

# Many-Objective Jaccard-Based Evolutionary Feature Selection for High-Dimensional Imbalanced Data Classification

H. Saadatmand<sup>ID</sup> and Mohammad-R. Akbarzadeh-T<sup>ID</sup>, *Senior Member, IEEE*

**Abstract**—Filters and wrappers represent two mainstream approaches to feature selection (FS). Although evolutionary wrapper-based FS outperforms filters in addressing real-world classification problems, extending these methods to high-dimensional, many-objective optimization problems with imbalanced data poses substantial challenges. Overcoming computational costs and identifying suitable performance metrics are vital for navigating search operation complexities. Here, we propose using the Jaccard similarity (JS) in a set-based evolutionary many-objective (JSEMO) FS search, addressing both evolutionary FS and imbalanced classifier choice concurrently. This study highlights the mutual influence between these aspects, impacting overall algorithm performance. JSEMO integrates JS into population initialization, reproduction, and elitism steps, enhancing diversity and avoiding duplicate solutions. The set-based variation operator utilizes intersection and union operators for compatibility with binary coding. We also introduce a double-weighted KNN (KNN2W) classifier with four supportive objectives as a many-objective FS problem to handle imbalanced distributions. Compared with 20 methods on 15 benchmark problems, JSEMO produces distinct optimal features, significantly improving overall accuracy, balance accuracy, and g-mean metrics with comparable feature set size and computational cost. The ablation study underscores the positive impact of all JSEMO components, highlighting the set-based variation operation with JS and KNN2W with relevant evaluation metrics as the most influential aspects.

**Index Terms**—Feature selection, high dimensionality, imbalanced data, multiobjective optimization, wrapper method.

## I. INTRODUCTION

MANY upcoming real-world classification problems are concerned with fundamental challenges such as class imbalance (CI) and high dimensionality (HD). Microarray gene expression [1], pattern recognition [2], medical image segmentation [3], and text classification, to name a few, form classes of such real-world problems. The microarray technology's cancer applications are one notable application domain in the healthcare industry that involves such challenges. For instance, in cancer scenarios with over a thousand features, an imbalanced class

Manuscript received 26 November 2023; revised 2 May 2024; accepted 5 June 2024. Date of publication 18 June 2024; date of current version 5 November 2024. (Corresponding author: Mohammad-R. Akbarzadeh-T.)

The authors are with the Department of Electrical Engineering, Center of Excellence on Soft Computing, and Intelligent Information Processing, Ferdowsi University of Mashhad, Mashhad 9177948974, Iran (e-mail: h.saadatmand@mail.um.ac.ir; akbazar@um.ac.ir).

Code Ocean: <https://codeocean.com/capsule/3381070/tree>  
Digital Object Identifier 10.1109/TPAMI.2024.3416196

distribution is evident, where the majority class significantly outnumbers the more critical minority class. Thus, CI and HD form vital aspects of real-world problems that often occur concurrently. However, the standard learning methods often ignore this aspect by considering only the overall accuracy [4] and supposing equal significance for all samples.

Feature selection (FS) is a vital preprocessing step in machine learning and data mining that aims to discern the most informative features from a large pool of available attributes [5]. By selecting relevant features, FS can effectively reduce the computational complexity and enhance the performance of the subsequent learning and decision-making processes [6]. FS can be categorized into three main strategies: wrapper-based, filter-based [7], [8], [9], and embedded [10]. These strategies involve a trade-off between achieving high accuracy and minimizing computational costs [11]. Wrapper excels over filter-based and embedded methods, as they consider the interactions among features and feature subsets' impact on the learning algorithm.

It is crucial to recognize that these interactions among features can profoundly impact the algorithmic efficacy. Notably, not all attributes selected in the optimal subset may individually exhibit substantial utility; some may even appear marginally irrelevant [12]. However, their decisive role may manifest when they are combined [13]. Wrappers, in this sense, exhibit a more accurate strategy in selecting feature subsets [14] compared to filters, which often rely on statistical measures. While wrappers are influential, they come with computational costs. As the training process is often executed once, mitigates concerns regarding computational costs.

Generally, FS is an NP-hard, multimodal, and combinatorial problem [5] that typically involves two primary objectives: 1) maximizing the accuracy and 2) minimizing the number of features [15], [16]. Nonetheless, many metaheuristic-based research have considered single-objective optimization [17], [18], [19], [20], due to its simplicity and performance. Hence, these works have often used a weighted sum to balance these two objectives [21].

An alternative approach is to treat FS as a multiobjective optimization problem [12], [14], [16], [22], [23], where the two objectives are optimized simultaneously without any weighting. This approach can preserve the diversity of the solutions and generate a set of optimal and non-dominated feature subsets, known as the Pareto front [24]. The Pareto front can provide more flexibility and choice for the decision-makers, who can

select the most suitable feature subset according to their preferences and constraints [24].

When designing optimal classifiers, moving beyond accuracy as the sole performance metric is crucial. Careful consideration is necessary for complex scenarios, such as medical diagnosis tasks with imbalanced class distributions [4]. While accuracy remains a central optimization criterion, such instances necessitate assessment across various metrics, including precision, recall, sensitivity, and specificity. This approach recognizes the multifaceted nature of automatic classification in real-world settings, where multiple criteria could hold relevance besides accuracy. It is pertinent to realize that not all errors carry equal weight. For instance, correctly identifying all patients in cancer screening is more crucial than occasionally misclassifying healthy individuals. Thus, FS should be formulated as a many-objective optimization task wherein the number of objectives surpasses three [25]. Alongside feature size, other objectives such as precision, recall, F-measure [26], G-mean [27] [28], and balanced accuracy [29] could be considered to gauge the classifier's performance across diverse dimensions comprehensively.

Many-objective optimization is a challenging task requiring advanced strategies to handle solutions' increased complexity and diversity [25]. Several many-objective evolutionary algorithms (MaOEAs) have been introduced and applied to FS, such as NSGA-III [25], NSGAII-SDR [30] and RVEA [31]. These MaOEAs excel at independently optimizing more than three objectives, resulting in a diverse and representative Pareto front. These methods employ additional effective population-based measures in the optimization process, including Pareto dominance and crowding estimation [25]. These measures collectively contribute towards optimal and non-dominated solution sets [30]. By doing so, they offer decision-makers an opportunity to address various aspects of the FS problems concurrently. Despite their potential, their efficacy in handling FS for HD and CI problems requires further investigation.

Numerous approaches have addressed CI databases [32]. The predominant methods involve re-sampling and enhancing learning techniques. However, re-sampling poses the risk of discarding valuable instances, potentially causing overfitting with the generation of artificial data points [32]. A better learning strategy could consider costs according to class distributions [33] and ensemble learning [34], but such a strategy is often inefficient for HD problems.

Evolutionary FS can effectively address CI challenges by identifying robust features [35]. Recent advancements in FS algorithms, such as PSOFS-FC [26] and SYMON [27], prioritize this perspective. These methods only consider a single objective and utilize a prevalent classifier, such as a support vector machine (SVM) and k-nearest neighbor (KNN). However, such standard classifiers do not specifically consider CI distributions [4]. For better handling of critical decision-making scenarios, an evolutionary many-objective FS approach proves more effective when concurrently considering two factors in its evolutionary processes: 1) a set of appropriate evaluation metrics [32] and 2) a classifier tailored for imbalanced data [33].

Technically, as the number of decisions and objectives increases, the efficiency of evolutionary algorithms for FS reduces.

In combinatorial search, striking a balance between diversity and convergence poses a significant challenge, particularly in the HD binary-coded space. Here, we depart from conventional strategies like single-point, double-point, and uniform crossovers by introducing set-based binary recombination operation and incorporating Jaccard similarity (JS) in the search process. Additionally, the duplication dilemma in the conventional evolutionary search processes, i.e., when the newly created solutions are copies of the previously generated individuals, threatens population diversity and leads to local optima and premature convergence [16]. Recently, efficient strategies have been developed to remove duplicated solutions using Manhattan and hamming distances [16], [12]. Expanding on this concept, we control the similarities of newly created solutions across generations, opting to recreate individuals rather than remove them, thereby preserving a more diversified population.

In short, we propose Jaccard similarity in a Set-based Evolutionary Many-Objective (JSEMO) for high dimensional class imbalanced feature selection. The primary contributions of this paper are as follows:

- 1) The proposed many-objective FS approach incorporates JS to ensure diversity during the search process, which is crucial in HD spaces. Initially, JS aids in creating a diverse archive of solutions during population initialization. As generations progress, JS monitors population diversity. If the newly created solutions by variation operation are duplicated or too similar to the population, they are reproduced. JS also aids in elitism by preserving dissimilar and non-dominated individuals for future generations.
- 2) A binary representation of set-based variation operations is introduced, which is more compatible with the FS binary coding, especially in its HD search space. Notably, 'union' and 'intersection' design the recombination operation. The union operator joins the selected parents' features and removes their redundancies. In contrast, the intersection operator serves as an efficient dimensionality reduction operation. An adaptive mutation is subsequently adopted to emphasize exploration. These operations differ from our previous research [36], shifting from an integer-coded form to a more appropriate binary representation. Additionally, the procedures are refined by incorporating JS to enhance diversity in the context of many objective FS problems.
- 3) A double-weighted KNN (KNN2W) classifier is offered, in which the first weight is the impact of the CI distribution, and the second is the degree of neighborhoods' similarities. Indeed, concurrently taking CI distributions and neighborhood similarities into account draws a primary distinction between KNN2W and the prevalent KNN variants, such as NWKNN [37], FKNN [38], and CCWKNN [39]. Moreover, we suggest four supportive evaluation metrics based on a many-objective strategy to further support our method in fronting real-world CI data.

The performance of JSEMO is rigorously evaluated against 20 contemporary FS approaches on 15 real-world problems, encompassing both filter and wrapper strategies. This comparison illustrates that JSEMO produces a unique set of optimal

features, leading to substantial improvements in accuracy, balance accuracy, and g-mean metrics. These enhancements are achieved while maintaining a comparable feature set size and computational cost.

The remainder of this paper is organized as follows. Section II outlines the relevant literature on evolutionary FS approaches, mainly focusing on multi- and many-objective imbalanced-based strategies. A many-objective FS problem is defined in Section III. Section IV elaborates on the proposed JSEMO in detail. Sections V and VI detail the experimental settings and results obtained by comparing algorithms with 15 real-world benchmark datasets. Finally, we conclude this paper by suggesting future research directions in Section VII.

## II. RELATED WORK

This section reviews the literature on evolutionary multiobjective FS and high-dimensional imbalanced data classification problems.

### A Multiobjective Feature Selection

Several metaheuristic algorithms (MAs) have been devised in the past decade to address multiobjective FS problems. Noteworthy among these algorithms are particle swarm optimization (PSO), genetic algorithm (GA), differential evolution (DE), and ant colony optimization (ACO). The subsequent review highlights some of the most representative models. Particle swarm optimization (PSO) [11] has emerged as a popular choice for tackling multiobjective FS due to its efficiency and ease of implementation. For instance, Zhou et al. [40] introduced a flexible cut-point PSO that generates new solutions using a Pareto multiobjective evolutionary algorithm for FS. In a different context, Zhang et al. [41] explored using a multiobjective PSO to address cost-based FS problems, aiming to identify and eliminate costly, redundant, and irrelevant features.

In addition, Zhang et al. [42] employed a binary DE for multiobjective FS, utilizing self-learning and binary mutation for effective solutions and elite mechanism promotion. Wang et al. [22], [13] addressed optimal feature subsets in HD problems using multiobjective DE, tackling cases where multiple subsets with different features yield the same objective values. Wang et al. [24] introduced an ACO-based algorithm for multiobjective FS, managing challenges from feature interactions and highly discontinuous Pareto fronts.

GA-based operations are commonly employed in multiobjective FS due to their inherent ability to handle binary coding without extra interpretation. Cheng et al. [43] devised a variable granular multiobjective evolutionary search algorithm, where a single bit (gene) represents a group of features, reducing computational time. Zhou et al. [44] introduced a problem-specific non-dominant sorting genetic algorithm (PS-NSGA) for FS, emphasizing accuracy-preferred domination and modified mutation operators. Xu et al. [16] proposed a duplication analysis-based evolutionary algorithm (DAEA) for multiobjective FS, using Manhattan distance to detect and remove duplicated solutions for enhanced diversity. More recently, Jiao et al. [12] developed the PRDH algorithm for HD classification problems, introducing

a handling duplication strategy considering both decision and objective space. DAEA and PRDH highlight the significance of duplicated challenges in multiobjective FS. Despite efficiently addressing the issue, they only focus on removing identical solutions, often incurring additional computational costs.

While the research mentioned above has found widespread application in the context of FS with only two objectives, the potential of many-objective optimization remains underexplored, particularly for handling imbalanced challenges. To date, there are only a limited number of research studies on many-objective FS, and those that do exist are primarily focused on specific case studies, such as motor imagery detection [45] and multi-label classification [46], [47]. Hence, a complete study in this framework is highly desirable.

### B. Imbalanced Feature Selection

Class-imbalanced feature selection (CIFS), consisting of non-evolutionary [48], [49] and evolutionary strategies [50], has been increasingly studied in recent years. Here, we focus on the relevant evolutionary CIFS algorithms. Zhang et al. [26] developed an efficient PSO-based FS process and fuzzy clustering (PSOFS-FC). PSOFS-FC offers an improved F-measure for handling missing values and CI issues as a single objective. Moayedikia et al. [27] established a CIFS method using a harmony search algorithm, which uses g-mean metric and symmetric uncertainty to weigh attributes due to their dependency on class labels. Li et al. [28] investigated a multiobjective optimization method with g-mean and selected feature rate objectives. Sun et al. [35] introduced fuzzy neighborhood rough sets and the binary whale optimization algorithm for imbalanced FS problems. Wei et al. [51] presented a multiobjective FS (FS-DOS) with a dynamic selection operation that utilizes balance accuracy as an evaluation metric. Also, an efficient survey in [4] investigates evolutionary computing in imbalanced data applications.

The above researches primarily rely on one of the other evaluation metrics, such as g-mean, f-measure, or balanced accuracy, for handling CI data. However, none of these metrics alone comprehensively address all aspects of real-world problems. Moreover, existing strategies often employ standard classifiers, which may not be well-suited to handle CI challenges effectively. These algorithms also face prohibitive computational costs with increasing classes and features. This is while many uprising real-world problems entail high dimensionality, multiple desirable performance measures and class imbalance.

We propose JSEMO in response to these challenges. JSEMO is a three-pronged approach to many-objective feature selection. JSEMO leverages Jaccard similarity, binary representations inspired by set operations, and a double-weighted KNN. Firstly, we view FS as a many-objective problem, incorporating multiple evaluation metrics to fulfill all requirements. Secondly, we introduce KNN2W, a tailored approach for CI databases. The choice of classifier for imbalanced data has mutual effects that influence each other and impact the overall algorithm's performance. Thirdly, we employ set-based binary operations coupled with Jaccard similarity to tackle the challenges posed by HD data.

TABLE I  
THE STANDARD EVALUATION METRICS FOR FS

$f_1: Precision = TP / (TP + FP)$ ,
$f_2: Recall \text{ (or Sensitivity)} = TP / (TP + FN)$
$f_3: Specificity = TN / (TN + FP)$
$f_4: Accuracy = (TP + TN) / (TP + FP + FN + TN)$
$f_5: F_\beta\text{-measure} = ((1 + \beta^2) * f_1 * f_2) / (\beta^2 * f_1 + f_2), \beta > 0$
$f_6: Weighted average accuracy = \alpha * f_2 + (1 - \alpha) * f_3, \alpha \in [0, 1]$
$f_7: Balance accuracy = 0.5 * f_2 + 0.5 * f_3$
$f_8: G\text{-mean}_1 = \sqrt{f_2 * f_3}$
$f_9: G\text{-mean}_2 = \sqrt{f_1 * f_2}$
$\text{Feature size ratio} = 1 -  z /D$

<sup>1</sup> TP = true positive, TN = true negative, FP = false positive, and FN = false negative.

