Adaptive Ensemble Multi-Objective Differential Evolution for High-Dimensional Gene Selection

Dhia Eddine Bouazizi
LIPSIC, Faculty of Science of Tunis, Tunis El Manar University
Tunis, Tunisia
dhiya.bouazizi@esen.tn

Asma Amdouni
LIPAH Research Lab, Faculty of Science of Tunis, Tunis El Manar University
Tunis, Tunisia
Mediterranean Institute of Technology, South Mediterranean University,
Tunis, Tunisia
asma.amdouni@fst.utm.tn

Daniel Rodriguez
Computer Science Department, University of Alcala
Madrid, Spain
daniel.rodriguezg@uah.es

Ghaith Manita
FST, University of Tunis El Manar
Tunis, Tunisia
ghaith.manita@fst.utm.tn

Abstract

This paper proposes *Adaptive Ensemble Multi-Objective Differential Evolution for Feature Selection* (AEMO-DEFS), a wrapper-based framework that simultaneously minimizes feature count and maximizes classification accuracy for high-dimensional gene expression data. AEMO-DEFS introduces two key adaptive mechanisms: (1) an ensemble of DE mutation strategies (rand/1/bin, best/1/bin, current-to-pbest/1/bin) with dynamic selection based on historical success, and (2) memory-based self-adaptation of the scaling factor *F* (using Cauchy distribution) and crossover rate *CR* (using Normal distribution). Non-dominated sorting combined with crowding distance maintains diverse Pareto-optimal solutions. Comprehensive experiments on five cancer microarray datasets demonstrate that AEMO-DEFS outperforms MOPSO, NSGA-III, and MOEA/D across all evaluation metrics, achieving superior Inverted Generational Distance (IGD), ϵ-indicator, classification accuracy, and significant feature reduction.

Keywords: Differential Evolution, Feature Selection, Multi-objective Optimization, Adaptive Ensemble, Parameter Adaptation

CCS Concepts

• **Computing methodologies → Evolutionary algorithms**; *Supervised learning*; **Feature selection**.

ACM Reference Format:
Dhia Eddine Bouazizi, Asma Amdouni, Daniel Rodriguez, and Ghaith Manita. 2026. Adaptive Ensemble Multi-Objective Differential Evolution for High-Dimensional Gene Selection. In *Genetic and Evolutionary Computation Conference (GECCO Companion '26), July 13–17, 2026, San Jose, Costa Rica.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3795095.3805170

GECCO Companion '26, San Jose, Costa Rica

ACM ISBN 979-8-4007-2488-6/2026/07
https://doi.org/10.1145/3795095.3805170

1 Introduction

High-dimensional datasets have become ubiquitous in modern computational biology, particularly in the field of genomics where microarray and next-generation sequencing technologies routinely generate datasets containing thousands to tens of thousands of features (genes) while sample sizes remain relatively small, often comprising only tens to hundreds of biological specimens [10]. This inherent imbalance between feature dimensionality and sample cardinality, commonly referred to as the "curse of dimensionality," presents substantial challenges for predictive modeling and knowledge discovery. Although these high-dimensional datasets are information-rich and contain valuable biological insights, they frequently harbor superfluous or non-essential features that can compromise the efficacy of predictive models, escalate computational demands, and induce overfitting phenomena where models memorize noise rather than learning meaningful patterns [6, 12].

Feature selection (FS) emerges as a crucial preprocessing strategy to mitigate these challenges by pinpointing and preserving only the most discriminative attributes, consequently boosting model precision, curtailing data dimensionality, and improving the interpretability and comprehensibility of results [1, 8]. The selection of relevant features not only reduces computational overhead during model training and inference but also facilitates biological interpretation by identifying key biomarkers associated with disease states, treatment responses, or other phenotypic characteristics of interest. In clinical contexts, this interpretability is particularly valuable as it enables domain experts to validate discovered patterns and potentially uncover novel biological mechanisms driving disease progression or therapeutic efficacy.

Methodologies for feature selection are generally classified into three main categories: filter, wrapper, and embedded techniques [3]. Filter approaches assess feature relevance independently of any specific learning algorithm, often relying on statistical metrics such as correlation coefficients, mutual information, or chi-square tests to evaluate the discriminative power of individual features. Despite their computational efficiency and independence from classifier choice, filter methods frequently overlook the interplay between features and their collective impact on classifier performance, potentially discarding informative feature combinations where individual features appear uninformative in isolation [10]. In contrast, wrapper methods evaluate potential feature subsets by employing the intended learning algorithm as part of the evaluation function, which typically results in superior predictive accuracy and accounts for feature interactions but incurs substantial computational overhead due to repeated model training and evaluation [12]. Embedded strategies integrate the feature selection process directly within the model construction phase, striking a practical compromise between predictive power and operational efficiency by learning which features are most informative during the training process itself [3].

Addressing the feature selection challenge inherently necessitates navigating multiple, often contradictory, goals: achieving maximal classification accuracy while concurrently minimizing the cardinality of the selected feature subset [2]. Conventional single-objective optimization strategies usually amalgamate these competing objectives through weighted summation or penalty functions, a process that demands meticulous parameter calibration to determine appropriate objective weights and frequently leads to solutions that are not globally optimal due to the arbitrary nature of weight assignment [11]. Multi-objective optimization paradigms overcome this fundamental limitation by pursuing the simultaneous optimization of all objectives without requiring a priori preference specification, generating a collection of Pareto-optimal solutions that delineate various trade-offs between the competing goals and enable decision-makers to select appropriate feature subsets based on domain-specific constraints and requirements [2].

