease development [1]. The World Health Organization reports that cardiovascular diseases (CVD), such as heart attacks and strokes, are among the leading causes of death globally, causing approximately 17.9 million fatalities annually. Cardiovascular diseases arise from disorders of the heart and blood vessels, and coronary heart disease primarily results from blockages that obstruct blood flow [2]. Thus, an automated system is necessary to predict heart disease in the early stages. The most pressing need in research is to enhance the accuracy of classification results and ensure their reliability in disease diagnosis [3]. For patients with early infection, treatment costs can be reduced through early prediction and the discovery of hidden patterns in patient data, rather than relying on traditional high-cost laboratory tests.

Manuscript received April 26, 2024; revised December 21, 2024.

F. A. F. Yahia is a postgraduate student of the College of Computer Sciences and Informatics, Amman Arab University, Amman 11953, Jordan. (e-mail: Fuadabusafia@gmail.com).

H. M. J. Mustafa is an Assistant Professor of the Faculty of Information Technology, University of Petra, Amman 921100, Jordan. (e-mail: hossam.mustafa@uop.edu.jo).

Therefore, machine learning algorithms (e.g., SVM) are chosen because of their effectiveness in predicting heart diseases [4]. The role of machine learning (ML) techniques in the medical field has increased through data analysis, which can be used to analyze X-ray images and predict chest or brain diseases. ML algorithms can be employed to detect diseases including cardiovascular diseases [5]. Thus, high-dimensional information on patient data may result in the "curse of dimensionality" in many situations. This is recognized as one of the key concerns that increases the difficulty of the classification process, which is commonly caused by noisy, redundant, or irrelevant data [6]. These features in the datasets may limit the performance of the learning algorithm. As a result, obtaining the best or nearly the best features is a critical challenge when designing a feature selection method (FS) technique and is an NP-hard and computationally expensive problem [7].

ML and FS are important preprocessing steps for many classification tasks. Feature selection (FS) techniques, such as the wrapper-based approach, utilize metaheuristic optimization algorithms to identify the appropriate subset of features. Additionally, the process of selecting a feature subset includes three key considerations. Initially, a supervised learning algorithm is employed to evaluate the efficiency of the selected feature subset. Subsequently, a search method, such as an optimization algorithm, is employed to seek an optimal subset of features within the original dimensions. Finally, evaluation criteria are applied to assess the efficiency of the selected feature subset. Moreover, the Flow Direction Optimization algorithm (FDA) is a novel optimization algorithm that demonstrates superior performance compared with other optimization algorithms [8]. Consequently, it is challenging to enhance the execution time and reduce the dimensional differences among the dataset features to improve the classification performance and accuracy [9].

In the proposed method, our objective is to predict cardiovascular disease (CVD) with high prediction accuracy. This is achieved by combining a Support Vector Machine (SVM) and K-nearest neighbours (KNN) with the Flow Direction Optimization (FDA) algorithm. The performance of the proposed algorithms is evaluated based on state-of-the-art metrics including accuracy, recall, and precision. The proposed algorithm is assessed by comparing it to previous research outcomes utilizing three datasets called Heart Failure Clinical [10], the standard Cleveland dataset [11] and the heart disease dataset [12] obtained from the UCI data set repository. The contributions of this study are fourfold:

1) A Flow Direction Optimization algorithm (FDA) combined Support Vector Machine (SVM) (named FDA-SVM) and K-Nearest Neighbors (named FDA-KNN) are offered for cardiovascular disease

classification.
2) Assess FDA-SVM and FDA-KNN with cardiovascular disease benchmark datasets.
3) Examine the performance metrics to determine the superior variant of the FDA algorithm, specifically by comparing the standard FDA with the OFDA.
4) Assess the performance metrics to ascertain the optimal classification algorithm for integration with the FDA algorithm, specifically by comparing SVM and KNN.

The remainder of this paper is organized as follows. Section II provides a review of the related literature concerning cardiovascular diseases and outlines the motivation behind this study. Section III provides the theoretical background and concepts relevant to this research, including the flow-direction optimization algorithm. In Section IV, we present our proposed approach, and the Flow Direction Optimization algorithm combined with the SVM and KNN. The evaluation metrics and experimental methodology are detailed in Section V. Section VI discusses and presents the results of the experiments along with a comparison with state-of-the-art algorithms. Finally, Section VII discusses future research and offers the conclusions drawn from this study.

II. Related Work

Due to the importance of predicting heart diseases in their early stages, several researchers have proposed new algorithms in the field of heart disease prediction, focusing on the most related features and reducing the dimensionality of the datasets to achieve high classification [3]. This section reviews the relevant research on cardiovascular disease classification and the techniques used in analyzing and extracting according to performance measures, such as accuracy, in addition to some measures that the researchers have set.

Some studies have proposed methods for detecting heart diseases using different machine learning techniques. For example, Jain et al. [13] used various feature selection techniques applied to classification algorithms for the detection of cardiovascular diseases. They discussed using machine learning techniques, including decision trees and SVM, for cardiovascular disease classification. Uddin et al. [14] used a hybrid model with a neural learning approach to forecast CVD and the Decision Tree Classifier performed better when ensembled with hard voting and ANN. Other researchers use different machine learning with feature selection techniques for cardiovascular disease classification such as [3], [7], [15], [16], [17], [18], [19]. Some other researchers conducted performance analyses of ML Algorithms to Predict Cardiovascular Disease such as [4], [5], [20].

