J. Electrical Systems 20-10s (2024): 3318-3330

[1]Anitha G.

[2]Rosiline Jeetha B.

Wrapper-Filter Feature Selection using Multi-Objective Forest Optimization Algorithm: A Fusion Methodology

Journal of Electrical Systems

Abstract: - The prevailing data upsurge has empowered the use of data for enhanced decision making process. Prior processing of data is a vital task to certify that the appropriate traits are measured during the study. To reduce the data extents, Feature Selection (FS) is an essential stage in the data pre-processing process along with stabilizing the significant data features. This paper introduces a Hybrid Wrapper Filter Multi-Objective Forest Optimization Algorithm with Local Search model (HWF-MOFOA-LS) algorithm for FS problem. The FS is performed using the hybrid wrapper-filter approach which are optimized concurrently. Initially a Multi-objective approach simultaneously optimizes the hybrid wrapper-filter fitness functions. The aim is to reduce the number of features and identify the related information to improve the classification accuracy. Next, the classified population with Pareto front solutions are enhanced by applying a local search selection strategy. Finally, the performance of the Multi-objective proposed technique are performed on 12 datasets from UCI repository. The results of the proposed approach is optimal when compared to the other multi-objective techniques. The proposed algorithms outperforms the other techniques by reducing the classification error along with selecting ¬minimum features.

Keywords: Feature selection; Wrapper; Filter; Forest optimization algorithm; Multi-objective Optimization; Local search.

1. Introduction

The growth of technologies like social networks, big data and internet of things creates a massive quantity of data that are collected for processing. Data mining and machine learning techniques are used in classification of these data (Al-Jarrah etal. 2015).Each instance present in the dataset is classified into various classes derived from its features. Practically, it is not possible to recognize the useful features from the collected data without previous knowledge. The features present in the high dimensional data are noisy, irrelevant and redundant. These features degrade the performance of the classification and diminish the feature set quality due to huge search space (Gheyas IA & Smith LS 2010). Hence FS is used to identify the relevant subset of features optimally by reducing the feature dimension space, execution time and enhanced classification performance (Dash M & Liu H 1997).

The objective of FS is to obtain an optimal feature subset consisting of a minimum features and maximum classification accuracy. In general, the conflicting plan among the accuracy and number creates a multi-objective problem. Only the applicable features are considered during the FS process for classification. The elimination of repeated and unconnected features aids in the FS procedure by streamlining the learning model and minimizing the number of features and maximizing the classification (Kohavi & John 1997).The challenging part is identifying an optimal feature subset in the huge search space consisting of all possible solutions (2n) for a dataset with 'n' features. The contemporary FS algorithms undergoes huge computational cost and local optimum due to huge search space (Kohavi & John 1997). There is a need for a competent global search method to address the FS tasks.

There are studies that uses evolutionary algorithms for solving problems with huge search space such as FS. Xue et.al (2015) discusses about evolutionary computing approach such as GA for FS. A group of evolutionary operators are involved in optimization algorithms (Hart et.al 2005) that has got the search capacity. To enhance the optimization process by hybridizing the GA with other algorithms such as memetic algorithms to develop the solutions (Zhu et.al 2010;Yang et.al 2008, November).These algorithms solves the local optimum convergence issues and identifies the best solutions with enough precisions. The three FS methods are filter, wrapper and hybrid approaches. Each of these approaches varies in their assessments (Kohavi & John 1997;Liu

[1 1*]Research Scholar, Department of Computer Science, Dr.N.G.P Arts and Science College, Coimbatore, Tamil Nadu, India.

E-mail: ganithamca@gmail.com

[2] Professor and Head, Department of Computer Science with AI, Nirmala College for Women, Coimbatore, Tamil Nadu, India.

E-mail: jeethasekar@gmail.com

et.al 2012).The filter approach uses essential characteristics of the training data to assess the features. The wrapper approach uses learning methods to assess the useful features. The wrapper approach involves high execution time along with high classification accuracy than the filter approaches. The hybrid approach optimizes the performance of the FS process. In the literature there are only few techniques which uses the hybrid approach for FS process (Zhu et.al 2010; Yang et.al 2008, November) for single objective algorithms. The use of hybrid FS approaches for multi-objective algorithms to optimize has not been studied in detail.

Various solutions have been proposed in the literature using the above specified approaches for FS process. The few drawbacks in these solutions are getting trapped in local optima, early convergence and high computational cost. Compared to greedy search techniques, the metaheuristic algorithms are efficient in handling these existing issues. The latter algorithm works on the population based solutions and the search ability is global.