Within the broad spectrum of evolutionary computation techniques, Differential Evolution (DE) has garnered considerable interest for tackling optimization tasks owing to its conceptual simplicity, minimal requirement for parameter tuning, and demonstrated robustness across diverse problem domains [14]. The DE algorithm employs a simple yet effective mutation operator that adds the scaled difference between two randomly selected population members to a third member, combined with binomial or exponential crossover to generate trial vectors. This straightforward mechanism has proven surprisingly effective across a wide range of optimization problems, from continuous function optimization to combinatorial problems such as feature selection [10]. Numerous multi-objective adaptations of DE have been developed, incorporating concepts from established algorithms such as NSGA-II, MOEA/D, and MOPSO to handle the multi-objective nature of complex optimization problems [4, 9, 13]. While these hybrid algorithms have shown encouraging outcomes in feature selection applications, they often struggle with effectively balancing the exploration of the search space against the exploitation of known good solutions, preserving sufficient population diversity, and preventing premature convergence to suboptimal regions, especially when confronted with datasets of extremely high dimensionality common in genomic applications.

This study introduces an innovative framework, designated as Adaptive Ensemble Multi-Objective Differential Evolution for Feature Selection (AEMO-DEFS), specifically tailored for high-dimensional feature selection tasks encountered in genomic analysis. The principal novel contributions presented herein encompass several interconnected aspects that collectively address the limitations of existing approaches. First, we propose an adaptive ensemble mechanism for DE operators (mutation and crossover) that dynamically selects the most effective strategies during the search based on their historical performance, enhancing the algorithm's adaptability to different problem landscapes and search stages without requiring manual strategy selection. Second, we introduce adaptive control mechanisms for the primary DE parameters, specifically the scaling factor (F) and crossover rate (CR), allowing the algorithm to self-tune these crucial parameters based on their effectiveness in generating successful offspring, thereby improving robustness and reducing the user effort required for parameter calibration. Third, we integrate these adaptive techniques within a standard multi-objective framework using nondominated sorting and crowding distance to effectively balance the dual objectives of minimizing feature subset size and maximizing classification accuracy while preserving solution diversity.

The remainder is organized as follows. Section 2 reviews prior research on multi-objective optimization and differential evolution applied to feature selection. Section 3 details the AEMO-DEFS framework, including its mathematical formulation, representation scheme, adaptive operators, and selection mechanisms. Section 4 describes the experimental configuration. Section 5 presents and analyzes the empirical results. Section 6 concludes the study.

2 Related Work

The challenge of feature selection in high-dimensional data, particularly in the context of genomic analysis, has attracted substantial research attention over the past two decades. This section provides a comprehensive review of relevant literature, examining feature selection methodologies from different perspectives and situating our proposed approach within the current research landscape.

The feature selection problem in bioinformatics contexts presents unique challenges arising from the characteristics of genomic data. Gene expression datasets typically contain thousands to tens of thousands of features (genes) while the number of available samples remains relatively small, often numbering only in the tens or hundreds. This extreme dimensionality-to-sample-size ratio creates a particularly challenging optimization landscape where traditional feature selection methods may fail to identify truly informative genes while eliminating redundant or noisy features [10]. The goal of feature selection in this context extends beyond mere dimensionality reduction to include the identification of biologically meaningful biomarker panels that can provide insights into disease mechanisms, support clinical decision-making, and potentially guide the development of targeted therapies.

Filter methods represent the most computationally efficient category of feature selection techniques, evaluating feature relevance using intrinsic data characteristics without reference to any specific learning algorithm. Common filter approaches include correlation-based methods that identify features highly correlated with class labels, mutual information-based approaches that measure statistical dependencies between features and target variables, and chi-square tests that assess statistical significance of feature-class associations [3]. While filter methods offer advantages in terms of computational efficiency and generalizability across different classifiers, they possess inherent limitations in that they evaluate features independently and cannot capture complex feature interactions that may be crucial for accurate classification. In genomic applications where genes function within interconnected biological pathways and regulatory networks, the independent evaluation of individual genes may fail to identify gene combinations that collectively provide strong discriminative power despite individual genes appearing uninformative in isolation [6].

Wrapper methods address the limitations of filter approaches by incorporating the learning algorithm directly into the feature selection process, evaluating candidate feature subsets based on their actual predictive performance rather than surrogate measures. This wrapper-based evaluation enables the discovery of synergistic feature combinations where features interact in complex ways to influence classification outcomes. However, wrapper methods incur substantial computational costs due to the repeated training and evaluation of the classifier for each candidate feature subset, making them computationally intensive for high-dimensional problems [12]. Various strategies have been proposed to mitigate this computational burden, including the use of efficient search algorithms such as genetic algorithms, particle swarm optimization, and differential evolution to explore the combinatorial space of feature subsets more effectively than exhaustive or greedy search procedures.