Due to the complexity involved in detecting cardiovascular diseases, effective optimization methods are in high demand [21]. The use of metaheuristics to solve feature selection problems is currently a field of research [22]. In general, discovering essential feature subsets requires evaluating the subsets and selecting the best. The performance of the FS method should examine features such as the classifiers (e.g., Support Vector Machine and k-nearest neighbours), and the performance of feature selection measurements such as the average number of picked features and the accuracy rate. For example, Krishna et al. [6] proposed a whale optimization algorithm, binary genetic algorithm, binary PSO, and Binary GWO (BGWO) to provide high performance in terms of accuracy and feature selection. Mohiddin [23] proposed a modified Grey Wolf Optimizer algorithm for feature selection to predict heart diseases using a Support Vector Machine (SVM) model (named MGWO-SVM). The research of Obayya et al. [24] proposed an Automated Cardiovascular Disease Diagnosis using Honeybadger optimization with a modified deep learning (ACVD-HBOMDL) model for feature selection and hyperparameter optimization in cardiovascular disease classification. Fajri et al. [25] introduced a hybrid model that combines the Q-learning with bee swarm optimization algorithm for feature selection in the classification of coronary heart disease. Zomorodi-Moghadam et al. [21] proposed a hybrid binary-real particle swarm optimization algorithm for feature selection in the classification of coronary artery disease. Al-Tashi et al. [22] proposed a feature selection method using Grey Wolf Optimization (GWO) for classifying coronary artery disease.

Feature selection is an optimization problem intended to improve the classification accuracy of a machine learning algorithm by utilizing a reduced set of features. Several metaheuristic methods have been proposed to investigate the solution space and determine the best or nearly optimal solution for the FS problem. Many related studies employed the standard datasets of the UCI repository (e.g., Heart Disease Prediction Dataset) to evaluate the performance of the introduced algorithm. Poor scalability is the main problem with the previously offered metaheuristics for the FS problem. The selection of a suitable FS algorithm and efficient parameter settings are critical. Several approaches have attempted to improve prediction accuracy according to performance metrics such as accuracy, one of the most important metrics, in addition to some measures that researchers have set such as Recall, precision, and F1-Score. The three datasets are used in this study, and many of the different techniques, methods, and methodologies used by the researchers are presented to provide an idea of the tools, methods, and methods used and the results that the researchers reached in the field of this study.

III. Background

This section briefly introduces the background of the FDA algorithm [8] with different variants including the Opposition Flow Direction Optimization Algorithm (OFDA) [26].

A. Flow Direction Optimization Algorithm (FDA)

The FDA algorithm [8] was influenced by the concept of the eight directions (D8), which aims to determine the direction of surface water runoff towards the drainage basin based on the principles of physical logic.

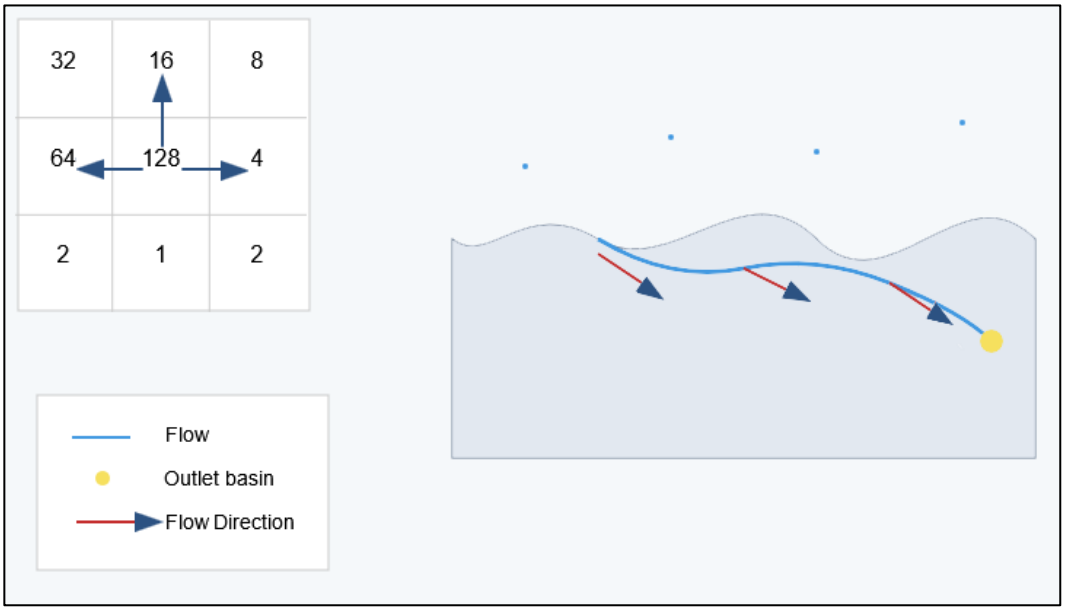

Fig. 1 Moving the flow scheme to the basin and the D8 method (redrawn based on [8])

Water flows from top to bottom with the property of collecting at low points to reach the outlet point or basin, as shown in Fig. 1. The FDA algorithm assumes the following in its operation:

1) All streams have locations and heights.
2) There are locations around each stream, each with an objective elevation or function.
3) The speed of flow movement is exactly associated with the slope.
4) The streams have a velocity of V and flow in the direction of minimum elevation.
5) The pelvic outlet point is the outflow position using the optimal objective function value.