Many studies for handling FS process involves hybrid GA (Oh Lee et.al 2004); PSO (Chuang LY et.al 2011); GP (Muni DP et.al 2006); WOA (Mafarja MM & Mirjalili S 2017); ABC (Hancer et.al 2015;); FOA (Ghaemi & Feizi-Derakhshi 2016;Ghaemi et.al 2014) ; FPA (AbdEl-Fattah et.al 2016); GWO(Al-Tashi Q et.al 2019) are few metaheuristic algorithms used to handle FS process. The FS process can be considered as a multi-objective optimization problem because of the contradictory objectives such as minimizing the feature count and maximizing the classification performance (Zhang et.al 2020; Al-Tashi et.al 2020). For the FS using multi-objective approaches there are only limited studies available like (Hamdani et.al 2007; Hancer et.al 2015, May; Miralles 2017;Mohapatra et.al 2018, February) when compared to single-objective approaches.

1.1 Studies on FS using filter-wrapper hybrid model

Hammami et.al (2019) integrated the filter and wrapper model in FS process using multi-objective GA. A local search strategy was used with two filter and one wrapper objectives. Emary et.al (2015) came up with a novel multi-objective GWO for FS using hybrid approach. The filter objective identifies the solution set with minimum redundancy and the wrapper objective improves the solutions obtained for optimal classification performance. This approach was considered as a multi-objective optimization in the filter and the wrapper stages for the FS process.

There are no appropriate studies to handle the FS process as multi-objective problem which involves a filter and wrapper fitness functions to optimize concurrently. Hence, a novel Multi-objective FOA with local search model is proposed in this study. The key objective of this study is to improve the FS process using a multi-objective FOA that concurrently minimizes the classification error and the features count along with a set of non-dominated solutions. In the proposed approach, a fusion of filter-wrapper is applied to optimize the search ability of the MOFOA-LS for FS problem. The qualities of filter and wrapper are fused to obtain an optimal performance. The primary aim is to reduce the feature count and enhance the relating features by minimizing the repeating features using filter function to find non-dominated subsets. The second aim is to use the wrapper function to maximize the classification accuracy. The third aim is to use the MOFOA with LS to overcome the existing limitations and to optimize the entire FS process.

The following are the goals of the proposed study:

i) A novel hybrid wrapper-filter approach for FS using Multi-objective forest optimization algorithm with local search called as HWF-MOFOA-LS which aims to provide a better performance.
ii) The performance analysis of the binary version of the proposed technique is compared with the existing multiple objective FS algorithms.
iii) The results are recorded and compared with the hybrid approaches without local search strategy to emphasis on the local search strategy.
iv) Comparing the performance of HWF-MOFOA-LS with single objective FS algorithms.

This paper is structured as follows. Section 2 provides the preliminary information. Section 3 presents the proposed methodology in detail. Section 4 gives the experimental results and the comparative study. Lastly, Section 5 concludes and delivers the future visions.

2. Preliminary Studies

In this segment, first, the notions of FS problem description and multi-objective optimization are presented. Second, the entropy and mutual information are discussed. At last, Multi objective forest optimization algorithm and its components are presented.

2.1 Feature Selection Problem

An attribute or a useful information is a feature that acts as a major characteristics of a dataset. The FS process selects a feature subset from the original feature. The feature space is minimized optimally based on certain criteria. Say there are N features and M dimensions in dataset, FS goal is to diminish M to M′ and M′ ≤ M. It is considered as a vital and widely used method to diminish the dimensions. Mathematically, the given original dataset is $(F_o) = (F_1, F_2, F_3, F_{4,...}F_n)$ and the objective is to extract subset $(F_{SE}) = (F_1, F_2, F_3, F_{4,...}F_m)$ where F_1, F_2,…, F_n are the features of the dataset. In binary representation of features, F_i=1 then the i^{th} attribute is selected, otherwise it is not selected. FS aids to enhance the learning capacity of a learning algorithm. The predictive accuracy is improved during classification.

2.2 Multi-Objective Optimization (MOO)

Multi-objective optimization problem has multiple objective functions that are differing. A set of solutions named as Pareto optimal solutions set represents a compromise between the objectives Mirjalili & Lewis (2015). Mathematically the formulation is as follows:

$$\begin{cases} \min imize: F(x) = \{ f_1(x), f_2(x), ..., f_o(x)\} \\ g_i(x) \geq 0,\ i = 1, 2, ..., m \\ h_1(x) = 0, i = 1, 2, ..., p \end{cases} \tag{1}$$

In $x \in X$, 'x' denotes the decision variable vector, 'X' denotes the search space, 'm' denotes the objective function count, 'g_i' and 'h_i' denotes the constraint function. A set of optimal solutions known as Pareto optimal solutions. As there are various different objectives, a specific link should be described for linking between solutions Coello CA (2009). For example, 'u' and 'v' two candidate solutions, 'u' is said to dominate 'v' (denoted by u ≺ v) if and only if:

$$\forall i \in \{1, 2, ..., m\}: f_i(u) \leq f_i(v)\ and\ \{\exists_j \in 1, 2, ...m\}: f_j(u) < f_j(v) \tag{2}$$

Where, 'm' denotes the objective function count.

Here, 'u' denotes Pareto-optimal solution and if the candidate solution is not dominated by some other feasible solution. The appearance of Pareto solutions values in the objective space is called Pareto front.