Evolutionary algorithms have proven particularly effective for multi-objective feature selection due to their ability to handle complex, non-convex search spaces and maintain diverse solution populations throughout the optimization process. Multi-objective evolutionary algorithms (MOEAs) have been successfully applied to feature selection across various domains, with researchers reporting competitive or superior performance compared to traditional single-objective and greedy approaches [2, 3]. The Non-dominated Sorting Genetic Algorithm II (NSGA-II) proposed by Deb and colleagues has been widely applied to multi-objective feature selection problems [5]. NSGA-II employs a fast nondominated sorting procedure to rank solutions based on Pareto dominance and uses crowding distance to preserve diversity among solutions with the same nondomination rank. Particle Swarm Optimization (PSO) has also been extensively adapted for multi-objective feature selection, with researchers developing various strategies to handle multiple objectives and maintain swarm diversity [4, 7]. The Multi-Objective Evolutionary Algorithm based on Decomposition (MOEA/D) represents an alternative paradigm that transforms the multi-objective problem into a set of scalar optimization subproblems, each associated with a weight vector representing a particular point on the Pareto front [13].

Differential Evolution has attracted substantial research attention due to its simplicity, effectiveness, and broad applicability across diverse optimization problems. One significant direction of research focuses on adaptive DE variants that automatically adjust algorithmic parameters or mutation strategies based on search performance. The Success-History based Adaptive DE (SHADE) algorithm maintains a memory of successful parameter settings and uses this history to generate parameter values for new individuals, achieving competitive performance without requiring manual parameter tuning. Similarly, the Adaptive DE with External Archive (JADE) algorithm employs an optional archive of historical solutions and adapts the mutation strategy and control parameters based on the characteristics of successful operations. Ensemble strategies in DE involve maintaining multiple mutation and crossover strategies and dynamically selecting among them based on their observed performance during the search process. The AEMO-DEFS framework proposed in this work builds upon these foundations by introducing an adaptive ensemble of DE operators combined with memory-based parameter tuning, all within a Pareto-based multi-objective selection environment.

3 Proposed Methodology: AEMO-DEFS Framework

This section presents the detailed formulation and implementation of the Adaptive Ensemble Multi-Objective Differential Evolution for Feature Selection (AEMO-DEFS) framework. The methodology is designed to address the bi-objective nature of high-dimensional feature selection: minimizing the number of selected features while maximizing classification accuracy. AEMO-DEFS introduces several novel mechanisms that collectively enhance the search capability and robustness of standard differential evolution for this challenging optimization problem.

The overall philosophy underlying AEMO-DEFS is to create an adaptive search framework that can automatically adjust its operational characteristics based on the problem landscape and search progress, eliminating the need for extensive manual parameter tuning while maintaining competitive performance across diverse problem instances. This adaptivity is achieved through two complementary mechanisms: an adaptive ensemble of mutation and crossover strategies that dynamically allocates search effort based on observed operator performance, and a memory-based parameter adaptation scheme that adjusts the scaling factor and crossover rate based on their historical effectiveness in generating successful offspring.

3.1 Problem Formulation

The feature selection problem addressed in this work is formulated as a multi-objective optimization problem (MOP) with two conflicting objectives. Given a dataset comprising N samples (instances) and D features (genes or attributes), the goal is to find a feature subset $S \subseteq \{1, 2, \ldots, D\}$ that optimizes the following objectives simultaneously:

$$\min \quad f_1(S) = |S| \tag{1}$$

$$\max \quad f_2(S) = \text{Accuracy}(S) \tag{2}$$

where $|S|$ denotes the cardinality (number of elements) of the selected feature subset, and $\text{Accuracy}(S)$ represents the classification accuracy obtained when training and evaluating a classifier using only the features contained in subset S. The objective $f_1(S)$ encourages parsimonious feature subsets that are easier to interpret and less prone to overfitting, while $f_2(S)$ promotes high predictive performance for downstream classification tasks.

The classification accuracy $\text{Accuracy}(S)$ is estimated using a k-Nearest Neighbor (k-NN) classifier with $k = 5$ and 5-fold cross-validation, following established practice in wrapper-based feature selection methods. In 5-fold cross-validation, the dataset is partitioned into five approximately equal folds; the classifier is trained on four folds and evaluated on the remaining fold, with this process repeated five times such that each fold serves as the test set exactly once. The reported accuracy represents the average classification accuracy across the five folds, providing a robust estimate of generalization performance that is less susceptible to the particularities of any single train-test split.

The solution to this multi-objective optimization problem is not a single optimal feature subset but rather a set of Pareto-optimal solutions representing different trade-offs between feature count and classification accuracy. A solution S_1 is said to Pareto-dominate another solution S_2 (denoted $S_1 \prec S_2$) if and only if:

$$\begin{aligned} S_1 \prec S_2 \iff & \big(f_1(S_1) \le f_1(S_2) \land f_2(S_1) \ge f_2(S_2)\big) \\ & \land \ \big(f_1(S_1) < f_1(S_2) \lor f_2(S_1) > f_2(S_2)\big). \end{aligned} \tag{3}$$

The set of nondominated solutions, known as the Pareto set, contains all solutions that are not dominated by any other feasible solution. The corresponding objective vectors constitute the Pareto front, which characterizes the achievable trade-offs between the competing objectives.

3.2 Solution Representation and Evaluation

Each potential solution (individual) in the differential evolution population is represented as a continuous vector $\mathbf{x} = (x_1, x_2, \dots, x_D)$, where D is the total number of features in the dataset and each component x_i is constrained to the interval $[0, 1]$. This continuous representation enables the application of standard differential evolution operators without modification, avoiding the combinatorial complexities associated with direct binary or integer representations.