The algorithm's initial parameters include the number of neighbours β, population number α, and neighbour radius Δ. The initial position of the streams in the FDA algorithm is manipulated using Eq. 1:

$$FlowX(i) = lb + rand \times (ub - lb) \quad (1)$$

FlowX(*i*) is the position of the *i*th flow, *ub* and *lb* are the upper and lower limits of the decision variables, and the *rand* is a random value with a uniform distribution between zero and one.

The FDA optimization algorithm undergoes several steps after fetching the dataset. The first step is (elevation), which starts representing the data according to a table where each column represents one feature in the dataset and the number of rows represents the total number of samples in the dataset, as shown in Fig. 1. The second step represents the eight directions (D8), where each number inside the cell is replaced by a code that symbolizes the direction representing the path of the smallest value of adjacent points. The third step is the flow direction, in which all the directions for each cell are known. The fourth step is flow accumulation, where the actual weight of each cell is collected by the amount of flow of the previous cells up to the downstream or outlet. Each cycle goes through these steps, trying different features in each random cycle.

B. *Opposition Flow Direction Algorithm Optimization*

As discussed in the previous section, the separate flow in the FDA update solution in the search space using random flow or neighbour. This may lead to a local optimal solution and a search trap. The possibility to avoid falling into the trap is by using pattern opposition learning (OBL) [26] that helps in the search of both directions. Consider that *FlowX(i) = {FlowX (i,1), FlowX (i,2),⋯ ,FlowX (i,d)}* is the flow in the *d*-dimensional space with range [*LB UB*]. The *UB = {ub(1),ub(2),⋯ ,ub(d)}* and *LB = {lb(1),lb(2),⋯ ,lb(d)}*. The opposite flow *OFlowX (i) = {OFlowX (i,1), OFlowX (i,2), ⋯, the OFlowX (i, d)}* is formulated in Eq. 2:

$$OFlowX(\mathrm{i,j}) = \mathrm{lb(j)} + \mathrm{ub(j)} - \mathrm{FlowX(i,j)}, \quad (2)$$

Where $\forall \mathrm{i} \in [1, \alpha]$ and $\forall \mathrm{j} \in [1, \mathrm{d}]$. The OFDA's updating the selection rule is shown in Eq. 3, where $\forall \mathrm{i} \in [1, \alpha]$:

$$FlowX(\mathrm{i}) = \begin{cases} \mathrm{OFlowX(i)}, & \mathrm{f(OFlowX(i))} < \mathrm{f(FlowX(i))} \\ \mathrm{FlowX(i)}, & \mathrm{Otherwise} \end{cases} \quad (3)$$

IV. Proposed Feature Selection Algorithm Based on FDA and OFDA

In this study, we employed a wrapper-based method that uses FDA algorithms to identify a suitable subset of features. As shown in Fig. 2, the selection of a subset of features has three primary considerations. First, the supervised learning algorithms (e.g. SVM and KNN) are used to classify the closest probabilities according to the special features of each sample. Second, the search method to improve the accuracy of classification by selecting features related to heart disease using an optimization algorithm (e.g. FDA and OFDA) is employed to seek an optimal subset of features. Third, evaluation criteria are used to assess the efficiency of the selected feature subset.

The FDA Algorithm aims to identify the most related features. Fig. 2 represents the algorithmic process necessary to accomplish this goal, utilizing three datasets: the Heart

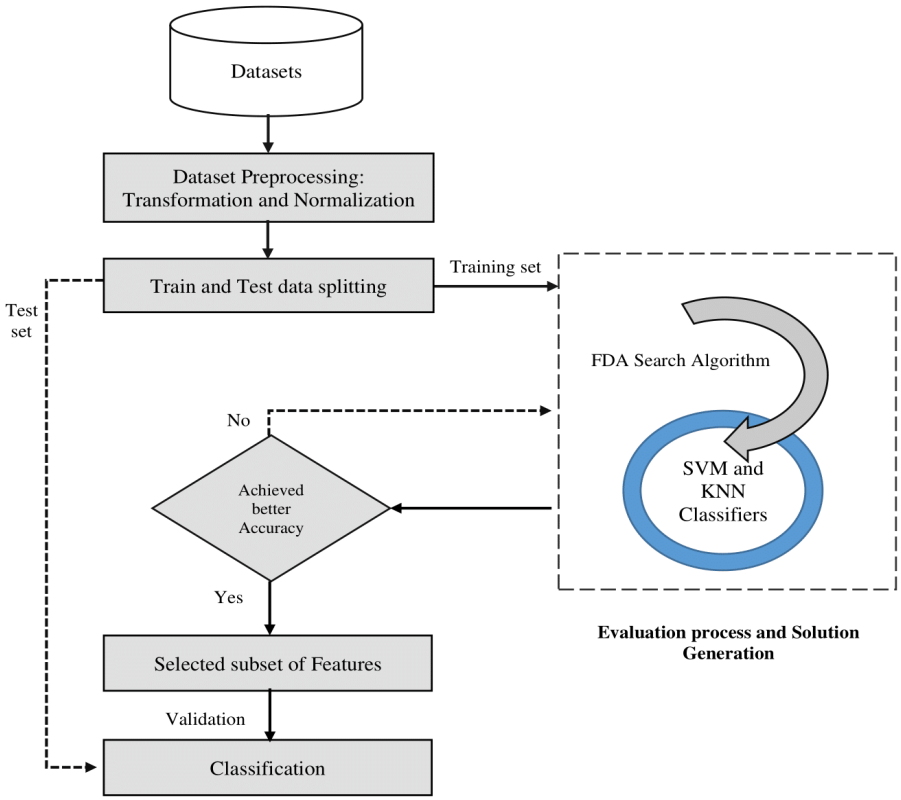

Fig. 2: Flowchart of the proposed algorithm

Failure Clinical Dataset (HFCD), the Heart Dataset (HD), and the Heart Disease Prediction Dataset (HDPD). This process involves processing, transforming, and normalizing the datasets, followed by splitting them into training and testing sets. Finally, SVM and KNN classifiers are employed to evaluate the accuracy of the algorithm's training:

1) Data collection: HFCD, HD, and HDPD datasets are chosen.
2) Dataset preprocessing: Data processing is essential for achieving high-quality results, as missing values are often encountered during data extraction or collection.
3) Splitting Data: Utilize most of the dataset for training the machine learning algorithm, reserving the remainder for testing after training. The FDA algorithm determines the most relevant features from the training process and discovers hidden connections within the data to enhance accuracy before the classification process.
4) SVM and KNN classification: This stage is determined by the features of the FDA optimization algorithm.
5) After completing the training, the classification results are compared, and the search is returned to the optimization algorithm to extract additional hidden features to enhance the accuracy.

The solutions in this study are denoted as a single-dimensional vector with a length equal to the number of features in the dataset. Each object in the vector can have one of two values, 1 or 0, where 0 shows that the equivalent feature is not chosen and 1 indicates that it is chosen. The values of 0 represent features that are excluded from the search cycle. The wrapper-based FS uses the FDA as a search algorithm and the SVM and KNN classifiers as evaluators. The feature selection solutions are denoted by a one-dimensional array, where a solution is denoted as a set of N features that are represented by set $X = \{x_0, x_1, \ldots, x_n\}$. As shown in Fig.3, features X_0, X_1, X_2, and X_9 are chosen. The accuracy of the KNN and SVM classifiers is used as a fitness function, which is calculated using the accuracy of five-fold cross-validation.

X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1	1	1	0	0	0	0	0	0	1

Fig. 3: Solution representation example of the proposed algorithm

V. Experimental Design

A. *Experimental setup*

In this study, the performance of the introduced algorithms is evaluated to determine the most reliable results in terms of classification accuracy. The SVM-FDA, SVM-OFDA, KNN-FDA, KNN-OFDA, SVM-OFDA, KNN-GA, SVM-GA, KNN-PSO, and SVM-PSO algorithms are employed in the experiments. These eight algorithms are applied to three different datasets related to heart diseases using performance measures to compare them and demonstrate the ability and strengths of the algorithms. All combined algorithms are implemented using MATLAB, and the results of the analysis are extracted according to the performance measures used, focusing on the accuracy of the performance metrics in the comparison process.

The average cost indicates the average accuracy achieved by the proposed approach based on a dataset used in all the runs of the algorithm. A higher average cost value denotes better quality of the proposed approach, falling within the range of [0, 1]. The average time indicates the average execution time of the proposed approach based on the dataset used in all runs of the algorithm, and a lower average time

indicates a better quality of the proposed approach. The hardware specifications used are an Intel (R) Core (TM) i7-3520M CPU, 290GHz (quad-core), 8192 MB Memory RAM equals 8192MB, and a Windows 10 Pro 64-bit Operating System. Table 1 demonstrates the parameter settings of the introduced approach, which is based on initial experiments that indicated that these values worked.

B. Dataset Description

The research methodology for evaluating the introduced algorithm involves a comparison of its results with the findings of previous studies using three datasets obtained from the UCI data repository: Heart Failure Clinical Records [10], the standard Cleveland dataset [11] and the heart disease dataset [12] obtained from the UCI data set repository.

TABLE 1
PARAMETER SETTINGS OF THE INTRODUCED APPROACH

Parameter Name	Value
Population size	45
Maximum iterations	65
No of runs	31
KNN Number of neighbourhoods	5
SVM kernel Function	Linear
Beta (B)	1

A summary of the datasets used is provided below:

- *Heart failure clinical records* [10]: The dataset includes 299 medical records of patients with heart conditions obtained from the UCI Machine Learning Repository. These records were collected during the follow-up period of the patients. The patient file contained 13 clinical features specific to 299 patients, of whom 194 were men and 105 were women, all aged over 40.
- *Heart dataset* [11]: The dataset contains 76 characteristics, which are also called Cleveland. These are the most used by researchers in the field of classification of heart disease using this group, in which they use a sub-dataset consisting of 14 characteristics. Where unnecessary data such as the patient's name and social security numbers were removed for patients and others, empty or incorrect data were also removed.
- *Heart_Disease_Prediction* [12]: The Heart Dataset is utilized for predicting heart disease. The patients were categorized into groups based on whether they exhibited signs of heart disease during cardiac catheterization. This dataset included 270 patients and 13 independent predictive variables.