2.3 Entropy and Mutual Information

The facts about the random variables can be estimated using entropy and mutual information (Shannon, 1948).The ambiguity of a discrete random variable X is estimated using entropy and its defined using equation (3):

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x) \tag{3}$$

Where $p(x) = \Pr(X = x)$ denotes the probability solidity of (X). The entropy works on the probability distribution of random variables and not depend on real values.

Conditional entropy is estimated after certain variable is identified and when other are unidentified. Let variable Y is provided, the conditional entropy H(X|Y) of X with respect to Y is defined using equation (4):

$$H(X \mid Y) = -\sum_{x \in X, y \in Y} p(x, y) \log_2 p(x \mid y) \tag{4}$$

If X is depends on Y, then H (X|Y)=0. Here there is no other information needed to describe X when Y is identified. Here, H (X|Y) = H(X) signifies that identifying Y will do nothing to notice X which means they are unrelated or completely independent.

The information shared between two random variables is known as the mutual information I (X;Y). Let X be the variable, the quantity of information gained about the variable Y is defined using equation (5):

$$I(X;Y) = H(X) - H(X \mid Y) = -\sum_{x \in X, y \in Y} p(x, y) \log_2 \frac{p(x, y)}{p(x), p(y)} \tag{5}$$

The mutual information I (X;Y) will be huge if X and Y variables are linked, then I (X;Y)=0 if X and Y are completely unlinked.

2.4 Forest Optimization Algorithm (FOA)

FOA is a metaheuristic technique proposed by (Ghaemi & Feizi-Derakhshi 2014) inspired by tree growing procedure in a forest. To grow further the trees compete with each other for the necessary resources available in the forest like sunlight, water and soil, like the other living creatures in a forest. Trees in the forest are watching for novel resources such as space for development through storm, water and other species. In general, trees grow and vanish when they get old and are replaced with new trees. Trees practice sowing technique for spreading their seeds which includes relocation and positioning of their seeds. FOA process is imitated to answer continuous search space issues. The two seeding strategies in FOA are local and global seeding. As the name indicates, in the local seeding process the seeds lie near the parent tree and begins to grow. While in the global seeding process the seeds are laid completely in a new place because of the external factors like wind, water and animals. The seeding processes are the inspiration for the FOA which uses the exploration and exploitation tools to identify the optimal environments for tree progression in the search space. The stages of FOA are as follows:

i) Initialize trees

Like the other evolutionary algorithms, FOA starts by randomly creating some trees to initialize a forest. Each tree is represented as a solution in the search space. Each tree in the forest will have an age, feature and fitness values represented as Tree= {age,y_1,y_2,…,y_n}vector. Here, y_1 to y_n are the search space variables. Feature value is a one dimensional vector consisting of '0' or '1' generated randomly. The feature vector length is equal to the number of features in the dataset. If the features count is n, the tree's feature value is a vector comprising of 0 or 1 digits. '0' indicates the deleted feature and '1' indicates the selected feature. Each tree's age is set to 0 at the time of initialization. During the learning process the fitness is evaluated and the features are selected. The lifetime parameter of a tree depends on the type of problem were old trees are omitted from the population and new trees are inserted into the candidate population during the global seeding stage.

ii) Local seeding

Local seeding happens naturally when the parent tree starts to separate its seeds and some seeds fall under the parent tree and transforms into young trees. During this process there will be a competition between the new trees and the neighboring trees for survival. Trees with optimal resources are likely to survive. In FOA, the local seeding technique simulate this process in the regional (forest) search space. In this stage, trees in the forest with age '0'are involved in the local seeding process which includes young trees in the locality and the remaining trees are not involved. The local seeding operator creates trees with age '0' and increases the age of the remaining trees by one unit. To control the population in the forest this mechanism is applied. Assume, the tree has a value of 2, each tree with age '0' will create two trees with the same feature value. The feature value is randomly created for a new tree and it's inverted. For example if the value is '0'then it becomes '1' and vice versa. The age of the trees are increased by 1 and the age value of a new tree is set to '0'.

iii) Population limiting

Naturally, the forest tree count will be excessive, so a population limitation operator is introduced with two parameters. First, the lifetime parameter works by eliminating the trees whose ages have surpassed the lifetime parameter value from the forest (main) population and are included in the candidate population. Next, the area limit parameter works by ordering the trees in decreasing order based on the fitness function and the top trees with optimal values are transferred to the next generation by specifying the area limit. The eliminated trees are included to the candidate population. This operator simulates the law of existence naturally.

iv) Global seeding

Global seeding happens naturally, when the tree seeds are scattered in remote places due to external agents like wind, animals or water. Hence, new trees grow in remote places from the parent trees. In FOA, the global seeding operator simulates this process in the process of avoiding local optimum. The transfer rate parameter operates by selecting some trees from the candidate population and changes few variables erratically in each tree. The exploration capability of the algorithm is enhanced using global seeding operator.

v) Update the optimal tree and stopping criteria

Choosing the tree with an optimal fitness value in the search space and setting the age value of these trees to '0' and placing these tree back in the search space (forest) is the process of best tree updating. Like the other evolutionary algorithms, the stopping criteria is i) pre-defined iteration count, ii) on attaining a pre-specified accuracy iii) fitness function evaluation count.