The continuous representation is mapped to a binary feature selection vector $\mathbf{b} = (b_1, b_2, \dots, b_D)$ through a thresholding operation. For each dimension i, the binary feature selection indicator b_i is computed as:

$$b_i = \begin{cases} 1 & \text{if } x_i > \theta \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where θ is a threshold parameter set to 0.5 in the current implementation. This thresholding approach is commonly used in evolutionary feature selection and allows the continuous search operators of DE to be applied to the combinatorial feature selection problem. The binary vector $\mathbf{b}$ directly indicates which features are selected ($b_i = 1$) and which are excluded ($b_i = 0$), and the resulting feature subset $S = \{i \mid b_i = 1\}$ is evaluated using the cross-validated k-NN classifier as described in the problem formulation.

The population initialization process generates random continuous vectors uniformly distributed in the hypercube $[0, 1]^D$. This random initialization provides diverse starting points for the evolutionary search, ensuring coverage of the feature space and preventing bias toward particular regions of the search landscape.

3.3 Adaptive Ensemble of DE Operators

At generation t, the population is $P^t = \{\mathbf{x}_i^t\}_{i=1}^N$. For each target vector $\mathbf{x}_i^t$, AEMO-DEFS samples one differential-evolution operator from an ensemble of $K = 3$ candidate strategies. The strategy index is drawn from a categorical distribution,

$$k_i^t \sim \mathrm{Cat}(\mathbf{p}^t), \qquad \mathbf{p}^t = (p_1^t, \dots, p_K^t), \quad p_k^t \ge 0, \quad \sum_{k=1}^{K} p_k^t = 1, \tag{5}$$

thereby turning operator choice into an explicit adaptive control variable rather than a fixed design commitment.

Mutation operators. Let r_1, r_2, r_3 be mutually distinct indices sampled from $\{1, \dots, N\} \setminus \{i\}$. Conditional on the sampled strategy k_i^t and parameter draws (F_i^t, CR_i^t) (specified in the subsequent subsection), the mutant vector $\mathbf{v}_i^t$ is generated according to one of the following prescriptions:

Strategy 1 (rand/1):

$$\mathbf{v}_i^t = \mathbf{x}_{r_1}^t + F_i^t\big(\mathbf{x}_{r_2}^t - \mathbf{x}_{r_3}^t\big), \tag{6}$$

Strategy 2 (best/1):

$$\mathbf{v}_i^t = \mathbf{x}_{\text{best}}^t + F_i^t\big(\mathbf{x}_{r_1}^t - \mathbf{x}_{r_2}^t\big), \tag{7}$$

Strategy 3 (current-to-pbest/1):

$$\mathbf{v}_i^t = \mathbf{x}_i^t + F_i^t\big(\mathbf{x}_{p\text{best}}^t - \mathbf{x}_i^t\big) + F_i^t\big(\mathbf{x}_{r_1}^t - \mathbf{x}_{r_2}^t\big). \tag{8}$$

The vector $\mathbf{x}_{\text{best}}^t$ is defined as the highest-ranked population member under the multi-objective ordering introduced below (lowest non-dominated rank; ties broken by highest crowding distance). The vector $\mathbf{x}_{p\text{best}}^t$ is sampled uniformly from the top $n_p = \lceil pN \rceil$ individuals under the same ordering, where $p = 0.1$ in this work. This design couples exploitation of a restricted elite set with a residual exploratory component.

Binomial crossover. Given the target $\mathbf{x}_i^t$ and mutant $\mathbf{v}_i^t$, the trial vector $\mathbf{u}_i^t$ is formed componentwise by binomial crossover:

$$u_{i,d}^t = \begin{cases} v_{i,d}^t, & \text{if } \mathrm{rand}(0,1) \le CR_i^t \text{ or } d = d_{\text{rand}}, \\ x_{i,d}^t, & \text{otherwise}, \end{cases} \tag{9}$$

where $d_{\text{rand}} \sim \mathcal{U}\{1, \dots, D\}$ enforces inheritance of at least one component from the mutant. To ensure all components remain within the feasible domain $[0, 1]$, a boundary repair is applied:

$$u_{i,d}^t \leftarrow \min\{1, \max\{0, u_{i,d}^t\}\}, \quad \forall d \in \{1, \dots, D\}. \tag{10}$$

Ranking for selection and elitism. Let $\mathrm{rank}(\mathbf{x}) \in \{1, 2, \dots\}$ denote the non-dominated sorting level (front number) and let $\mathrm{cd}(\mathbf{x})$ denote the crowding distance computed within each front. A total order convenient for defining $\mathbf{x}_{\text{best}}^t$ and the top-p set is

$$\mathbf{a} \prec_R \mathbf{b} \iff \begin{cases} \mathrm{rank}(\mathbf{a}) < \mathrm{rank}(\mathbf{b}), & \text{or} \\ \mathrm{rank}(\mathbf{a}) = \mathrm{rank}(\mathbf{b}) \ \land \ \mathrm{cd}(\mathbf{a}) > \mathrm{cd}(\mathbf{b}). \end{cases} \tag{11}$$

Environmental selection. Let $U^t = \{\mathbf{u}_i^t\}_{i=1}^N$ be the set of trial vectors and define the union

$$Q^t = P^t \cup U^t. \tag{12}$$

The next generation is obtained by an NSGA-II-type environmental selection operator,

$$P^{t+1} = \mathrm{Select}_N(Q^t), \tag{13}$$

which fills successive non-dominated fronts $\mathcal{F}_1, \mathcal{F}_2, \dots$ until the population size N is reached. If a front cannot be fully included, individuals within that front are ranked by descending crowding distance, and those with the largest distances are selected.