### III. PROBLEM DEFINITION

A many-objective high-dimensional optimization problem for feature selection can be defined as follows:

$$\text{Maximize } \mathcal{F}(z) = \{f_1(z), f_2(z), \dots, f_M(z)\} \quad (1)$$

where  $f_j$  is the  $j^{\text{th}}$  objective,  $j = 1, \dots, M$ , and  $z = (z_1, z_2, \dots, z_d, \dots, z_D) \in \Omega$ ,  $z_d \in \{0, 1\}$ , represents a candidate solution (feature subset) of the search space  $\Omega$ , and  $D > 1000$  is the number of features (dimension). Generally, if  $M$  is two or three, it is known as a multiobjective problem, and if  $M$  exceeds three, it is known as a many-objective problem [25]. FS strategies aim to discern  $z^* \subseteq \Omega$ , in which  $z^* = \{z_1^*, z_2^*, \dots, z_V^*\}$  is the smallest and best-performing feature subset, where  $V$  is the optimal (unknown) number of features.

Table I lists the prevalent and standard classification metrics and the feature size ratio widely utilized as FS objectives. The true positive (TP) and true negative (TN) represent cases where the model accurately predicts the positive and negative classes. Conversely, the false positive (FP) and false negative (FN) denote instances where the model incorrectly predicts the positive and negative classes. The overall *accuracy* (or misclassification rate) is extensively used for FS [24], [12], and is suitable only for balanced data.  $F_\beta\text{-measure}$  is a combination of *precision* ( $f_1$ ) and *recall* ( $f_2$ ), often recommended for imbalanced data [26], [52], where  $\beta$  is the vital and challenging parameter to balance the significance of precision and recall. *Weighted average accuracy* is defined as the sum of the weighted values of  $f_2$  and  $f_3$  for a specific  $\alpha$ , like *balance accuracy* with  $\alpha = 0.5$  [29]. Geometric mean (g-mean) [53] includes  $g\text{-mean}_1$  and  $g\text{-mean}_2$ , which calculates the degree of inductive bias in terms of  $f_2$  and  $f_3$ , and  $f_1$  and  $f_2$ , respectively.

How does one choose a proper evaluation metric? We should recognize that no single metric may fit all scenarios. The choice depends heavily on the unique requirements and characteristics of the problem [54]. A review of the common class-imbalanced metrics, however, reveals that they are often derived from three measures of precision ( $f_1$ ), recall ( $f_2$ ), and specificity ( $f_3$ ). These three metrics capture the essential aspects of model performance and are particularly relevant in scenarios with imbalanced class distributions, especially those encountered in medicine. By simultaneously optimizing these independent objectives alongside the number of features, the proposed approach

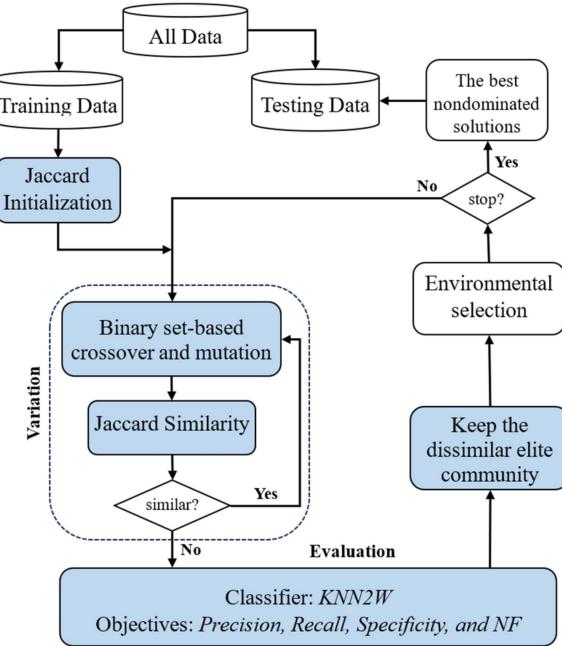


Fig. 1. The general components of proposed JSEMO, in which our contributions are shown by the blue blocks.

aims to address the diverse performance requirements of FS for imbalanced datasets. It should be mentioned that the approach is general and can be trivially extended to include other metrics like  $F_\beta\text{-measure}$ , *balanced accuracy*, and *g-mean*, as listed in Table I, providing additional problem-specific insights.

### IV. THE PROPOSED APPROACH JSEMO

The JSEMO framework, which integrates several essential components, is shown in Fig. 1. It begins with a database that is randomly partitioned into training and testing portions. Population initialization creates an initial set of solutions. The core operations—variation, evaluation, and environmental selection (ES) [12]—shape each generation. ES, commonly used in many-objective approaches [22], selects the best individuals for the next generation. Strategies like Pareto-dominance and crowding distance estimation enhance ES. Additionally, an elitist strategy maintains an archive of dissimilar, nondominated individuals, improving overall functionality.

JSEMO is further detailed in Algorithm 1. It comprises five principal steps. A diverse population is initially established by considering JS, thereby regulating the degree of similarity within the population. The subsequent steps involve set-based variation operations and JS mechanisms to generate novel and non-repetitive offspring. Following this, the offspring undergo evaluation through the proposed KNN2W method, yielding four objective functions that strategically address imbalanced data. In the fourth step, an archive is maintained, housing dissimilar and non-dominated individuals in a community of elites. This archival function augments diversity and prevents the loss of helpful feature subsets with many members. Finally, the population for the next generation is obtained by ES based on NSGAII-SDR [30]. This strategy incorporates a strengthened dominance

**Algorithm 1:** The Proposed JSEMO.

---

Inputs population size (N), maximum iteration (T), similarity parameter  $\vartheta_0$ , and Training data (X, Y). dimension (D)

Output return the first Pareto front of  $P$

Step 1  $P \leftarrow \text{JIM}(N, \vartheta_0)$ ; % Jaccard initialization method  
Alg. 3  
 $P \leftarrow \text{Evaluation}(P, X, Y);$   
 $Q \leftarrow P$  % An archive of dominant and dissimilar solutions while termination criteria are not met, do

Step 2  $O \leftarrow \text{SetbasedVariationJS}(P, Q)$ ; Alg. 4

Step 3  $O \leftarrow \text{Evaluation}(O, X, Y);$

Step 4  $Q \leftarrow Q \cup O;$

Step 5  $P \leftarrow \text{EnvironmentalSelection}(Q, N);$   
If  $|Q| > 2N$  then remove the worst individuals  
end while

---

**Algorithm 2:** Jaccard Similarity (JS).

1. Inputs: Archive/population ( $Q$ ) and current solution ( $z$ )
  2. Outputs: the most similar individual with its similarity level
  3. for  $i = 1 : |Q|$
  4.  $S(i) \leftarrow \frac{|Q(i) \cap z|}{|Q(i) \cup z|}$ ; Similarity of  $i^{th}$  individual in  $Q$  and  $z$
  5. end
  6. Find the most similar individual to  $z$  with its similarity level
- 

relation and crowding distance, meticulously fostering a refined balance between diversity and convergence. Detailed explanations of the four proposed primary contributions of JSEMO are detailed in the following subsections.

*A. Jaccard Initialization Method (JIM)*

Evolutionary algorithms typically initiate their search from a randomly generated population within the search space. The FS context also follows this common strategy. However, relying solely on random initialization proves ineffective, particularly in HD decision spaces, as prior research indicates [16]. In such scenarios, a randomly generated population may exhibit close similarities or repetitions in certain regions while other expansive areas remain unexplored. To address this limitation, we suggest incorporating a Jaccard Similarity (JS) coefficient [55] in the initialization step and evolutionary processes. The JS coefficient is a measure of similarity between two sets, calculated as the size of their intersection divided by the size of their union. A higher JS indicates more similarity between sets. JS is used in data mining and information retrieval to measure the similarity between two sets of items, documents, or features. This metric is particularly suitable for measuring the similarity between two binary feature sets (solutions) utilized in evolutionary algorithms. In Algorithm 2, the JS measure quantifies the similarity between a solution  $z$  and the  $i^{th}$  solution in set  $Q$ ,

**Algorithm 3:** Jaccard Initialization Method (JIM).

---

1. Inputs: population size (N), degree of similarity ( $\vartheta$ )
2. Outputs: population ( $P$ )
3.  $i \leftarrow 1$ ,  $P \leftarrow$  create a binary random solution
4. while  $i \leq N$  do
5.      $z \leftarrow$  create a binary random solution with limitation selected features in [16].
6.      $s \leftarrow JS(P, z)$  % Jaccard Similarity
7.     if  $s < \vartheta$  then
8.          $P \leftarrow P \cup z$
9.     end
10. end while

---

is expressed mathematically as:

$$S(i) = \frac{|Q(i) \cap z|}{|Q(i) \cup z|}. \quad (2)$$

JIM is also delineated in Algorithm 3, in which the first member of the population  $P$  is randomly created. It is imperative to emphasize that the maximum number of selected features limits all randomly generated genes within the context of JIM. Subsequently, additional solutions are introduced into  $P$  in a random manner if their similarity level remains below a pre-determined threshold  $\vartheta$ . This iterative process persists until  $N$  dissimilar members are successfully produced.

*B. Set-Based Variation Operation With JS*

The pivotal mechanism for generating a new population constitutes the central processing unit within MAs and represents an indispensable component of their functionality. The imperative lies in adopting methodologies that are more aligned with the intricacies of the problem at hand and its respective representation, thereby ensuring high-quality solutions. In this context, we introduce a variation operation predicated on set-based binary coding, wherein the crossover design is formed using intersection and union operators. The intersection operator assumes a pivotal role, serving as a dimensionality reduction operation by retaining only the mutually shared genes between parents, thereby significantly reducing the feature size. On the other hand, the union operator preserves all relevant attributes from both parents while concurrently eliminating any unnecessary attributes.

Notably, the intersection operator adheres to a contractionary (greedy) policy, strategically facilitating the attainment of a concise feature subset. In contrast, the *union* follows an expansionary policy strategically aimed at harnessing all pertinent information to create novel combinations in successive generations. This approach also contributes to the evasion of local optima. Subsequently, the two offspring generated through intersection and union undergo a mutation process, thereby introducing additional diversity into the population. Moreover, the similarity level of the produced solutions is checked using JS to avoid closely similar or duplicate solutions, thereby enhancing diversity. To strike the delicate balance between convergence and diversity, the similarity threshold  $\vartheta_t$  gradually adjusted during

**Algorithm 4:** Set-Based Variation With JS.

---

Inputs Population  $P$ , archive  $Q$ , similarity boundaries ( $\vartheta_{LB}, \vartheta_{UB}$ )  
Output Offspring  $O$

Step 1     $\vartheta_t \leftarrow \vartheta_{LB} + (t/T) * (\vartheta_{UB} - \vartheta_{LB})$   
 $n_m \leftarrow n_{LB} + (1 - t/T) * (n_{UB} - n_{LB})$   
while  $|O| = N$   
     $\{p_1, p_2\} \leftarrow$  randomly select two individuals from  $P$   
Step 2     $O_1 \leftarrow p_1 \cap p_2$  % the first child of intersection operation  
 $O_1 \leftarrow \text{mutate}(O_1, n_m)$  % randomly alter  $n_m$  genes in  $O_1$   
 $s_1 \leftarrow JS(Q, O_1)$  % Jaccard Similarity  
if  $s_1 \leq \vartheta_t$  then  
     $O \leftarrow O \cup O_1$   
else  
     $p_1 \leftarrow$  randomly select a new individual from  $P - p_1$   
    go to Step 2  
end  
Step 3  
Step 4     $O_2 \leftarrow p_1 \cup p_2$  % the second child of union operation  
 $O_2 \leftarrow \text{mutate}(O_2, n_m)$  % randomly alter  $n_m$  genes in  $O_2$   
 $s_2 \leftarrow JS(Q, O_2)$  % Jaccard Similarity  
if  $s_2 \leq \vartheta_t$  then  
     $O \leftarrow O \cup O_2$   
else  
     $p_2 \leftarrow$  randomly select a new individual from  $P - p_2$   
    go to Step 4  
end  
end while

---

the search process. Initially a lower bound ( $\vartheta_{LB} < 1$ ) provides an opportunity to explore new regions of the decision space while still excluding very similar solutions in the early stages. As the search progresses, the threshold is gradually increased towards unity ( $\vartheta_{UB}=1$ ), to avoid only duplicate solutions and promote convergence towards optimal solutions in the last generations. Simultaneously, the number of mutated genes  $n_m$  decreases, ensuring an effective equilibrium between exploration and exploitation strategies.

Algorithm 4 shows the production process of distinct offspring populations in five simple steps.  $\vartheta_t$  and  $n_m$  are updated in Step 1. After selecting two parents from the mating selection population  $P'$ , the first child is produced in Step 2, in which the intersection and mutation operations are applied to two parents, respectively. Then, the similarity level  $s_1$  of the created individual with the archive  $Q$  is determined. If  $s_1 < \vartheta_t$ , the offspring is then added to the new population  $O$ . Otherwise, a new parent is chosen from  $P'$ , and the process returns to Step 2. Similarly, all processes are the same for the second child  $O_2$  in Steps 4 and 5, the difference is in using the union operation instead of the intersection.

**C. KNN2W and Many-Objective FS Metrics**

KNN [56] is renowned for its simplicity, low computational cost, and effectiveness [11], [24], [29], making it a preferred choice compared to other classifiers such as SVM and random forest [57], especially in the context of wrapper-based evolutionary-based FS algorithms. However, standard KNN faces challenges when encountering imbalanced data, and fails to consider the similarity of neighbors, which can significantly

**Algorithm 5:** The Proposed KNN2W for Imbalanced Data.

---

Inputs: Training data (X, Y), Test query ( $x$ ), and  $K$   
Output: Label of  $x$

1.  $w_c = 1/H$ ; %  $H$  is the histogram of Y per class
2. [Negs, dis]  $\leftarrow$  find  $K$  nearest neighbors to  $x$  and their distance
3.  $w_s \leftarrow 1/(1 + dis)$ ; % the Negs's weight
4.  $w \leftarrow w_c(\text{Negs}) + w_s$  % the overall weight
5. for  $c = 1 : NC$
6.     $vote(c) \leftarrow \sum_{\forall I \in Y(\text{Negs} = c)} w(I)$
7. end for
8. Label  $\leftarrow \text{argmax}(vote)$

---

impact its performance [38]. Notably, when the number of neighbors ( $k$ ) is even, standard KNN struggles to predict the correct label due to the lack of a majority class among the neighbors. However, leveraging neighbor similarity can be particularly beneficial in these scenarios, providing additional context for decision-making.

To this end, Algorithm 5 introduces a double-weighted k-nearest neighbor (KNN2W). Two weights are employed: one for balancing and another to account for the neighborhoods' similarity. To mitigate the impact of class-imbalanced data, we assign a weight of  $w_c = 1/H$  for each sample  $x$ , where  $H$  is the number of samples with  $x$ 's class in the problem. This weight decreases as the number of instances in a class increases, effectively addressing imbalances. Thus, we assign  $w_s$  to  $x$ 's neighbors based on their distance using the Euclidean metric, with larger weights assigned to more similar neighbors. The final weight ( $w$ ) is determined as the sum of  $w_c$  and  $w_s$ . Subsequently, there are  $K$  weights for the  $K$  neighbors involved in a competition to determine the final label. The competition's winner is the class with the highest weight concerning its own class-based neighbor(s). This process is repeated to predict all labels in the validation set within a k-fold cross-validation framework. The KNN2W's effectiveness is compared with four competing variants of KNN in Table XV, highlighting its superiority in handling imbalanced data and leveraging neighbor similarity for improved classification performance.

For clarity, Fig. 2 provides a simple example demonstrating KNN2W's effectiveness over KNN with imbalanced data. Fig. 2.A shows the data distribution, while 2.B illustrates neighborhoods of a test point with varying  $k$  values (1 to 5) and associated overall weights ( $w$ ). Table II then compares the performance of KNN and KNN2W with five  $k$  values. KNN2W consistently predicts all scenarios accurately, regardless of even or odd  $k$  values. In contrast, KNN only achieves correct predictions when  $k = 1$ . For instances where KNN fails, “~” denotes that the final label should be chosen randomly.

For evaluating the KNN2W classifier, the many-objective optimization problem in (1) is redefined to minimize the number of features. Hence, the four suggested objectives for imbalanced

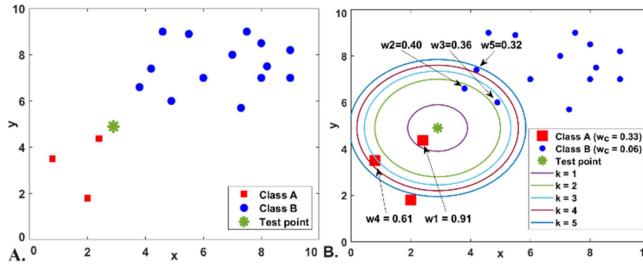


Fig. 2. The application of KNN and KNN2W on an imbalanced classification dataset is illustrated. The data distribution is visualized in (Left), while the neighborhoods of the test point with varying values of  $k = 1, \dots, 5$  are shown in (Right). Each neighbor is depicted with an overall weight  $w$ .