3. The Proposed Approach : HWF-MOFOA-LS

This section presents the proposed hybrid wrapper filter feature selection using Multi objective forest optimization algorithm with local search strategies and fitness formulation for Multi objective FS problem.

In general, FS is a multiple objective problem were at least dual objectives are considered like minimizing the feature count and classification error. In this study FOA is applied to solve the FS problem by hybridizing the filter and wrapper approaches. To the best of our knowledge there are no concrete studies for the FS problem which uses Multi objectives to optimize concurrently. The proposed HWF-MOFOA-LS technique solves the FS problem by applying multi-objective FOA with local search strategy are developed. The FS happens in a discrete search space with solutions that are limited to 0 or 1 values. Hence a binary variant of the proposed HWF-MOFOA-LS technique is developed. A novel Multi-objective FOA with local search model which uses hybrid approach for FS are proposed in this study.

3.1 Tree Representation and Initialization

In every metaheuristic optimization algorithms solution representation is the primary step. In this study, the FS problem is represented as a candidate feature subset. It is selected and encoded as a chromosome in a binary bit string with each bit encoded as a single feature. Each bit in the string is represented as $S = F_1\ F_2\ F_3 \ldots F_i$, where $i = 1,2,\ldots,n$. In FOA, each individual in the tree population is represented by a vector of n bits. The value of each bit is 1or 0. The total count of available features in the dataset is denoted as n and it represents the dimension of the search space (Yang CS et.al 2008). Figure 1 illustrates the tree representation in binary mode of selected and excluded feature. In the binary mode representation, each variable can take a value of 0 or 1. The value 1 denotes selecting the feature and 0 denotes unselecting the feature. The dataset is allocated into training and test datasets. Initially the search begins by randomly initializing the population size 'p'.

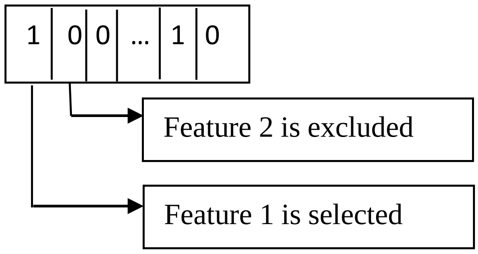


Figure 1 Tree representation in binary mode

3.2 Evaluate the initial forest

The initialized trees are evaluated as per the objective function. The key element to handle the FS problem is selecting the suitable objective function. The proposed HWF-MOFOA-LS technique chains the filter and wrapper approach into a single approach. Combining filter and wrapper approach concurrently during the optimization process is the key element. This means all the iterations are qualified to evaluate and enhance the objectives simultaneously. There are other existing FS algorithms that combines filter and wrapper approaches but there is no technique that combines filter and wrapper approach concurrently. There is a possibility of eliminating few relevant features in the filter and wrapper approaches are performed separately. Hence, a combined approach implemented in this study.

3.3 Multi Objective FS Problem formulation

To formulate the objective function, the Mutual Information (MI) in the filter approach and the classification accuracy in the wrapper approach are utilized. To measure the relevance of any feature and measure the redundancy in a feature set, MI is used as a filter metrics. Hence, a combined MI based objective function is specified in equation (6) which maximizes the relevance of the chosen feature subset and maximizes the redundancy inside the designated subset.

$$Fit_1 = Avg(I(x;c)) - Avg(I(x_i;x_j)) \tag{6}$$

Where $x, x_i, x_j \in S$, 'c' denotes the target class and S denotes the designated feature subset. A maximization function Fit_1 aims to maximize $Avg(I(x;c))$ which specifies the average significance of certain subset to the objective class. Diminishes the feature count by minimizing $Avg(I(x_i;x_j))$ which specifies the average redundancy amongst the selected feature subset. The relevance and the redundancy has the same weight in Fit_1 . The classification accuracy is vital than the features count. Hence, Fit_1 is given weights as specified in equation (7)

$$Fit_1 = \alpha.Avg(I(x;c)) - (1-\alpha).Avg(I(x_i;x_j)) \tag{7}$$

where, α is a constant number [0.5,1].

