

Received June 28, 2021, accepted July 11, 2021, date of publication July 14, 2021, date of current version July 21, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3097206

A Robust Multi-Objective Feature Selection Model Based on Local Neighborhood Multi-Verse Optimization

IBRAHIM ALJARAH¹, HOSSAM FARIS^{1,2}, ALI ASGHAR HEIDARI^{3,4}, MAJDI M. MAFARJA⁵,
ALA' M. AL-ZOUBI^{1,6}, PEDRO A. CASTILLO⁷, AND JUAN J. MERELO⁷

¹King Abdullah II School for Information Technology, The University of Jordan, Amman 11942, Jordan

²School of Computing and Informatics, Al Hussein Technical University, Amman 11831, Jordan

³School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran 1439957131, Iran

⁴Department of Computer Science, School of Computing, National University of Singapore, Singapore 117417

⁵Department of Computer Science, Birzeit University, Birzeit 72439, Palestine

⁶School of Science, Technology and Engineering, University of Granada, 18011 Granada, Spain

⁷Department of Computer Architecture and Technology, University of Granada, 18071 Granada, Spain

Corresponding author: Ibrahim Aljarah (i.aljarah@ju.edu.jo)

This work was supported in part by the Ministerio de Economía y Competitividad under Grant TIN2017-85727-C4-2-P.

ABSTRACT Classification tasks often include, among the large number of features to be processed in the datasets, many irrelevant and redundant ones, which can even decrease the efficiency of classifiers. Feature Selection (FS) is the most common preprocessing technique utilized to overcome the drawbacks of the high dimensionality of datasets and often has two conflicting objectives: The first function aims to maximize the classification performance or reduce the error rate of the classifier. In contrast, the second function is designed to minimize the number of features. However, the majority of wrapper FS techniques are developed for single-objective scenarios. Multi-verse optimizer (MVO) is considered as one of the well-regarded optimization approaches in recent years. In this paper, the binary multi-objective variant of MVO (MOMVO) is proposed to deal with feature selection tasks. The standard MOMVO suffers from local optima stagnation, so we propose an improved binary MOMVO to deal with this issue using the memory concept and personal best of the universes. The experimental results and comparisons indicate that the proposed binary MOMVO approach can effectively eliminate irrelevant and/or redundant features and maintain a minimum classification error rate when dealing with different datasets compared with the most popular feature selection techniques. Furthermore, the 14 benchmark datasets showed that the proposed approach outperforms the state-of-the-art multi-objective optimization algorithms for feature selection.

INDEX TERMS Wrapper feature selection, multi-verse algorithm, optimization, classification.

I. INTRODUCTION

Data mining is the process of extracting valuable knowledge, and interesting patterns embedded in different data sources (e.g., databases and data warehouses) [1]. Data mining techniques are mainly classified into supervised (e.g., classification) and unsupervised (e.g., clustering) techniques [2]. Supervised learning techniques, such as kernel extreme learning [3], k-nearest neighborhood (KNN) [4], and support vector machines (SVM) [5], [6] tend to learn a model to be able map a data instance to a specific category or class. Unsupervised learning techniques, such as clustering, on the

other hand, infer the structure of the data without having a piece of prior knowledge about their categories of classes [7].

Classification methods have been widely used in different real-world applications such as health informatics [5], [8], medical systems [9], [10], image processing [11], [12], protein classification [13], and feature fusion [14]. The main challenge with these applications is that the datasets become very large due to the advancements in data collection tools [15]. The high-dimensional datasets may include, in addition to the valuable features, some irrelevant and redundant features that may reduce the efficiency of the learning algorithms [16]. Therefore, preprocessing and preparing the datasets became a crucial step in determining the success or failure of the learning algorithms [17].

The associate editor coordinating the review of this manuscript and approving it for publication was Hisao Ishibuchi¹.

Dimensionality reduction (i.e., Feature Selection (FS) and Feature Extraction (FE)) is one of the most common pre-processing techniques used to overcome the challenges of high-dimensional datasets [18]–[20]. This paper focuses on FS for classification tasks, where FS methods aim to determine the most informative features in a dataset during a reasonable training time for a specific classifier, simplify the learned models, and improve the performance of the searching and classification engines [21]. However, searching for the most informative features is challenging due to the large feature space as there are 2^n possible feature subsets in a dataset with n features. A specific feature may be considered an important and beneficial one for the classification model, yet it might be considered redundant when combined with other features. By contrast, a feature may be classified as irrelevant when considered individually while being relevant and beneficial for the learning performance in conjunction with other features. Therefore, for a large n , it is impractical to exhaustively evaluate all feature subsets to get the best performing one. Under those circumstances, FS is considered as an NP-hard combinatorial problem [22]–[24].

Different search strategies have been used for FS, such as greedy-based strategies (i.e., sequential forward selection (SFS) and sequential backward selection (SBS) [25]). However, those methods typically have high computational complexity or suffer from premature convergence problems [26]. To overcome these problems, metaheuristics have been widely applied as search strategies in FS methods [27]. Evolutionary computation (EC) algorithms are population-based metaheuristics that have been successfully applied to tackle FS problems due to their superior global search ability. The most popular EC algorithms commonly used for dealing with FS are Genetic Algorithms (GA) [28], [29] and Particle Swarm Optimization (PSO) [30].

Paying attention to the number of objectives considered when evaluating a solution for the optimization problem, metaheuristics can be classified into two categories. As the name implies, single-objective methods deal with one objective, while multi-objective techniques deal with two or more objectives, which are often in conflict [31]–[34]. The critical point in the EC techniques is that they manipulate a set of solutions at each iteration of the optimization process. In other words, EC can produce multiple trade-off solutions in a single run, which enables them to show good efficacy on multi-objective optimization [35].

FS methods can be categorized into two main categories when considering how they evaluate the generated feature subsets [36], [37]: *Filter* approaches consider the correlations between features and the class without evolving any learning algorithm. In contrast, *wrapper* approaches consider the performance of a learning algorithm (e.g., classification) in the evaluation process. Wrapper FS methods try to optimize two contradictory objectives when evaluating a feature subset: to obtain the minimum number of features and the minimum classification error rate. Hence, FS problems can be treated as multi-objective problems with contradictory cost functions.

Most of the existing FS methods in the literature deal with a single objective, while a few multi-objective FS studies have been reported.

Multi-Verse optimization (MVO) is a new swarm-based approach that has shown its exploratory and exploitative performance in dealing with several real-life engineering and science problems [38]. MVO algorithm was proposed by Mirjalili *et al.* [39] to mathematically model the philosophy of multi-verse in astrophysics. However, most binary problems such as feature selection normally have more variables than continuous variables, which requires more efficient optimization approaches to handle this challenge. This paper proposes an efficient binary Multi-objective MVO optimizer with personal best to improve the efficacy of the basic MVO to handle the feature selection tasks for the first time in literature.

In this paper, we have made the following key contributions:

- Two enhancements of Multi-objective MVO are proposed. In the first approach, a binary approach using an efficient transfer function is developed. In the second approach, the personal best location and “local best” is embedded in MOMVO.
- The hybrid MOMVO and personal best is proposed for the first time to solve the feature selection tasks.
- The proposed approaches have been tested on fourteen real benchmarks datasets with different settings and characteristics to show their efficiency for feature selection tasks.
- The efficacy and qualitative results of the proposed technique are compared to the several well-regarded and state-of-the-art multi-objective optimizers in the FS field from different aspects. The multi-objective versions of the PSO (MOPSO), non-dominated sorting genetic algorithm (NSGA-II), multiobjective evolutionary algorithm based on decomposition (MOEAD), improved strength Pareto evolutionary algorithm (SPEA2), and Pareto envelop-based selection algorithm (PESA2).

The rest of this paper is organized as follows: Section 2 presents the review of the related works about multi-objective feature selection algorithms. Section 3 describes the preliminaries of the feature selection and multi-objective optimization, and the MVO algorithm. Section 4 presents the details of the proposed approaches. The experiments and results are presented in Section 5. Finally, Section 6 discusses the concluding remarks and future works.

II. REVIEW OF RELATED WORKS

Feature selection techniques have been widely used in different computational applications including but not limited to medical science [40]–[42], sales forecasting [43], face recognition [44], and customer churn prediction [45]. When designing a machine learning technique [46], reducing the number of features in a dataset contributes to decreasing the required learning time by removing the redundant features. Also, it enhances the performance of the employed learning

technique by removing the irrelevant, misleading, and inappropriate features [25].

Many works applied FS methods as an improvement of the machine learning models. One example for the FS study is the work in [47], where the authors discussed various types of evaluation measures for feature selection. Another example investigated the FS for supervised classification approaches [48], whereas, [49] proposed an FS model based on Shapley value embedded genetic algorithms and support vector machine. While the recent work [50] presented an improved feature selection using Harris Hawks optimizer for gene expression data. [51] introduced An efficient FS approach using Modified Social Spider Optimization (MSSO) algorithm and [52] investigated the unsupervised feature selection for enhancing the performance of classification models. Further, the authors of [53] applied FS combined with CGA-NN classifier for the optimal solution. Also, more various studies employed the FS methods in the literature such as [54]–[58].

As FS methods tend to improve the learning performance of the algorithm (e.g., classifier) by using the minimal number of features, the use of multi-objective optimization methods to tackle the FS methods has significantly grown in recent years [59], [60]. In this section, we explain the most crucial multi-objective FS approaches.

Recently, multi-objective EC algorithms (e.g., GA and DE) have been utilized to address the FS problem. In this sense, [61] proposed a multi-objective micro GA to form an ensemble optimizer that optimizes the FS problem in addition to optimizing the neural network classifier. Nondominated sorting-based multi-objective GA II (NSGA-II), is a GA variant that was initially proposed by [62] to solve multi-objective optimization problems. NSGA-II was used in [45] as a multi-objective FS approach, with Decision Tree C4.5 classifier as an evaluator to design a customer churn predictor. [63] proposed a novel multi-objective FS approach that considers both the feature weights and a number of selected features as two objectives to be achieved for a facial recognition application. Another NSGA-II based multi-objective FS approach was proposed in [64], where the classification accuracy and the number of selected features were treated as two objectives, and the user is allowed to choose a subset in the Pareto-front. NSGA-III is another variant of the multi-objective GA used as a search strategy in several FS methods. [65] proposed an improved NSGA-III with niche preservation procedure for multi-objective FS problem, where the number of selected features and the sum weight were used as two different objectives to be achieved in the selected subset.

Besides, more works utilized the multi-objective FS technique; for instance, a hyperparameter tuning based on the multi-objective FS method is also proposed by [66]. [67] presented a survey about multi-Objective FS and their applications. While, [68] employed multi-objective FS for bacterial foraging optimization.

Concerning the DE algorithm, a multi-objective FS approach was proposed in [69], where maximizing

the classification accuracy and minimizing the number of selected features were considered as two opposing objectives. In addition, [44] proposed a multi-objective FS approach based on the DE algorithm. The authors used their approach for Facial Expression Recognition (FER) method. Therefore, they applied the modified multi-objective DE to select the best subset of features and the support vector machine classifiers for emotion recognition accuracy.

Recently, many multi-objective PSO-based FS approaches were proposed in the literature. [70] introduced the use of a multi-objective PSO algorithm for the FS problem. In that paper, two variants of PSO were proposed to generate the Pareto front of non-dominated feature subsets. [71] proposed a multi-objective PSO FS approach, called RFPSOFS, where the features are ranked based on their frequencies in the archive set, and then they are used to guide the particles and for the archive refinement. [72] proposed an enhanced multi-objective PSO-based FS approach by employing an adaptive uniform mutation operator to enhance the exploration capability of the PSO algorithm, in addition to adopting a local learning strategy to enhance the algorithm's exploitation capability. A similar multi-objective PSO approach was proposed by [73] to optimize both the parameters of the SVM classifier and the number of selected features. In [74], an enhanced multi-objective PSO was used to search for the Pareto front feature subsets that satisfy different objectives. Another multi-objective PSO-based FS approach was proposed in [75], where the reliability and the classification accuracy were considered as two objectives to be achieved, and a bare-bones-based PSO with reinforced memory strategy and a hybrid mutation operator was used to search the Pareto-front feature subsets.

Moreover, other multi-objective swarm-based metaheuristic algorithms were used to tackle the FS problem. For example, a multi-objective variant of Artificial Bee Colony (ABC) was used as a searching strategy in a multi-objective FS method in [26] and [76]. In addition, a multi-objective version of the Gravitational Search Algorithm (GSA) was proposed and used to tackle the FS algorithm in [77].

Therefore, our work differs from other techniques in proposing two different versions of Multi-objective MVO. In the first version, we applied a binary method utilizing the transfer function-based approach, while in the second version, the local neighborhood local space is discovered, as well as the personal best location in this space is utilized with the MOMVO.