C. Evaluation matrices

The algorithms' performance is evaluated using metrics like recall, precision, F-score, and accuracy which are widely recognized in the literature for assessing classification models. Moreover, the standard Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) [27] [28] combined with SVM and KNN are employed for comparison to assess the classification performance against the proposed OFDA and FDA algorithms, investigating their significance. The performance analysis is performed using the accuracy of the eight algorithms, KNN-FDA, SVM-FDA, KNN-OFDA, SVM-OFDA, KNN-GA, SVM-GA, KNN-PSO, and SVM-PSO. The average accuracy across multiple test runs is calculated to identify the algorithm that consistently delivers the highest classification accuracy. Furthermore, the average execution time is reported to assess the computational efficiency of each algorithm. The mathematical formulation of the performance metrics is presented in Equations 4-8, providing an accurate base for the comparative analysis of the algorithms.

$$Accuracy = \frac{Number\ of\ correctly\ classified\ predictions}{Total\ Predictions} \quad (4)$$

$$F-Score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (6)$$

$$Recall = \frac{\text{True Positive}}{\text{False Negative} + \text{True Positive}} \quad (7)$$

$$\text{selected features ratio} = \frac{\text{Number of selected feature}}{Total\ Number\ of\ features\ in\ dataset} \quad (8)$$

To assess the performance of machine learning algorithms, various important metrics are typically employed to gain a comprehensive understanding of their effectiveness. Accuracy is a primary measure that generally indicates how well the model performs across all classes but may not assess performance in imbalanced datasets. Therefore, Precision and Recall are employed to provide more detailed insights. The F-Score incorporates Precision and Recall into a single metric by calculating their harmonic mean, providing a balanced measure. In addition, in feature selection contexts, the ratio of the selected feature is an important metric that evaluates the proportion of features selected by the model relative to the total number of features in the dataset. This ratio helps assess the effectiveness of feature selection techniques and their impact on model performance.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

The performances of the SVM and KNN algorithms are compared, and the proposed SVM-FDA, SVM-OFDA, SVM-GA, and SVM-PSO algorithms are compared with KNN-FDA, KNN-OFDA, KNN-GA, and KNN-PSO. All comparisons are performed according to the three datasets used.

A. Results without Optimization Algorithms

Table 2 presents the results of the performance measures for the SVM and KNN classification algorithms before combining them with FDA Optimization and OFDA Optimization. We aim to use optimization algorithms to select the features most related to heart disease and whether this is reflected in increasing the classification accuracy in the three datasets used. The best results of the classification algorithms are highlighted in bold font.

TABLE 2
PERFORMANCE MEASURES MACHINE LEARNING ALGORITHMS WITHOUT OPTIMIZATION ALGORITHMS

Technique	Dataset	Acc.	F1	Pre	Rec	Test time (s)
SVM	Heart Failure Clinical Records Dataset	**84.0**	**75.9**	**76.5**	**77.3**	0.031
	heart dataset	**89.3**	**80.8**	**80.9**	**80.9**	0.030
	Heart Disease Prediction dataset	**88.4**	**81.5**	**81.5**	**81.5**	0.027
KNN	Heart failure clinical records dataset.	44.1	56.1	54.3	60.2	0.104
	heart dataset	69.3	67.0	67.2	67.3	0.061
	Heart Disease Prediction dataset	70.4	67.5	67.6	67.8	0.56

TABLE 3
RESULTS OF THE COMPETITIVE ALGORITHMS BASED ON ACCURACY, AVERAGE COST, AND AVERAGE EXECUTION TIME

	Heart failure clinical records			Heart dataset			Heart Disease Prediction dataset		
Algorithm	**Acc.**	**Avg. Cost**	**Avg. Time (s)**	ACC.	**Avg. Cost**	**Avg. Time (s)**	**Acc.**	**Avg. Cost**	**Avg. Time (s)**
SVM-FDA	89.16	88.07	190.7	88.01	86.82	210.8	**90.70**	**89.49**	183.8
SVM-OFDA	87.86	87.16	**52.90**	88.06	87.3	279.6	89.81	88.42	***117.8***
KNN-FDA	**90.75**	**89.64**	220.3	**89.66**	88.27	305.2	89.35	89.00	237.0
KNN-OFDA	90.00	88.00	*229.1*	89.10	**88.75**	**171.5**	88.80	88.22	503.6
SVM-GA	86.76	86.33	120.6	84.44	83.97	185.3	82.14	83.61	120.5
SVM-PSO	85.01	85.12	105.8	83.73	84.45	184.2	83.10	82.03	122.7
KNN-GA	85.01	84.18	198.4	84.25	83.52	210.1	82.95	83.25	175.3
KNN-PSO	86.51	85.95	185.7	84.01	84.20	195.2	83.65	82.10	165.4

Table 2 shows that the highest accuracy is 84%, 89.3%, and 88.4% when using the SVM algorithm, which indicates the effectiveness of the SVM algorithm over the KNN algorithm for the three datasets used. The SVM algorithm showed better results than the KNN algorithm based on all performance measures because of its ability to handle outliers in the datasets used. The results in Table 2 emphasize the clear improvement of the SVM algorithm over KNN across all three heart disease datasets, as reflected in superior accuracy, F1-score, precision, recall, and significantly faster execution times. SVM gained the highest accuracy, reaching 89.3% on the Heart dataset, while KNN results fall behind, particularly on the Heart Failure Clinical Records dataset, with an accuracy of only 44.1%. This disparity is likely due to SVM's robustness in handling outliers and its ability to maintain strong decision boundaries even in noisy or complex data environments, which is crucial in medical datasets. In contrast, KNN's performance suffered due to its sensitivity to the choice of *K* and distance metrics.