FS are as follows:

$$\text{Maximize } \mathcal{F}(z) = \{f_1(z), f_2(z), f_3(z), 1 - f_{NF}(z)\} \quad (3)$$

where  $f_1$ ,  $f_2$ , and  $f_3$ , are *precision*, *recall*, and *specificity*, respectively. The fourth objective  $f_{NF} = |z|/D$  is the number of features (NF), and  $z$  is as defined in (1). Indeed, the KNN2W method assumes a pivotal role in classifying problems in the many-objective evolutionary FS process using the generated feature subsets. Its application is instrumental in achieving the three key objectives of precision, recall, and specificity. The fourth (last) objective is the number of selected feature sets.

#### D. Dissimilar Elite Community

Maintaining an elite community serves to mitigate the risk of loss during the evolutionary process and bolsters the subsequent phase of environmental selection. An illustrative scenario involves the potential loss of numerous informative feature subsets [22] due to an excessively large objective pertaining to feature size. Nevertheless, upon evaluating the offspring, these subsets are incorporated into the archive  $Q$ , comprising predominant and dissimilar individuals, forming an elite society. This approach diverges from conventional elitist strategies, which typically focus solely on performance. The discernible disparity lies in the creation of an archive of more heterogeneous and dissimilar solutions, marking a fundamental distinction from traditional elitist methods. The environmental selection mechanism then identifies the  $N$  best individuals from the archive for propagation into the succeeding generation. In the event that the archive surpasses twice the total population, additional weaker individuals are systematically excluded, guided by the average of the initial three objectives, independent of the feature size ratio. This ensures that even relevant and helpful subsets, which may be omitted due to their large size, are considered in the selection process.

## V. THE EXPERIMENTAL DESIGNS

### A. Datasets

This research selects 15 real-world medical classification problems to include balanced and imbalanced datasets to cover a wide range of the imbalance rate  $I_r$  (from 1.02 to 23.7).

TABLE II  
KNN VS. KNN2W ON IMBALANCED DATA

Neighbors	KNN		KNN2W	
	Class A	Class B	Class A ( $w$ )	Class B ( $w$ )
$k = 1$	✓		✓ (0.91)	(-)
$k = 2$	~	~	✓ (0.91)	(0.40)
$k = 3$		✓	✓ (0.91)	(0.76)
$k = 4$	~	~	✓ (1.52)	(0.76)
$k = 5$		✓	✓ (1.52)	(1.09)
Correct N.	1/5		5/5	

TABLE III  
DATABASES IN THIS STUDY

No.	Name	$I_r$	$D$	$N_s$	$N_c$
1	Prostate-GE	1.02	5966	102	2
2	MLL	1.40	12583	72	3
3	Colon Tumor	1.82	2000	62	2
4	CNS	1.85	7129	60	2
5	Leukemia_2	1.88	7129	72	2
6	Breast Cancer	1.89	22283	52	2
7	GLI	2.27	22283	85	2
8	Covid-19	2.44	15979	234	3
9	DLBCL	3.05	5469	77	2
10	Leukemia_3	4.22	7129	72	3
11	11 Tumors	4.50	12533	174	11
12	Lung Harvard	4.84	12533	181	2
13	Lymphoma	5.11	4026	66	3
14	Brain Tumor	15.0	5920	90	5
15	Lung Cancer	23.7	12600	203	5

The imbalance rate  $I_r$  is determined by dividing the number of observations of the majority class by the number of samples of the minority class. They are also chosen to consider a wide range of feature numbers  $D$  (from 2000 to 22283), the number of classes  $N_c$ , and the number of samples  $N_s$ . These datasets are sorted according to increasing  $I_r$  in Table III. The datasets are available on the UCI repository [58] and a standard source of microarray gene expression datasets [59]. These data sets have been frequently used in previous studies on high dimensional and imbalanced FS, such as in [20], [21], [26], [27], [36], [44], [60]. The datasets are randomly partitioned into training (70%) and testing (30%). The training data is processed by five-fold cross-validation to avoid overfitting [12].

### B. Competing Algorithms

Twenty benchmark algorithms are selected to serve as peer FS algorithms, encompassing a wide range of single-, multi-, many-objective, and filter strategies designed for addressing challenges associated with HD and CI problems. Notably, DAEA [16], NSGAII-SDR [30], RVEA [31], PS-NSGA [44], and FS-DOS [51] algorithms are implemented based on PlatEMO [61] and the provided code by the corresponding authors. FS-DOS, PS-NSGA, and DAEA are multiobjective state-of-the-art (SOTA) FS methods. FS-D OS and PS-NSGA predominantly employ balanced accuracy as their performance metric. DAEA is also an efficient HD FS that handles duplicate challenges like the

TABLE IV  
IGD METRIC WITH AVERAGE AND STANDARD DEVIATION (MEAN  $\pm$  STD) FOR THE SIX COMPETING ALGORITHMS

No.	NSGAII-SDR [30]	FS-DOS [51]	PS-NSGA [44]	DAEA [16]	RVEA [31]	JSEMO
1	6.144e-2 $\pm$ 1.860e-2 ↓	6.093e-2 $\pm$ 2.298e-2 ↓	3.927e-2 $\pm$ 1.839e-2 ↓	3.204e-2 $\pm$ 1.386e-2 ↓	1.016e-1 $\pm$ 1.350e-2 ↓	<b>2.353e-2</b> $\pm$ 1.831e-2
2	4.291e-2 $\pm$ 3.289e-2 ↓	4.563e-2 $\pm$ 3.467e-2 ↓	<b>3.681e-3</b> $\pm$ 6.876e-3 ≈	4.705e-3 $\pm$ 1.296e-4≈	7.208e-2 $\pm$ 2.323e-2 ↓	4.087e-3 $\pm$ 8.807e-3
3	2.861e-1 $\pm$ 4.328e-2 ↓	1.779e-1 $\pm$ 5.168e-2 ↓	<b>5.180e-2</b> $\pm$ 3.621e-2 ↑	1.069e-1 $\pm$ 4.166e-2 ↓	2.949e-1 $\pm$ 6.067e-2 ↓	5.758e-2 $\pm$ 2.802e-2
4	2.492e-1 $\pm$ 2.227e-2 ↓	3.764e-1 $\pm$ 9.653e-2 ↓	2.312e-1 $\pm$ 2.988e-2 ↓	2.346e-1 $\pm$ 3.660e-2 ↓	3.106e-1 $\pm$ 6.798e-2 ↓	<b>2.065e-1</b> $\pm$ 5.731e-3
5	2.879e-2 $\pm$ 3.289e-2 ↓	5.418e-2 $\pm$ 4.372e-2 ↓	3.408e-3 $\pm$ 1.018e-2 ↓	<b>4.544e-4</b> $\pm$ 1.092e-4≈	8.224e-2 $\pm$ 2.309e-2 ↓	7.387e-4 $\pm$ 6.610e-4
6	1.232e-1 $\pm$ 5.276e-2 ↓	9.966e-2 $\pm$ 6.067e-2 ↓	6.601e-2 $\pm$ 5.707e-2 ↓	<b>1.777e-4</b> $\pm$ 3.772e-5 ↑	1.558e-1 $\pm$ 2.219e-2 ↓	2.543e-4 $\pm$ 1.108e-4
7	4.245e-2 $\pm$ 2.560e-2 ↓	1.113e-1 $\pm$ 3.705e-2 ↓	4.889e-3 $\pm$ 1.425e-2 ↓	2.274e-3 $\pm$ 9.943e-3 ↓	9.190e-2 $\pm$ 1.576e-2 ↓	<b>5.265e-4</b> $\pm$ 2.181e-4
8	4.231e-1 $\pm$ 2.728e-2 ↓	5.520e-1 $\pm$ 5.359e-2 ↓	4.352e-1 $\pm$ 4.949e-2 ↓	2.914e-1 $\pm$ 2.393e-2 ↓	4.109e-1 $\pm$ 1.828e-2 ↓	<b>2.882e-1</b> $\pm$ 2.213e-2
9	5.455e-2 $\pm$ 3.579e-2 ↓	3.398e-2 $\pm$ 4.303e-2 ↓	7.021e-4 $\pm$ 3.407e-4 ≈	<b>5.266e-4</b> $\pm$ 8.039e-5≈	1.197e-1 $\pm$ 3.108e-2 ↓	6.948e-4 $\pm$ 1.576e-4
10	1.673e-2 $\pm$ 8.593e-3 ↓	4.932e-2 $\pm$ 3.119e-2 ↓	1.442e-3 $\pm$ 1.210e-3 ≈	<b>9.482e-4</b> $\pm$ 4.328e-4≈	8.722e-2 $\pm$ 1.952e-2 ↓	1.851e-3 $\pm$ 9.457e-4
11	1.881e-1 $\pm$ 3.510e-2 ↓	6.572e-1 $\pm$ 1.691e-1 ↓	3.233e-1 $\pm$ 1.161e-1 ↓	2.712e-1 $\pm$ 1.262e-1 ↓	5.831e-1 $\pm$ 2.096e-1 ↓	<b>1.335e-1</b> $\pm$ 1.293e-2
12	1.728e-2 $\pm$ 1.330e-2 ↓	7.661e-3 $\pm$ 1.515e-2 ↓	3.606e-4 $\pm$ 1.956e-4 ≈	3.914e-4 $\pm$ 4.606e-5≈	4.350e-2 $\pm$ 3.458e-2 ↓	<b>3.138e-4</b> $\pm$ 1.330e-4
13	1.595e-2 $\pm$ 9.201e-3 ↓	4.281e-3 $\pm$ 1.039e-2 ↓	6.756e-4 $\pm$ 1.684e-4 ≈	5.067e-4 $\pm$ 4.967e-5≈	3.155e-2 $\pm$ 1.392e-2 ↓	<b>4.213e-4</b> $\pm$ 1.257e-2
14	6.232e-1 $\pm$ 1.085e-1 ↓	9.858e-1 $\pm$ 8.647e-2 ↓	4.426e-1 $\pm$ 1.753e-1 ↓	2.944e-1 $\pm$ 2.936e-1 ↓	7.612e-1 $\pm$ 1.440e-1 ↓	<b>7.944e-2</b> $\pm$ 7.656e-2
15	3.007e-1 $\pm$ 1.094e-1 ↓	1.384e-1 $\pm$ 9.311e-2 ↓	5.553e-2 $\pm$ 2.525e-2 ↓	2.449e-1 $\pm$ 3.230e-3 ↓	2.953e-1 $\pm$ 1.614e-2 ↓	<b>4.052e-2</b> $\pm$ 1.614e-2
↑/≈/↓	0/0/15	0/0/15	1/5/9	1/6/8	0/0/15	

proposed JSEMO. Furthermore, RVEA is a SOAT evolutionary many-objective optimization algorithm based on a reference vector-guided evolutionary strategy. NSGAII-SDR represents an effective many-objective technique that employs a strengthened dominance relation method and crowding distance. This technique has found widespread application in the FS field, as evident in [16] and [24].

In addition to the algorithms above, 15 contemporary FS methods have been considered. These include six SOTA filter and embedded methods, Inf-FS-S [8], Inf-FS-U [8], MRRM [9], ReliefF [62], Correlation-based FS (CFS) [63], Random Forest (RF) [57], and two imbalanced approaches, PSOFS-FC [26] and SYMON [27], along with four single-objective methods: PPSO [11], HFS-C-P [20], SIFE [36], and VLPSO-LS [60]. Additionally, three multiobjective algorithms, namely SM-MOEA [21], PRDH [12], and INSA [24], have been specifically designed to address classification problems.

### C. Performance Metrics

Seven performance metrics, namely hypervolume (HV) [64], inverted generational distance (IGD) [65] that measure both the convergence and diversity of nondominated feature subsets, number of features (NF), computational time, balanced accuracy (BAC) [26], geometric mean (GM) [53], and overall accuracy (AC), are employed for empirical comparisons among competing algorithms. HV and IGD, recognized as key indicators in many-objective optimization, gauge the convergence and diversity of the non-dominated solution set. Smaller IGD and larger HV values denote superior performance, respectively. As the optimal Pareto front is unknown [16], [12], the ideal point is considered in terms of classification metrics, with at least one feature defining the optimal feature set. The utilized reference point,  $[1, 1, 1, 1-1/D]$ , encapsulates the four objectives: precision, recall, specificity, and feature size. Conversely, for calculating IGD, we use the  $[0, 0, 0, 1/D]$ . BAC and GM are the two recommended imbalanced metrics to validate the ability of competing algorithms against imbalanced problems.

### D. Parameter Settings

The population size is 100. The number of function evaluations is the stopping criterion for all competing algorithms, which is set depending on  $N$  and  $T$ . The maximum iteration  $T$  is set from 100 to 300 depending on the number of features, i.e., more features mean a higher dimensional search space and more time to search. For example, for the first and seventh test problems with 2000 and 22284 features, T is set to 100 and 300, respectively. The above parameters try to make a fair condition, especially in the number of function evaluations for all competing methods. Crossover and mutation probabilities in JSEMO and other GA-based algorithms are set at one and  $1/D$ , respectively. The mutated genes' number adaptively decreases per iteration from  $n_{UB} = 0.5$  to  $n_{LB} = 0.1$  percent of  $D$ . The similarity level threshold  $\vartheta_t$  linearly increases between  $\vartheta_{LB} = 0.98$  to  $\vartheta_{UB} = 1$  at each iteration. In the early generations, besides disregarding directly copied solutions, individuals that exhibit a high degree of similarity (greater than 98%) to previously generated solutions are also reproduced to uphold population diversity. As generations progress, the similarity threshold is gradually raised to 1. At this point, only entirely duplicated individuals are reproduced, facilitating convergence in the population. This adaptive approach to reproduction aims to balance exploration and exploitation throughout the evolutionary process.