III. PRELIMINARIES

A. MULTI-OBJECTIVE OPTIMIZATION

In Multi-objective (MO) problems, it is required to deal with two or more opposing objectives to obtain the best set of solutions. In MO optimization, the purpose is to optimize several conflicting objective functions to attain the optimum solutions. The mathematical formulation of a MO minimization is as follows:

$$\text{minimize } F(\vec{x}) = [f_1(\vec{x}), f_2(\vec{x}), \dots, f_k(\vec{x})] \quad (1)$$

subject to

$$g_i(\vec{x}) \leq 0, \quad i = 1, 2, \dots, m \quad (2)$$

$$h_i(\vec{x}) = 0, \quad i = 1, 2, \dots, l \quad (3)$$

where \vec{x} shows the decision variables, $f_i(\vec{x})$ denotes the function of \vec{x} , k show how many functions is to be minimized, and $g_i(\vec{x})$ and $h_i(\vec{x})$ are the constraints for the intended problem. In MO optimization, we measure the quality of solutions according to trade-offs between considered objective functions. If x_1 and x_2 be two solutions of the above k-objective problem, when the conditions in Eq. (4) are satisfied, it means that x_1 dominates x_2 :

$$\forall i : f_i(x_1) \leq f_i(x_2) \text{ and } \exists j : f_j(x_1) < f_j(x_2), \quad i, j \in \{1, 2, 3, \dots, k\} \quad (4)$$

In the case that the rest of the solutions cannot dominate a solution, it is recognized as a *Pareto-optimal* solution. These solutions generate a trade-off surface, which is called the *Pareto front*. In a MO optimization technique, it is intended to find a set of non-dominated solutions.

B. MULTI-VERSE OPTIMIZER (MVO)

Multi-Verse optimization (MVO) is a kind of swarm-based approach that has shown its exploratory and exploitative performance in dealing with several real-world engineering and science problems.

MVO algorithm was inspired to mathematically simulate the multi-verse in astrophysics [39]. This physical theory describes the role of the big bangs in forming multiple universes. It also explains that universes can interact with other peers based on hypothetical classes of holes such as white, black, and wormholes. The black and white holes can interconnect using a tunnel, indicating a transmission between paired universes. The black holes can attract other masses, while the white holes can emit other objects. Wormholes also can create tunnels for connecting paired universes in line with the time dimension. Each universe is matched with an inflation rate, which assists it in expanding over space.

These concepts are inspired by a population-based algorithm, which we have called MVO, to develop and efficient exploration and exploitation mechanisms. For this purpose, some initial random universes (search agents) are generated inside the search space. In MVO, each variable/feature in the solution vector corresponds to an object in that universe. Furthermore, each solution has an inflation rate (fitness value) to measure the quality of solutions. Like other metaheuristic methods [78], [79], MVO obtains the fitness values based on the corresponding objective function of the problem. For example, a better fitness value is assigned to a search agent when white holes are observed, whereas an inferior objective value is given to an agent if the black holes are generated. Furthermore, if more communications between white and black holes happen, the variable values of better agents are sent to poorer agents.

The core mathematical formulation of MVO is obtained based on Eqs. (5) and (6):

$$X_i^j = \begin{cases} X_k^j, & r_1 < NI(U_i) \\ X_i^j, & r_1 \geq NI(U_i) \end{cases} \quad (5)$$

where X_i^j is the j^{th} object of i^{th} agent (universe), r_1 indicates a random value inside (0,1), $NI(U_i)$ is the normalized fitness value (inflation rate) of the i^{th} agent (universe) and X_k^j is the j^{th} object of the k^{th} universe selected by a roulette wheel selection mechanism.

Another operation used to provide local changes for each universe is given as follows:

$$X_i^j = \begin{cases} \begin{cases} ((X_j + TDR \times (ub - lb) \times r_4 + lb), \\ r_3 < 0.5 \\ (X_j - TDR \times (ub - lb) \times r_4 + lb), \\ r_3 \geq 0.5, \end{cases} & r_2 < WEP \\ X_i^j, & r_2 \geq WEP \end{cases} \quad (6)$$

where X_j is the j^{th} element of the fittest universe attained so far, ub is the superior limit, lb is the inferior limit, Traveling Distance Rate (*TDR*) plays the role of a coefficient, Wormhole Existence Probability (*WEP*) is another coefficient, r_2 , r_3 and r_4 are random values inside (0, 1). *WEP* and *TDR* are adaptive parameters, which *WEP* is utilized in MVO to boost the exploitation power, and *TDR* is used to improve exploitation in the vicinity of the best agent found so far. The adaptive rule for *WEP* and *TDR* coefficients can be calculated as follows:

$$WEP = min + l \times \left(\frac{max - min}{L} \right) \quad (7)$$

$$TDR = 1 - \frac{l^{1/p}}{L^{1/p}} \quad (8)$$

where p shows the exploitation factor, *min* and *max* indicate the minimum and maximum, respectively, l is iteration, and L denotes the maximum bound of iterations.

In the MVO, the user first sets the maximum iteration number and number of population. The optimization process is initialized using a set of randomly distributed universes inside the upper and lower bounds. At each iteration, variables in the universes with higher fitness values (higher inflation rates) will update their locations toward the universes with lower inflation rates using white/black holes. Temporarily, each universe runs into a random transfer in its objects over wormholes and in the direction of the fittest universe, which has the minimum fitness value. The whole steps are repeated until a termination condition is satisfied. MVO keeps the best agent during iterations and employs it to guide other universes toward the optimum. The pseudocode of MVO is represented in Algorithm 1. Note in this algorithm, *SU* denotes the sorted universes, *NI* shows the normalized inflation rate, i denotes the black hole index, m denotes the white hole index,

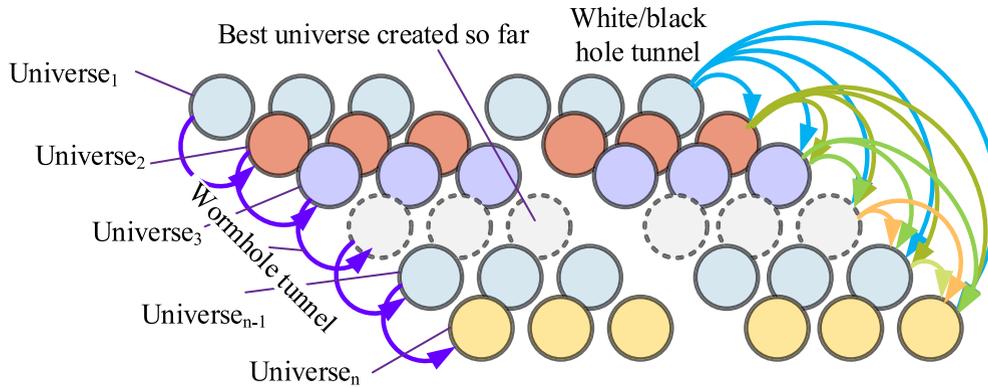


FIGURE 1. Conceptual design of the MVO ($I(U_1) > I(U_2) > \dots > I(U_{n-1}) > I(U_n)$).

RWS denotes the procedure of Roulette Wheel Selection, and r_1, r_2, r_3, r_4 are random numbers inside interval (0, 1).

Algorithm 1 Pseudo-Code of MVO Algorithm

```

Input: Total number of universes and number of iterations (L).
Output: The best universe and the corresponding inflation rate.
Generate the initial random universes  $x_i(i = 1, 2, \dots, n)$ , WEP, TDR, and best universe.
while (Termination condition is not true) do
    Calculate the fitness of current universes.
    for (each Universei) do
        Update WEP and TDR using Eqs. (7) and (8)
        for (each Objectj) do
            if  $r_1 < NI(U_i)$  then
                 $m = RWS(-NI)$ ,  $U(i, j) = SU(m, j)$  ▷ Call Algorithm 2
            end if
        end for
        Update universes using Eq. (15).
    end for
end while
Return: The best universe
    
```

1) MULTI-OBJECTIVE MVO (MOMVO)

In MOMVO, similarly to multi-objective Particle Swarm Optimization (MOPSO) [80] and Pareto Archived Evolution Strategy (PAES) [81], an archive is utilized to keep the best non-dominated universes found so far. The exploration and exploitation processes in MOMVO are very similar to the core processes in MVO, in which all candidate universes are evolved based on the interaction between white holes, black holes, and wormholes [82]. Because of several best universes, the white holes and particularly wormholes can be selected from the archive. A leader selection scheme is utilized in MOMVO to select solutions from the archive and open tunnels between universes. For this purpose, the crowding distance between universes in the repository is measured, and

Algorithm 2 Pseudo-Code of RWS Operator

```

function RWS
     $P(\text{choice} = i) = DEF \frac{F(i)}{\sum_{j=1}^n F(j)}$ 
    S := 0;
    for  $i = 1, i < N, i++$  do
        S := S + P(choice = i)
        if rand < i then return i
        end if
    end for
end function
    
```

the number of universes in the locality is calculated to obtain the coverage or diversity. In MOMVO, universes are selected based on a roulette wheel scheme from the less colonized areas of the archive. For this purpose, Eq. (9) is utilized to enhance the distribution of universes in the repository through all objectives:

$$P_i = \frac{c}{N_i} \tag{9}$$

where $c > 1$ is a constant, which can be regarded as a strategy for fitness sharing, and N_i is the number of universes located in the neighborhood of the i -th universe. The archive can store a limited number of non-dominated universes hence, Equation (10) is used to assign higher chances to undesired universes (one with many neighboring agents) to be eliminated from the archive by the MOMVO.

$$P'_i = \frac{N_i}{c} \tag{10}$$

The MOMVO can store Pareto's best universes in the archive and evolve them during iterations with these operators. This technique satisfies the following rules for comparing and addition of a universe to the archive:

- When a new search agent can dominate any agent selected from the archive, the algorithm replaces it with that in the archive immediately.
- When a new search agent cannot dominate any agent selected from the archive, the algorithm discards it, and it will not be permitted to be inserted into the archive.

- When a new search agent is non-dominated concerning all agents inside the archive, the algorithm adds it to the archive.
- When the archive is full, the algorithm deletes an undesired search agent and adds a new non-dominated agent to the archive.

In this method, the ideas of *Pareto optimality* and *Pareto optimal solution* are employed to be able to compare all universes. In this algorithm, exchanging variables between universes can happen between a universe and an archive universe or two non-dominated universes in the feature space. This rule can enhance the algorithm's exploration tendency, which may also undesirably affect the algorithm's convergence behavior. An equal chance of picking up an archive universe or a non-dominated universe in the feature space is assigned in this method to enhance the trade-off between the exploration and exploitation proclivities.

The MOMVO starts the searching process using a number of random universes and approximates the true Pareto optimal front for the target problem. Every universe is related to some objective values. Initially, the algorithm selects all the non-dominated universes to insert them into the archive. By the first iteration, MOMVO evolves the universes using Eq. (15). Based on the rule in Eq. (15), we have equal chances to exchange variables with an archived agent or one of the non-dominated universes in the up-to-date swarm. The former operation deepens the intensification of the fittest Pareto optimal universes found so far. The latter mechanism improves the diversification of universes inside the search space. The optimization by evolving the universes will be continued in MOMVO to satisfy a termination condition. In addition, the coverage of universes across all objectives will be enhanced by choosing universes from the less colonized areas of the archive. The source code of MOMVO is publicly available .

C. FEATURE SELECTION FOR CLASSIFICATION

A training set often includes some rows, which are also known as objects, and some columns, known as features. These rows and columns are associated with a number of specific classes called decision features. Classification is a well-studied and highly demanded task in machine learning and data mining researches. Based on [21], the main mission of classification is devoted to predicting the possible class of an unidentified object.

Referring to [83], the redundant and irrelevant features can negatively affect the quality of classification. The core reason is that when we face more features in the dataset, we need to add more instances, which raises the learning time of the classifier. In addition, learning from irrelevant features decreases the accuracy of a classifier compared to the same classifier that only deals with relevant features in a more reasonable time. Furthermore, the irrelevant features in the dataset can mislead the classifier, and then, we will face the over-fitting problem. Another remark, the redundant and irrelevant features in a dataset may upsurge the complexity of

the main classifier, which can make it complicated to realize the learning outcomes.

Feature selection approaches aim to efficiently determine the irrelevant and redundant features and eliminate them from the dataset to improve the efficacy of the main classifier in terms of the time consumed for the learning process, the accuracy of classification results, and the clearness of the output data. As a fact confirmed by literature, it is very important to choose an efficient searching strategy in FS methods to augment the efficacy of the learning model. By applying an efficient FS method and determining the most informative features, and eliminating the redundant records, the dimensionality of the searching space will be decreased. Then, the performance and convergence rate of the learning algorithm can be boosted [84].

1) SINGLE-OBJECTIVE OPTIMIZATION FOR FS

Feature selection usually is tackled in literature as a single-objective optimization with two cost functions, including accuracy of classification and number of selected features [85]. According to the weight of each component and the importance of each measure, some weighting factors are carefully selected by the user/practitioner before the FS process. Usually, more weights are assigned to the accuracy of the FS rather than the selected features, and accuracy should be maximized while we seek to find the minimized value of the second part. To deal with such a fitness function, every single-objective optimizer can be utilized to detect the minimum fitness value.

2) MULTI-OBJECTIVE OPTIMIZATION FOR FS

The nature of FS can also be studied as a multi-objective task. In this case, the cost functions for accuracy and number of selected features are evolved together, and this allows feature sets to be assessed concerning various dimensions simultaneously. The multi-objective formulation can assist us in negating some of the pitfalls observed when dealing with fitness-based exploration and exploitation phases, such as convergence to sub-optimal solutions and early stagnation. In a multi-objective scenario, multiple factors can be simultaneously involved in the cost function, in which some of them can be potentially more complicated than other parts.