Success indicator for strategy adaptation. A crisp and implementation faithful success indicator is survival into the next generation:

$$s_i^t = \begin{cases} 1, & \text{if } \mathbf{u}_i^t \in P^{t+1}, \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

Aggregating by strategy yields usage counts and successes,

$$N_k^t = \sum_{i=1}^{N} \mathbb{I}[k_i^t = k], \qquad S_k^t = \sum_{i=1}^{N} \mathbb{I}[k_i^t = k]\, s_i^t, \qquad \hat{\rho}_k^t = \frac{S_k^t}{\max\{1, N_k^t\}}. \tag{15}$$

When finer-grained credit assignment is desired, survival can be replaced by a nonnegative contribution score Δ_i^t. One conservative choice, compatible with rank-based selection, is a rank improvement measured on the union set:

$$\Delta_i^t = \max\Big\{0,\ \mathrm{rank}_{Q^t}(\mathbf{x}_i^t) - \mathrm{rank}_{Q^t}(\mathbf{u}_i^t)\Big\}, \qquad S_k^t = \sum_{i=1}^{N} \mathbb{I}[k_i^t = k]\, \Delta_i^t. \tag{16}$$

Probability update. Let ρ_k^t denote a smoothed estimate of strategy quality. A stable update is obtained by exponential smoothing followed by normalization with a positivity floor $\varepsilon = 0.01$:

$$\rho_k^{t+1} = (1-\alpha)\rho_k^t + \alpha\hat{\rho}_k^t, \quad \alpha = 0.1, \tag{17}$$

$$p_k^{t+1} = \frac{\rho_k^{t+1} + \varepsilon}{\sum_{j=1}^{K}(\rho_j^{t+1} + \varepsilon)}. \tag{18}$$

This update both amplifies recently effective strategies and prevents premature exclusion, which is particularly important in multi-objective landscapes where the locally useful operator can shift across generations and across regions of the Pareto set.

3.4 Adaptive Control of DE Parameters

Operator adaptation alone does not determine the effective search dynamics: the scaling factor F regulates step magnitudes, while the crossover rate CR controls how strongly the mutant perturbs the target. Because the utility of (F, CR) depends not only on the problem structure but also on the chosen mutation scheme, AEMO-DEFS attaches a strategy-conditional memory state to each operator. Specifically, each strategy k maintains

$$\Theta_k^t = \big(\mu_{F,k}^t, \gamma_{F,k}^t, \mu_{CR,k}^t, \sigma_{CR,k}^t\big), \tag{19}$$

from which parameter values are sampled, and which is updated using only the parameter values associated with successful offspring generated by that strategy. The memories are initialized to $\mu_{F,k}^0 = 0.5$, $\gamma_{F,k}^0 = 0.1$, $\mu_{CR,k}^0 = 0.5$, and $\sigma_{CR,k}^0 = 0.1$ for all k.

Sampling with truncation. For an offspring produced under strategy k:

$$F_i^t \sim \mathrm{Cauchy}\big(\mu_{F,k}^t, \gamma_{F,k}^t\big), \quad F_i^t \leftarrow \min\{1, \max\{\delta, F_i^t\}\}, \tag{20}$$

$$CR_i^t \sim \mathcal{N}\big(\mu_{CR,k}^t, \sigma_{CR,k}^t\big), \quad CR_i^t \leftarrow \min\{1, \max\{0, CR_i^t\}\}, \tag{21}$$

where $\delta = 0.1$ prevents degenerate mutation steps. The heavy-tailed Cauchy draw for F preserves a small but persistent probability of large exploratory jumps, whereas the Gaussian draw for CR concentrates around a learned mean while allowing controlled variability.

Successful samples and weights. Define, for each strategy k, the successful parameter multisets

$$\mathcal{F}_k^t = \{F_i^t : k_i^t = k,\ s_i^t = 1\}, \qquad C_k^t = \{CR_i^t : k_i^t = k,\ s_i^t = 1\}. \tag{22}$$

To bias updates toward offspring with greater measured contribution, we associate nonnegative weights. Using the contribution signal Δ_i^t (or $\Delta_i^t \equiv 1$ under pure survival credit),

$$\tilde{w}_i^t = \begin{cases} \Delta_i^t, & \text{if } s_i^t = 1, \\ 0, & \text{otherwise,} \end{cases} \qquad w_i^t = \frac{\tilde{w}_i^t}{\sum_{j:\, k_j^t = k} \tilde{w}_j^t + \eta}, \tag{23}$$

with $\eta = 10^{-6}$ ensuring well-defined normalization. If a strategy produces no successful offspring in a generation, its memory state is left unchanged, which prevents spurious drift.