The customizable parameters for main compared algorithms are selected based on the findings presented in their original papers, which have shown their effectiveness:

- DAEA [16]: the local mating rate is set to 0.8, the global mating rate is set to 0.2, and the modified mutation ratio is set to 0.8.
- PS-NSGA [44]: divisions along each objective axis are set to 27, the number of mutations and crossover match the population size with the mutation probability of 0.1.
- FS-DOS [51]: the minimum elimination probability is fixed at 0.15, while three control parameters, namely F,  $\alpha$ , and  $\beta$ , are set to 0.5, 0.3, and 0.2, respectively.
- RVEA [31]: this algorithm incorporates two specific parameters: the index  $\alpha$ , which is configured to 2 to regulate the rate of change of the penalty function, and the frequency

TABLE V  
HV METRIC WITH AVERAGE AND STANDARD DEVIATION (MEAN  $\pm$  STD) FOR THE SIX COMPETING ALGORITHMS

No.	NSGAII-SDR [30]	FS-DOS [51]	PS-NSGA [44]	DAEA [16]	RVEA [31]	JSEMO
1	9.200e-1 $\pm$ 2.729e-2 ↓	9.289e-1 $\pm$ 2.336e-2 ↓	9.522e-1 $\pm$ 2.144e-2 ↓	9.597e-1 $\pm$ 1.786e-2 ↓	8.514e-1 $\pm$ 1.772e-2 ↓	<b>9.707e-1</b> $\pm$ 2.435e-2
2	9.407e-1 $\pm$ 5.159e-2 ↓	9.332e-1 $\pm$ 5.311e-2 ↓	9.967e-1 $\pm$ 8.200e-3 ≈	<b>9.998e-1</b> $\pm$ 1.387e-5 ↑	8.842e-1 $\pm$ 3.759e-2 ↓	9.967e-1 $\pm$ 1.138e-2
3	6.725e-1 $\pm$ 4.468e-2 ↓	7.769e-1 $\pm$ 6.035e-2 ↓	9.578e-1 $\pm$ 4.519e-2 ↓	9.000e-1 $\pm$ 6.098e-2 ↓	6.337e-1 $\pm$ 6.455e-2 ↓	<b>9.876e-1</b> $\pm$ 1.749e-2
4	8.271e-1 $\pm$ 2.825e-2 ↓	5.853e-1 $\pm$ 1.121e-1 ↓	8.733e-1 $\pm$ 4.151e-2 ↓	8.476e-1 $\pm$ 5.106e-2 ↓	6.483e-1 $\pm$ 1.158e-1 ↓	<b>8.830e-1</b> $\pm$ 1.903e-2
5	9.707e-1 $\pm$ 4.665e-2 ↓	9.369e-1 $\pm$ 5.796e-2 ↓	9.971e-1 $\pm$ 1.114e-2 ≈	<b>9.998e-1</b> $\pm$ 1.604e-5 ≈	8.967e-1 $\pm$ 4.169e-2 ↓	9.997e-1 $\pm$ 5.337e-5
6	8.483e-1 $\pm$ 6.924e-2 ↓	9.049e-1 $\pm$ 8.448e-2 ↓	9.391e-1 $\pm$ 6.524e-2 ↓	9.999e-1 $\pm$ 4.898e-6 ≈	7.814e-1 $\pm$ 3.586e-2 ↓	<b>9.999e-1</b> $\pm$ 2.038e-6
7	9.691e-1 $\pm$ 3.393e-2 ↓	8.913e-1 $\pm$ 4.566e-2 ↓	9.977e-1 $\pm$ 7.887e-3 ≈	9.981e-1 $\pm$ 9.088e-3 ↓	8.589e-1 $\pm$ 2.953e-2 ↓	<b>9.998e-1</b> $\pm$ 2.517e-5
8	4.826e-1 $\pm$ 2.826e-2 ↓	3.726e-1 $\pm$ 4.349e-2 ↓	4.783e-1 $\pm$ 4.523e-2 ↓	<b>6.355e-1</b> $\pm$ 2.419e-2 ↓	4.710e-1 $\pm$ 1.947e-2 ↓	6.334e-1 $\pm$ 2.406e-2
9	9.336e-1 $\pm$ 4.916e-2 ↓	9.621e-1 $\pm$ 4.992e-2 ↓	9.997e-1 $\pm$ 4.623e-5 ≈	<b>9.997e-1</b> $\pm$ 1.216e-5 ≈	8.343e-1 $\pm$ 4.231e-2 ↓	9.996e-1 $\pm$ 5.088e-5
10	9.897e-1 $\pm$ 7.208e-3 ↓	9.307e-1 $\pm$ 4.642e-2 ↓	9.994e-1 $\pm$ 3.053e-4 ≈	<b>9.997e-1</b> $\pm$ 4.156e-5 ≈	9.213e-1 $\pm$ 1.791e-2 ↓	9.995e-1 $\pm$ 1.829e-4
11	7.640e-1 $\pm$ 3.567e-2 ↓	3.464e-1 $\pm$ 1.279e-1 ↓	6.450e-1 $\pm$ 1.103e-1 ↓	7.121e-1 $\pm$ 1.220e-1 ↓	4.144e-1 $\pm$ 1.800e-1 ↓	<b>8.297e-1</b> $\pm$ 1.534e-2
12	9.857e-1 $\pm$ 1.535e-2 ↓	9.956e-1 $\pm$ 1.142e-2 ↓	9.997e-1 $\pm$ 1.431e-4 ≈	9.994e-1 $\pm$ 4.873e-5 ≈	5.988e-1 $\pm$ 4.256e-2 ↓	<b>9.998e-1</b> $\pm$ 1.922e-5
13	9.893e-1 $\pm$ 7.751e-3 ↓	9.967e-1 $\pm$ 9.949e-3 ↓	9.991e-1 $\pm$ 1.030e-4 ≈	9.987e-1 $\pm$ 2.034e-5 ↓	9.724e-1 $\pm$ 1.273e-2 ↓	<b>9.993e-1</b> $\pm$ 1.125e-2
14	3.696e-1 $\pm$ 8.600e-2 ↓	1.128e-1 $\pm$ 4.974e-2 ↓	5.357e-1 $\pm$ 1.253e-1 ↓	7.357e-1 $\pm$ 2.413e-1 ↓	2.572e-1 $\pm$ 9.878e-2 ↓	<b>9.062e-1</b> $\pm$ 7.457e-2
15	6.957e-1 $\pm$ 9.582e-2 ↓	8.275e-1 $\pm$ 9.186e-2 ↓	9.058e-1 $\pm$ 3.065e-2 ↓	7.159e-1 $\pm$ 8.076e-3 ↓	6.797e-1 $\pm$ 1.345e-1 ↓	<b>9.529e-1</b> $\pm$ 1.695e-2
↑/≈/↓	0/0/15	0/0/15	0/7/8	1/6/8	0/0/15	

$f_r$ , set to 0.1 to facilitate the utilization of reference vector adaptation.

- **SFC** [63]: SFC is a filter-based subset selection algorithm, which is too time-consuming, particularly for datasets with more than 1000 attributes. Thus, FN is initially reduced to 1000 using reliefF, and SFC is then utilized.

## VI. RESULTS AND DISSECTIONS

The results obtained from the study assess peer algorithms based on two primary aspects: their ability in optimizing the objective function by IGD, HV, FN, and CPU runtime metrics, and their efficacy on unseen data, gauged by AC, BAC, and GM metrics. Additionally, an ablation study is conducted to elucidate the impact of key components in JSEMO. Lastly, the proposed KNN2W undergoes comparison with four standard KNN-based methods, with presented results representing the averages of 25 independent runs along with standard deviations. The superior achievements among the algorithms are denoted in bold. Symbols “↑,” “↓,” and “≈” signify that the corresponding benchmark technique is significantly better than, worse than, or has no significant difference from JSEMO, respectively. The Wilcoxon sign-ranked test, with a significance level of 0.05, is also employed.

### A. Comparisons Regarding the Optimization View

Tables IV and V validate the six competing algorithms from the optimization view in terms of IGD, HV, computational time, and NF, respectively.

Table IV provides the mean and standard deviation of IGD. JSEMO excels among the six algorithms in IGD, particularly in addressing HD imbalanced data. Notably, JSEMO significantly outperforms its peers in 8 out of 15 cases, with no significant differences in four additional cases. In only Colon Tumor and Breast cancer scenarios does JSEMO concede to PS-NSGA and DAEA, respectively. The second- and third-ranking algorithms are DAEA and PS-NSGA, the former being proficient with balanced data and achieving results comparable to JSEMO's. RVEA and NSGAII-SDR exhibit equivalent performance, securing the lowest positions in this competition.

As indicated in Table V, HV performance aligns well with the observed pattern in IGD. JSEMO attains the best HV compared with its peers. Generally, JSEMO outperforms the other five competing algorithms in 8 out of 15 problems. Moreover, in six cases, JSEMO closely reaches HV achieved by DAEA and FS-NSGA, the subsequent best algorithms. JSEMO loses to DAEA on MLL data.

Table VI analyses the computational time of each peer method using CPU runtime in seconds. MATLAB codes are executed in a system with 16 GB RAM, and its CPU is AMD Ryzen 9 5900HS. From this Table, JSEMO exhibits reasonable computational runtime, ranking as the fastest in three instances: Colon Tumor, Brain Tumor, and Lung Cancer. Overall, RVEA is the fastest algorithm and significantly performs better than JSEMO regarding runtime. In five cases, DAEA and PS-NSGA take more time than JSEMO. FS-DOS is the loser in the computational time competition.

Table VII gives the average of NF, in which JSEMO achieves a smaller size of NF in comparison to the other competitors. DAEA is closely similar to JSEMO in most cases, while it works better than JSEMO in Lymphoma and Covid-19 databases. PS-NSGA is better than JSEMO only in one case (Lymphoma). FS-DOS, RVEA, and NSGAII-SDR cannot defeat JSEMO in terms of finding the smallest NF.

In summary, from an optimization perspective, SEMO excels in IGD and HV metrics, while DAEA and JSEMO perform similarly in identifying optimal feature sizes. RVEA stands out in computational efficiency. JSEMO's strengths lie in its many-objective approach, which comprehensively addresses imbalanced data, and its integration of JS divergence to maintain population diversity to avert premature convergence and avoid local optima. More details are presented in Section VI-D (Ablation study). Additionally, the second-best algorithm, DAEA, emphasizes the role of duplicated solutions during the search process.

### B. Comparisons Regarding the Classification Metrics

Tables VIII-X compare the effectiveness of six algorithms—NSGAII-SDR, FS-DOS, PS-NSGA, DAEA, RVEA, and

**TABLE VI**  
COMPUTATIONAL TIME METRIC WITH AVERAGE AND STANDARD DEVIATION (MEAN  $\pm$  STD) FOR THE SIX COMPETING ALGORITHMS

No.	NSGAI-II-SDR [30]	FS-DOS [51]	PS-NSGA [44]	DAEA [16]	RVEA [31]	JSEMO
1	7.556e+1 $\pm$ 1.529e+0↓	8.601e+1 $\pm$ 2.047e+0↓	<b>6.600e+1</b> $\pm$ 1.312e+0↑	8.122e+1 $\pm$ 1.285e+0↓	7.326e+1 $\pm$ 3.151e+0≈	7.239e+1 $\pm$ 1.510e+0
2	4.009e+1 $\pm$ 1.447e+0↑	6.390e+1 $\pm$ 1.120e+0↓	4.343e+1 $\pm$ 8.881e-1≈	4.215e+1 $\pm$ 3.227e+0↑	<b>3.341e+1</b> $\pm$ 1.211e+0↑	4.762e+1 $\pm$ 5.446e-01
3	3.026e+1 $\pm$ 7.431e-1↓	3.071e+1 $\pm$ 2.398e-1↓	2.822e+1 $\pm$ 3.819e-1≈	3.011e+1 $\pm$ 9.245e-1↓	2.718e+1 $\pm$ 5.367e-1≈	<b>2.646e+1</b> $\pm$ 1.660e+0
4	9.542e+1 $\pm$ 1.618e+0↑	1.316e+2 $\pm$ 1.040e+0↓	<b>9.052e+1</b> $\pm$ 2.714e+0↑	1.034e+2 $\pm$ 2.630e+0↑	9.269e+1 $\pm$ 1.455e+0↑	1.124e+2 $\pm$ 4.871e+0
5	3.532e+1 $\pm$ 1.007e+0≈	4.540e+1 $\pm$ 4.172e-1↓	3.265e+1 $\pm$ 2.579e-1↑	3.804e+1 $\pm$ 1.127e+0≈	<b>3.001e+1</b> $\pm$ 9.285e-1↑	3.971e+1 $\pm$ 1.018e+0
6	7.472e+1 $\pm$ 1.854e+0↑	1.659e+2 $\pm$ 9.248e+0↓	7.091e+1 $\pm$ 1.832e+0↑	8.688e+1 $\pm$ 2.404e+0↓	<b>6.673e+1</b> $\pm$ 4.034e+0↑	1.040e+2 $\pm$ 4.548e+0
7	<b>8.824e+1</b> $\pm$ 2.633e+0↑	1.687e+2 $\pm$ 1.404e+0↓	8.355e+1 $\pm$ 2.305e+0↑	1.018e+2 $\pm$ 4.291e-1↑	8.843e+1 $\pm$ 7.586e+0↑	1.181e+2 $\pm$ 4.539e+1
8	2.178e+2 $\pm$ 2.140e+1↑	3.926e+2 $\pm$ 8.625e+1↓	<b>1.857e+2</b> $\pm$ 8.471e+0↑	2.851e+2 $\pm$ 1.723e+0↓	2.632e+2 $\pm$ 4.861e+1↓	2.289e+2 $\pm$ 4.539e+0
9	6.644e+1 $\pm$ 1.248e+0≈	7.970e+1 $\pm$ 5.280e-1↓	<b>6.242e+1</b> $\pm$ 2.387e+0≈	7.110e+1 $\pm$ 1.100e+0≈	6.253e+1 $\pm$ 1.582e+0≈	6.624e+1 $\pm$ 1.234e+0
10	2.659e+1 $\pm$ 1.069e+0↑	4.716e+1 $\pm$ 5.986e-1↓	3.635e+1 $\pm$ 6.614e-1↓	3.012e+1 $\pm$ 5.584e-1↓	<b>2.299e+1</b> $\pm$ 7.717e-1↑	2.857e+1 $\pm$ 4.417e-01
11	1.702e+2 $\pm$ 6.009e+0↓	3.531e+2 $\pm$ 8.342e+1↓	2.297e+2 $\pm$ 2.051e+1↓	2.124e+2 $\pm$ 8.909e+0↓	<b>1.184e+2</b> $\pm$ 2.770e+0↑	1.262e+2 $\pm$ 3.770e+0
12	6.307e+1 $\pm$ 4.059e+0≈	6.260e+1 $\pm$ 1.226e+0≈	7.0289e+1 $\pm$ 3.898e+0↓	<b>5.624e+1</b> $\pm$ 2.826e+0↑	5.810e+1 $\pm$ 1.001e+1↑	6.453e+1 $\pm$ 1.850e+0
13	2.421e+1 $\pm$ 4.436e-1↑	3.777e+1 $\pm$ 3.223e-1↓	3.438e+1 $\pm$ 5.317e+0↓	2.681e+1 $\pm$ 1.722e+0↑	<b>2.234e+1</b> $\pm$ 3.887e-1↑	3.043e+1 $\pm$ 8.331e-01
14	1.262e+2 $\pm$ 6.805e+0↓	2.167e+0 $\pm$ 8.456e+0↓	1.836e+2 $\pm$ 6.610e+0↓	2.083e+2 $\pm$ 6.352e+0↓	1.143e+2 $\pm$ 4.115e+0↓	<b>1.036e+2</b> $\pm$ 9.073e-01
15	1.481e+2 $\pm$ 7.437e+0↓	2.838e+2 $\pm$ 1.649e+2↓	1.669e+2 $\pm$ 9.034e+0↓	1.434e+2 $\pm$ 1.266e+0≈	1.414e+2 $\pm$ 8.582e+0≈	<b>1.407e+2</b> $\pm$ 7.635e+0
↑/≈/↓	7/3/5	0/1/14	5/3/7	5/3/7	9/4/2	

**TABLE VII**  
NF METRIC WITH AVERAGE AND STANDARD DEVIATION (MEAN  $\pm$  STD) FOR THE SIX COMPETING ALGORITHMS

No.	NSGAI-II-SDR [30]	FS-DOS [51]	PS-NSGA [44]	DAEA [16]	RVEA [31]	JSEMO
1	48.22 $\pm$ 36.20 ↓	<b>5.775</b> $\pm$ 3.455 ↓	8.518 $\pm$ 1.871 ↓	3.441 $\pm$ 0.296 ≈	433.8 $\pm$ 132.0 ↓	<b>3.103</b> $\pm$ 1.133
2	153.7 $\pm$ 101.0 ↓	96.73 $\pm$ 46.70 ↓	26.74 $\pm$ 9.550 ↓	<b>4.095</b> $\pm$ 0.508 ≈	631.5 $\pm$ 183.0 ↓	7.503 $\pm$ 4.329
3	38.65 $\pm$ 13.72 ↓	32.84 $\pm$ 12.45 ↓	8.320 $\pm$ 2.261 ↓	<b>3.623</b> $\pm$ 0.488 ↓	174.0 $\pm$ 42.84 ↓	5.560 $\pm$ 2.148
4	77.75 $\pm$ 46.80 ↓	70.81 $\pm$ 21.02 ↓	18.88 $\pm$ 5.385 ↓	<b>3.167</b> $\pm$ 0.622 ≈	565.8 $\pm$ 88.32 ↓	6.718 $\pm$ 2.044
5	91.94 $\pm$ 64.55 ↓	62.70 $\pm$ 27.66 ↓	19.50 $\pm$ 7.952 ↓	3.341 $\pm$ 0.187 ≈	504.7 $\pm$ 146.2 ↓	<b>2.767</b> $\pm$ 1.422
6	195.2 $\pm$ 163.3 ↓	7.451 $\pm$ 5.506 ↓	5.069 $\pm$ 1.337 ≈	3.004 $\pm$ 0.247 ≈	36.90 $\pm$ 43.50 ↓	<b>2.790</b> $\pm$ 0.797
7	200.2 $\pm$ 111.5 ↓	101.7 $\pm$ 48.48 ↓	29.13 $\pm$ 11.51 ↓	<b>3.804</b> $\pm$ 0.404 ≈	504.1 $\pm$ 414.4 ↓	5.789 $\pm$ 3.632
8	207.9 $\pm$ 147.6 ↓	74.83 $\pm$ 135.8 ↓	7.899 $\pm$ 3.363 ↓	<b>6.424</b> $\pm$ 2.342 ↑	78.40 $\pm$ 22.87 ↓	13.43 $\pm$ 6.787
9	68.51 $\pm$ 42.42 ↓	19.92 $\pm$ 7.511 ↓	6.212 $\pm$ 1.609 ↓	3.027 $\pm$ 0.218 ≈	318.1 $\pm$ 101.7 ↓	<b>2.158</b> $\pm$ 0.434
10	92.57 $\pm$ 58.71 ↓	96.43 $\pm$ 38.01 ↓	14.49 $\pm$ 6.814 ↓	<b>3.921</b> $\pm$ 0.526 ≈	646.9 $\pm$ 145.2 ↓	5.661 $\pm$ 2.747
11	157.5 $\pm$ 60.70 ↓	162.9 $\pm$ 38.75 ↓	54.55 $\pm$ 9.695 ↓	<b>11.85</b> $\pm$ 2.750 ≈	269.7 $\pm$ 135.4 ↓	13.03 $\pm$ 6.096
12	155.5 $\pm$ 103.7 ↓	4.124 $\pm$ 2.014 ≈	6.974 $\pm$ 3.718 ↓	3.152 $\pm$ 0.172 ≈	540.3 $\pm$ 302.0 ↓	<b>2.251</b> $\pm$ 0.798
13	54.80 $\pm$ 35.19 ↓	25.67 $\pm$ 13.62 ↓	4.128 $\pm$ 1.554 ↑	<b>2.363</b> $\pm$ 0.285 ↑	239.8 $\pm$ 78.51 ↓	9.224 $\pm$ 4.867
14	66.94 $\pm$ 22.19 ↓	42.48 $\pm$ 9.650 ↓	13.70 $\pm$ 3.298 ↓	<b>5.274</b> $\pm$ 0.704 ↑	248.9 $\pm$ 40.66 ↓	8.167 $\pm$ 3.740
15	130.3 $\pm$ 79.85 ↓	23.97 $\pm$ 5.087 ↓	15.80 $\pm$ 5.149 ↓	<b>5.673</b> $\pm$ 0.825 ≈	459.1 $\pm$ 139.5 ↓	9.836 $\pm$ 2.975
↑/≈/↓	0/0/15	0/1/14	1/1/13	3/11/1	0/0/15	