D. K-NEAREST NEIGHBOR (K-NN) CLASSIFIER

The k -NN algorithm is a well-regarded non-parametric and instance-based classification technique which works based on categorizing unlabeled instances. k -NN method can evaluate the distance between a specified instance and the related neighboring k instances (k neighbors) [86]. The core logic behind k -NN method is that the label assigned to an object in the feature space can be probably more similar to its nearby objects. To measure distance, there are many rules employed in previous literature. Often, Euclidean distance is used along with k -NN, which can be obtained by Eq. (11):

$$DIST(P_1, P_2) = \left(\sum_{i=1}^n (x_{1,i} - x_{2,i})^2 \right)^{\frac{1}{2}} \quad (11)$$

where P_1 and P_2 show two points with n dimensions. The well-known KNN is one of the simplest and most recommended approaches to wrapper-based FS methods compared to other learning models.

IV. THE PROPOSED APPROACH

This subsection describes the proposed BMOMVO-pbest algorithm. The motivation is to introduce a new feature selection technique based on a multi-objective MVO algorithm, which not only has a good classification performance on solving the feature selection problem but also is minimizing the number of selected features. The BMOMVO-pbest algorithm is summarized in the following subsections.

A. BINARY MOMVO FEATURE SELECTION (BMOMVO)

The MOMVO algorithm was initially intended to deal with complex features of problems in continuous spaces. Due to FS problems' nature, the solutions in MOMVO are planned to change in limited directions within the binary space (0 and 1 values). We have applied transfer functions (TF) as a valid work for converting the version to a binary variant [87]. If the feature is selected, we see 1 in the element, otherwise, it will be zero. In this paper, we used the most popular TF that suggested by Kennedy and Eberhart in [88] to convert the continuous MOMVO version to a binary version as in Eq. (12).

$$T(X_i^j(t)) = \frac{1}{1 + e^{-2X_i^j(t)}} \quad (12)$$

where X_i^j is the j^{th} dimension in the i^{th} universe, and t is the current iteration. The transfer function T is depicted in Fig. 2.

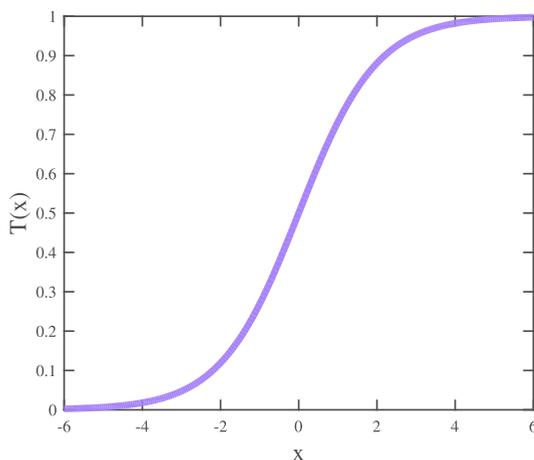


FIGURE 2. Transfer function.

Depending on the produced probability from Eq. (12), the universe in the next iteration can be updated using Eq. (13):

$$X_i^j(t+1) = \begin{cases} 0 & \text{If } rand < T(X_i^j(t)) \\ 1 & \text{If } rand \geq T(X_i^j(t)) \end{cases} \quad (13)$$

where $X_i^j(t+1)$ is the j^{th} dimension in the i -th universe, and X_i^j is given in the following rule:

$$X_i^j(t+1) = \begin{cases} \begin{cases} (X_j + TDR \times r_4), & r_3 < 0.5 \\ (X_j - TDR \times r_4), & r_3 \geq 0.5, \end{cases} & r_2 < WEP \\ X_i^j, & r_2 \geq WEP \end{cases} \quad (14)$$

where X_j is the j^{th} element of the fittest universe attained so far that selected by ranking mechanism that used in the standard MOMVO.

B. BMOMVO FEATURE SELECTION WITH PERSONAL BEST (BMOMVO-PBEST)

This subsection discusses a new approach to FS using binary multi-objective MVO to explore the Pareto front of feature subsets. The standard MVO and MOMVO use the fittest universe/best universe to update all universe's positions. This operation supports more exploration behavior of the algorithm. To add more exploitation capability for the universes, a new term is added to the Eq. (14). This term is called *personal fittest*, which is the best position achieved by the universe itself so far. It can be viewed as the universe's memory. The new $X_i^j(t+1)$ is given in the following rule:

$$X_i^j(t+1) = \begin{cases} \begin{cases} ((X_j + TDR \times (P_j^i - X_i^j(t)) \times r_4), & r_3 < 0.5 \\ (X_j - TDR \times (P_j^i - X_i^j(t)) \times r_4), & r_3 \geq 0.5, \end{cases} & r_2 < WEP \\ X_i^j, & r_2 \geq WEP \end{cases} \quad (15)$$

where P_j^i is the personal best of the i -th universe. The P_j^i is updated based on the dominance relationship, such as if the new P_j^i dominates the old one, then the new one is employed; otherwise, the old one continues to be used.

In this approach, universes explore the space and store the best position they have achieved so far. Each universe has momentum that allows the universe to explore more areas in the search space. Furthermore, the universe is also attracted to its personal best location in its memory. The main strength is that a universe can balance its exploration and exploitation ability; moreover, diversity can be maintained.

C. FITNESS FORMULATION

The proposed BMOMVO-pbest uses two main objectives, namely; classification error rate (1- Accuracy) and number of features. The first objective is the Classification Error Rate (CER), which is given by Eq. (16):

$$CER = \frac{FP + FN}{TP + TN + FP + FN} \times 100\% \quad (16)$$

where FP , FN , TP , and TN are false positives, false negatives, true positives, and true negatives, respectively.

The second objective considers the number of selected features (*NSF*), which is given in the following equation:

$$NSF = \frac{l}{A} \quad (17)$$

where l is the number of selected features and A is the total number of features in the given dataset.

Algorithm 3 represents the pseudo-code of the proposed BMOMVO-*pbest* wrapper method for FS problems.

Algorithm 3 Pseudo-Code of BMOMVO-*pbest*-Based Wrapper FS Approach

Input: Total number of universes, number of iterations (L), and divided datasets into training and testing sets.

Output: The best universe and the corresponding inflation rate.

Generate the initial random universes $x_i (i = 1, 2, \dots, n)$, WEP , TDR , and best universe.

Archive = { }

Obtain the inflation rates of all universes

Update the Archive using Eq. 4

Remark the best solution (X_j)

Remark the best personal solution for each universe (P^i)

while (Termination condition is not true) **do**

for (each $Universe_i$) **do**

 Calculate two objective CER and NSF for current universes by Eqs. (16) and (17) for each Universe

 Update the Archive using Eq. 4

 Update the best personal solution (P^i)

 Update WEP and TDR using Eqs. (7) and (8)

for (each $Object_j$) **do**

if $r_1 < NI(U_i)$ **then**

$m = RWS(-NI)$, $U(i, j) = SU(m, j) \triangleright$ Call

Algorithm 2

end if

 Update universes using Eq. (15)

end for

end for

end while

Return: The best feature sets

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. SYSTEM DETAILS

Table 1 reveals the properties of the computing system and utilized testing environments.

B. DATASETS AND PARAMETER SETTINGS

Table 2 tabulates 14 used datasets for comparative experiments. These datasets are publicly available in UCI machine learning repository [89]. These well-studied datasets are chosen to have different numbers of features, classes, and instances. These representative samples can show how the proposed MVO-based multi-objective techniques can address the optimal feature subsets of FS problems.

TABLE 1. The settings of the system.

Name	Setting
Hardware	
CPU	Intel Core(TM) i5 processor
Frequency	3.1GHz
RAM	4GB
Hard drive	500 GB
Software	
Operating system	Windows 7
Language	MATLAB R2018a

All evaluations and plans of tests are performed based on fair comparison rules in deep learning [90]–[92]. The compared methods are all wrapper-based multi-objective algorithms, i.e., requiring a learning method to be used within the training stage to assess the resulting feature subset's classification efficacy. The well-known KNN is one of the simplest and most recommended methods used within wrapper FS techniques compared to other learning models. In these experiments, we employed KNN with $K = 5$ to simplify the evaluation procedure.

As a training/testing methodology, we randomly have managed all of the instances in each dataset to be processed under two sets: 70% are considered inside the training set and 30% for the test set. Then, we have utilized 5-folds cross-validation, where the training data set is divided into five equal parts. Note that 5-fold cross-validation is achieved as an internal loop in the training procedure inside the fitness function to assess the classification error of nominated features on the training set. After this process, the selected features are assessed on the test set to attain the testing classification error proportion.

The experiments are repeated for 30 runs to minimize random effects and test the results if they statistically have significant differences compared with other methods. Each run is set to 100 iterations as stopping criteria with a 30 population size. For the proposed BMOMVO-based approaches, we used the same parameters and setting that used in [82].

C. RESULTS AND DISCUSSIONS

1) RESULTS OF BMOMVO AND BMOMVO-PBEST

Table 3 compares the average (AVG), standard deviation (STD), best (BEST), and worst (WORST) error rate results for the proposed BMOMVO-*pbest* and BMOMVO on all datasets. As per AVG results in Table 3, we see that the BMOMVO-*pbest* can outperform the BMOMVO on 85.71 % of datasets. According to STD, BEST, and WORST values, the proposed BMOMVO-*pbest* provides competitive and better classification error rate results compared to the basic binary BMOMVO on several datasets.

Experimental results of BMOMVO and the proposed BMOMVO-*pbest* in tackling BreastEW, Exactly, HeartEW, SonarEW, CongressEW, and KrvskpEW, Tic-tac-toe, Vote, WineEW, Zoo, Semeion, and Leukemia cases are

TABLE 2. List of used datasets from UCI repository.

#	Dataset	No. of Features	No. of instances	No. of classes
1	BreastEW	30	569	2
2	Exactly	13	1000	2
3	HeartEW	13	270	2
4	SonarEW	60	208	2
5	CongressEW	16	435	2
6	KrvskpEW	36	3196	2
7	Tic-tac-toe	9	958	2
8	Vote	16	300	2
9	WineEW	13	178	3
10	Zoo	16	101	6
11	Semeion	256	1593	2
12	Leukemia	7129	72	2
13	GLIOMA	4434	50	4
14	Nci9	9712	60	9

TABLE 3. Error rate results obtained by MOMVO-*pbest* and BMOMVO.

Measure	BMOMVO				BMOMVO- <i>pbest</i>			
	AVG	STD	BEST	WORST	AVG	STD	BEST	WORST
BreastEW	7.510	1.095	5.044	9.795	9.472	0.835	8.070	11.436
Exactly	31.627	1.922	28.933	37.000	21.768	4.149	15.333	31.889
HeartEW	27.494	1.786	22.222	30.159	23.593	1.357	21.164	26.543
SonarEW	22.600	2.594	18.413	28.118	23.805	2.653	14.286	30.159
CongressEW	8.628	3.663	6.107	23.282	4.392	2.136	2.290	12.595
KrvskpEW	10.573	2.314	4.901	14.937	10.506	2.837	5.869	17.901
Tic-tac-toe	26.855	0.629	25.738	29.253	26.413	0.630	25.174	28.704
Vote	6.000	1.712	4.000	10.556	5.635	1.543	4.444	10.000
WineEW	13.084	1.798	8.995	16.667	12.040	1.799	8.951	15.556
Zoo	19.294	3.311	14.516	25.806	16.281	2.842	11.290	22.581
semeion	3.001	0.383	2.022	3.882	2.283	0.243	1.674	2.836
Leukemia	17.402	1.149	13.636	18.182	16.088	1.862	11.364	18.182
GLIOMA	31.243	2.129	26.667	33.333	8.028	2.747	3.333	16.667
Nci9	63.931	2.207	60.185	66.667	41.674	2.682	37.500	47.222

Note: The values in bold indicate the best results.

demonstrated in Fig. 3. Figure 6 also shows the experimental results for BMOMVO-*pbest* and BMOMVO algorithms in dealing with GLIOMA and Nci9 datasets. In these figures, each sub-figure corresponds to one of the studied datasets. Please note that the numbers in the brackets located at the top of each sub-figure indicate the number of available features and the related classification error values based on all features. The horizontal axis in these charts shows the number of selected features, while the vertical axis indicates the related error values. The curves in these figures show the average Pareto front obtained by BMOMVO-*pbest* and BMOMVO algorithms over 30 independent runs. Note that, in some test cases, the conventional and improved optimizers may optimize an identical subset in several runs, and the same points in the plots can be observed. Consequently, while 30 results are shown, we can see less than 30 separate points in the plots of average Pareto fronts.

As shown in Figs. 3 and 6, we see that the BMOMVO-*pbest* can efficiently explore the Pareto front and find feature subsets, which covers a smaller number of features and reveal better classification rates than MOMVO. For almost all datasets, except BreastEW and SonarEW, BMOMVO-*pbest* includes two or more subsets, which effectively obtained a smaller number of features and attained a better error value than the rate obtained based on all features. For KrvskpEW and Tic-tac-toe cases, we see the classification efficacy of both methods are very competitive, while there still is a slight superiority in the results of BMOMVO-*pbest* algorithm.

Table 4 compares the computational time recorded for the BMOMVO and the MOMVO-*pbest* in dealing with all datasets. As per results in Table 4, in the same condition, the BMOMVO uses relatively less time than the proposed BMOMVO-*pbest* approach.