Moreover, SVM not only delivers higher accuracy but also shows greater efficiency with much lower test time, which make it more suitable for real-time applications where speed is critical. For instance, on the Heart Disease Prediction dataset, SVM's test time is 0.027 seconds, compared to KNN's 0.56 seconds. This efficiency, combined with its superior handling of classification tasks, highlights SVM as the preferred algorithm for heart disease prediction. These findings suggest that in situations where both accuracy and computational speed are vital, SVM should be the algorithm of choice.

B. Results with FDA and OFDA Optimization Algorithms

This section demonstrates the results of classification algorithms using the FDA and OFDA algorithms based on accuracy, average cost, and execution time. Table 3 demonstrates the results of the comparison of the competitive algorithms using SVM-FDA, SVM-OFDA, KNN-FDA, KNN-OFDA, KNN-GA, SVM-GA, KNN-PSO, and SVM-PSO based on the performance metrics mentioned above. Thus, Figures 4-6 reveal that using the FDA and OFDA FS algorithms improved the classification results based on different heart disease datasets.

Moreover, Table 4 shows the results of Friedman's test ranking using accuracy [29]. The results of Friedman's test show the significance of the KNN-FDA algorithm compared with other competing algorithms. Friedman's test shows an EB-HBO algorithm significance with a 0.0147 *p*-value below the ($\alpha = 0.05$) significance level.

Table 5 provides a comparison between the proposed KNN-FDA and other competing algorithms based on the selected features ratio. It is evident that in all three datasets, the KNN-FDA outperformed SVM-FDA, KNN-OFDA, SVM-OFDA, KNN-GA, SVM-GA, KNN-PSO, and SVM-PSO. These results indicate that KNN-FDA enhances the ability to identify and remove unnecessary features for describing a dataset. The best values are highlighted in bold font in each column.

C. Discussion

The results of our experiments indicate the competitive performance of SVM and KNN algorithms when combined with FDA and OFDA optimization algorithms across different heart disease datasets, as shown in Figures 4-6. The KNN-FDA achieved the highest accuracy (90.75%) on the Heart Failure Clinical Records dataset, indicating its effectiveness in handling this specific type of data. However, SVM-FDA is close behind with a 90.7% accuracy on the Heart Disease Prediction dataset, implying that SVM might be more suited for this dataset due to its essential structure and the nature of the optimization provided by the FDA.

TABLE 4: FRIEDMAN TESTS BASED ON THE ACCURACY OF THE COMPETING ALGORITHMS

Algorithm	Ranking
KNN-FDA	1.66
SVM-FDA	2.33
KNN-OFDA	2.66
SVM-OFDA	3.33
SVM-GA	5.66
KNN-PSO	6.33
SVM-PSO	6.66
KNN-GA	7.33

The cost metric is relatively consistent across all algorithms, with slight variations. Remarkably, SVM-FDA and KNN-FDA both showed competitive accuracy results, but SVM-OFDA achieved a slightly lower average cost on the Heart Disease Prediction dataset, which might be attributed to the more effective feature optimization provided by OFDA. Moreover, the SVM-OFDA algorithm is notably faster (52.9 seconds) on the Heart Failure Clinical Records dataset, making it the most efficient algorithm in this specific case. This suggests that the OFDA optimization not only maintains accuracy but also significantly reduces

computational time, which is crucial in real-time applications.

TABLE 5: A COMPARISON OF THE SELECTED FEATURES RATIO BETWEEN KNN-FDA AND OTHER ALGORITHMS

Algorithm	Heart failure clinical records	Heart dataset	Heart Disease Prediction dataset
KNN-FDA	**51.5%**	**50.2%**	**54.5%**
SVM-FDA	61.7%	55.7%	59.7%
KNN-OFDA	53.2%	55.6%	58.6%
SVM-OFDA	56.1%	52.3%	64.3%
VM-GA	58.4%	54.1%	60.2%
SVM-PSO	59.1%	53.8%	61.5%
KNN-GA	54.8%	52.5%	57.3%
KNN-PSO	55.4%	53.1%	56.7%

In addition, both GA and PSO demonstrated reasonable accuracy, for instance, KNN-GA and SVM-GA recorded lower accuracy levels, with KNN-GA achieving 85.01% and 82.95% accuracy on the Heart Failure Clinical Records and Heart Disease Prediction datasets, respectively. When evaluating execution time, both GA and PSO exhibited mixed performance. However, both algorithms are outperformed by the OFDA and FDA based on computational efficiency.

The Friedman test results, presented in Table 4, further demonstrate the differences in ranking among the algorithms. KNN-FDA ranked highest and showed its superior performance across the datasets. However, SVM-OFDA's slightly lower ranking (3.33) compared to SVM-FDA (2.33) suggests that while OFDA optimizes for speed, it might trade off some accuracy compared to FDA.

Table 5 shows the feature selection ratios between KNN-FDA and other algorithms. SVM-FDA selected a higher percentage of features across all datasets, which might explain its slightly better accuracy in some cases. However, KNN-FDA's more conservative feature selection (around 50-54%) suggests that it achieves high accuracy with fewer features. The higher feature selection ratio in SVM-OFDA on the Heart Disease Prediction dataset (64.3%) indicates that OFDA might be selecting more relevant features, which could explain its competitive accuracy despite its faster computation time.

When comparing these results to baseline SVM and KNN algorithms without FDA/OFDA, the optimization technique significantly enhances both accuracy and efficiency. For instance, while a standard SVM might achieve sufficient accuracy, the addition of FDA or OFDA results in more optimized feature sets, leading to improved performance metrics across the results.