**TABLE VIII**  
AC METRIC WITH AVERAGE AND STANDARD DEVIATION (MEAN  $\pm$  STD) FOR THE SIX COMPETING ALGORITHMS

No.	NSGAI-II-SDR [30]	FS-DOS [51]	PS-NSGA [44]	DAEA [16]	RVEA [31]	JSEMO
1	82.68 $\pm$ 3.579 ↓	86.83 $\pm$ 1.703 ↓	89.06 $\pm$ 2.137 ≈	89.13 $\pm$ 1.288 ↓	82.58 $\pm$ 2.936 ↓	<b>90.32</b> $\pm$ 1.140
2	95.50 $\pm$ 2.352 ↑	96.66 $\pm$ 2.794 ↑	97.22 $\pm$ 1.854 ↑	<b>97.50</b> $\pm$ 1.636 ↑	92.50 $\pm$ 2.112 ↓	94.00 $\pm$ 0.912
3	85.00 $\pm$ 2.300 ↓	92.59 $\pm$ 3.555 ↓	91.11 $\pm$ 2.342 ↓	90.18 $\pm$ 2.316 ↓	86.67 $\pm$ 2.614 ↓	<b>95.00</b> $\pm$ 1.242
4	<b>81.37</b> $\pm$ 5.053 ≈	68.15 $\pm$ 5.579 ↓	73.33 $\pm$ 5.965 ↓	70.98 $\pm$ 3.491 ↓	81.37 $\pm$ 5.952 ≈	79.60 $\pm$ 2.973
5	97.19 $\pm$ 2.325 ↓	95.36 $\pm$ 2.572 ↓	98.24 $\pm$ 1.875 ↓	<b>100.0</b> $\pm$ 0.000 ≈	<b>100.0</b> $\pm$ 0.000 ≈	<b>100.0</b> $\pm$ 0.000
6	74.76 $\pm$ 3.688 ↓	80.28 $\pm$ 4.682 ↑	85.07 $\pm$ 4.359 ↑	<b>87.38</b> $\pm$ 2.285 ↑	75.23 $\pm$ 4.953 ↓	78.48 $\pm$ 1.476
7	92.46 $\pm$ 2.153 ↑	92.11 $\pm$ 3.228 ↑	<b>95.36</b> $\pm$ 2.638 ↑	91.59 $\pm$ 2.153 ↑	88.55 $\pm$ 2.390 ↑	86.42 $\pm$ 3.912
8	67.62 $\pm$ 3.050 ↓	64.26 $\pm$ 3.538 ↓	70.30 $\pm$ 3.351 ↓	80.10 $\pm$ 2.263 ↓	60.35 $\pm$ 2.017 ↓	<b>88.57</b> $\pm$ 2.654
9	91.11 $\pm$ 3.651 ≈	91.42 $\pm$ 2.788 ≈	95.13 $\pm$ 3.086 ↑	<b>97.14</b> $\pm$ 2.414 ↑	92.22 $\pm$ 2.104 ≈	91.90 $\pm$ 3.610
10	73.16 $\pm$ 9.423 ↓	91.40 $\pm$ 3.601 ↓	<b>96.00</b> $\pm$ 2.011 ≈	71.16 $\pm$ 4.711 ↓	75.33 $\pm$ 4.711 ↓	95.00 $\pm$ 2.500
11	81.34 $\pm$ 3.631 ↓	76.12 $\pm$ 5.588 ↓	83.26 $\pm$ 2.727 ↓	75.27 $\pm$ 4.573 ↓	79.23 $\pm$ 3.715 ↓	<b>84.35</b> $\pm$ 5.234
12	98.40 $\pm$ 0.736 ↓	99.25 $\pm$ 0.753 ≈	98.75 $\pm$ 0.868 ≈	97.00 $\pm$ 0.707 ↓	98.33 $\pm$ 0.723 ↓	<b>99.86</b> $\pm$ 0.351
13	97.83 $\pm$ 1.858 ↓	<b>100.0</b> $\pm$ 0.000 ≈	98.55 $\pm$ 2.076 ↓	83.00 $\pm$ 5.606 ↓	<b>100.0</b> $\pm$ 0.000 ≈	<b>100.0</b> $\pm$ 0.000
14	73.58 $\pm$ 9.062 ↓	70.36 $\pm$ 5.091 ↓	79.49 $\pm$ 9.388 ↓	75.67 $\pm$ 8.914 ↓	70.36 $\pm$ 5.091 ↓	<b>85.63</b> $\pm$ 5.639
15	90.11 $\pm$ 1.780 ↓	91.40 $\pm$ 2.481 ↓	93.02 $\pm$ 3.293 ↓	78.39 $\pm$ 4.681 ↓	91.95 $\pm$ 1.159 ↓	<b>96.72</b> $\pm$ 1.542
↑/≈/↓	2/2/11	3/3/9	4/3/8	4/1/10	1/4/10	

JSEMO—against unseen testing data using prevalent classification metrics AC, BAC, and GM. Notably, metrics are computed based on the optimal Pareto front of the last population generated by peer algorithms. For practical purposes, these solutions are considered as potential candidates for decision-making. However, in our reported results, we calculate the average performance metrics across the last non-dominated solutions.

In Table **VIII**, JSEMO consistently outperforms other techniques, especially in Colon Tumor, Covid-19, 11 Tumors, Brain Tumor, and Lung Cancer cases. It matches other best algorithms on Prostate-GE, CNS, Leukemia\_2, Leukemia\_3, Lung Harvard, and Lymphoma, but falters on only 4 of 15 primarily balanced databases.

Tables **IX** and **X** similarly evaluate JSEMO and its peers by BAC and GM. Notably, JSEMO outperforms its peers on

TABLE IX  
BAC METRIC WITH AVERAGE AND STANDARD DEVIATION (MEAN  $\pm$  STD) FOR THE SIX COMPETING ALGORITHMS

No.	NSGAII-SDR [30]	FS-DOS [51]	PS-NSGA [44]	DAEA [16]	RVEA [31]	JSEMO
1	82.77 $\pm$ 3.526 ↓	86.76 $\pm$ 1.653 ↓	89.50 $\pm$ 2.122 ≈	89.04 $\pm$ 1.293 ≈	82.68 $\pm$ 2.958 ↓	<b>90.27 <math>\pm</math> 1.081</b>
2	96.73 $\pm$ 1.678 ↑	97.42 $\pm$ 2.221 ↑	97.88 $\pm$ 1.415 ↓	<b>98.12 <math>\pm</math> 1.174 ↑</b>	94.64 $\pm$ 1.509 ↓	95.46 $\pm$ 0.499
3	80.97 $\pm$ 2.843 ↓	90.71 $\pm$ 4.381 ↓	89.26 $\pm$ 3.061 ↓	87.98 $\pm$ 2.606 ↓	82.85 $\pm$ 3.361 ↓	<b>94.35 <math>\pm</math> 1.583</b>
4	78.66 $\pm$ 7.855 ↓	74.03 $\pm$ 4.368 ↓	75.43 $\pm$ 4.277 ↓	74.41 $\pm$ 3.555 ↓	<b>83.20 <math>\pm</math> 4.432 ↑</b>	81.21 $\pm$ 1.928
5	97.38 $\pm$ 2.149 ↓	94.74 $\pm$ 2.673 ↓	98.41 $\pm$ 1.705 ↓	<b>100.0 <math>\pm</math> 0.000 ≈</b>	<b>100.0 <math>\pm</math> 0.000 ≈</b>	<b>100.0 <math>\pm</math> 0.000</b>
6	71.92 $\pm$ 4.941 ↓	80.22 $\pm$ 4.551 ↓	85.43 $\pm$ 4.288 ↑	<b>87.07 <math>\pm</math> 2.146 ↑</b>	73.03 $\pm$ 4.986 ↓	81.79 $\pm$ 1.649
7	93.93 $\pm$ 1.717 ↑	91.81 $\pm$ 3.551 ↑	<b>94.17 <math>\pm</math> 3.010 ↑</b>	91.22 $\pm$ 1.586 ↑	91.22 $\pm$ 1.833 ↑	86.77 $\pm$ 3.296
8	72.33 $\pm$ 2.369 ↓	71.51 $\pm$ 2.421 ↓	75.93 $\pm$ 2.297 ↓	81.60 $\pm$ 1.926 ↓	66.55 $\pm$ 1.830 ↓	<b>85.85 <math>\pm</math> 2.935</b>
9	85.77 $\pm$ 6.550 ↓	86.53 $\pm$ 4.611 ↓	94.81 $\pm$ 3.140 ↑	<b>97.83 <math>\pm</math> 1.776 ↑</b>	86.55 $\pm$ 3.791 ↓	92.83 $\pm$ 2.981
10	79.71 $\pm$ 10.00 ↓	94.53 $\pm$ 2.186 ↓	96.80 $\pm$ 1.177 ≈	70.32 $\pm$ 2.595 ↓	85.42 $\pm$ 2.721 ↓	<b>97.08 <math>\pm</math> 1.437</b>
11	82.32 $\pm$ 2.425 ↓	77.37 $\pm$ 3.267 ↓	83.18 $\pm$ 1.898 ↑	74.20 $\pm$ 2.456 ↓	81.50 $\pm$ 3.068 ↓	<b>84.38 <math>\pm</math> 3.026</b>
12	95.55 $\pm$ 2.046 ↓	98.21 $\pm$ 1.866 ↓	96.63 $\pm$ 2.349 ↓	98.63 $\pm$ 0.977 ≈	95.37 $\pm$ 2.010 ↓	<b>99.70 <math>\pm</math> 0.267</b>
13	96.50 $\pm$ 3.002 ↓	<b>100.0 <math>\pm</math> 0.000 ≈</b>	98.67 $\pm$ 2.247 ↓	75.34 $\pm$ 8.059 ↓	<b>100.0 <math>\pm</math> 0.000 ≈</b>	<b>100.0 <math>\pm</math> 0.000</b>
14	83.73 $\pm$ 3.674 ↓	82.48 $\pm$ 1.637 ↓	83.54 $\pm$ 3.723 ↓	83.00 $\pm$ 3.284 ↓	82.48 $\pm$ 1.637 ↓	<b>88.25 <math>\pm</math> 4.626</b>
15	84.48 $\pm$ 1.252 ↓	91.91 $\pm$ 3.993 ↓	94.05 $\pm$ 3.873 ↓	69.77 $\pm$ 4.447 ↓	85.65 $\pm$ 3.041 ↓	<b>96.34 <math>\pm</math> 1.135</b>
↑/≈/↓	2/0/13	2/1/12	3/2/10	4/3/8	2/2/11	

TABLE X  
GM METRIC WITH AVERAGE AND STANDARD DEVIATION (MEAN  $\pm$  STD) FOR THE SIX COMPETING ALGORITHMS

No.	NSGAII-SDR [30]	FS-DOS [51]	PS-NSGA [44]	DAEA [16]	RVEA [31]	JSEMO
1	82.64 $\pm$ 3.620 ↓	86.67 $\pm$ 1.602 ↓	89.37 $\pm$ 2.131 ↓	88.95 $\pm$ 1.319 ↓	82.56 $\pm$ 2.958 ↓	<b>90.02 <math>\pm</math> 0.959</b>
2	96.60 $\pm$ 1.770 ↓	97.31 $\pm$ 2.347 ↑	97.81 $\pm$ 1.465 ↑	<b>98.06 <math>\pm</math> 1.212 ↑</b>	94.44 $\pm$ 1.577 ↓	95.28 $\pm$ 0.492
3	78.74 $\pm$ 3.521 ↓	90.16 $\pm$ 4.783 ↓	88.60 $\pm$ 3.449 ↓	87.23 $\pm$ 2.960 ↓	80.91 $\pm$ 4.141 ↓	<b>94.13 <math>\pm</math> 1.635</b>
4	74.49 $\pm$ 13.06 ↓	70.57 $\pm$ 5.697 ↓	73.51 $\pm$ 4.838 ↓	72.74 $\pm$ 3.470 ↓	<b>82.37 <math>\pm</math> 4.608 ↑</b>	79.95 $\pm$ 1.812
5	97.30 $\pm$ 2.212 ↓	94.56 $\pm$ 2.774 ↓	98.37 $\pm$ 1.754 ↓	<b>100.0 <math>\pm</math> 0.000 ≈</b>	<b>100.0 <math>\pm</math> 0.000 ≈</b>	<b>100.0 <math>\pm</math> 0.000</b>
6	70.34 $\pm$ 6.335 ↓	79.57 $\pm$ 4.867 ≈	85.02 $\pm$ 4.120 ↑	<b>86.72 <math>\pm</math> 2.077 ↑</b>	72.38 $\pm$ 4.969 ↓	80.98 $\pm$ 2.020
7	93.74 $\pm$ 1.808 ↑	91.72 $\pm$ 3.569 ↑	<b>94.00 <math>\pm</math> 3.073 ↑</b>	91.02 $\pm$ 1.647 ↑	90.77 $\pm$ 1.998 ↑	86.45 $\pm$ 3.254
8	69.40 $\pm$ 2.616 ↓	69.86 $\pm$ 2.466 ↓	74.75 $\pm$ 2.506 ↓	79.93 $\pm$ 2.417 ↓	61.45 $\pm$ 2.530 ↓	<b>82.02 <math>\pm</math> 3.061</b>
9	84.33 $\pm$ 7.746 ↓	85.51 $\pm$ 5.162 ↓	94.61 $\pm$ 3.292 ↑	<b>97.78 <math>\pm</math> 1.823 ↑</b>	85.40 $\pm$ 4.409 ↓	92.60 $\pm$ 3.155
10	72.85 $\pm$ 16.79 ↓	94.27 $\pm$ 2.315 ↓	96.78 $\pm$ 1.205 ≈	52.66 $\pm$ 2.886 ↓	83.15 $\pm$ 3.725 ↓	<b>96.96 <math>\pm</math> 1.545</b>
11	80.20 $\pm$ 3.123 ↓	76.15 $\pm$ 4.411 ↓	81.93 $\pm$ 2.319 ↓	74.55 $\pm$ 3.217 ↓	79.16 $\pm$ 3.550 ↓	<b>84.15 <math>\pm</math> 2.681</b>
12	95.42 $\pm$ 2.106 ↓	98.15 $\pm$ 1.937 ↓	96.54 $\pm$ 2.417 ↓	94.60 $\pm$ 1.014 ↓	95.23 $\pm$ 2.069 ↓	<b>99.61 <math>\pm</math> 1.006</b>
13	96.14 $\pm$ 3.385 ↓	<b>100.0 <math>\pm</math> 0.000 ≈</b>	98.58 $\pm$ 2.428 ↓	59.16 $\pm$ 13.36 ↓	98.66 $\pm$ 2.374 ↓	<b>100.0 <math>\pm</math> 0.000</b>
14	80.01 $\pm$ 4.883 ↓	78.34 $\pm$ 2.298 ↓	80.33 $\pm$ 4.955 ↓	79.58 $\pm$ 4.056 ↓	78.34 $\pm$ 2.298 ↓	<b>87.53 <math>\pm</math> 5.413</b>
15	74.20 $\pm$ 1.412 ↓	89.04 $\pm$ 7.535 ↓	92.00 $\pm$ 6.265 ↓	51.54 $\pm$ 8.454 ↓	76.30 $\pm$ 5.741 ↓	<b>96.18 <math>\pm</math> 1.745</b>
↑/≈/↓	1/0/14	2/2/11	4/1/10	4/1/10	2/1/12	

datasets with high imbalanced rates. For instance, in the case of Lung Cancer data, which exhibits the highest imbalanced rate, JSEMO attains BAC and GM scores of 96.34% and 96.18%, respectively. In comparison, the second-best algorithm on this dataset, PS-NSGA, achieves 94.05% and 92.00% for BAC and GM, respectively. According to the Wilcoxon sign-ranked test, the overall second-best algorithms are DAEA and PS-NSGA. However, their performance often deteriorates with the increasing data imbalance. RVEA wins the third place, and RVEA is better than JSEMO on CNS problem. FS-DOS is slightly better than NSGAII-SDR. At higher data imbalance, JSEMO performs better than its peers regarding BAC and GM. Consequently, PS-NSGA and FS-DOS are marginally better than DAEA, RVEA, and NSGAII-SDR in handling imbalanced data. In summary, Tables VIII and IX reveal that JSEMO is well capable of facing imbalanced databases.