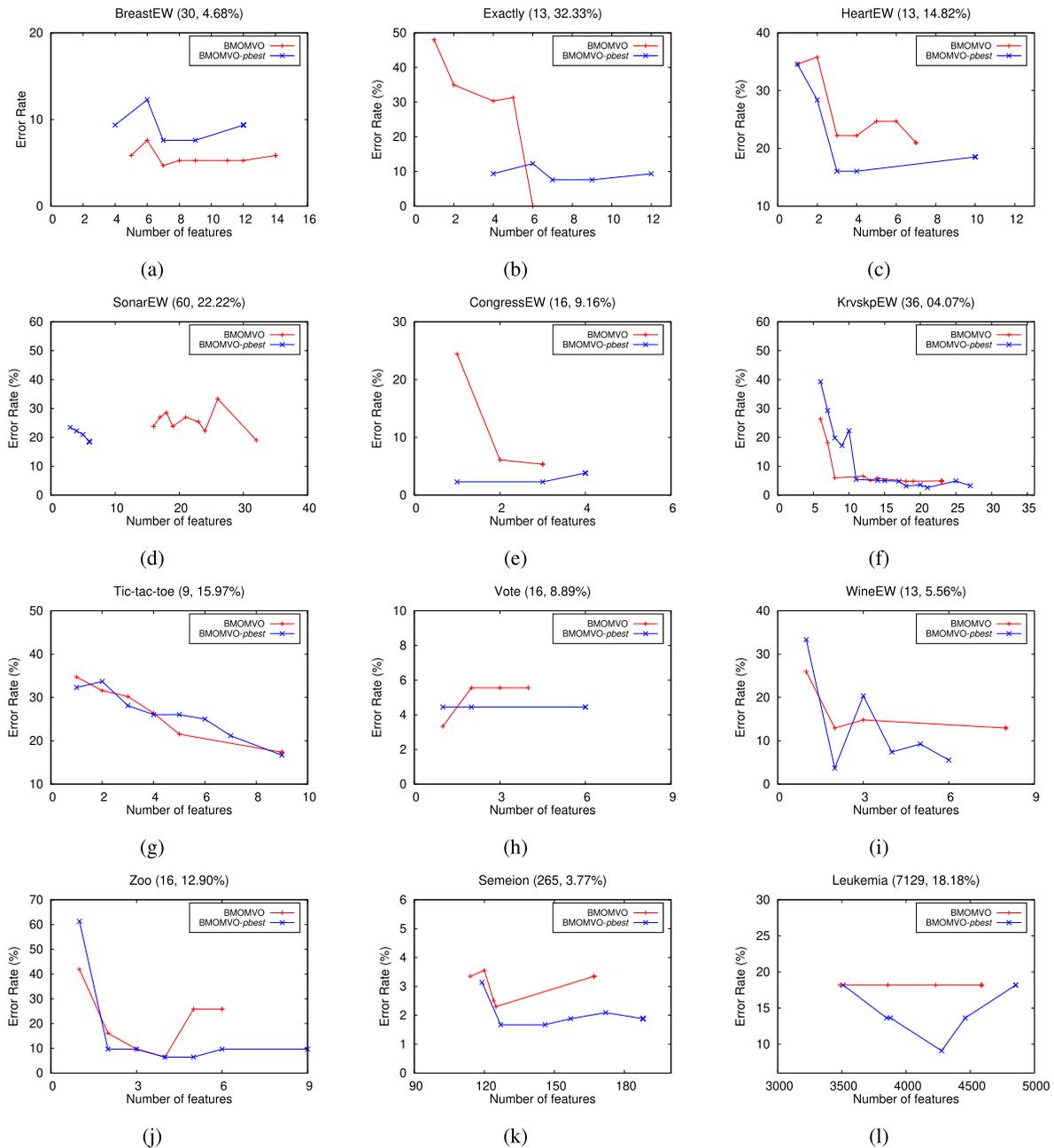


FIGURE 3. Comparison of BMOMVO-pbest and BMOMVO on BreastEW, Exactly, HeartEW, SonarEW, CongressEW, and KrvskepEW, Tic-tac-toe, Vote, WineEW, Zoo, Semeion, and Leukemia datasets.

2) COMPARISON WITH EVOLUTIONARY MULTI-OBJECTIVE ALGORITHMS

In this section, the efficacy of the proposed approach in terms of classification error rates, number of features, and computational time are compared with other popular multi-objective techniques. Namely, Nondominated Sorting-based Multi-objective Genetic Algorithm II (NSGAI) [93], Strength Pareto Evolutionary Algorithm 2 (SPEA2) [94], Pareto Archived Evolutionary Strategy (PAES2) [95], Multiobjective Evolutionary Algorithm Based

on Decomposition (MOEA/D) [96], and MOPSO [97]. These well-studied algorithms have shown excellent performance in dealing with many multiobjective problems in literature. Hence, we compared the performance of BMOMVO-pbest with these well-known methods in solving all datasets. Furthermore, all settings and parameters that are used in the experiments for all popular multi-objective techniques are obtained from [93]–[97]

Table 5 compares the error rates returned by the proposed BMOMVO-pbest with those obtained by other

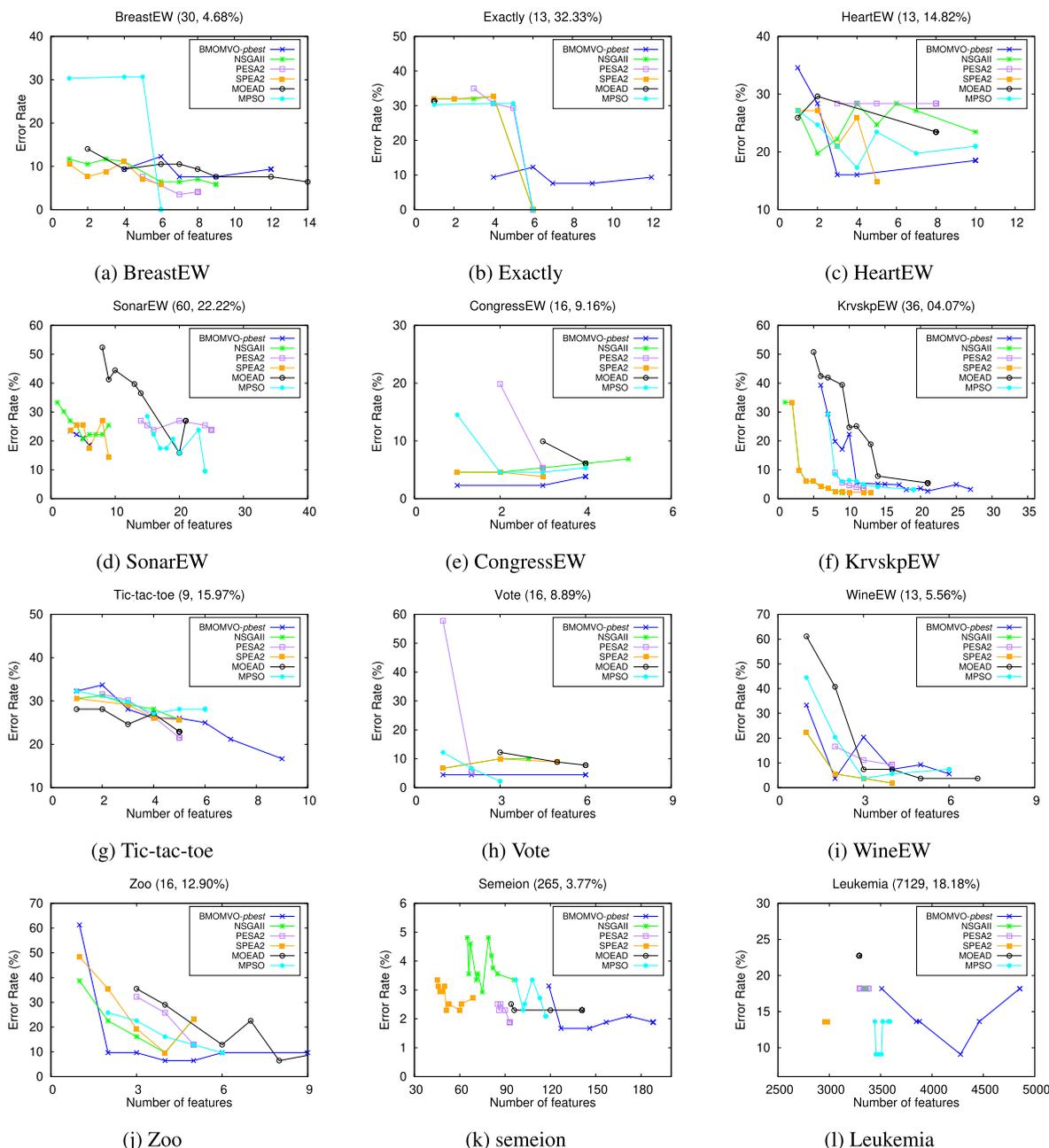


FIGURE 4. Comparisons between BMOMVO-pbest and other methods for BreastEW, Exactly, HeartEW, SonarEW, CongressEW, and KrvskpEW, Tic-tac-toe, Vote, WineEW, Zoo, Semeion, and Leukemia benchmark datasets.

compared methods. The F-test statistic is also provided at the last row of Table 5. Experimental results of the proposed BMOMVO-pbest in solving BreastEW, Exactly, HeartEW, SonarEW, CongressEW, and KrvskpEW, Tic-tac-toe, Vote, WineEW, Zoo, Semeion, and Leukemia cases are compared to other peers in Fig. 4. Figure 7 also shows the experimental results of all multi-objective methods in realizing GLIOMA and Nci9 datasets. Boxplots of error rates are also shown in Figs. 5 and 8. Note that in these figures, BMOMVO-pbest is denoted by MVOpb due to the limited space.

As per AVG results in Table 5, the BMOMVO-pbest can archives better error rates in dealing with 64.28 % of datasets. The minimum value of BMOMVO-pbest for F-test results also supports this observation. According to STD rates, we see the BMOMVO-pbest has shown a relatively more stable performance than other competitors in dealing with the majority of cases. A similar pattern in the superiority of BMOMVO-pbest can also be observed according to BEST and WORST results. According to the F-test results in Table 5, we see that NSGAI, SPEA2, MOPSO,

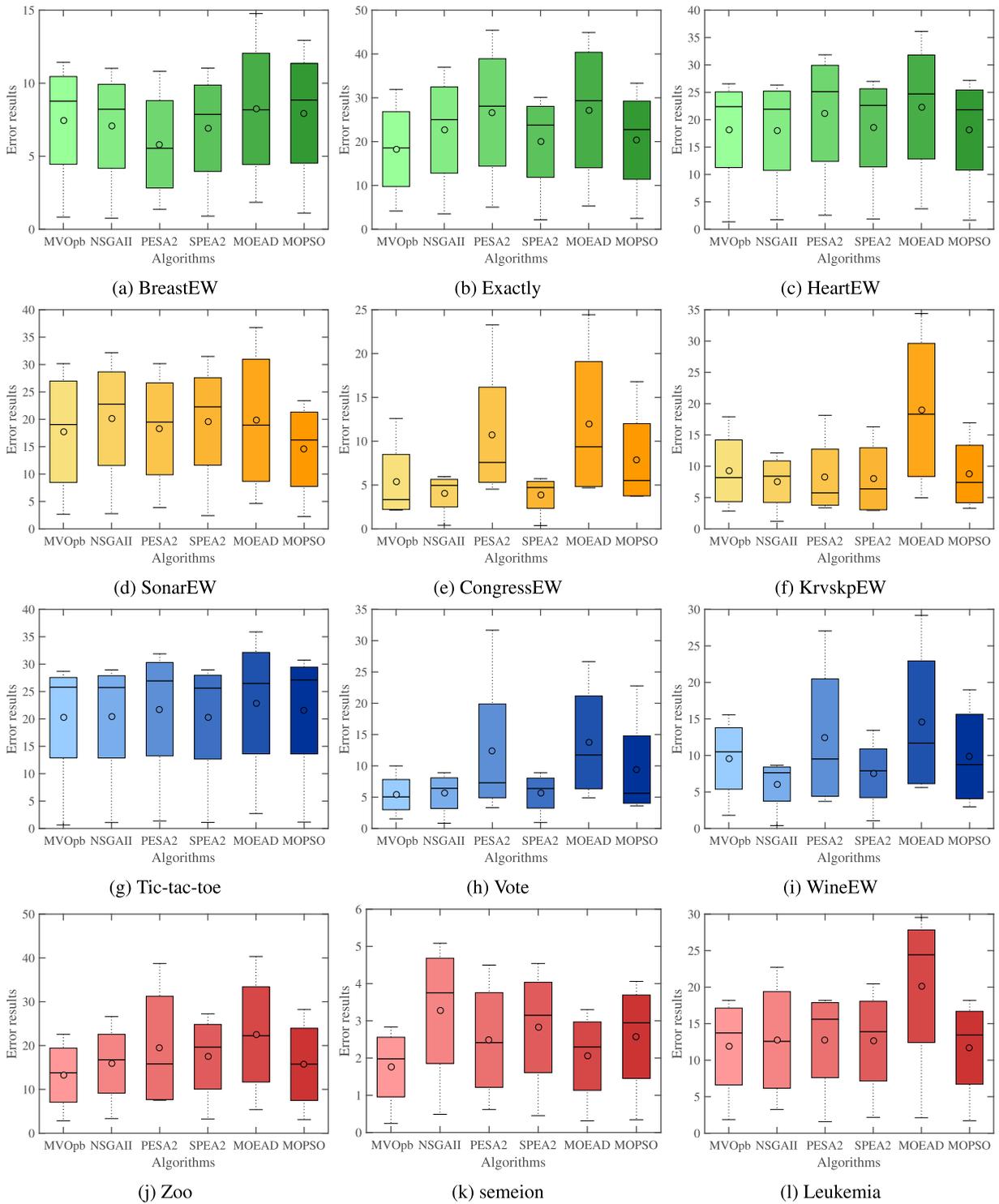


FIGURE 5. Boxplot of error results for BreastEW, Exactly, HeartEW, SonarEW, CongressEW, and KrsvkpEW, Tic-tac-toe, Vote, WineEW, Zoo, Semeion, and Leukemia benchmark datasets.