Memory updates. A robust update for F uses a weighted Lehmer mean, which accentuates larger successful F values when they are systematically beneficial:

$$\widehat{\mu}_{F,k}^t = \frac{\sum_{i:\, k_i^t = k,\, s_i^t = 1} w_i^t (F_i^t)^2}{\sum_{i:\, k_i^t = k,\, s_i^t = 1} w_i^t F_i^t}. \tag{24}$$

For CR, the weighted arithmetic mean is typically adequate:

$$\widehat{\mu}_{CR,k}^t = \sum_{i:\, k_i^t = k,\, s_i^t = 1} w_i^t CR_i^t. \tag{25}$$

The location parameters are then updated by exponential smoothing with learning rate $c = 0.1$:

$$\mu_{F,k}^{t+1} = (1-c)\mu_{F,k}^t + c\,\widehat{\mu}_{F,k}^t, \tag{26}$$

$$\mu_{CR,k}^{t+1} = (1-c)\mu_{CR,k}^t + c\,\widehat{\mu}_{CR,k}^t. \tag{27}$$

To adapt dispersion, the normal standard deviation is updated via a weighted variance estimate,

$$\widehat{\sigma}_{CR,k}^t = \sqrt{\sum_{i:\, k_i^t = k,\, s_i^t = 1} w_i^t \big(CR_i^t - \widehat{\mu}_{CR,k}^t\big)^2}, \tag{28}$$

followed by smoothing:

$$\sigma_{CR,k}^{t+1} = (1-c)\sigma_{CR,k}^t + c\,\widehat{\sigma}_{CR,k}^t. \tag{29}$$

For the Cauchy scale $\gamma_{F,k}^t$, a robust alternative is a weighted median absolute deviation (wMAD) around $\widehat{\mu}_{F,k}^t$,

$$\widehat{\gamma}_{F,k}^t = \kappa\, \mathrm{wMAD}\Big(\{F_i^t\}, \widehat{\mu}_{F,k}^t, \{w_i^t\}\Big), \tag{30}$$

where $\kappa = 1.4826$ calibrates the scale to match the standard deviation of a Gaussian, and wMAD denotes the weighted median of $\{|F_i^t - \widehat{\mu}_{F,k}^t|\}$ under weights $\{w_i^t\}$, followed by smoothing:

$$\gamma_{F,k}^{t+1} = (1-c)\gamma_{F,k}^t + c\,\widehat{\gamma}_{F,k}^t. \tag{31}$$

Coupled adaptation dynamics. Because (F, CR) are sampled conditional on the strategy index, operator adaptation and parameter adaptation form a coupled learning process. The probabilities $\mathbf{p}^t$ regulate the frequency with which each memory state Θ_k^t is queried, while the memory states evolve only when their associated operator produces offspring that survive environmental selection or yield measurable rank-based contribution. The resulting mechanism is closely aligned with the intuition of reward-weighted reinforcement: operator families and parameter regimes that demonstrably support progress toward a well-distributed approximation of the Pareto set are amplified, but the positivity floor in p_k^t and the non-negligible sampling variance preserve the algorithm's capacity to reallocate search effort when the optimization dynamics shift.

3.5 External Archive Management

To guarantee elitism and preserve the best solutions discovered during evolution, AEMO-DEFS maintains an external archive A that stores all nondominated solutions identified throughout the optimization process. At each generation, the archive is updated by adding any newly found nondominated solutions from P^{t+1}:

$$A^{t+1} \leftarrow \text{Nondominated}\big(A^t \cup P^{t+1}\big). \tag{32}$$

If the archive size exceeds a predefined capacity $|A|_{\max}$, solutions with the smallest crowding distance are iteratively removed until $|A^{t+1}| \le |A|_{\max}$, maintaining a diverse representative set covering the Pareto front.

The complete algorithmic procedure is summarized in Algorithm 1.

4 Experimental Setup

This section describes the experimental configuration used to evaluate the AEMO-DEFS framework, including the benchmark datasets, baseline algorithms for comparison, parameter settings, and performance evaluation metrics.

To assess the performance of AEMO-DEFS, five gene expression datasets representing different cancer types and classification challenges were selected from the literature. These datasets vary in terms of dimensionality (number of features), sample size, and number of classes, providing a comprehensive testbed for evaluating feature selection algorithms. Table 1 presents the characteristics of each dataset.

Table 1: Characteristics of the benchmark gene expression datasets (MC: multi-class, BC: binary-class)

Dataset	Instances	Features	Classes	Class Distribution
11-Tumors	174	12,533	11 (MC)	[27, 26, 25, 23, 14, 14, 12, 11, 8, 7, 7]
Brain-Tumor	50	10,367	4 (MC)	[14, 6, 14, 16]
Lung Cancer	203	12,600	5 (MC)	[139, 6, 21, 20, 17]
MLL	72	12,582	3 (MC)	[24, 24, 24]
Prostate Tumor	102	10,509	2 (BC)	[50, 52]

The 11-Tumors dataset contains gene expression profiles from 174 samples covering 11 different tumor types, representing a challenging multi-class classification problem with a large feature space relative to sample size. The Brain-Tumor dataset comprises expression data from 50 samples across 4 tumor classes, with a particularly severe sample-to-feature ratio. The Lung Cancer dataset includes 203 samples from 5 types of lung cancer and normal tissues, characterized by a high-dimensional feature space and moderate class imbalance. The MLL (Mixed-Lineage Leukemia) dataset contains 72 samples from 3 leukemia subtypes with balanced class distribution. The Prostate Tumor dataset represents a binary classification problem with 102 samples distinguishing between prostate tumor and normal tissue.

For all datasets, the features (genes) were preprocessed following standard practices for gene expression data analysis. Missing values were imputed using the k-nearest neighbor algorithm with $k = 5$, and all features were standardized to zero mean and unit variance to ensure comparable scales across features. No feature preprocessing specific to AEMO-DEFS was performed, maintaining consistency with the wrapper-based paradigm where the classifier handles the raw features.