Table XI presents a comparative analysis of JSEMO against nine SOTA algorithms as reported in their associated references, including two imbalanced FS techniques (PSOFC-FC and SYMON), four single-objective FS algorithms (PPSO, HFS-C-P, VLPSO-LS, and SIFE), and three multiobjective FS methods (INSA, PRHD, and SM-MOEA). The effectiveness of these competitors is evaluated through three metrics: NF, AC, GM. The absence of a metric is denoted by the symbol “-.” Regarding

GM as the recommended imbalanced classification metric, this table indicates that JSEMO outperforms its peer algorithms in all cases, except for Lymphoma, in which PSOFC-FC similarly performs to JSEMO. JSEMO is considerably better regarding GM than SYMON as another evolutionary FS algorithm for imbalanced data. Furthermore, the proposed JSEMO achieves significantly better overall accuracy than most of its peers, except for HFS-C-P, which performs better for CNS and DLBCL cases. Additionally, JSEMO is the best algorithm for reducing feature size on DLBCL, Leukemia\_3, 11 Tumors, and Lung Cancer problems.

### C. JSEMO Against Filter-Based and Embedded FS Methods

Table XII delineates the efficacy of the JSEMO method when compared against various filter-based techniques. The presented data underscores the superiority of JSEMO as a wrapper approach in contrast to representative filter methods. This comparative analysis encompasses both the supervised and unsupervised versions of Inf-FS [8], max-dependency max-relevance (MRMR) [9], ReliefF [62], CFS [63], Random Forest [57], and JSEMO, focusing on accuracy metrics. Notably, CFS proves effective when dealing with low-dimensional data due to its computational cost. In this regard, we initially reduce

TABLE XI  
NF, AC, AND GM METRICS FOR TEN FS ALGORITHMS

Data (Size)	Algorithm	NF	AC	GM
Prostate-GE (D=5966)	JSEMO	6.65	<b>90.32</b>	<b>90.02</b>
	PRDH [12]	<b>2.53</b>	48.50	-
Colon Tumor (D=2000)	JSEMO	10.9	<b>95.00</b>	<b>94.13</b>
	HFS-C-P [20]	5.9	92.47	-
	SYMON [27]	-	-	71.50
CNS (D=7130)	SIFE [36]	<b>5.72</b>	77.06	-
	JSEMO	12.5	79.65	<b>79.95</b>
	HFS-C-P [20]	7.40	<b>85.91</b>	-
Leukemia_2 (D=7130)	SYMON [27]	-	-	79.00
	SIFE [36]	<b>6.68</b>	63.66	-
	JSEMO	4.40	<b>100</b>	<b>100</b>
Breast Cancer (D=22283)	HFS-C-P [20]	6.12	<b>100</b>	-
	INSA [24]	30	99.9	-
	SIFE [36]	<b>2.2</b>	89.14	-
DLBCL (D=5470)	JSEMO	4.80	<b>78.48</b>	<b>80.98</b>
	SIFE [36]	<b>1.28</b>	75.10	-
	JSEMO	<b>3.85</b>	91.90	<b>92.60</b>
Leukemia_3 (D=7130)	HFS-C-P [20]	5.83	<b>100</b>	-
	VLPSO-LS [60]	8.8	96.13	-
	PSOFS-FC [26]	156	-	89.50
11 Tumors (D=12533)	SM-MOEA [21]	11.2	96.22	-
	SYMON [27]	-	-	79.00
	PRDH [12]	4.53	90.49	-
Lymphoma (D=4027)	PPSO [11]	44	86.22	-
	JSEMO	<b>7.80</b>	<b>95.00</b>	<b>96.96</b>
	VLPSO-LS [60]	7.91	93.75	-
Brain Tumor (D=5920)	PSOFS-FC [26]	55	-	89.05
	PPSO [11]	80	94.37	-
	JSEMO	<b>8.16</b>	<b>84.35</b>	<b>84.15</b>
Lung Cancer (D=12601)	VLPSO-LS [60]	367	82.81	-
	SM-MODA [21]	25.2	76.70	-
	PPSO [11]	167	76.83	-
JSEMO	JSEMO	9.23	<b>100</b>	<b>100</b>
	PSOFS-FC [26]	46	-	100
	SIFE [36]	<b>2.32</b>	98.76	-
VLPSO-LS [60]	JSEMO	<b>13.1</b>	85.63	<b>87.53</b>
	SM-MODA [21]	22.5	<b>90.28</b>	-
	PPSO [11]	73.4	74.40	-
VLPSO-LS [60]	JSEMO	<b>9.83</b>	<b>96.72</b>	<b>96.18</b>
	PSOFS-FC [26]	307	90.17	-
	PPSO [11]	360	-	79.34
SIFE [36]	JSEMO	203	79.38	-
	VLPSO-LS [60]	19.1	88.29	-
	PPSO [11]	-	-	-

the feature set to 1000 using ReliefF and subsequently apply CSF. This table elucidates that JSEMO is consistently competitive or better than its peers. Specifically, in 12 out of 15 instances, JSEMO demonstrates a significant advantage over the filters under consideration. In the remaining three cases, JSEMO exhibits a marginally inferior performance compared to the alternative methods. In this competition, JSEMO requires a longer computational time than filter methods to find the optimal subset.

#### D. Ablation Study on Principle JSEMO's Elements

In the proposed JSEMO algorithm, the set-based JS variation measure and KNN2W work together to contribute to improving the classification performance of JSEMO in handling HD and CI data sets. To investigate their effectiveness, we compare the performance of JSEMO, SEMO, JSEMO-RI, JSEMO-KNN, and NSGAII-SDR-KNN in Table XIII. SEMO is JSEMO without the JS measure, i.e., without the duplication handling strategy. JSEMO-RI shows the contribution of JSEMO without JIM; with JIM replaced by a random population initialization. To illustrate

the positive impact of KNN2W in our contributions, we utilize JSEMO-KNN, in which KNN is interchangeably replaced with KNN2W in JSEMO. Finally, NSGAII-SDR-KNN is an efficient many-objective optimization algorithm in [30] that is used as the basic algorithm in the above algorithms for FS with the standard KNN classifier.

Table XIII shows that JSEMO is the superior algorithm, and NSGAII-SDR-KNN is the worst among the five competitors. Comparing JSEMO and SEMO reveals that JS significantly improves the proposed algorithm's performance regarding IGD. Indeed, JS works as the duplication handling strategy in the set-based variation operation to enhance diversity and avoid premature convergence. The comparison between JSEMO and JSEMO-IR shows that JIM positively affects our approach, in which JIM initializes the starting population with dissimilar and more diverse members. For example, JSEMO outperforms the JSEMO-IR in seven cases, including CNS, GLI, Covid-19, Lymphoma, 11 Tumors, Brain Tumor, and Lung databases. JSEMO and JSEMO-IR also perform similarly in seven cases.

JSEMO is inferior only in Colon Tumor against JSEMO-IR. Furthermore, JSEMO performs considerably better than JSEMO-KNN, particularly in imbalanced problems. For instance, in the Lymphoma and Lung Cancer data with the highest imbalanced rate, JSEMO outperforms JSEMO-KNN. This comparison reveals that the KNN2W significantly contributes to our proposed method. To sum up, Table XIII reveals that each of JSEMO's components positively affects its final performance. Among them, the set-based variation operation integrated with JS contributes the most to JSEMO.

Moreover, Table XIV illustrates the impact of incorporating different objectives during the evolutionary process of JSEMO. In this Table  $f_1$ ,  $f_2$ ,  $f_3$ , and  $f_4$  represent precision, recall, specificity, and NF, respectively. Additionally, the “mean” objective represents the average of  $f_1$ ,  $f_2$ , and  $f_3$ . Therefore, JSEMO competes with four different approaches, including X1, X2, X3 (which involves NF besides precision, recall, and specificity, respectively), and X4 (which considers the average of the first three objectives besides NF). Notably, JSEMO, which optimizes all four basic objectives concurrently in a single run, achieves the best results in terms of the IGD metric. In addition, by considering all aspects of the FS problems, JSEMO provides decision-makers with a comprehensive range of choices. Specifically, the optimal Preto front obtained by JSEMO allows decision-makers to access detailed information, facilitating decisive decisions. The second-best approach, X4, employs an aggregation method. While aggregation can be beneficial in certain cases, it falls short of providing the granularity needed to address all aspects of the problem during decision-making.

In summary, JSEMO's holistic optimization approach yields superior results, offering decision-makers a richer set of options, whereas aggregation, represented by X4, has limitations in providing detailed information for decision-making.

#### E. KNN2W

Table XV compares the proposed KNN2W for imbalanced data with the standard KNN and three efficient variants of KNN,

TABLE XII  
JSEMO AGAINST FILTER-BASED AND EMBEDDED FS METHODS IN TERMS OF ACCURACY (AC)

No.	Inf-FS-S [8]		Inf-FS-U [8]		MRMR [9]		ReliefF [62]		Random Forest [57]		CFS [63]		JSEMO	
	NF=10	NF=100	NF=10	NF=100	NF=10	NF=100	NF=10	NF=100	NF=10	NF=100	AC	NF	AC	NF
1	70.96	83.87	41.93	45.16	61.29	61.29	83.87	83.87	80.64	74.19	83.87	23	<b>90.32</b>	<b>3</b>
2	80.95	93.23	93.23	93.23	66.66	71.42	93.23	93.23	93.23	93.71	85.71	39	<b>94.00</b>	<b>8</b>
3	78.94	84.21	57.89	73.68	89.47	84.21	73.68	84.21	84.21	84.21	78.94	9	<b>95.00</b>	<b>6</b>
4	44.44	66.67	61.11	61.11	55.55	77.77	44.44	55.55	44.44	61.11	61.11	26	<b>79.60</b>	<b>6</b>
5	100.0	100.0	66.66	85.71	90.47	100.0	95.23	95.23	85.71	100.0	95.23	35	<b>100.0</b>	<b>3</b>
6	80.00	73.33	66.67	80.00	86.66	80.00	80.00	73.33	60.00	<b>93.33</b>	80.00	16	78.48	<b>3</b>
7	80.76	85.46	30.76	80.76	84.61	85.46	84.61	85.46	76.92	80.76	<b>88.46</b>	25	86.42	6
8	75.71	58.57	47.14	54.28	64.28	64.28	65.71	62.85	67.14	67.14	61.42	<b>5</b>	<b>88.57</b>	13
9	69.56	91.30	69.56	86.95	78.26	95.65	95.65	86.95	82.61	91.30	<b>100.0</b>	28	91.90	<b>2</b>
10	90.47	94.23	66.67	80.95	80.95	80.95	80.95	85.71	76.19	85.71	80.95	25	<b>95.00</b>	<b>6</b>
11	69.23	80.53	55.76	63.46	57.69	73.07	48.07	75.00	59.62	61.53	71.15	37	<b>84.35</b>	13
12	98.14	98.26	66.67	98.14	94.44	96.29	98.14	98.14	98.00	96.29	<b>100.0</b>	72	99.86	<b>2</b>
13	90.00	100.0	75.00	90.00	75.00	100.0	90.00	95.00	90.00	95.00	100.0	21	<b>100.0</b>	<b>9</b>
14	80.00	<b>92.00</b>	60.00	88.00	88.00	84.00	68.00	84.00	76.00	76.00	80.00	24	85.63	<b>8</b>
15	90.16	90.16	77.04	86.88	86.88	86.88	90.16	95.08	88.52	86.88	88.52	36	<b>96.72</b>	<b>10</b>
↑/≈/↓	1/2/12	1/0/14	1/2/12	0/0/15	1/2/12	2/2/11								

TABLE XIII  
ABLATION STUDY ON THE MAIN ELEMENTS OF JSEMO (IGD)

No.	NSGAI-II- SDR-KNN	JSEMO- KNN	JSEMO-RI	SEMO	JSEMO
1	6.144e-2 ↓	3.612e-2 ↓	3.449e-2 ≈	3.545e-2 ↓	<b>2.353e-2</b>
2	4.291e-2 ↓	5.967e-3 ≈	<b>3.842e-3</b> ≈	1.244e-2 ↓	4.087e-3
3	2.861e-1 ↓	1.174e-1 ↓	<b>5.488e-2</b> ↑	4.773e-2 ≈	5.758e-2
4	2.492e-1 ↓	2.136e-1 ↓	2.075e-1 ↓	2.119e-1 ↓	<b>2.065e-1</b>
5	2.879e-2 ↓	7.153e-4 ≈	<b>5.961e-4</b> ≈	8.696e-4 ≈	7.387e-4
6	1.232e-1 ↓	7.753e-3 ↓	2.647e-4 ≈	4.267e-3 ↓	<b>2.543e-4</b>
7	4.245e-2 ↓	6.518e-3 ↓	9.619e-3 ↓	1.753e-2 ↓	<b>5.265e-4</b>
8	4.231e-1 ↓	3.406e-1 ↓	3.093e-1 ↓	3.541e-1 ↓	<b>2.882e-1</b>
9	5.455e-2 ↓	7.314e-4 ≈	<b>5.302e-4</b> ≈	6.034e-4 ≈	6.948e-4
10	1.673e-2 ↓	2.167e-3 ≈	<b>1.339e-3</b> ≈	1.662e-3 ≈	1.851e-3
11	1.881e-1 ↓	1.734e-1 ↓	1.389e-1 ↓	1.493e-1 ↓	<b>1.335e-1</b>
12	1.728e-2 ↓	2.832e-4 ≈	3.391e-4 ≈	3.890e-3 ↓	<b>3.138e-4</b>
13	1.595e-2 ↓	1.502e-3 ↓	4.570e-3 ↓	6.706e-4 ≈	<b>4.213e-4</b>
14	6.232e-1 ↓	3.103e-1 ↓	1.525e-1 ↓	2.912e-1 ↓	<b>7.944e-2</b>
15	4.097e-1 ↓	3.502e-1 ↓	5.484e-2 ↓	4.852e-2 ↓	<b>4.052e-2</b>
↑/≈/↓	0/0/15	0/5/10	1/7/7	0/5/10	

TABLE XIV  
ABLATION STUDY ON DIFFERENT OBJECTIVES IN JSEMO (IGD)