PESA2, and MOEAD have obtained the next overall ranks.

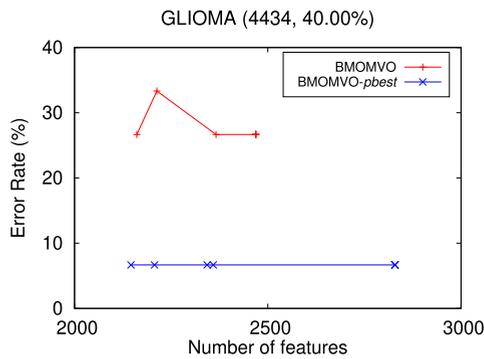
According to average Pareto front in Figs 4 and 7, we can detect that the proposed BMOMVO-*best* can efficiently explore the Pareto front and obtain subsets on most of the

cases, which contain a smaller number of features and show better error rates than other peers. Inspecting the Pareto fronts of NSGAI and SPEA2, we observe that these methods have shown a similar classification performance in most test datasets such as BreastEW, Exactly, SonarEW, KrsvkpEW,

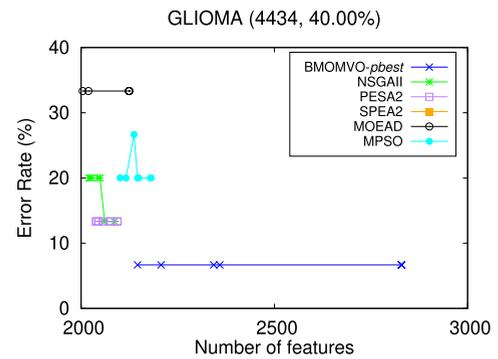
TABLE 4. Comparisons of computational time (seconds) obtained by BMOMVO-*pbest* and BMOMVO method for all datasets.

Measure	BMOMVO				BMOMVO- <i>pbest</i>			
	AVG	STD	BEST	WORST	AVG	STD	BEST	WORST
BreastEW	99.755	1.594	96.202	101.961	119.851	3.349	108.269	125.074
Exactly	117.945	0.531	116.351	119.108	127.048	2.954	119.180	132.315
HeartEW	76.325	0.429	75.485	77.330	78.670	0.406	78.049	79.779
SonarEW	74.790	0.300	74.151	75.673	92.275	2.447	85.216	96.564
CongressEW	84.022	0.315	83.538	84.742	92.076	0.861	89.539	93.733
KrvskpEW	530.506	8.388	515.713	548.602	645.533	17.341	596.584	679.468
Tic-tac-toe	105.262	0.622	104.018	106.808	114.598	1.744	108.607	117.164
Vote	77.452	0.328	76.699	78.133	81.035	0.483	80.052	82.010
WineEW	72.480	0.299	71.937	73.389	76.477	0.422	75.648	77.574
Zoo	79.547	0.358	78.890	80.475	80.328	0.499	79.419	81.460
semeion	923.988	9.329	896.818	937.850	1025.448	17.083	983.794	1052.187
Leukemia	352.138	3.125	346.238	359.054	436.758	16.283	384.381	467.675
GLIOMA	174.191	0.719	172.166	175.419	495.267	122.815	275.425	616.384
Nci9	413.306	1.680	409.043	417.716	480.315	6.570	467.186	494.330

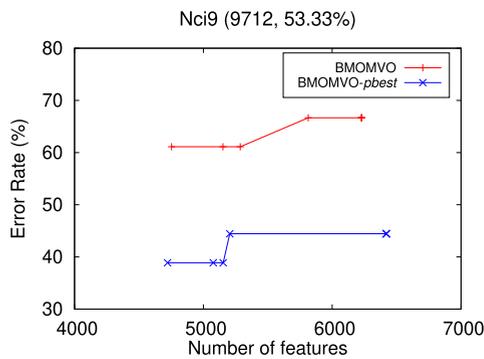
Note: The values in bold indicate the best results.



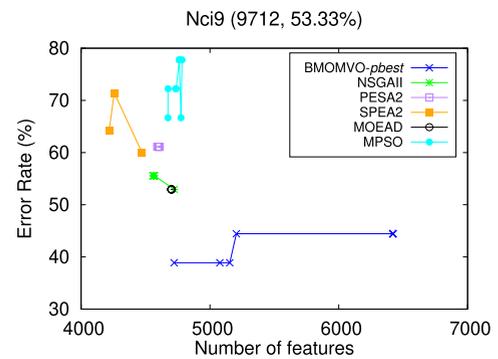
(a) GLIOMA



(a) GLIOMA



(b) Nci9



(b) Nci9

FIGURE 6. Comparisons of BMOMVO-*pbest* with the BMOMVO on GLIOMA and Nci9 datasets.

FIGURE 7. Comparisons between BMOMVO-*pbest* and other methods for GLIOMA and Nci9 benchmark datasets.

Tic-tac-toe, Vote, and WineEW. This observation is also confirmed by the same F-test's results for both algorithms, which can be seen in Table 5. For most of the datasets, BMOMVO-*pbest* contains two or more subsets, which successfully nominated a smaller number of features and maintained or attained a better error value than that revealed using all features and other competitors.

The computational time of BMOMVO-*pbest* is compared to those recorded for NSGAI, PESA2, SPEA2, MOEAD, and MOPSO approaches in Table 7. As per records in Table 7, we see that the BMOMVO-*pbest* has shown the fastest performance on HeartEW, Vote, WineEW, and Zoo. We observed that the PESA2 is the fastest method on six datasets, including Exactly, SonarEW, CongressEW, Leukemia, GLIOMA,

TABLE 5. Comparison of error rate results for BMOMVO-*pbest* versus NSGAI, PESA2, SPEA2, MOEAD, and MOPSO methods for all datasets.

Benchmark	Measure	BMOMVO- <i>pbest</i>	NSGAI	PESA2	SPEA2	MOEAD	MOPSO
BreastEW	AVG	9.472	8.833	6.799	8.709	9.342	9.757
	STD	0.835	0.749	1.359	0.900	1.845	1.110
	BEST	8.070	7.602	4.289	7.018	7.018	7.937
	WORST	11.436	11.014	10.819	11.028	14.766	12.939
Exactly	AVG	21.768	27.927	32.417	25.975	35.893	25.162
	STD	4.149	3.522	5.044	2.160	5.285	2.469
	BEST	15.333	22.111	23.750	21.556	22.778	20.333
	WORST	31.889	37.000	45.417	30.111	44.889	33.333
HeartEW	AVG	23.593	24.073	28.008	24.298	27.493	23.702
	STD	1.357	1.748	2.560	1.853	3.736	1.658
	BEST	21.164	19.753	22.222	20.910	21.914	19.929
	WORST	26.543	26.337	31.852	26.984	36.111	27.160
SonarEW	AVG	23.805	25.173	23.129	23.710	25.168	19.226
	STD	2.653	2.769	3.896	2.432	4.659	2.263
	BEST	14.286	20.370	15.873	20.833	12.698	13.228
	WORST	30.159	32.143	30.159	31.481	36.735	23.413
CongressEW	AVG	4.392	5.335	9.046	5.107	13.764	7.212
	STD	2.136	0.413	4.527	0.375	4.688	3.712
	BEST	2.290	4.580	6.107	4.326	4.962	3.817
	WORST	12.595	5.954	23.282	5.725	24.427	16.794
KrvskpEW	AVG	10.506	9.586	7.313	9.608	24.868	9.779
	STD	2.837	1.220	4.181	2.945	4.971	3.301
	BEST	5.869	7.238	3.354	3.163	11.783	5.065
	WORST	17.901	12.135	18.144	16.302	34.393	16.934
Tic-tac-toe	AVG	26.413	26.798	28.730	27.002	28.404	28.203
	STD	0.630	1.079	1.361	1.100	2.715	1.171
	BEST	25.174	24.653	25.139	24.236	24.537	26.042
	WORST	28.704	28.958	31.875	28.958	35.880	30.729
Vote	AVG	5.635	7.278	8.111	7.210	15.686	6.816
	STD	1.543	0.841	6.488	0.944	4.907	3.619
	BEST	4.444	5.556	3.333	5.556	7.778	4.444
	WORST	10.000	8.889	31.667	8.889	26.667	22.778
WineEW	AVG	12.040	8.167	13.923	8.377	16.697	12.290
	STD	1.799	0.396	5.111	1.049	5.600	2.968
	BEST	8.951	7.099	3.704	7.407	6.667	5.185
	WORST	15.556	8.642	27.037	13.426	29.167	18.981
Zoo	AVG	16.281	18.558	23.797	22.417	26.515	19.695
	STD	2.842	3.325	7.832	3.248	5.376	3.105
	BEST	11.290	14.952	7.527	16.852	17.972	11.828
	WORST	22.581	26.602	38.710	27.247	40.323	28.226
semeion	AVG	2.283	4.283	3.016	3.532	2.642	3.330
	STD	0.243	0.489	0.616	0.453	0.315	0.343
	BEST	1.674	3.222	1.813	2.766	1.953	2.563
	WORST	2.836	5.083	4.498	4.538	3.301	4.059
Leukemia	AVG	16.088	16.061	17.576	15.676	26.131	15.202
	STD	1.862	3.240	1.572	2.173	2.112	1.713
	BEST	11.364	9.091	13.636	12.121	22.727	11.688
	WORST	18.182	22.727	18.182	20.455	29.545	18.182
GLIOMA	AVG	8.028	13.525	13.439	20.048	36.870	21.251
	STD	2.747	2.540	0.324	0.185	3.569	1.386
	BEST	3.333	6.667	13.333	20.000	26.667	18.667
	WORST	16.667	20.000	14.444	20.833	40.000	24.444
Nci9	AVG	41.674	52.727	61.286	66.962	50.716	72.633
	STD	2.682	1.551	0.645	2.480	4.071	2.045
	BEST	37.500	47.794	61.111	62.031	47.059	67.593
	WORST	47.222	55.556	64.444	71.376	58.824	76.389
F-test		2.1429	3.2143	3.9286	3.2143	5.0714	3.4286

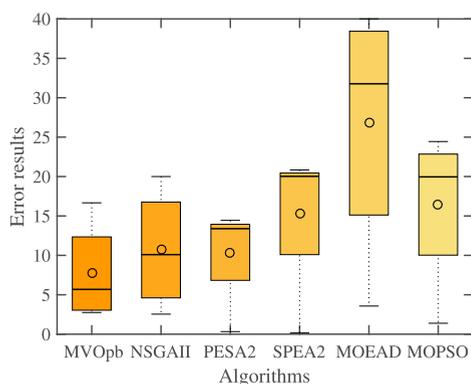
and Nci9. The computational time results show that the MOPSO has the slowest exploratory and exploitative trend compared to other methods, including BMOMVO-*pbest*.

The experimental results on 14 datasets show that the proposed BMOMVO-*pbest* approach can effectively

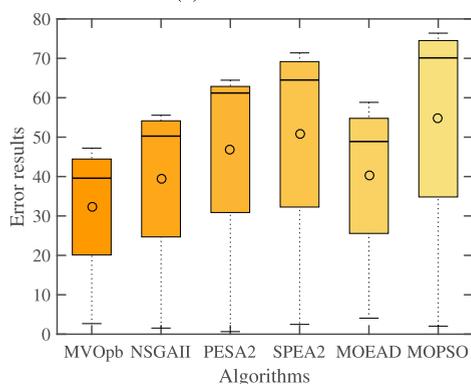
eliminate irrelevant and/or redundant features with a particular classification rate compared to other competitors. Furthermore, the proposed BMOMVO-*pbest* improves a set of non-dominated feature subsets that contain better error rates and smaller feature subsets than employing the full features.

TABLE 6. P-values of the Wilcoxon test of BMOMVO_pbest vs other algorithms (p ≥ 0.05 are underlined).