Algorithm 1 AEMO-DEFS Algorithm

Input: Feature dimension D, population size N, max generations $T_{\max}$, archive size $|A|_{\max}$
Output: Archive A of Pareto-optimal feature subsets
Initialize: $P = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ with $\mathbf{x}_i \sim \mathcal{U}[0,1]^D$
Initialize $\mathbf{p}^0 = (1/K, \ldots, 1/K)$ and Θ_k^0 for all k
Initialize $A \leftarrow \emptyset$
for each $\mathbf{x}_i \in P$ **do**
  $\mathbf{b}_i \leftarrow \mathbb{I}[\mathbf{x}_i > 0.5]$
  Evaluate $f_1(\mathbf{b}_i), f_2(\mathbf{b}_i)$ via 5-fold CV with k-NN
end for
Perform nondominated sorting and crowding distance on P
$A \leftarrow \text{Nondominated}(P)$
for $t = 1$ to $T_{\max}$ **do**
  for $i = 1$ to N **do**
    Sample strategy $k_i \sim \text{Cat}(\mathbf{p}^{t-1})$
    Sample F_i, CR_i from $\Theta_{k_i}^{t-1}$ using Eqs. (20)–(21)
    Generate $\mathbf{v}_i$ via Eq. (6), (7), or (8)
    Generate $\mathbf{u}_i$ via binomial crossover and repair
    $\mathbf{b}_i' \leftarrow \mathbb{I}[\mathbf{u}_i > 0.5]$
    Evaluate $f_1(\mathbf{b}_i'), f_2(\mathbf{b}_i')$
  end for
  $U^t \leftarrow \{\mathbf{u}_1, \ldots, \mathbf{u}_N\}$
  $Q^t \leftarrow P \cup U^t$
  Perform nondominated sorting and crowding distance on Q^t
  $P \leftarrow \text{Select}_N(Q^t)$ {NSGA-II selection}
  $A \leftarrow \text{Nondominated}(A \cup P)$
  if $|A| > |A|_{\max}$ **then**
    Prune A by removing solutions with smallest crowding distance
  end if
  Compute success indicators $\{s_i\}$ and contribution scores $\{\Delta_i\}$
  Update $\mathbf{p}^t$ via Eqs. (17)–(18)
  Update $\{\Theta_k^t\}$ using successful parameter samples
end for
Return: Archive A

To provide a comprehensive evaluation, AEMO-DEFS was compared against three established multi-objective optimization algorithms that have been widely used for feature selection tasks:

MOPSO (Multi-Objective Particle Swarm Optimization) [4]: A particle swarm optimization algorithm adapted for multi-objective optimization, using an external archive to store nondominated solutions and selecting leaders from the archive based on crowding distance.

NSGA-III (Non-dominated Sorting Genetic Algorithm III) [5]: A reference-point-based NSGA-II variant designed for many-objective optimization, using a predefined set of reference points to maintain diversity in high-dimensional objective spaces.

MOEA/D (Multi-Objective Evolutionary Algorithm based on Decomposition) [13]: A decomposition-based algorithm that transforms the multi-objective problem into a set of scalar optimization subproblems.

These baseline algorithms were selected to represent different paradigms within multi-objective evolutionary computation: swarm intelligence (MOPSO), domination-based selection with reference points (NSGA-III), and decomposition-based optimization (MOEA/D). All algorithms were implemented with standard settings and were configured to use the same fitness evaluation procedure (5-NN with 5-fold cross-validation) to ensure a fair comparison.

To ensure a fair and meaningful comparison, consistent parameter settings were used across all algorithms where applicable. The population size of 30 was selected as a balance between exploration capability and computational cost, considering the expensive fitness evaluation required for wrapper-based feature selection. The maximum of 200 generations was chosen to allow sufficient optimization time while maintaining manageable computational requirements for the 30 independent runs performed for each algorithm-dataset combination.

The performance of AEMO-DEFS and the baseline algorithms was assessed using multiple evaluation metrics that capture different aspects of solution quality in multi-objective optimization:

Inverted Generational Distance (IGD): The IGD metric measures both the convergence and diversity of an approximated Pareto front by computing the average distance from each point on the reference front to its nearest point in the approximated front [15]. A lower IGD value indicates better performance.

Epsilon Indicator (ϵ-indicator): The additive ϵ-indicator measures the minimum value by which the approximated front must be translated in objective space to dominate the reference front [15]. This metric captures the worst-case quality of solutions in the approximated front, with lower values indicating better performance.

Classification Accuracy: The average classification accuracy achieved by solutions in the approximated Pareto front, evaluated using 5-fold cross-validation with a 5-NN classifier. Higher accuracy indicates better predictive performance.

Number of Selected Features: The average cardinality of feature subsets in the approximated Pareto front, representing the degree of dimensionality reduction achieved. Lower values indicate more parsimonious feature subsets.

The true Pareto front is unknown for these real-world feature selection problems. To address this challenge, a reference front was constructed by aggregating all nondominated solutions found across all runs and all algorithms, then filtering to retain only globally nondominated points.

5 Results and Analysis

This section presents the experimental results comparing AEMO-DEFS against the baseline algorithms across the five benchmark datasets. The analysis examines multiple dimensions of performance, including Pareto front quality metrics, classification accuracy, and feature subset cardinality.