No.	X1( $f_1, f_4$ )	X2( $f_2, f_4$ )	X3( $f_3, f_4$ )	X4 (mean, $f_4$ )	JSEMO
1	2.937e-1 ↓	4.149e-1 ↓	2.594e-1 ↓	<b>4.273e-3</b> ↑	2.353e-2
2	3.165e-2 ↓	<b>2.714e-3</b> ↑	1.235e-2 ↓	3.185e-3 ≈	4.087e-3
3	1.003e-1 ↓	8.564e-2 ↓	<b>2.833e-2</b> ↑	7.266e-2 ↓	5.758e-2
4	2.281e-1 ↓	1.440e-1 ↑	1.889e-3 ↑	<b>1.814e-3</b> ↑	2.065e-1
5	1.459e-1 ↓	1.213e-1 ↓	1.671e-1 ↓	1.277e-1 ↓	<b>7.387e-4</b>
6	4.456e-2 ↓	7.526e-2 ↓	1.005e-1 ↓	6.432e-4 ≈	<b>2.543e-4</b>
7	1.675e-2 ↓	2.127e-1 ↓	2.067e-1 ↓	1.373e-3 ↓	<b>5.265e-4</b>
8	3.459e-2 ↓	3.345e-1 ↓	3.643e-1 ↓	3.077e-1 ↓	<b>2.882e-1</b>
9	2.123e-2 ↓	2.886e-2 ↓	1.415e-2 ↓	9.020e-3 ↓	<b>6.948e-4</b>
10	1.438e-2 ↓	6.536e-2 ↓	4.121e-2 ↓	9.912e-3 ↓	<b>1.851e-3</b>
11	1.947e-1 ↓	1.310e-1 ≈	<b>2.074e-4</b> ↑	4.521e-4 ↑	1.335e-1
12	5.404e-4 ≈	8.000e-4 ≈	8.106e-4 ≈	5.722e-4 ≈	<b>3.138e-4</b>
13	8.776e-4 ≈	8.445e-4 ≈	8.776e-4 ≈	8.610e-4 ≈	<b>4.213e-4</b>
14	4.570e-1 ↓	9.414e-1 ↓	2.541e-1 ↓	8.260e-2 ↓	<b>7.944e-2</b>
15	8.980e-1 ↓	4.899e-2 ↓	4.814e-2 ↓	<b>2.551e-3</b> ↑	4.052e-2
↑/≈/↓	0/2/13	2/3/10	3/2/10	4/4/7	

including NWKNN [37], FKNN [38], and CCWKNN [39]. NWKNN and CCWKNN stress the imbalanced data, and FKNN considers the neighborhood's similarities. BAC is utilized in this table to show the significance of five KNN-based algorithms

TABLE XV  
BAC FOR KNN2W AND FOUR VARIANTS OF KNN

No.	KNN [56]	FKNN [38]	NWKNN [37]	CCWKNN [39]	Proposed KNN2W
1	<b>65.24</b>	50.00	65.98	61.01	59.88
2	84.40	88.85	87.23	88.85	<b>88.86</b>
3	87.07	94.53	94.53	94.53	<b>94.54</b>
4	65.34	68.59	66.83	<b>68.59</b>	67.87
5	60.02	<b>61.38</b>	60.02	61.38	60.02
6	73.60	79.85	<b>81.86</b>	79.85	<b>81.86</b>
7	57.67	68.66	72.50	68.67	<b>72.50</b>
8	83.61	84.62	83.60	84.62	<b>85.98</b>
9	63.70	67.29	61.87	67.30	<b>68.24</b>
10	<b>83.72</b>	83.55	83.21	83.56	82.21
11	77.07	81.10	73.77	<b>81.10</b>	77.14
12	81.94	80.38	91.69	80.38	<b>92.13</b>
13	83.41	<b>100.0</b>	98.70	<b>100.0</b>	<b>100.0</b>
14	72.38	80.52	80.52	80.42	<b>83.06</b>
15	72.65	79.20	81.94	79.20	<b>82.04</b>
Mean	74.12	77.90	78.95	78.63	<b>79.75</b>

in confronting imbalanced problems. Overall, the proposed KNN2W is better than its peers in 10 out of 15 cases regarding BAC. This table also reveals that KNN2W is comparable with its peers for the data with balanced class-based distributions. Mainly, as the imbalanced rate increases, KNN2W significantly performs better than those of the compared algorithms. For instance, in the last nine applications with an imbalanced rate larger than two, KNN2W only loses DLBLC and Leukemia\_3 data against KNN and CCWKNN methods, respectively. Furthermore, CWKNN and CCWKNN carry out similarly and are better than FKNN; and FKNN is better than KNN.

#### F. Further Illustration of Evolutionary Processes

For an intuitive analysis, Fig. 3 illustrates the evolutionary processes of the six competing algorithms on three challenging cases, including CNS, Brain Tumor, and Lung Cancer, regarding IGD and HV. All trends are obtained by averaging 25 independent runs. Brain Tumor and Lung Cancer are this study's two most imbalanced databases. The first and second columns of the figure show the optimization process regarding IGD and HV, respectively. The efficiency of JSEMO is better than its

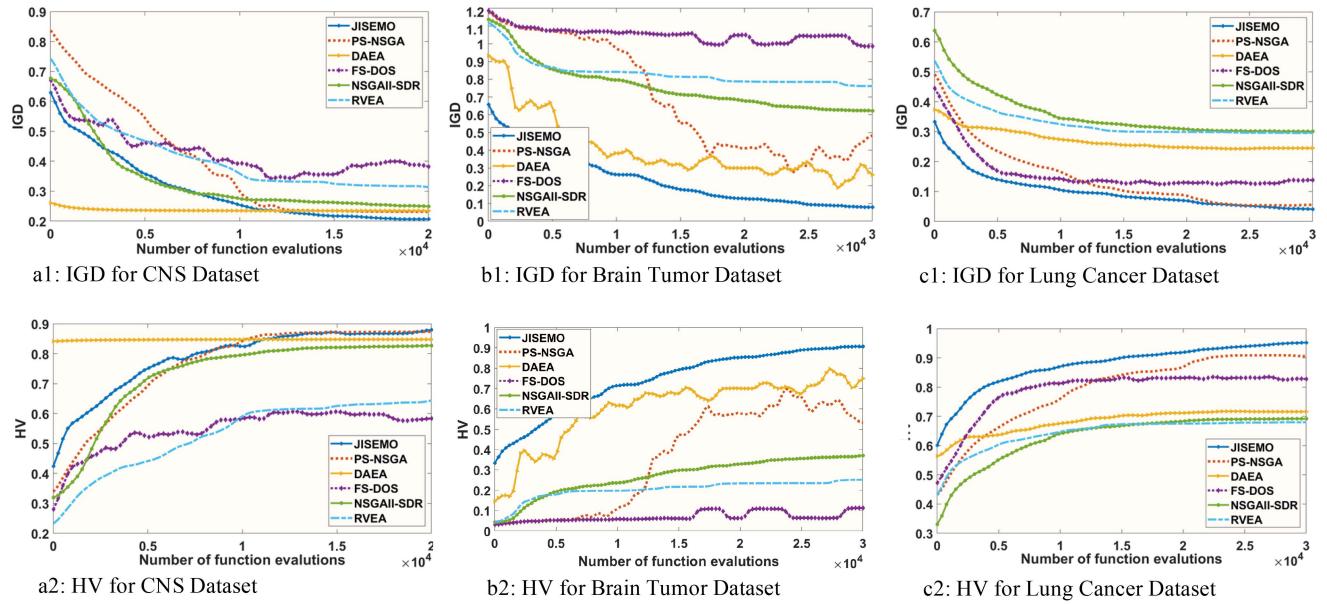


Fig. 3. The illustration of evolution processes regarding IGD and HV on CNS, Brain Tumor, and Lung Cancer data.

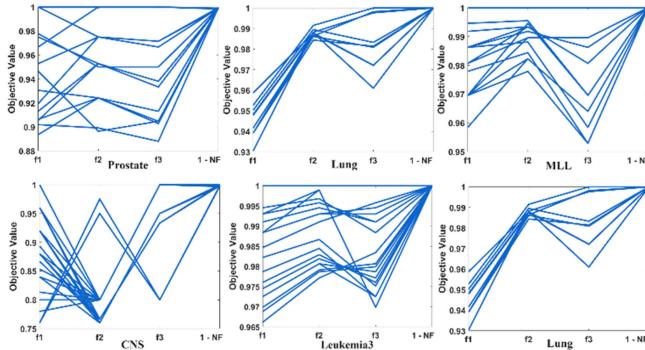


Fig. 4. The parallel coordinates plot of six datasets (Prostate-GE, Colon, MLL, CNS, Leukemia\_3, and Lung Tumor), obtained by JISEMO across the four involved objectives.

peers in most cases. Notably, JSEMO is strictly decreasing and increasing with the passing generation regarding IGD and HV, respectively, while its peers often reach a constant point or even fluctuate during evolution. Hence, JSEMO better balances the exploration and exploitation concepts during the evolution process with the proposed set-based variation operation for FS and handles the duplicated challenges with JS to promote diversity, particularly in creating the initial generations, aiming to escape the local optima.

Fig. 4 displays a parallel coordinates plot of six datasets acquired by JSEMO across the four involved objectives, which include Prostate-GE, Colon, MLL, CNS, Leukemia, and Lung Tumor. These plots visually represent how solutions perform across multiple objectives simultaneously. By observing the distribution of lines (representing solutions in objective space) across the parallel axes, JSEMO achieves a well-diversified Pareto front in the last generated population, with the JS measure

playing a significant role by avoiding duplicated solutions. This visualization aids in understanding the relationships between objectives, facilitating informed decisions about the best compromises or trade-offs among competing objectives.

## VII. CONCLUSION

High-dimensional imbalanced data classification comprises an increasingly significant class of real-world modern problems. The proposed JSEMO in this paper addresses such problems through an evolutionary many-objective FS algorithm, contributing to three key aspects. Firstly, Jaccard similarity is integrated into critical algorithmic steps—population initialization, reproduction operations, and the elitist mechanism—to consistently enhance diversity throughout the search process and prevent duplicated solutions. Secondly, a set-based variation operation, leveraging intersection and union operators compatible with binary coding, is proposed for more effective offspring generation. Thirdly, the KNN2W classifier, which is appropriate for imbalanced data classification, coupled with a set of practical objectives (precision, recall, specificity, and feature size) to view the FS problem as a many-objective paradigm, bolsters JSEMO in addressing imbalanced databases.

The proposed JSEMO is verified using 15 real-world classification benchmark datasets with a wide range of feature numbers and class-imbalanced rates. Compared with 20 up-to-date state-of-the-art techniques, the experimental results confirm that JSEMO considerably outperforms its peers, particularly on HD and imbalanced problems, in terms of IGD, HV, accuracy, balance accuracy, and g-mean metrics with small and optimal set of feature set and competitive computational cost. Furthermore, JSEMO surpasses representative filter-based methods regarding classification metrics. Finally, an ablation study confirms the significance of the Jaccard similarity measure in the initialization,

variation, and elitism processes. The effectiveness of KNN2W with four appropriate evaluation metrics is shown particularly for imbalanced problems.

In the future, we hope to extend JSEMO to databases with missing values and multi-label classification. In addition, we plan to apply ensemble learning strategies to JSEMO and study the classification performance of combining different feature subsets with similar classification accuracy.