The next objective function Fit_2 of the wrapper approach aims to maximize the classification accuracy as specified in equation (8):

$$Fit_2 = Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

where, TP and TN denotes true positive and true negative. FP and FN denotes false positive and false negative. This study uses k-nearest neighbor (KNN) classifierby Altman, N. S. (1992) to evaluate Fit_2 .Finally based on the equations (7) and (8), FS problem is formulated as a multi-objective optimization problem as specified in equation (9):

$$\begin{cases} Maximize\ Fit_1 = \alpha.Avg(I(x;c)) - (1-\alpha).Avg(I(x_i;x_j)) \\ Maximize\ Fit_2 = \dfrac{TP+TN}{TP+TN+FP+FN} \\ Subject\ to\ x_i \in \{0,1\} \end{cases} \tag{9}$$

Figure 2 outlines the pseudo code of HWF-MOFOA. In this study, the non-dominated solutions (feature subsets) are included in the archive and the solutions are arranged in descending order based on the crowding distance values. To lead the optimization process and to obtain a Pareto front, a lead tree is selected from the sorted archive. The position of the present population is updated by performing local seeding, population limitation and global seeding operations. Then, the present population and the present archive are united to form a repository which is temporary. At last, a new non-dominated solutions are recognized to formulate the archive to the subsequent iteration. When the archive reaches its maximum size, based on the crowding distance value, the solutions are removed from the archive. The final archive consists of optimal feature subset (Pareto front).According to the proposed HWF-MOFOA-LS algorithm structure, the archive population is used to maintain the Pareto front.

Divide dataset into training and test set

Initialize forest by generating trees randomly. Tree age is set to 0.

Evaluate each individual using equation (9) on the training set

For iteration $\leftarrow$ 1 to maximum number of iterations Do

Extract non-dominated trees from the forest (NDT)

Add NDT to the archive

Delete dominated trees from the archive

While (archive length >archive max_size) do

//Local seeding $\{T_1, T_2, \ldots T_{\text{area limit}}\}$= local seeding(Forest)

// Population limiting for (j=1 to Forest Size) do

If (Tree age_j> life time)

Candidate population = candidate population + $Tree_j$; // transfer old trees

End if

End for

Forest= sort(forest);

If(Forest size > area Limit)

Candidate population= candidate population + $\{T_{\text{area limit}+1}, \ldots, Tree_{end}\}$;

Forest=$\{Tree_1, \ldots, T_{\text{area limit}}\}$;

End if

// global seeding Forest =global-seeding(forest);

// update the best tree ${}_{age}Tree_{Best}$=0;

End while

Evaluate the objective function

Remove duplicate from the collection

Return $Tree_{Best}$;

Figure 2 Pseudo code of HWF-MOFOA

3.4 Applying Local Search (LS) Strategy to HWF-MOFOA

The local search strategy is adopted to HWF-MOFOA at the end of each iteration to enhance the current best solutions. Initially, LS begins by storing the best solution received from HWF-MOFOA at the end of HWF-MOFOA iteration in ‘temp’ variable. LS iterates to develop the current best solution. During the iteration process, LS chooses three features randomly from the ‘temp’ and this strategy sets or resets the chosen feature based on its value. LS determines the fitness value based on the objective function. If the value is better than the fitness value of ‘F’ then ‘F’ is set to ‘temp’ else ‘F’ remains unaffected. The pseudo code of LS Strategy is given in figure 3.

Temp=F, here F denotes the current best solution at the end of HWF-MOFOA iteration

Lt=1 here Lt denotes current iteration of LS

While (Lt < maximum number of iterations)

Choose three features randomly from temp

If chosen feature==1

Chosen feature =0 else chosen feature=1

Endif

Calculate the fitness value of temp

If f(temp)<f(F)

F=temp

Endif

Lt=Lt+1

End while

Return F

Figure 3 Pseudo code of LS Strategy

4. Experimental Results

This section discusses the test results obtained to validate the performance of the proposed HWF-MOFOA-LS algorithm for the FS problem. The benchmark datasets and experiment setup are presented. Then, the test results obtained by HWF-MOFOA-LS algorithm are discussed and compared with the existing algorithms and it’s analyzed. Lastly, the efficacy of the proposed approach is assessed with respect to the computational time.

4.1 Setups for Experiment, UCI benchmark datasets and Parameter tuning

The experiments are conducted on Intel Core i7 processor, 2.7 GHz, 500GB hard disk, 16 GB RAM, and Microsoft Windows 10 operating system. The proposed model was coded using Python 3.9 language and its additional libraries. This study uses 12 benchmark datasets selected from the UC Irvine Machine Learning Repository (Blake and Merz 1998) to perform the experiments. The dataset information are presented in Table 1. These datasets are taken from chemistry, physics and mainly from health care field with diverse dimensions. Each algorithm is independently executed 10 times and the average fitness value is considered for evaluation.

Parameter values aids in obtaining optimal solutions in optimization algorithms. The parameter tunings for the proposed algorithms are presented in Table 2.