Dataset	NSGAI	PESA2	SPEA2	MOEAD	MOPSO
BreastEW	0.0022619	3.0378e-09	0.001404	0.1579	0.28708
Exactly	<u>2.7346e-07</u>	<u>1.1645e-09</u>	4.21e-06	<u>2.1213e-10</u>	<u>7.8522e-05</u>
HeartEW	0.12937	<u>5.8405e-08</u>	0.086145	<u>1.0393e-05</u>	0.75594
SonarEW	0.18075	0.53433	0.22248	0.1037	<u>1.9213e-08</u>
CongressEW	<u>0.0003341</u>	<u>7.3325e-09</u>	<u>0.00083871</u>	<u>2.8489e-10</u>	<u>0.00022084</u>
KrvskpEW	0.14127	<u>9.2113e-05</u>	0.42896	<u>7.3891e-11</u>	0.34783
Tic-tac-toe	0.13909	<u>2.5792e-09</u>	0.0079234	<u>0.0010542</u>	<u>1.0579e-07</u>
Vote	<u>1.1483e-05</u>	0.19808	<u>1.4278e-05</u>	<u>7.839e-11</u>	<u>0.022032</u>
WineEW	<u>9.2771e-12</u>	0.088865	<u>1.5849e-10</u>	<u>0.00036519</u>	0.95871
Zoo	<u>0.015534</u>	<u>8.6783e-06</u>	<u>2.7318e-08</u>	<u>5.25e-10</u>	<u>6.0796e-05</u>
Semeion	<u>3.0104e-11</u>	<u>5.9984e-07</u>	<u>3.6851e-11</u>	<u>1.1593e-05</u>	<u>5.1996e-11</u>
Leukemia	0.98783	<u>0.00020419</u>	0.41009	<u>1.561e-11</u>	0.054116
GLIOMA	<u>8.9219e-09</u>	<u>1.8636e-10</u>	<u>1.7285e-12</u>	<u>1.56e-11</u>	<u>2.2478e-11</u>
Nci9	<u>2.6774e-11</u>	<u>3.108e-12</u>	<u>2.9785e-11</u>	<u>8.5306e-11</u>	<u>2.9229e-11</u>



(a) GLIOMA



(b) Nci9

FIGURE 8. Boxplots of error results for GLIOMA and Nci9 benchmark datasets.

The comparative results emphasize the validity and enhanced efficacy of the proposed multi-objective wrapper FS model.

To measure the overall efficacy, we have utilized a Wilcoxon statistical examination with a 5% significance level. We performed this test on the gained average classification error results. Table 6 exposes Wilcoxon’s test P-values between the BMOMVO_pbest and other approaches.

The P-values less than 0.05 designate that the results of the BMOMVO_pbest have statistically significant variances compared with other methodologies. Values greater than 0.05 (underlined) are not significant concerning the other peers. According to this table, the results of the BMOMVO_pbest are statistically significant in dealing with most of the data sets.

The memory term in the proposed method can facilitate the exchange of search information more coherently and preserves a more stable balance between exploration and exploitation trends in the MOMVO-based FS method.

In addition, some performance merits of the proposed BMOMVO_pbest compared to NSGAI, PESA2, SPEA2, MOEAD, and MOPSO techniques are because of the exploratory and exploitative advantages of conventional MOMVO. For instance, using wormholes, some variables of agents can be re-spanned around the best universe attained so far over the optimization stages. This can guarantee sufficient exploitation around the promising zones of the feature space. Furthermore, adaptive WEP values can smoothly emphasize exploitation trends within more iterations, and adaptive TDR values increase the accuracy of exploitative tendencies over the iterations while abrupt changes also help the algorithm to resolve LO stagnation.

3) COMPARISON WITH CONVENTIONAL FILTER FS METHODS

In this part of experiments, we compared the performance of the proposed BMOMVO_pbest in terms of error rates to well-established filter-based approaches [98]: correlation-based (Correlation) [99], ReliefF, InfoGain [100], and symmetrical [101]. Filter methods can select features independent of the used classification algorithm. However, the main drawback of these methods is that it methodically disregards the impacts of the obtained feature subset on the efficacy of the induction engine. The best subset depends

TABLE 7. Comparisons of computational time (seconds) obtained by BMOMVO-*pbest* with those recorded for NSGAI, PESA2, SPEA2, MOEAD, and MOPSO methods on all datasets.

Benchmark	Measure	BMOMVO_ <i>pbest</i>	NSGAI	PESA2	SPEA2	MOEAD	MOPSO
BreastEW	AVG	119.851	133.573	114.268	124.657	104.889	157.362
	STD	3.349	1.252	2.685	0.979	2.865	1.805
	BEST	108.269	129.932	106.252	122.811	100.437	153.340
	WORST	125.074	135.425	119.056	126.143	109.510	160.906
Exactly	AVG	127.048	151.683	112.830	140.030	115.443	141.450
	STD	2.954	3.106	1.412	3.203	3.318	1.868
	BEST	119.180	145.489	110.208	132.874	108.300	137.644
	WORST	132.315	157.048	115.888	143.953	120.163	145.288
HeartEW	AVG	78.670	117.840	86.501	115.130	86.495	379.106
	STD	0.406	0.865	0.708	0.541	3.509	106.811
	BEST	78.049	115.532	85.177	113.824	80.584	144.921
	WORST	79.779	119.030	87.565	116.135	91.733	472.226
SonarEW	AVG	92.275	113.647	83.843	108.715	88.783	161.811
	STD	2.447	0.805	0.557	0.813	2.551	1.730
	BEST	85.216	111.818	82.826	107.230	85.256	157.003
	WORST	96.564	114.926	84.882	110.130	95.518	163.719
CongressEW	AVG	92.076	120.557	90.184	111.501	91.063	117.918
	STD	0.861	0.667	0.884	1.212	2.403	1.351
	BEST	89.539	119.103	88.173	107.722	87.682	113.774
	WORST	93.733	121.926	91.660	113.147	98.516	119.885
KrvskpEW	AVG	645.533	508.060	399.165	406.094	372.800	625.808
	STD	17.341	22.687	38.096	21.048	54.882	10.587
	BEST	596.584	462.537	325.703	381.441	266.820	602.971
	WORST	679.468	543.807	454.428	461.784	467.000	645.376
Tic-tac-toe	AVG	114.598	152.871	113.592	134.916	107.781	125.750
	STD	1.744	1.682	1.575	1.967	4.107	1.934
	BEST	108.607	149.732	109.886	131.522	101.890	121.563
	WORST	117.164	155.965	118.922	140.791	115.612	128.278
Vote	AVG	81.035	113.497	83.536	105.890	298.940	101.956
	STD	0.483	0.377	0.573	0.779	80.935	0.876
	BEST	80.052	112.568	81.973	103.774	183.399	100.677
	WORST	82.010	114.118	84.557	107.451	404.279	105.780
WineEW	AVG	76.477	109.957	81.772	105.225	84.378	93.786
	STD	0.422	0.434	0.775	0.551	3.385	0.449
	BEST	75.648	109.120	79.887	104.205	80.663	92.910
	WORST	77.574	110.699	84.539	106.379	90.651	95.312
Zoo	AVG	80.328	113.334	90.482	125.571	90.262	100.242
	STD	0.499	0.490	0.489	1.028	3.800	0.451
	BEST	79.419	111.891	89.585	123.364	84.627	99.411
	WORST	81.460	114.249	91.483	127.581	95.796	102.011
semeion	AVG	1025.448	914.917	663.784	662.001	755.691	1074.858
	STD	17.083	21.567	30.307	14.648	44.250	13.470
	BEST	983.794	878.434	596.159	625.965	661.211	1045.683
	WORST	1052.187	958.875	715.050	688.606	819.110	1104.238
Leukemia	AVG	436.758	408.313	299.533	329.817	347.249	7239.272
	STD	16.283	13.078	3.216	9.157	11.702	85.296
	BEST	384.381	384.044	290.480	308.033	327.390	6926.832
	WORST	467.675	427.125	306.724	346.134	365.748	7420.000
GLIOMA	AVG	495.267	186.143	168.688	179.555	205.567	413.765
	STD	122.815	2.832	6.175	5.934	5.938	6.416
	BEST	275.425	179.207	148.458	164.437	188.864	400.102
	WORST	616.384	190.429	174.134	188.303	213.405	424.431
Nci9	AVG	480.315	438.787	343.928	357.847	378.667	810.332
	STD	6.570	3.312	5.665	5.581	6.613	5.847
	BEST	467.186	430.944	326.044	344.692	358.630	795.950
	WORST	494.330	443.598	350.473	364.997	386.634	821.645

on some biased terms and the induction method. Concerning this assumption, wrapper approaches employ a classifier to assess the excellence of nominated features. Regardless of the used learning machine, wrappers aim to propose a simple and influential way to tackle FS tasks.

The error rates of BMOMVO-*pbest* versus other filter methods are compared in Table 8.

As per the rates in Table 8, it is observed that the BMOMVO-*pbest* optimizer can beat other filter methods on 64.28 % of datasets, whereas the ReliefF and symmetrical

TABLE 8. Comparison between BMOMVO-*pbest* error rates and filter based feature selection methods.

Datasets	ReliefF	InfoGain	Correlation	Symmetrical	BMOMVO- <i>pbest</i>
BreastEW	5.724	6.521	7.171	6.528	9.472
Exactly	32.165	31.313	32.042	31.313	21.768
HeartEW	22.395	18.494	19.679	18.198	23.593
SonarEW	23.756	26.034	26.930	26.034	23.805
CongressEW	4.452	4.452	4.452	4.452	4.392
KrvskpEW	4.828	5.733	6.204	5.778	10.506
Tic-tac-toe	19.523	19.523	19.523	19.523	26.413
Vote	5.720	5.720	5.720	5.720	5.635
WineEW	20.680	16.195	15.816	18.071	12.040
Zoo	26.707	20.836	26.707	20.711	16.281
semeion	3.312	2.729	2.720	2.743	2.283
Leukemia	11.241	8.647	6.907	8.602	16.088
GLIOMA	22.956	24.200	22.795	24.057	8.028
Nci9	55.041	63.702	55.058	63.702	41.674

techniques have attained the best rates only for 3 and 2 datasets, respectively. For the Tic-tac-toe case, we see that all filter methods have obtained the same error. These results indicate that wrapper-based BMOMVO-*pbest* can offer more improvements in the rates compared to filter-based techniques. The reason is that the BMOMVO-*pbest*-based wrapper can consider both labels and dependencies throughout the selection process of related subsets. The results show that wrapper FS can enhance the quality of feature sets compared to filter-based methods. The main reason is that the wrapper methods can use either labels or dependencies when selecting the relevant subsets. Based on these results, it can be determined that the developed MOMVO-based wrapper shows performance merits compared to other well-known methods and outperforms/competes with studied filter methods as well.

VI. CONCLUSION AND FUTURE WORKS

In this work, a binary multi-objective variant of Multi-Verse Optimizer (MOMVO) was proposed for the feature selection task in machine learning. The MOMVO algorithm was designed as a wrapper-based feature selection approach based on utilizing three cosmology concepts: white hole, black hole, and wormhole. In addition, a variant of the MOMVO that incorporates the personal best solution in its updating was proposed as well. Unlike most of the evolutionary wrapper approaches, the proposed MOMVO-based approaches dealt with an accuracy of the model and the reduction in the dimensional as a multi-objective optimization problem. The results of the conducted experiments based on 14 benchmark datasets showed that the BMOMVO-*pbest* approach outperforms BMOMVO in the majority of the datasets. Moreover, BMOMVO-*pbest* showed superior results when compared with the state-of-art multi-objective optimization algorithms for feature selection. We are considering employing more objectives such as parameter optimization and fitness selection and employing different metaheuristic algorithms for the multi-objective problem in future work.

REFERENCES

- [1] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Mag.*, vol. 17, no. 3, p. 37, 1996.
- [2] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. Amsterdam, The Netherlands: Elsevier, 2011.
- [3] H.-L. Chen, G. Wang, C. Ma, Z.-N. Cai, W.-B. Liu, and S.-J. Wang, "An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease," *Neurocomputing*, vol. 184, pp. 131–144, Apr. 2016.
- [4] M. Mafarja, I. Aljarah, A. A. Heidari, A. I. Hammouri, H. Faris, A.-Z. Ala'M, and S. Mirjalili, "Evolutionary population dynamics and grasshopper optimization approaches for feature selection problems," *Knowl.-Based Syst.*, vol. 145, pp. 25–45, Apr. 2018.
- [5] A. M. Al-Zoubi, M. A. Hassonah, A. A. Heidari, H. Faris, M. Mafarja, and I. Aljarah, "Evolutionary competitive swarm exploring optimal support vector machines and feature weighting," *Soft Comput.*, vol. 25, no. 4, pp. 3335–3352, Feb. 2021.
- [6] L. Shen, H. Chen, Z. Yu, W. Kang, B. Zhang, H. Li, B. Yang, and D. Liu, "Evolving support vector machines using fruit fly optimization for medical data classification," *Knowl.-Based Syst.*, vol. 96, pp. 61–75, Mar. 2016.
- [7] L. Al-Qaisi, M. A. Hassonah, M. M. Al-Zoubi, and A.-Z. Ala'M, "A review of evolutionary data clustering algorithms for image segmentation," in *Evolutionary Data Clustering: Algorithms and Applications*. Singapore: Springer, 2021, p. 201.
- [8] C. Yu, M. Chen, K. Cheng, X. Zhao, C. Ma, F. Kuang, and H. Chen, "SGOA: Annealing-behaved grasshopper optimizer for global tasks," *Eng. Comput.*, pp. 1–28, Jan. 2021.
- [9] L. Hu, G. Hong, J. Ma, X. Wang, and H. Chen, "An efficient machine learning approach for diagnosis of paraquat-poisoned patients," *Comput. Biol. Med.*, vol. 59, pp. 116–124, Apr. 2015.
- [10] W. Shan, Z. Qiao, A. A. Heidari, H. Chen, H. Turabieh, and Y. Teng, "Double adaptive weights for stabilization of moth flame optimizer: Balance analysis, engineering cases, and medical diagnosis," *Knowl.-Based Syst.*, vol. 214, Feb. 2021, Art. no. 106728.
- [11] D. Zhao, L. Liu, F. Yu, A. A. Heidari, M. Wang, G. Liang, K. Muhammad, and H. Chen, "Chaotic random spare ant colony optimization for multi-threshold image segmentation of 2D Kapur entropy," *Knowl.-Based Syst.*, vol. 216, Mar. 2021, Art. no. 106510.
- [12] Q. Jiang, F. Shao, W. Lin, K. Gu, G. Jiang, and H. Sun, "Optimizing multistage discriminative dictionaries for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2035–2048, Aug. 2018.
- [13] L. Xu, S. Jiang, J. Wu, and Q. Zou, "An *in silico* approach to identification, categorization and prediction of nucleic acid binding proteins," *Briefings Bioinf.*, vol. 22, no. 3, May 2021, Art. no. bbaa171.
- [14] X.-F. Wang, P. Gao, Y.-F. Liu, H.-F. Li, and F. Lu, "Predicting thermophilic proteins by machine learning," *Current Bioinf.*, vol. 15, no. 5, pp. 493–502, Oct. 2020.