## REFERENCES

- [1] M. Shah, M. Marchand, and J. Corbeil, "Feature selection with conjunctions of decision stumps and learning from microarray data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 174–186, Jan. 2012, doi: [10.1109/TPAMI.2011.82](https://doi.org/10.1109/TPAMI.2011.82).
- [2] Q. Dong, S. Gong, and X. Zhu, "Imbalanced deep learning by minority class incremental rectification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 6, pp. 1367–1381, Jun. 2019, doi: [10.1109/TPAMI.2018.2832629](https://doi.org/10.1109/TPAMI.2018.2832629).
- [3] D. O. Medley, C. Santiago, and J. C. Nascimento, "CyCoSeg: A cyclic collaborative framework for automated medical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 8167–8182, Nov. 2022, doi: [10.1109/TPAMI.2021.3113077](https://doi.org/10.1109/TPAMI.2021.3113077).
- [4] W. Pei, B. Xue, M. Zhang, L. Shang, X. Yao, and Q. Zhang, "A survey on unbalanced classification: How can evolutionary computation help?," *IEEE Trans. Evol. Computation*, vol. 28, no. 2, pp. 353–373, Apr. 2024, doi: [10.1109/TEVC.2023.3257230](https://doi.org/10.1109/TEVC.2023.3257230).
- [5] B. H. Nguyen, B. Xue, and M. Zhang, "A survey on swarm intelligence approaches to feature selection in data mining," *Swarm Evol. Computation*, vol. 54, May 2020, Art. no. 100663, doi: [10.1016/j.swevo.2020.100663](https://doi.org/10.1016/j.swevo.2020.100663).
- [6] Y. Sun, S. Todorovic, and S. Goodison, "Local-learning-based feature selection for high-dimensional data analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1610–1626, Sep. 2010, doi: [10.1109/TPAMI.2009.190](https://doi.org/10.1109/TPAMI.2009.190).
- [7] C. Lazar, J. Taminau, S. Meganck, D. Steenhoff, and A. Coletta, "A survey on filter techniques for feature selection in gene expression microarray analysis," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 9, no. 4, pp. 1106–1119, Jul./Aug. 2012, doi: [10.1109/TCBB.2012.33](https://doi.org/10.1109/TCBB.2012.33).
- [8] G. Roffo, S. Melzi, U. Castellani, A. Vinciarelli, and M. Cristani, "Infinite feature selection: A graph-based feature filtering approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4396–4410, Dec. 2021, doi: [10.1109/TPAMI.2020.3002843](https://doi.org/10.1109/TPAMI.2020.3002843).
- [9] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005, doi: [10.1109/TPAMI.2005.159](https://doi.org/10.1109/TPAMI.2005.159).
- [10] B. Cancela, V. Bolón-Canedo, and A. Alonso-Betanzos, "E2E-FS: An end-to-end feature selection method for neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 7, pp. 8311–8323, Jul. 2023, doi: [10.1109/TPAMI.2022.3228824](https://doi.org/10.1109/TPAMI.2022.3228824).
- [11] B. Tran, B. Xue, and M. Zhang, "A new representation in PSO for discretization-based feature selection," *IEEE Trans. Cybern.*, vol. 48, no. 6, pp. 1733–1746, Jun. 2018, doi: [10.1109/TCYB.2017.2714145](https://doi.org/10.1109/TCYB.2017.2714145).
- [12] R. Jiao, B. Xue, and M. Zhang, "Solving multiobjective feature selection problems in classification via problem reformulation and duplication handling," *IEEE Trans. Evol. Computation*, Oct. 2022, doi: [10.1109/TEVC.2022.3215745](https://doi.org/10.1109/TEVC.2022.3215745).
- [13] P. Wang, B. Xue, and M. Zhang, "Multiobjective differential evolution for feature selection in classification," *IEEE Trans. Cybern.*, vol. 53, no. 7, pp. 4579–4593, Jul. 2023, doi: [10.1109/TCYB.2021.3128540](https://doi.org/10.1109/TCYB.2021.3128540).
- [14] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimization for feature selection in classification: A multiobjective approach," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1656–1671, Dec. 2013, doi: [10.1109/TCYB.2012.2227469](https://doi.org/10.1109/TCYB.2012.2227469).
- [15] H. Dong, T. Li, R. Ding, and J. Sun, "A novel hybrid genetic algorithm with granular information for feature selection and optimization," *Appl. Soft Comput.*, vol. 65, pp. 33–46, Apr. 2018, doi: [10.1016/j.asoc.2017.12.048](https://doi.org/10.1016/j.asoc.2017.12.048).
- [16] H. Xu, B. Xue, and M. Zhang, "A duplication analysis-based evolutionary algorithm for biobjective feature selection," *IEEE Trans. Evol. Computation*, vol. 25, no. 2, pp. 205–218, Apr. 2021, doi: [10.1109/TEVC.2020.3016049](https://doi.org/10.1109/TEVC.2020.3016049).
- [17] S. Kashef and H. Nezamabadi-pour, "An advanced ACO algorithm for feature subset selection," *Neurocomputing*, vol. 147, pp. 271–279, Jan. 2015, doi: [10.1016/j.neucom.2014.06.067](https://doi.org/10.1016/j.neucom.2014.06.067).
- [18] R. Yaghobzadeh et al., "A binary grasshopper optimization algorithm for feature selection," *Int. J. Eng. Res. Technol.*, vol. 9, pp. 761–775, 2020, doi: [10.17577/IJERTV9IS030420](https://doi.org/10.17577/IJERTV9IS030420).
- [19] I.-S. Oh, J.-S. Lee, and B.-R. Moon, "Hybrid genetic algorithms for feature selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1424–1437, Nov. 2004, doi: [10.1109/TPAMI.2004.105](https://doi.org/10.1109/TPAMI.2004.105).
- [20] L. Qu et al., "Explicit and size-adaptive PSO-based feature selection for classification," *Swarm Evol. Computation*, vol. 77, Mar. 2023, Art. no. 101249, doi: [10.1016/j.swevo.2023.101249](https://doi.org/10.1016/j.swevo.2023.101249).
- [21] F. Cheng, F. Chu, Y. Xu, and L. Zhang, "A steering-matrix-based multiobjective evolutionary algorithm for high-dimensional feature selection," *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 9695–9708, Sep. 2022, doi: [10.1109/TCYB.2021.3053944](https://doi.org/10.1109/TCYB.2021.3053944).
- [22] P. Wang, B. Xue, J. Liang, and M. Zhang, "Differential evolution-based feature selection: A niching-based multiobjective approach," *IEEE Trans. Evol. Computation*, vol. 27, no. 2, pp. 296–310, Apr. 2023, doi: [10.1109/TEVC.2022.3168052](https://doi.org/10.1109/TEVC.2022.3168052).
- [23] R. Jiao, B. H. Nguyen, B. Xue, and M. Zhang, "A survey on evolutionary multiobjective feature selection in classification: Approaches, applications, and challenges," *IEEE Trans. Evol. Computation*, early access, Jul. 5, 2023, doi: [10.1109/TEVC.2023.3292527](https://doi.org/10.1109/TEVC.2023.3292527).
- [24] Z. Wang, S. Gao, M. Zhou, S. Sato, J. Cheng, and J. Wang, "Information-theory-based nondominated sorting ant colony optimization for multiobjective feature selection in classification," *IEEE Trans. Cybern.*, vol. 8, no. 9, pp. 5276–5289, Aug. 2023, doi: [10.1109/TCYB.2022.3185554](https://doi.org/10.1109/TCYB.2022.3185554).
- [25] K. Deb and H. Jain, "An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, Part I: Solving problems with box constraints," *IEEE Trans. Evol. Computation*, vol. 18, no. 4, pp. 577–601, Aug. 2014, doi: [10.1109/TEVC.2013.2281535](https://doi.org/10.1109/TEVC.2013.2281535).
- [26] Y. Zhang, Y.-H. Wang, D.-W. Gong, and X.-Y. Sun, "Clustering-guided particle swarm feature selection algorithm for high-dimensional imbalanced data with missing values," *IEEE Trans. Evol. Computation*, vol. 26, no. 4, pp. 616–630, Aug. 2022, doi: [10.1109/TEVC.2021.3106975](https://doi.org/10.1109/TEVC.2021.3106975).
- [27] A. Moayedikia, K.-L. Ong, Y. L. Boo, W. G. S. Yeoh, and R. Jensen, "Feature selection for high dimensional imbalanced class data using harmony search," *EAAI*, vol. 57, pp. 38–49, Jan. 2017, doi: [10.1016/j.engappai.2016.10.008](https://doi.org/10.1016/j.engappai.2016.10.008).
- [28] A. Li, Z. He, Q. Wang, and Y. Zhang, "Key quality characteristics selection for imbalanced production data using a two-phase bi-objective feature selection method," *Eur. J. Oper. Res.*, vol. 274, no. 3, pp. 978–989, May 2019, doi: [10.1016/j.ejor.2018.10.051](https://doi.org/10.1016/j.ejor.2018.10.051).
- [29] L. Li, M. Xuan, Q. Lin, M. Jiang, Z. Ming, and K. C. Tan, "An evolutionary multitasking algorithm with multiple filtering for high-dimensional feature selection," *IEEE Trans. Evol. Computation*, vol. 27, no. 4, pp. 802–816, Aug. 23, doi: [10.1109/TEVC.2023.3254155](https://doi.org/10.1109/TEVC.2023.3254155).
- [30] Y. Tian, R. Cheng, X. Zhang, Y. Su, and Y. Jin, "A strengthened dominance relation considering convergence and diversity for evolutionary many-objective optimization," *IEEE Trans. Evol. Computation*, vol. 23, no. 2, pp. 331–345, Apr. 2019, doi: [10.1109/TEVC.2018.2866854](https://doi.org/10.1109/TEVC.2018.2866854).
- [31] R. Cheng, Y. Jin, M. Olhofer, and B. Sendhoff, "A reference vector guided evolutionary algorithm for many-objective optimization," *IEEE Trans. Evol. Computation*, vol. 20, no. 5, pp. 773–791, Oct. 2016, doi: [10.1109/TEVC.2016.2519378](https://doi.org/10.1109/TEVC.2016.2519378).
- [32] Y. Liu, Y. Wang, X. Ren, H. Zhou, and X. Diao, "A classification method based on feature selection for imbalanced data," *IEEE Access*, vol. 7, pp. 81794–81807, Jun. 2019, doi: [10.1109/ACCESS.2019.2923846](https://doi.org/10.1109/ACCESS.2019.2923846).
- [33] Y. Tang, Y.-Q. Zhang, N. V. Chawla, and S. Krasser, "SVMs modeling for highly imbalanced classification," *IEEE Trans. Syst., Man, Cybern. B Cybern.*, vol. 39, no. 1, pp. 281–288, Feb. 2009, doi: [10.1109/TSMCB.2008.2002909](https://doi.org/10.1109/TSMCB.2008.2002909).
- [34] Z. Sun et al., "A novel ensemble method for classifying imbalanced data," *Pattern Recognit.*, vol. 48, no. 5, pp. 1623–1637, Jan. 2015, doi: [10.1016/j.patcog.2014.11.014](https://doi.org/10.1016/j.patcog.2014.11.014).
- [35] L. Sun et al., "TSFNFR: Two-stage fuzzy neighborhood-based feature reduction with binary whale optimization algorithm for imbalanced data classification," *Knowl.-Based Syst.*, vol. 256, Nov. 2022, Art. no. 109849, doi: [10.1016/j.knosys.2022.109849](https://doi.org/10.1016/j.knosys.2022.109849).
- [36] H. Saadatmand and M.-R. Akbarzadeh-T, "Set-based integer-coded fuzzy granular evolutionary algorithms for high-dimensional feature selection," *Appl. Soft Comput.*, vol. 142, Apr. 2023, Art. no. 110240, doi: [10.1016/j.asoc.2023.110240](https://doi.org/10.1016/j.asoc.2023.110240).

- [37] S. Tan, "Neighbor-weighted K-nearest neighbor for unbalanced text corpus," *Expert Syst. Appl.*, vol. 28, no. 4, pp. 667–671, May 2005, , doi: [10.1016/j.eswa.2004.12.023](https://doi.org/10.1016/j.eswa.2004.12.023).
- [38] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy K-nearest neighbor algorithm," *IEEE Trans. Syst., Man, Cyber.*, vol. SMC-15, no. 4, pp. 580–585, Jul./Aug. 1985, doi: [10.1109/TSMC.1985.6313426](https://doi.org/10.1109/TSMC.1985.6313426).
- [39] W. Liu and S. Chawla, "Class confidence weighted kNN algorithms for imbalanced data sets," in *Pacific-Asia Conf. Knowl. Discov. Data Mining*, Berlin, Heidelberg, 2011, pp. 345–356, doi: [10.1007/978-3-642-20847-8\\_29](https://doi.org/10.1007/978-3-642-20847-8_29).
- [40] Y. Zhou, J. Kang, S. Kwong, X. Wang, and Q. Zhang, "An evolutionary multiobjective optimization framework of discretization-based feature selection for classification," *Swarm Evol. Computation*, vol. 60, Feb. 2021, Art. no. 100770, , doi: [10.1016/j.swevo.2020.100770](https://doi.org/10.1016/j.swevo.2020.100770).
- [41] Y. Zhang, D. -w. Gong, and J. Cheng, "Multiobjective 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, doi: [10.1109/TCBB.2015.2476796](https://doi.org/10.1109/TCBB.2015.2476796).
- [42] Y. Zhang et al., "Binary differential evolution with self-learning for multiobjective feature selection," *Inf. Sci.*, vol. 507, pp. 67–85, Jan. 2020, , doi: [10.1016/j.ins.2019.08.040](https://doi.org/10.1016/j.ins.2019.08.040).
- [43] Cheng, J. C., Q. Wang, and L. Zhang, "A variable granularity search-based multiobjective feature selection algorithm for high-dimensional data classification," *IEEE Trans. Evol. Computation*, vol. 27, no. 2, pp. 266–280, Apr. 2023, doi: [10.1109/TEVC.2022.3160458](https://doi.org/10.1109/TEVC.2022.3160458).
- [44] Y. Zhou et al., "A problem-specific non-dominated sorting genetic algorithm for supervised feature selection," *Inf. Sci.*, vol. 547, pp. 841–859, Feb. 2021, doi: [10.1016/j.ins.2020.08.083](https://doi.org/10.1016/j.ins.2020.08.083).
- [45] M. Pal and S. Bandyopadhyay, "Many-objective feature selection for motor imagery EEG signals using differential evolution and support vector machine," in *Proc. Int. Conf. Microelectronics Comput. Commun.*, Durgapur, India, 2016, pp. 1–6, doi: [10.1109/MicroCom.2016.7522574](https://doi.org/10.1109/MicroCom.2016.7522574).
- [46] K. Demir, B. Nguyen, B. Xue, and M. Zhang, "Co-operative co-evolutionary many-objective embedded multi-label feature selection with decomposition-based PSO," in *Proc. Genet. Evol. Computation Conf.*, New York, NY, USA, USA, 2023, pp. 438–444, doi: [10.1145/3583131.3590373](https://doi.org/10.1145/3583131.3590373).
- [47] H. Dong et al., "A many-objective feature selection for multi-label classification," *Knowl.-Based Syst.*, vol. 208, Nov. 2020, Art. no. 106456, , doi: [10.1016/j.knosys.2020.106456](https://doi.org/10.1016/j.knosys.2020.106456).
- [48] H. Chen, T. Li, X. Fan, and C. Luo, "Feature selection for imbalanced data based on neighborhood rough sets," *Inf. Sci.*, vol. 483, pp. 1–20, May 2019, , doi: [10.1016/j.ins.2019.01.041](https://doi.org/10.1016/j.ins.2019.01.041).
- [49] P. Zhou et al., "Online feature selection for high-dimensional class-imbalanced data," *Knowl.-Based Syst.*, vol. 136, pp. 187–199, Nov. 2017, , doi: [10.1016/j.knosys.2017.09.006](https://doi.org/10.1016/j.knosys.2017.09.006).
- [50] E. S. Hosseini and M. H. Moattar, "Evolutionary feature subsets selection based on interaction information for high dimensional imbalanced data classification," *Appl. Soft Comput.*, vol. 82, Sep. 2019, Art. no. 105581, , doi: [10.1016/j.asoc.2019.105581](https://doi.org/10.1016/j.asoc.2019.105581).
- [51] W. Wei et al., "Multiobjective optimization algorithm with dynamic operator selection for feature selection in high-dimensional classification," *Appl. Soft Comput.*, vol. 143, Aug. 2023, Art. no. 110360, , doi: [10.1016/j.asoc.2023.110360](https://doi.org/10.1016/j.asoc.2023.110360).
- [52] M. Liu et al., "Cost-sensitive feature selection by optimizing F-measures," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1323–1335, Mar. 2018, doi: [10.1109/TIP.2017.2781298](https://doi.org/10.1109/TIP.2017.2781298).
- [53] H. He and Y. Ma, "Assessment metrics for imbalanced learning," in *Imbalanced Learning: Foundations, Algorithms, and Applications*, Piscataway, NJ, USA: IEEE, 2013, pp. 187–206, doi: [10.1002/9781118646106.ch8](https://doi.org/10.1002/9781118646106.ch8).
- [54] U. Bhowan, M. Johnston, and M. Zhang, "Developing new fitness functions in genetic programming for classification with unbalanced data," *IEEE Trans. Syst., Man, Cyber. B Cyber.*, vol. 42, no. 2, pp. 406–421, Apr. 2012, doi: [10.1109/TSMCB.2011.2167144](https://doi.org/10.1109/TSMCB.2011.2167144).
- [55] N. Chung et al., "Jaccard/Tanimoto similarity test and estimation methods for biological presence-absence data," *BMC Bioinf.*, vol. 20, 2019, Art. no. 644, doi: [10.1186/s12859-019-3118-5](https://doi.org/10.1186/s12859-019-3118-5).
- [56] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967, doi: [10.1109/TIT.1967.1053964](https://doi.org/10.1109/TIT.1967.1053964).
- [57] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998, doi: [10.1109/34.709601](https://doi.org/10.1109/34.709601).
- [58] M. Lichman, "UCI machine learning repository Irvine," Univ. California, School Inf. Comput. Sci., Berkeley, CA, 2013. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [59] Z. Zhu, Y. S. Ong, and M. Dash, "Markov blanket-embedded genetic algorithm for gene selection," *Pattern Recognit.*, vol. 49, no. 11, pp. 3236–3248, Nov. 2007, doi: [10.1016/j.patcog.2007.02.007](https://doi.org/10.1016/j.patcog.2007.02.007).
- [60] B. Tran, B. Xue, and M. Zhang, "Variable-length particle swarm optimization for feature selection on high-dimensional classification," *IEEE Trans. Evol. Computation*, vol. 23, no. 3, pp. 473–487, Jun. 2019, doi: [10.1109/TEVC.2018.2869405](https://doi.org/10.1109/TEVC.2018.2869405).
- [61] Y. Tian, R. Cheng, X. Zhang, and Y. Jin, "PlatEMO: A MATLAB platform for evolutionary multiobjective optimization [Educational Forum]," *IEEE Comput. Intell. Mag.*, vol. 12, no. 4, pp. 73–87, Nov. 2017, doi: [10.1109/MCI.2017.2742868](https://doi.org/10.1109/MCI.2017.2742868).
- [62] M. Robnik-Šikonja and I. Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, no. 1-2, pp. 23–69, 2003.
- [63] M. A. Hall, "Correlation-based feature selection of discrete and numeric class machine learning," Ph.D. dissertation, Dept. Comp. Sci., Univ. Waikato, Waikato, Hamilton, New Zealand, 1999.
- [64] P. A. Bosman and D. Thierens, "The balance between proximity and diversity in multiobjective evolutionary algorithms," *IEEE Trans. Evol. Computation*, vol. 7, no. 2, pp. 174–188, Apr. 2003, doi: [10.1109/TEVC.2003.810761](https://doi.org/10.1109/TEVC.2003.810761).
- [65] L. While, P. Hingston, L. Barone, and S. Huband, "A faster algorithm for calculating hypervolume," *IEEE Trans. Evol. Computation*, vol. 10, no. 1, pp. 29–38, Feb. 2006, doi: [10.1109/TEVC.2005.851275](https://doi.org/10.1109/TEVC.2005.851275).



**Hassan Saadatmand** received the MS degree in control engineering from the Ferdowsi University of Mashhad, Iran, in 2017. He is currently a MATLAB instructor, specializing in meta-heuristic optimization, deep learning, data mining, and machine learning. His research interests include deep learning, soft computing, data mining, multi-objective optimization, signal analysis, control, and robotics.



**Mohammad-R. Akbarzadeh-T.** (Senior Member, IEEE) received the PhD degree in evolutionary optimization and fuzzy control of complex systems from the Electrical and Computer Engineering Department, University of New Mexico (UNM), in 1998. He is currently a professor and founding member of the Center of Excellence on Soft Computing and Intelligent Information Processing (SCIIP), Department of Electrical Engineering, Ferdowsi University of Mashhad. He is the founding president of the Intelligent Systems Scientific Society of Iran and the founding councilor representing the Iranian Coalition on Soft Computing in IFSA. He has received several awards, including the IEEE Iran Section's 2021 Caro Lucas Award, the Ferdowsi Academic Foundation's Outstanding Faculty Award in 2021, and the National Outstanding Faculty Award in 2019. His research interests include bio-inspired computing/optimization, fuzzy logic and control, soft computing, multi-agent systems, complex systems, robotics, cognitive sciences, and medical informatics. He has published over 450 peer-reviewed articles in these and related research fields.