Table 1. Description of benchmark datasets from UCI data repository

Dataset Name	**Number of attributes**	**Number of objects**
WineEW	13	178
IonosphereEW	34	351
Vehicle	846	18
Wholesale customers	440	7
Healthcare datasets		
Breastcancer	9	699
Lymphography	18	148
Parkinsons	754	756
Lung Cancer	56	32
Hepatitis	19	155
ILPD	10	583
SRBCT	2308	83
Cervical cancer	35	858

Table 2. Parameter Tuning

Parameter Name	**Value**
Population size	20
Number of iterations	150
Archive size	50
Lifetime	15
Local seeding	2 (Depends on feature count on each dataset)
Global seeding	7 (Depends on feature count on each dataset)
Dimension	Number of features
Number of runs for each technique	10
K-NN Classifier (K)	K=3 and K= 5

4.2 Results and Analysis

Table 3 compares the average classification accuracy attained by the proposed HWF-MOFOA and HWF-MOFOA-LS approaches with the existing approaches on the 12 datasets. The results show that HWF-MOFOA-LS attains better classification accuracy on most datasets compared to HWF-MOFOA. The performance of

HWF-MOFOA-LS approach outperformed in 10 out of 12 datasets. Furthermore, for 2 datasets the average accuracy is same for both the proposed approaches. HWF-MOFOA-LS proves better robustness than HWF-MOFOA, which specifies the capacity to handle datasets from different sectors effectively. The optimum results are emphasized in bold. The results of HWF-MOFOA-LS are compared with the existing FW-NSGA-II, FW-PGAWOA and BMOChOA approaches which uses hybrid FS based approaches. Multi-objective FOA algorithms with local search methods proved to be enhance the search ability. Therefore, HWF-MOFOA-LS achieves better results by employing the local search strategies, which supplements the algorithm's search ability and provides solutions to avoid the local optima trap.

Table 3. Comparison of average classification accuracy of HWF-MOFOA&HWF-MOFOA-LS approaches with existing approaches

Dataset Name	**HWF-MOFOA**	**HWF-MOFOA-LS**	**Existing approaches**	
WineEW	**1**	**1**	0.989 **FW-NSGA-II (Hammami M et.al 2019)**	
IonosphereEW	0.987	**0.998**	0.927 **FW-NSGA-II (Hammami M et.al 2019)**	
Vehicle	0.685	**0.774**	0.714 **FW-PGAWOA (Got A et.al2021)**	
Wholesale customers	0.872	**0.947**	0.921 **FW-PGAWOA (Got A et.al 2021)**	
Health care datasets				
Breastcancer	**1**	**1**	0.96	**BMOChOA Piri J et.al (2021)**
Lymphography	0.991	**0.997**	0.825	
Parkinsons	**0.893**	0.892	0.847	
Lung Cancer	0.697	**0.772**	0.675	
Hepatitis	**0.98**	0.96	0.85	
ILPD	0.796	**0.799**	0.721	
SRBCT	0.679	**0.699**	0.692	
Cervical cancer	0.971	**0.983**	0.964	

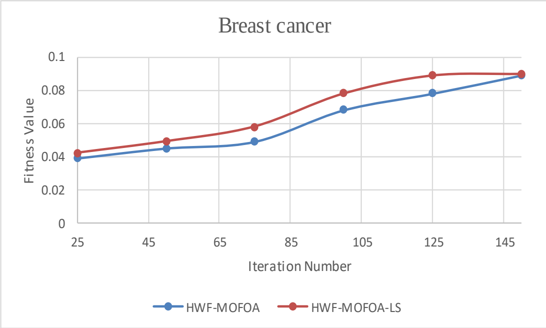
Table 4 compares the selected features count attained by HWF-MOFOA and HWF-MOFOA-LS approaches with the existing approaches on the 12 datasets. The optimum results emphasized in bold. Table 4 specify that HWF-MOFOA-LS progressively choses lesser features than HWF-MOFOA across 10 out of 12 datasets. Furthermore, HWF-MOFOA-LS proves a lesser features count compared to HWF-MOFOA approach. The HWF-MOFOA-LS results are compared with the existing FW-NSGA-II, FW-PGAWOA and BMOChOA approaches. For only one dataset the average selected features of FW-PGAWOA approach is better than the

HWF-MOFOA and HWF-MOFOA-LS approaches. The capacity of the HWF-MOFOA-LS minimize the chosen features count effectively in comparison to HWF-MOFOA approach hence eliminating noisy or unrelated features.

Table 4. Comparison of average selected features of HWF-MOFOA&HWF-MOFOA-LS approaches with existing approaches

<table>
<tr><th>Dataset Name</th><th>HWF-MOFOA</th><th>HWF-MOFOA-LS</th><th colspan="2">Existing approaches</th></tr>
<tr><td>WineEW</td><td>3.92</td><td>3.13</td><td colspan="2">6.3
FW-NSGA-II
(Hammami M et.al 2019)</td></tr>
<tr><td>IonosphereEW</td><td>6.4</td><td>5.6</td><td colspan="2">15.15
FW-NSGA-II
(Hammami M et.al 2019)</td></tr>
<tr><td>Vehicle</td><td>9.27</td><td>8.6</td><td colspan="2">9.05
FW-PGAWOA
(Got A et.al2021)</td></tr>
<tr><td>Wholesale customers</td><td>3.27</td><td>2.86</td><td colspan="2">2.49
FW-PGAWOA
(Got A et.al 2021)</td></tr>
<tr><td colspan="5">Health care datasets</td></tr>
<tr><td>Breastcancer</td><td>3.3</td><td>2.7</td><td>3</td><td rowspan="8">BMOChOA
Piri J et.al (2021)</td></tr>
<tr><td>Lymphography</td><td>4.7</td><td>3.5</td><td>4</td></tr>
<tr><td>Parkinsons</td><td>5.92</td><td>4.81</td><td>5</td></tr>
<tr><td>Lung Cancer</td><td>2.3</td><td>2.53</td><td>2</td></tr>
<tr><td>Hepatitis</td><td>3.8</td><td>3.62</td><td>3.8</td></tr>
<tr><td>ILPD</td><td>2.68</td><td>2.1</td><td>2.5</td></tr>
<tr><td>SRBCT</td><td>153.6</td><td>150.68</td><td>151.14</td></tr>
<tr><td>Cervical cancer</td><td>4.10</td><td>4.49</td><td>4.66</td></tr>
</table>