- [15] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *Proc. 20th Int. Conf. Mach. Learn. (ICML)*, 2003, pp. 856–863.
- [16] Y. Zhang, R. Liu, A. A. Heidari, X. Wang, Y. Chen, M. Wang, and H. Chen, "Towards augmented kernel extreme learning models for bankruptcy prediction: Algorithmic behavior and comprehensive analysis," *Neurocomputing*, vol. 430, pp. 185–212, Mar. 2021.
- [17] J. Hu, H. Chen, A. A. Heidari, M. Wang, X. Zhang, Y. Chen, and Z. Pan, "Orthogonal learning covariance matrix for defects of grey wolf optimizer: Insights, balance, diversity, and feature selection," *Knowl.-Based Syst.*, vol. 213, Feb. 2021, Art. no. 106684.
- [18] Y. Zhang, R. Liu, X. Wang, H. Chen, and C. Li, "Boosted binary harris hawks optimizer and feature selection," *Eng. Comput.*, pp. 1–30, May 2020.
- [19] J. Zhang and B. Liu, "A review on the recent developments of sequence-based protein feature extraction methods," *Current Bioinf.*, vol. 14, no. 3, pp. 190–199, Mar. 2019.
- [20] X. Zhao, D. Li, B. Yang, C. Ma, Y. Zhu, and H. Chen, "Feature selection based on improved ant colony optimization for online detection of foreign fiber in cotton," *Appl. Soft Comput.*, vol. 24, pp. 585–596, Nov. 2014.
- [21] M. Dash and H. Liu, "Feature selection for classification," *Intell. Data Anal.*, vol. 1, nos. 1–4, pp. 131–156, 1997.
- [22] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Jan. 2003.
- [23] J. Li, S. Fong, R. K. Wong, and V. W. Chu, "Adaptive multi-objective swarm fusion for imbalanced data classification," *Inf. Fusion*, vol. 39, pp. 1–24, Jan. 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1566253517302087>
- [24] V. Bolón-Canedo and A. Alonso-Betanzos, "Ensembles for feature selection: A review and future trends," *Inf. Fusion*, vol. 52, pp. 1–12, Dec. 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1566253518303440>
- [25] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, vol. 454. USA: Springer, 2012.
- [26] E. Hancer, B. Xue, M. Zhang, D. Karaboga, and B. Akay, "Pareto front feature selection based on artificial bee colony optimization," *Inf. Sci.*, vol. 422, pp. 462–479, Jan. 2017.
- [27] X. Zhang, Y. Xu, C. Yu, A. A. Heidari, S. Li, H. Chen, and C. Li, "Gaussian mutational chaotic fruit fly-built optimization and feature selection," *Expert Syst. Appl.*, vol. 141, Mar. 2020, Art. no. 112976. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417419306943>
- [28] M. L. Raymer, W. F. Punch, E. D. Goodman, L. A. Kuhn, and A. K. Jain, "Dimensionality reduction using genetic algorithms," *IEEE Trans. Evol. Comput.*, vol. 4, no. 2, pp. 164–171, Jul. 2000.
- [29] D. A. Al-Qudah, A. M. Al-Zoubi, P. A. Castillo-Valdivieso, and H. Faris, "Sentiment analysis for e-payment service providers using evolutionary eXtreme gradient boosting," *IEEE Access*, vol. 8, pp. 189930–189944, 2020.
- [30] M. Mafarja, R. Jarrar, S. Ahmad, and A. A. Abusnaina, "Feature selection using binary particle swarm optimization with time varying inertia weight strategies," in *Proc. 2nd Int. Conf. Future Netw. Distrib. Syst.*, Amman, Jordan, vol. 2, Jun. 2018.
- [31] M. Rong, D. Gong, Y. Zhang, Y. Jin, and W. Pedrycz, "Multidirectional prediction approach for dynamic multiobjective optimization problems," *IEEE Trans. Cybern.*, vol. 49, no. 9, pp. 3362–3374, Sep. 2019.
- [32] H. Yu, W. Li, C. Chen, J. Liang, W. Gui, M. Wang, and H. Chen, "Dynamic Gaussian bare-bones fruit fly optimizers with abandonment mechanism: Method and analysis," *Eng. Comput.*, pp. 1–29, Oct. 2020.
- [33] J. Tu, H. Chen, J. Liu, A. A. Heidari, X. Zhang, M. Wang, R. Ruby, and Q.-V. Pham, "Evolutionary biogeography-based whale optimization methods with communication structure: Towards measuring the balance," *Knowl.-Based Syst.*, vol. 212, Jan. 2021, Art. no. 106642.
- [34] X. Wang, X. Wang, L. Jin, R. Lv, B. Dai, M. He, and T. Lv, "Evolutionary algorithm-based and network architecture search-enabled multiobjective traffic classification," *IEEE Access*, vol. 9, pp. 52310–52325, 2021.
- [35] Y. Tian, X. Zhang, R. Cheng, C. He, and Y. Jin, "Guiding evolutionary multiobjective optimization with generic front modeling," *IEEE Trans. Cybern.*, vol. 50, no. 3, pp. 1106–1119, Mar. 2020.
- [36] G. Kou, Y. Xu, Y. Peng, F. Shen, Y. Chen, K. Chang, and S. Kou, "Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection," *Decis. Support Syst.*, vol. 140, Jan. 2021, Art. no. 113429.
- [37] M. Wang and H. Chen, "Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis," *Appl. Soft Comput.*, vol. 88, Mar. 2020, Art. no. 105946.
- [38] H. Chen, A. A. Heidari, H. Chen, M. Wang, Z. Pan, and A. H. Gandomi, "Multi-population differential evolution-assisted harris hawks optimization: Framework and case studies," *Future Gener. Comput. Syst.*, vol. 111, pp. 175–198, Oct. 2020.
- [39] S. Mirjalili, S. M. Mirjalili, and A. Hatamlou, "Multi-verse optimizer: A nature-inspired algorithm for global optimization," *Neural Comput. Appl.*, vol. 27, no. 2, pp. 495–513, 2016.
- [40] M. Wang, H. Chen, B. Yang, X. Zhao, L. Hu, Z. Cai, H. Huang, and C. Tong, "Toward an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses," *Neurocomputing*, vol. 267, pp. 69–84, Dec. 2017.
- [41] J. Xia, H. Chen, Q. Li, M. Zhou, L. Chen, X. Cai, Y. Fang, and H. Zhou, "Ultrasound-based differentiation of malignant and benign thyroid Nodules: An extreme learning machine approach," *Comput. Methods Programs Biomed.*, vol. 147, pp. 37–49, Oct. 2017.
- [42] X. Zhao, X. Zhang, Z. Cai, X. Tian, X. Wang, Y. Huang, H. Chen, and L. Hu, "Chaos enhanced grey wolf optimization wrapped ELM for diagnosis of paraquat-poisoned patients," *Comput. Biol. Chem.*, vol. 78, pp. 481–490, Feb. 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1476927118307965>
- [43] F. Jiménez, G. Sánchez, J. M. García, G. Sciacicco, and L. Miralles, "Multi-objective evolutionary feature selection for online sales forecasting," *Neurocomputing*, vol. 234, pp. 75–92, Apr. 2017.
- [44] U. Mlakar, I. Fister, J. Brest, and B. Potočnik, "Multi-objective differential evolution for feature selection in facial expression recognition systems," *Expert Syst. Appl.*, vol. 89, pp. 129–137, Dec. 2017.
- [45] B. Huang, B. Buckley, and T.-M. Kechadi, "Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications," *Expert Syst. Appl.*, vol. 37, no. 5, pp. 3638–3646, May 2010.
- [46] A. Sheta, H. Faris, A. Rodan, E. Kovač-Andrić, and A. M. Al-Zoubi, "Cycle reservoir with regular jumps for forecasting ozone concentrations: Two real cases from the east of croatia," *Air Qual., Atmos. Health*, vol. 11, no. 5, pp. 559–569, Jun. 2018.
- [47] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, Jul. 2018.
- [48] S. Sasikala, S. A. A. Balamurugan, and S. Geetha, "A novel adaptive feature selector for supervised classification," *Inf. Process. Lett.*, vol. 117, pp. 25–34, Jan. 2017.
- [49] S. Sasikala and S. Geetha, "A novel memetic algorithm for discovering knowledge in binary and multi class predictions based on support vector machine," *Appl. Soft Comput.*, vol. 49, pp. 407–422, Dec. 2016.
- [50] C. Qu, L. Zhang, J. Li, F. Deng, Y. Tang, X. Zeng, and X. Peng, "Improving feature selection performance for classification of gene expression data using harris hawks optimizer with variable neighborhood learning," *Briefings Bioinf.*, Apr. 2021.
- [51] S. A. A. Balamurugan and S. G. Nancy, "An efficient feature selection and classification using optimal radial basis function neural network," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 26, no. 5, pp. 695–715, Oct. 2018.
- [52] D. A. B. Singh, S. A. A. Balamurugan, and E. J. Leavline, "An unsupervised feature selection algorithm with feature ranking for maximizing performance of the classifiers," *Int. J. Autom. Comput.*, vol. 12, no. 5, pp. 511–517, Oct. 2015.
- [53] S. G. Nancy and S. Appavu, "Optimal feature selection for classification using rough set-based CGA-NN classifier," *Int. J. Bus. Intell. Data Mining*, vol. 11, no. 4, pp. 357–378, 2016.
- [54] Q. Al-Tashi, S. J. Abdulkadir, H. M. Rais, S. Mirjalili, H. Alhussian, M. G. Ragab, and A. Alqushaibi, "Binary multi-objective grey wolf optimizer for feature selection in classification," *IEEE Access*, vol. 8, pp. 106247–106263, 2020.
- [55] A.-Z. Ala'M, A. A. Heidari, M. Habib, H. Faris, I. Aljarah, and M. A. Hassonah, "Salp chain-based optimization of support vector machines and feature weighting for medical diagnostic information systems," in *Evolutionary Machine Learning Techniques*. Singapore: Springer, 2020, pp. 11–34.
- [56] A. S. Sadiq, H. Faris, A.-Z. Ala'M, S. Mirjalili, and K. Z. Ghafoor, "Fraud detection model based on multi-verse features extraction approach for smart city applications," in *Smart Cities Cybersecurity and Privacy*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 241–251.