VII. CONCLUSIONS AND FUTURE WORK

In this study, we investigated the effectiveness of combining FDA and OFDA optimization algorithms with SVM and KNN classifiers to improve the accuracy of cardiovascular disease prediction. The findings confirm the superiority of these optimization techniques in improving classification performance across multiple datasets. Notably, the KNN-FDA algorithm achieved the highest accuracy on the Heart Failure Clinical Records dataset, while the SVM-FDA algorithm excelled on the Heart Disease Prediction dataset. The close performance between the FDA and OFDA in terms of accuracy implies that both are robust methods for feature selection. However, OFDA demonstrated a significant advantage in reducing computational time, making it a preferable choice in scenarios where efficiency is paramount.

Future research will explore the integration of deep learning techniques and neural networks with FDA and OFDA to further enhance the prediction accuracy of cardiovascular diseases. Additionally, the study will expand to include more diverse and larger datasets and address challenges such as class imbalance and real-time prediction. Investigating the application of these optimized algorithms in other medical domains, such as diabetes or cancer prediction, could also provide valuable insights and extend the utility of the proposed approach.

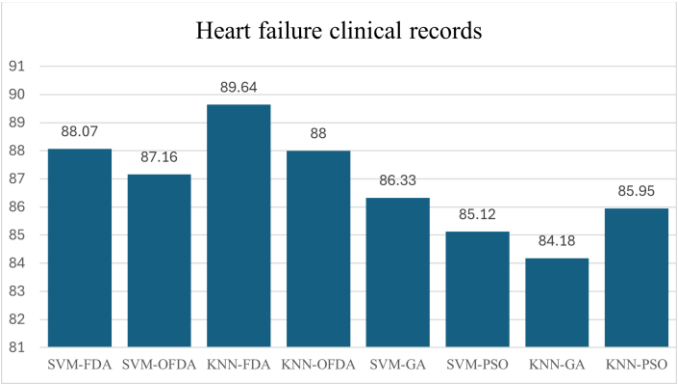

Fig. 4. Heart failure clinical records dataset results based on the accuracy between the competing algorithms

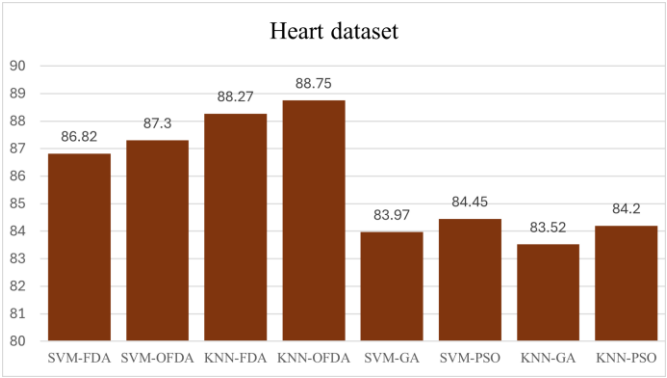

Fig. 5. Heart dataset results based on the accuracy between the competing algorithms

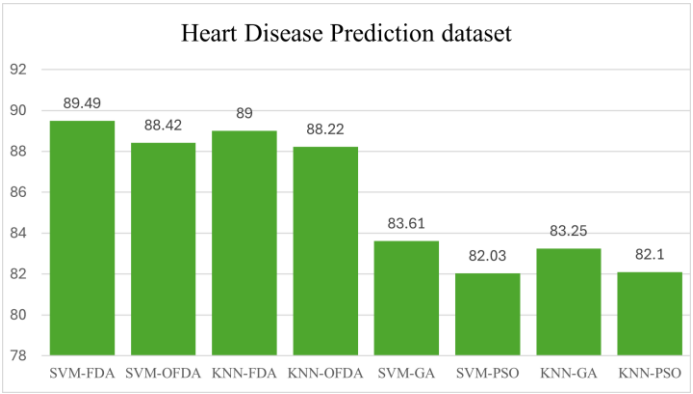

Fig. 6. Heart disease prediction dataset results based on the accuracy between the competing algorithms

REFERENCES

[1] A. Dabir, P. Khedkar, L. Panch, T. Thakare, and D. M. A. Pradhan, "Analysis of Cardiovascular Disease using Machine Learning Techniques," *Int J Res Appl Sci Eng Technol*, vol. 11, no. 5, pp. 4901–4905, May 2023.

[2] Y. Ruan et al., "Cardiovascular disease (CVD) and associated risk factors among older adults in six low-and middle-income countries: Results from SAGE Wave 1," BMC Public Health, vol. 18, no. 1, pp. 1–13, 2018.

[3] F. and C. Z. and K. M. and A. M. Manaa Abderzak and Brahimi, "Cardiovascular Diseases Prediction Based on Dense-DNN and Feature Selection Techniques," in Modelling and Implementation of Complex Systems, G. and A. A. and C. A. and S. D. E. and K. M. K. Chikhi Salim and Diaz-Descalzo, Ed., Cham: Springer International Publishing, 2023, pp. 333–347.

[4] H. V Ramesh and R. K. Pathinarupothi, "Performance Analysis of Machine Learning Algorithms to Predict Cardiovascular Disease," in 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), 2023, pp. 1–8.

[5] P. S. Mattas and I. Nadaan, "Optimizing Cardiovascular Disease Diagnosis with Machine Learning: An Analysis," International Journal of Research Publication and Reviews, vol. 04, no. 02, pp. 430–434, 2023.