Figure 4 illustrates the convergence rates of HWF-MOFOA and HWF-MOFOA-LS based algorithms on two datasets. The x-axis and y-axis represents the epoch count and the fitness function value. The convergence rate is similar to the convergence performance on all most all datasets.



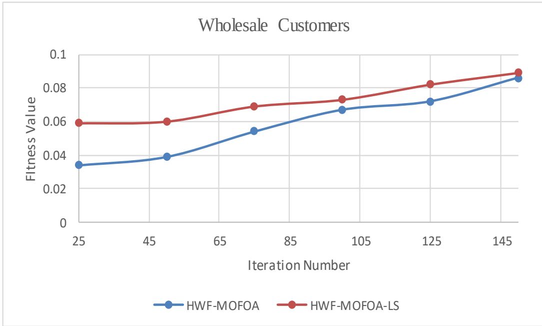


Figure 4 Convergence rates for HWF-MOFOA and HWF-MOFOA-LS

5. Conclusion

In this study, multi-objective FS problem is considered with the goal to maximize the classification accuracy and minimize the number of preferred features by excluding the noise, inappropriate and redundant features. A hybrid wrapper filter approach using multi-objective forest optimization algorithm with local search is proposed for FS problem. The FS is performed using the hybrid wrapper-filter approach which are optimized in parallel. HWF-MOFOA and HWF-MOFOA-LS are the proposed approaches for solving the FS problem. The objective to diminish the feature count and categorize the relevant information are achieved. A strategy is employed to enhance the search and to evade the local optima trap which supplements the proposed approach. The performance of the proposed technique are performed on 12 datasets from UCI repository. A Pareto optimal solutions are the results of the proposed HWF-MOFOA and HWF-MOFOA-LS approaches. The results are satisfactory when compared with the existing multi-objective approaches. The proposed HWF-MOFOA-LS approach overtakes the other techniques by reducing the classification error, maximizing the classification accuracy and decreasing the preferred number of features.

References

[1] Al-Jarrah, O. Y., Yoo, P. D., Muhaidat, S., Karagiannidis, G. K., & Taha, K. (2015). Efficient machine learning for big data: A review. Big Data Research, 2(3), 87-93.

[2] Gheyas, I. A., & Smith, L. S. (2010). Feature subset selection in large dimensionality domains. Pattern recognition, 43(1), 5-13.

[3] Dash, M., & Liu, H. (1997). Feature selection for classification. Intelligent data analysis, 1(1-4), 131-156.

[4] Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. Artificial intelligence, 97(1-2), 273-324.

[5] Xue, B., Zhang, M., Browne, W. N., & Yao, X. (2015). A survey on evolutionary computation approaches to feature selection. IEEE Transactions on evolutionary computation, 20(4), 606-626.

[6] Hart, W. E., Krasnogor, N., & Smith, J. E. (2005). Memetic evolutionary algorithms. In Recent advances in memetic algorithms (pp. 3-27). Berlin, Heidelberg: Springer Berlin Heidelberg.

[7] Zhu, Z., Jia, S., & Ji, Z. (2010). Towards a memetic feature selection paradigm [application notes]. IEEE Computational Intelligence Magazine, 5(2), 41-53.

[8] Yang, C. S., Chuang, L. Y., Chen, Y. J., & Yang, C. H. (2008, November). Feature selection using memetic algorithms. In 2008 third international conference on convergence and hybrid information technology (Vol. 1, pp. 416-423). IEEE.

[9] Liu, H., & Zhao, Z. (2012). Manipulating data and dimension reduction methods: Feature selection. In Computational Complexity: theory, techniques, and applications (pp. 1790-1800). Springer New York.

[10] Oh, I. S., Lee, J. S., & Moon, B. R. (2004). Hybrid genetic algorithms for feature selection. IEEE Transactions on pattern analysis and machine intelligence, 26(11), 1424-1437.

[11] Chuang, L. Y., Tsai, S. W., & Yang, C. H. (2011). Improved binary particle swarm optimization using catfish effect for feature selection. Expert Systems with Applications, 38(10), 12699–12707.