- [57] A. Al-Zoubi, J. Alqatawna, H. Faris, and M. A. Hassonah, "Spam profiles detection on social networks using computational intelligence methods: The effect of the lingual context," *J. Inf. Sci.*, vol. 47, no. 1, 2021, Art. no. 0165551519861599.
- [58] W. Hijawi, H. Faris, J. Alqatawna, A. M. Al-Zoubi, and I. Aljarah, "Improving email spam detection using content based feature engineering approach," in *Proc. IEEE Jordan Conf. Appl. Electr. Eng. Comput. Technol. (AEECT)*, Oct. 2017, pp. 1–6.
- [59] E. de la Hoz, E. de la Hoz, A. Ortiz, J. Ortega, and A. Martínez-Álvarez, "Feature selection by multi-objective optimisation: Application to network anomaly detection by hierarchical self-organising maps," *Knowl.-Based Syst.*, vol. 71, pp. 322–338, Nov. 2014.
- [60] P. Martín-Smith, J. Ortega, J. Asensio-Cubero, J. Q. Gan, and A. Ortiz, "A supervised filter method for multi-objective feature selection in EEG classification based on multi-resolution analysis for BCI," *Neurocomputing*, vol. 250, pp. 45–56, Aug. 2017.
- [61] C. J. Tan, C. P. Lim, and Y. Cheah, "A multi-objective evolutionary algorithm-based ensemble optimizer for feature selection and classification with neural network models," *Neurocomputing*, vol. 125, pp. 217–228, Feb. 2014.
- [62] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II," in *Proc. Int. Conf. Parallel Problem Solving Nature*. Berlin, Germany: Springer, 2000, pp. 849–858.
- [63] H. Soyel, U. Tekguc, and H. Demirel, "Application of NSGA-II to feature selection for facial expression recognition," *Comput. Electr. Eng.*, vol. 37, no. 6, pp. 1232–1240, Nov. 2011.
- [64] A. Khan and A. R. Baig, "Multi-objective feature subset selection using non-dominated sorting genetic algorithm," *J. Appl. Res. Technol.*, vol. 13, no. 1, pp. 145–159, Feb. 2015.
- [65] Y. Zhu, J. Liang, J. Chen, and Z. Ming, "An improved NSGA-III algorithm for feature selection used in intrusion detection," *Knowl.-Based Syst.*, vol. 116, pp. 74–85, Jan. 2017.
- [66] M. Binder, J. Moosbauer, J. Thomas, and B. Bischl, "Multi-objective hyperparameter tuning and feature selection using filter ensembles," in *Proc. Genetic Evol. Comput. Conf.*, 2020, pp. 471–479.
- [67] Q. Al-Tashi, S. J. Abdulkadir, H. Md Rais, S. Mirjalili, and H. Alhussian, "Approaches to multi-objective feature selection: A systematic literature review," *IEEE Access*, vol. 8, pp. 125076–125096, 2020.
- [68] B. Niu, W. Yi, L. Tan, S. Geng, and H. Wang, "A multi-objective feature selection method based on bacterial foraging optimization," *Natural Comput.*, vol. 20, pp. 1–14, Jul. 2019.
- [69] B. Xue, W. Fu, and M. Zhang, "Multi-objective feature selection in classification: A differential evolution approach," in *Proc. Asia-Pacific Conf. Simulated Evol. Learn.* Berlin, Germany: Springer, 2014, pp. 516–528.
- [70] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimization for feature selection in classification: A multi-objective approach," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1656–1671, Dec. 2013.
- [71] M. Amoozegar and B. Minaei-Bidgoli, "Optimizing multi-objective PSO based feature selection method using a feature elitism mechanism," *Expert Syst. Appl.*, vol. 113, pp. 499–514, Dec. 2018.
- [72] Y. Zhang, D.-W. Gong, X.-Y. Sun, and Y.-N. Guo, "A PSO-based multi-objective multi-label feature selection method in classification," *Sci. Rep.*, vol. 7, no. 1, p. 376, Dec. 2017.
- [73] I. Behravan, O. Dehghantanha, and S. H. Zahiri, "An optimal SVM with feature selection using multi-objective PSO," in *Proc. 1st Conf. Swarm Intell. Evol. Comput. (CSIEC)*, Mar. 2016, pp. 76–81.
- [74] Y. Zhang, D.-W. Gong, and J. Cheng, "Multi-objective particle swarm optimization approach for cost-based feature selection in classification," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 14, no. 1, pp. 64–75, Jan./Feb. 2017.
- [75] Z. Yong, G. Dun-Wei, and Z. Wan-Qiu, "Feature selection of unreliable data using an improved multi-objective PSO algorithm," *Neurocomputing*, vol. 171, pp. 1281–1290, Jan. 2016.
- [76] E. Hancer, B. Xue, M. Zhang, D. Karaboga, and B. Akay, "A multi-objective artificial bee colony approach to feature selection using fuzzy mutual information," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, May 2015, pp. 2420–2427.
- [77] B. B. Dickson, S. Wang, R. Dong, and C. Wen, "A feature selection method based on multi-objective optimisation with gravitational search algorithm," in *Geo-Informatics in Resource Management and Sustainable Ecosystem*. Berlin, Germany: Springer, 2015, pp. 549–558.
- [78] Y. Xu, H. Chen, J. Luo, Q. Zhang, S. Jiao, and X. Zhang, "Enhanced moth-flame optimizer with mutation strategy for global optimization," *Inf. Sci.*, vol. 492, pp. 181–203, Aug. 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0020025519303299>
- [79] X. Xu and H.-L. Chen, "Adaptive computational chemotaxis based on field in bacterial foraging optimization," *Soft Comput.*, vol. 18, no. 4, pp. 797–807, Apr. 2014.
- [80] C. A. C. Coello, G. T. Pulido, and M. S. Lechuga, "Handling multiple objectives with particle swarm optimization," *IEEE Trans. Evol. Comput.*, vol. 8, no. 3, pp. 256–279, Jun. 2004.
- [81] J. Knowles and D. Corne, "The Pareto archived evolution strategy: A new baseline algorithm for Pareto multiobjective optimisation," in *Proc. Congr. Evol. Comput. (CEC)*, vol. 1, 1999, pp. 98–105.
- [82] S. Mirjalili, P. Jangir, S. Z. Mirjalili, S. Saremi, and I. N. Trivedi, "Optimization of problems with multiple objectives using the multi-verse optimization algorithm," *Knowl.-Based Syst.*, vol. 134, pp. 50–71, Oct. 2017.
- [83] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*. Boston, MA, USA: Kluwer, 1998.
- [84] P. Moradi and M. Gholampour, "A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy," *Appl. Soft Comput.*, vol. 43, pp. 117–130, Jun. 2016.
- [85] C. Li, L. Hou, B. Y. Sharma, H. Li, C. Chen, Y. Li, X. Zhao, H. Huang, Z. Cai, and H. Chen, "Developing a new intelligent system for the diagnosis of tuberculous pleural effusion," *Comput. Methods Programs Biomed.*, vol. 153, pp. 211–225, Jan. 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169260717309288>
- [86] F. Pernkopf, "Bayesian network classifiers versus selective k-NN classifier," *Pattern Recognit.*, vol. 38, no. 1, pp. 1–10, Jan. 2005.
- [87] S. Mirjalili and A. Lewis, "S-shaped versus V-shaped transfer functions for binary particle swarm optimization," *Swarm Evol. Comput.*, vol. 9, pp. 1–14, Apr. 2013.
- [88] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. Comput. Simulation*, vol. 5, Oct. 1997, pp. 4104–4108.
- [89] M. Lichman. (2013). *UCI Machine Learning Repository*. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [90] Z. Lv, L. Qiao, A. K. Singh, and Q. Wang, "Fine-grained visual computing based on deep learning," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 17, no. 1, pp. 1–19, Apr. 2021.
- [91] Z. Lv, L. Qiao, J. Li, and H. Song, "Deep-learning-enabled security issues in the Internet of Things," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 9531–9538, Jun. 2021.
- [92] Z. Lv, A. K. Singh, and J. Li, "Deep learning for security problems in 5G heterogeneous networks," *IEEE Netw.*, vol. 35, no. 2, pp. 67–73, Mar. 2021.
- [93] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.
- [94] M. Kim, T. Hiroyasu, M. Miki, and S. Watanabe, "SPEA2+: Improving the performance of the strength Pareto evolutionary algorithm 2," in *Proc. Int. Conf. Parallel Problem Solving Nature*. Berlin, Germany: Springer, 2004, pp. 742–751.
- [95] J. Knowles and D. Corne, "Approximating the nondominated front using the Pareto archived evolution strategy," *Evol. Comput.*, vol. 8, no. 2, pp. 149–172, Jun. 2000.
- [96] S. Özdemir, B. A. Attea, and Ö. A. Khalil, "Multi-objective evolutionary algorithm based on decomposition for energy efficient coverage in wireless sensor networks," *Wireless Pers. Commun.*, vol. 71, no. 1, pp. 195–215, Jul. 2013.
- [97] S. Mostaghim and J. Teich, "Strategies for finding good local guides in multi-objective particle swarm optimization (MOPSO)," in *Proc. IEEE Swarm Intell. Symp. (SIS)*, Apr. 2003, pp. 26–33.
- [98] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.
- [99] M. A. Hall and L. A. Smith, "Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper," in *Proc. FLAIRS Conf.*, 1999, pp. 235–239.
- [100] T. M. Cover and J. A. Thomas, *Elements of information theory*. Hoboken, NJ, USA: Wiley, 2012.
- [101] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, nos. 1–2, pp. 23–69, Oct. 2003.



IBRAHIM ALJARAH received the bachelor's degree in computer science from Yarmouk University, Jordan, in 2003, the master's degree in computer science and information systems from the Jordan University of Science and Technology, Jordan, in 2006, and the Ph.D. degree in computer science from North Dakota State University, Fargo, ND, USA, in 2014. He is currently an Associate Professor of big data mining and computational intelligence with the Department of

Information Technology, The University of Jordan, Jordan. He contributed in many projects in USA, such as Vehicle Class Detection System (VCDS), Pavement Analysis Via Vehicle Electronic Telemetry (PAVVET), and Farm Cloud Storage System (CSS). He participated in many conferences in the field of data mining, machine learning, and Big data, such as CEC, GECCO, NITIT, CSIT, IEEE NABIC, CASON, and BIGDATA Congress. He has published over 80 publications in refereed international conferences and first and second quartile journals with over 3500 citations and an H-index of 32. His research interests include data mining, data science, machine learning, opinion mining, sentiment analysis, big data, MapReduce, Hadoop, swarm intelligence, evolutionary computation, and large-scale distributed algorithms. He ranked amongst the top 2% scientists in the world and the one of top ten scientists in Jordan in artificial intelligence and image processing field, as per Stanford University Global Ranking 2020.



HOSSAM FARIS received the B.A. degree in computer science from Yarmouk University, in 2004, the M.Sc. degree (Hons.) in computer science from Al-Balqa' Applied University, in 2008, and the Ph.D. degree in e-business from the University of Salento, Italy, in 2011. He was awarded a full-time, competition-based Ph.D. scholarship from the Italian Ministry of Education and Research to pursue his Ph.D. degree. In 2016, he worked as a Postdoctoral Researcher

with the GeNeura Team, Information and Communication Technologies Research Center (CITIC), University of Granada, Spain. He is currently a Full Professor with the Information Technology Department, King Abdullah II School for Information Technology, The University of Jordan, Jordan. His research interests include applied computational intelligence, evolutionary computation, knowledge systems, data mining, semantic web, and ontologies.



ALI ASGHAR HEIDARI received the B.Sc. and M.Sc. degrees from the University of Tehran. He is currently an Exceptionally Talented Researcher with the University of Tehran and a Researcher with the School of Computing, National University of Singapore (NUS). He is also an Elite Researcher of the Iran's National Elites Foundation (INEF). He has authored more than 100 research articles with over 5020 citations (i10-index of 69 and H-index of 37) in prestigious international journals, such as IEEE INTERNET OF THINGS, *Information Fusion*, *Information Sciences*, *Future Generation Computer Systems*, *Renewable and Sustainable Energy Reviews*, the *Journal of Cleaner Production*, *Energy*, *Energy Conversion and Management*, *Applied Soft Computing*, *Knowledge-Based Systems*, IEEE ACCESS, and *Expert Systems with Applications*. He has several highly cited and hot cited articles. His research interests include advanced machine learning, evolutionary computation, optimization, prediction, solar energy, information systems, and mathematical modeling. He was awarded and recognized by the Iran's National Elites Foundation (INEF). He has been recognized by (publons.com/a/1397783) as the top 1% peer reviewer in computer science and cross-field. He ranked among world's top 2% scientists list of Stanford University as an Outstanding Researcher with an impressive record of cooperation on many international research projects with different top researchers from the optimization and artificial intelligence community.



MAJDI M. MAFARJA received the B.Sc. degree in software engineering from Philadelphia University, Jordan, in 2005, the M.Sc. degree in computer information systems from The Arab Academy for Banking and Financial Sciences, Jordan, in 2007, and the Ph.D. degree in computer science from the National University of Malaysia (UKM). He was a member of the Data Mining and Optimization Research Group (DMO). He is currently an Associate Professor with the Department of Computer

Science, Birzeit University. His research interests include evolutionary computation, meta-heuristics, and data mining.



ALA' M. AL-ZOUBI received the B.Sc. degree in software engineering from Al-Zaytoonah University, in 2014, and the M.Sc. degree in web intelligence from The University of Jordan, Jordan, in 2017. He is currently pursuing the Ph.D. degree with the School of Science, Technology and Engineering, University of Granada. He worked as a Teacher and a Research Assistant on several projects funded by The University of Jordan. During his graduate studies, he has published several

publications in well-recognized journals and conferences. His research interests include evolutionary computation, machine learning, and security in social network analysis and other research fields. He is a member of two research groups, such as the JISDF Research Group, that focus on bridge the gap between the academic and industry mechanisms in security and the Evo-ML Research Group, where the group focuses on evolutionary algorithms, machine learning, and their applications for solving important problems in different areas.



PEDRO A. CASTILLO was born in Granada, in 1974. He received the B.Sc. degree in computer science and the Ph.D. degree from the University of Granada, Spain, in 1997 and 2000, respectively. He has worked as a Teaching Assistant with the Computer Science Department, University of Jaén, Spain, and a Visiting Researcher with Edinburgh Napier University, Edinburgh, U.K., in July 1998, and the Santa Fe Institute, Santa Fe, NM, USA, in September 2000. He is currently an Associate

Professor with the Department of Computer Architecture and Technology, University of Granada. He has led several research projects and directed five Ph.D. students. His main research interests include bio-inspired systems, hybrid systems, and the combination of evolutionary algorithms and neural networks.



JUAN J. MERELO received the B.Sc. degree in theoretical physics and the Ph.D. degree from the University of Granada, Granada, Spain, in 1988 and 1994, respectively. He has been a Visiting Researcher with the Santa Fe Institute, Santa Fe, NM, USA, Politecnico Torino, Turin, Italy, RISC-Linz, Linz, Austria, the University of Southern California, Los Angeles, CA, USA, and Université Paris-V, Paris, France. He is currently a Full Professor with the Department of Computer

Architecture and Technology, University of Granada. His research interests include evolutionary algorithms, artificial life, unconventional distributed computing, and games.