[6] E. R. Krishna and N. Devarakonda, "Feature Selection Method Based On GWO-PSO for Coronary Artery Disease Classification," in *2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, 2023, pp. 1–8.

[7] P. Singh, G. K. Pal, and S. Gangwar, "Prediction of Cardiovascular Disease Using Feature Selection Techniques," International Journal of Computer Theory and Engineering, vol. 14, no. 3, pp. 97–103, Aug. 2022.

[8] H. Karami, M. V. Anaraki, S. Farzin, and S. Mirjalili, "Flow Direction Algorithm (FDA): A Novel Optimization Approach for Solving Optimization Problems," Comput Ind Eng, vol. 156, p. 107224, 2021.

[9] K. Dissanayake, "Comparative Study on Heart Disease Prediction Using Feature," Hindawi Applied Computatuional Intelligence and Soft Computing, vol. 2021, p. 17, 2021.

[10] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," BMC Med Inform Decis Mak, vol. 20, 2020.

[11] R. C. Detrano et al., "International application of a new probability algorithm for the diagnosis of coronary artery disease.," Am J Cardiol, vol. 64 5, pp. 304–10, 1989.

[12] S. W. P. M. Janosi Andras and R. Detrano, "Heart Disease," 1988.

[13] A. K. Jain, K. Kumar, R. G. Tiwari, N. Jain, V. Gautam, and N. K. Trivedi, "Machine Learning-Based Detection of Cardiovascular Disease using Classification and Feature Selection," in 2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT), 2023, pp. 214–219.

[14] M. I. Uddin, A. A. A. Ismom, Z. H. Anik, and M. N. Arefin, "Cardio Vascular Disease prediction using machine learning with improved feature selection," in 2022 12th International Conference on Electrical and Computer Engineering (ICECE), 2022, pp. 336–339.

[15] P. Kumar and A. Kumar, "Heart Disease Classification and Recommendation by Optimized Features and Adaptive Boost Learning," 2023.

[16] K. N. Devi, S. Suruthi, and S. Shanthi, "Coronary Artery Disease prediction using Machine Learning Techniques," in 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), 2022, pp. 1029–1034.

[17] T. Y. Rashme, L. Islam, S. Jahan, and A. A. Prova, "Early Prediction of Cardiovascular Diseases Using Feature Selection and Machine Learning Techniques," in 2021 6th International Conference on Communication and Electronics Systems (ICCES), 2021, pp. 1554–1559.

[18] A. H. Shahid, M. P. Singh, B. Roy, and A. Aadarsh, "Coronary Artery Disease Diagnosis Using Feature Selection Based Hybrid Extreme Learning Machine," in 2020 3rd International Conference on Information and Computer Technologies (ICICT), 2020, pp. 341–346.

[19] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques," Algorithms, vol. 16, no. 2, Feb. 2023.

[20] A. Ekong, "Evaluation of Machine Learning Techniques Towards Early Detection of Cardiovascular Diseases," American Journal of Artificial Intelligence, vol. 7, no. 1, pp. 6–16, Apr. 2023.

[21] M. Zomorodi-moghadam, M. Abdar, Z. Davarzani, X. Zhou, P. Pławiak, and U. R. Acharya, "Hybrid particle swarm optimization for rule discovery in the diagnosis of coronary artery disease," Expert Syst, vol. 38, no. 1, p. e12485, Jan. 2021.

[22] Q. Al-Tashi, H. Rais, and S. Jadid, "Feature Selection Method Based on Grey Wolf Optimization for Coronary Artery Disease Classification," in Recent Trends in Data Science and Soft Computing, F. Saeed, N. Gazem, F. Mohammed, and A. Busalim, Eds., Cham: Springer International Publishing, 2019, pp. 257–266.

[23] S. P. ,Tummala S. , T. V. V. H. K. , V. H. Shaik Khaja Mohiddin, "A Modified Grey Wolf Optimizer algorithm for feature selection to predict heart diseases," international journal of food and nutritional sciences, vol. 11, no. 12, Apr. 2023.

[24] M. Obayya, J. M. Alsamri, M. A. Al-Hagery, A. Mohammed, and M. A. Hamza, "Automated Cardiovascular Disease Diagnosis Using Honey Badger Optimization With Modified Deep Learning Model," IEEE Access, vol. 11, pp. 64272–64281, 2023.

[25] Y. A. Z. A. Fajri, W. Wiharto, and E. Suryani, "Hybrid Model Feature Selection with the Bee Swarm Optimization Method and Q-Learning on the Diagnosis of Coronary Heart Disease," Information (Switzerland), vol. 14, no. 1, Jan. 2023.

[26] R. Panda, M. Swain, M. K. Naik, S. Agrawal, and A. Abraham, "A Novel Practical Decisive Row-Class Entropy-Based Technique for Multilevel Threshold Selection Using Opposition Flow Directional Algorithm," IEEE Access, vol. 10, pp. 110473–110484, 2022.

[27] J. H. Holland, "Genetic Algorithms," Sci Am, vol. 267, pp. 66–72, 1992.

[28] J. Kennedy, R. Eberhart, and bls gov, "Particle Swarm Optimization," in Proceedings of ICNN'95 - International Conference on Neural Networks, Perth, WA, Australia: IEEE, 1995, pp. 1942–194.

[29] D. G. Pereira, A. Afonso, and F. M. Medeiros, "Overview of Friedmans Test and Post-hoc Analysis," Commun Stat Simul Comput, vol. 44, no. 10, pp. 2636–2653, 2015.