[12] Muni, D. P., Pal, N. R., & Das, J. (2006). Genetic programming for simultaneous feature selection and classifier design. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 36(1), 106-117.

[13] Mafarja, M. M., & Mirjalili, S. (2017). Hybrid whale optimization algorithm with simulated annealing for feature selection. Neurocomputing, 260, 302-312.

[14] Hancer, E., Xue, B., Karaboga, D., & Zhang, M. (2015). A binary ABC algorithm based on advanced similarity scheme for feature selection. Applied Soft Computing, 36, 334-348.

[15] Ghaemi, M., & Feizi-Derakhshi, M. R. (2016). Feature selection using forest optimization algorithm. Pattern Recognition, 60, 121-129.

[16] AbdEl-Fattah Sayed, S., Nabil, E., & Badr, A. (2016). A binary clonal flower pollination algorithm for feature selection. Pattern Recognition Letters, 77, 21–27.
[17] Al-Tashi, Q., Kadir, S. J. A., Rais, H. M., Mirjalili, S., & Alhussian, H. (2019). Binary optimization using hybrid grey wolf optimization for feature selection. IEEE Access, 7, 39496-39508.
[18] Zhang, Y., Gong, D. W., Gao, X. Z., Tian, T., & Sun, X. Y. (2020). Binary differential evolution with self-learning for multi-objective feature selection. Information Sciences, 507, 67-85.
[19] Al-Tashi, Q., Abdulkadir, S. J., Rais, H. M., Mirjalili, S., & Alhussian, H. (2020). Approaches to multi-objective feature selection: a systematic literature review. IEEE Access, 8, 125076-125096.
[20] Hamdani, T. M., Won, J. M., Alimi, A. M., & Karray, F. (2007). Multi-objective feature selection with NSGA II. In Adaptive and Natural Computing Algorithms: 8th International Conference, ICANNGA 2007, Warsaw, Poland, April 11-14, 2007, Proceedings, Part I 8 (pp. 240-247). Springer Berlin Heidelberg.
[21] Hancer, E., Xue, B., Zhang, M., Karaboga, D., & Akay, B. (2015, May). A multi-objective artificial bee colony approach to feature selection using fuzzy mutual information. In 2015 IEEE congress on evolutionary computation (CEC) (pp. 2420-2427). IEEE.
[22] Miralles, L. (2017). Multi-objective evolutionary feature selection for online sales forecasting. OPENAIRE.
[23] Ghaemi, M., & Feizi-Derakhshi, M. R. (2014). Forest optimization algorithm. Expert systems with applications, 41(15), 6676-6687.
[24] Mohapatra, S., Aryendu, I., Panda, A., & Padhi, A. K. (2018, February). A modern approach for load balancing using forest optimization algorithm. In 2018 Second International Conference on Computing Methodologies and Communication (ICCMC) (pp. 85-90). IEEE.
[25] Hammami, M., Bechikh, S., Hung, C. C., & Ben Said, L. (2019). A multi-objective hybrid filter-wrapper evolutionary approach for feature selection. Memetic Computing, 11, 193-208.
[26] Emary, E., Yamany, W., Hassanien, A. E., & Snasel, V. (2015). Multi-objective gray-wolf optimization for attribute reduction. Procedia Computer Science, 65, 623-632.
[27] Mirjalili, S., & Lewis, A. (2015). Novel performance metrics for robust multi-objective optimization algorithms. Swarm and Evolutionary Computation, 21, 1-23.
[28] Coello, C. A. (2009). Evolutionary multi-objective optimization: some current research trends and topics that remain to be explored. Frontiers of Computer Science in China, 3, 18-30.
[29] Shannon C, Weaver W (1948) The mathematical theory of communication. The University of Illinois Press, Urbana, p 144. ISBN:978-0252725487.
[30] Ghaemi, M., & Feizi-Derakhshi, M. R. (2014). Forest optimization algorithm. Expert Syst Appl 41 (15): 6676–6687.
[31] Yang C-S, Chuang L-Y, Chen Y-J, Yang C-H (2008) Feature selection using memetic algorithms. In: Proceedings of the third international conference on convergence and hybrid information technology (ICCIT '08), pp 416–423.
[32] Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician, 46(3), 175-185.
[33] Got, A., Moussaoui, A., & Zouache, D. (2021). Hybrid filter-wrapper feature selection using whale optimization algorithm: A multi-objective approach. Expert Systems with Applications, 183, 115312.
[34] Hammami, M., Bechikh, S., Hung, C. C., & Ben Said, L. (2019). A multi-objective hybrid filter-wrapper evolutionary approach for feature selection. Memetic Computing, 11, 193-208.
[35] Piri, J., Mohapatra, P., Pradhan, M. R., Acharya, B., & Patra, T. K. (2021). A binary multi-objective chimp optimizer with dual archive for feature selection in the healthcare domain. IEEE Access, 10, 1756-1774.
[36] C. L. Blake and C. J. Merz. (1998, 1 June). UCI Repository of machine learning databases. Available: http://www.ics.uci.edu/~mlearn/