Table 2 presents the Inverted Generational Distance (IGD) results for all algorithms across the five datasets. Lower IGD values indicate better performance, signifying both closer approximation to the reference front and better distribution of solutions.

Table 2: Inverted Generational Distance (IGD) comparison (mean ± std); lower is better

Dataset	AEMO-DEFS	MOPSO	NSGA-III	MOEA/D
11-Tumors	**0.0214 ± 0.0098**	0.0592 ± 0.0200	0.0522 ± 0.0139	0.0297 ± 0.0132
Brain-Tumor	**0.0157 ± 0.0215**	0.0381 ± 0.0314	0.0165 ± 0.0247	0.0221 ± 0.0259
Lung Cancer	**0.0092 ± 0.0047**	0.0199 ± 0.0070	0.0175 ± 0.0065	0.0106 ± 0.0033
MLL	**0.0087 ± 0.0193**	0.0140 ± 0.0262	0.0164 ± 0.0280	0.0128 ± 0.0267
Prostate Tumor	**0.0083 ± 0.0176**	0.0140 ± 0.0262	0.0164 ± 0.0280	0.0128 ± 0.0267

The results demonstrate that AEMO-DEFS achieves the lowest mean IGD on all five datasets, with particularly pronounced improvements on the 11-Tumors dataset where the mean IGD of 0.0214 represents a reduction of approximately 28% compared to MOEA/D (0.0297) and over 64% compared to MOPSO (0.0592). This strong performance indicates that AEMO-DEFS produces Pareto front approximations that are both closer to the reference front and better distributed across the range of achievable trade-offs.

Table 3 presents the ϵ-indicator results, which provide additional evidence of AEMO-DEFS's superior Pareto front quality, particularly in terms of dominance-based metrics.

Table 3: Epsilon Indicator (ϵ-indicator) comparison (mean ± std); lower is better

Dataset	AEMO-DEFS	MOPSO	NSGA-III	MOEA/D
11-Tumors	**0.1876 ± 0.0298**	0.2549 ± 0.0322	0.2480 ± 0.0275	0.2275 ± 0.0401
Brain-Tumor	**0.0643 ± 0.0218**	0.0907 ± 0.0342	0.0667 ± 0.0269	0.0759 ± 0.0272
Lung Cancer	**0.0412 ± 0.0118**	0.0475 ± 0.0120	0.0483 ± 0.0130	0.0483 ± 0.0160
MLL	**0.0098 ± 0.0215**	0.0141 ± 0.0261	0.0165 ± 0.0279	0.0130 ± 0.0266
Prostate Tumor	**0.0042 ± 0.0153**	0.0154 ± 0.0234	0.0163 ± 0.0207	0.0089 ± 0.0164

The ϵ-indicator results reinforce the findings from the IGD analysis, with AEMO-DEFS achieving the best (lowest) values on all datasets. The improvements are particularly notable on the 11-Tumors dataset (0.1876 vs. 0.2275 for MOEA/D) and the Prostate Tumor dataset (0.0042 vs. 0.0089 for MOEA/D).

Table 4 and 5 present the classification accuracy and number of selected features, respectively. AEMO-DEFS achieves the best classification accuracy on all five datasets while simultaneously selecting the smallest number of features.

The results demonstrate that AEMO-DEFS achieves the best classification accuracy on all five datasets while simultaneously selecting the smallest number of features. On the 11-Tumors dataset,

Table 4: Classification accuracy (higher is better)

Dataset	AEMO-DEFS	MOPSO	NSGA-III	MOEA/D
11-Tumors	**0.9358**	0.9103	0.9011	0.9154
Brain-Tumor	**0.8892**	0.8769	0.8538	0.8615
Lung Cancer	**0.9683**	0.9606	0.9508	0.9557
MLL	**0.9724**	0.9630	0.9506	0.9630
Prostate Tumor	**0.9529**	0.9314	0.9216	0.9363

Table 5: Number of selected features (lower is better)

Dataset	AEMO-DEFS	MOPSO	NSGA-III	MOEA/D
11-Tumors	**7.83**	12.37	15.22	10.18
Brain-Tumor	**5.21**	8.21	10.45	7.32
Lung Cancer	**5.94**	9.84	12.19	8.53
MLL	**4.35**	7.35	9.62	6.41
Prostate Tumor	**4.78**	7.68	9.94	6.83

AEMO-DEFS selects an average of 7.83 features compared to 10.18 for MOEA/D (the next-best baseline), representing a reduction of approximately 23%. This reduction is even more pronounced on the MLL dataset, where AEMO-DEFS selects only 4.35 features on average compared to 6.41 for MOEA/D, a 32% reduction.

The superior performance of AEMO-DEFS can be attributed to its core adaptive mechanisms working in concert to address the fundamental challenges of high-dimensional feature selection. First, the algorithm's adaptive ensemble of DE operators dynamically selects among multiple mutation and crossover strategies based on their recent success rates, thereby intrinsically balancing exploration and exploitation without requiring manual strategy selection. Second, AEMO-DEFS employs adaptive control of the scaling factor F and crossover rate CR via memory-based distributions, updating these distributions according to the parameters that produced successful offspring in previous generations. This self-tuning mechanism enhances robustness by obviating the need for user-specified parameter settings while enabling the algorithm to discover parameter values that are effective for the specific problem landscape.

Despite the strong performance demonstrated across the benchmark datasets, several limitations of the current study should be acknowledged. First, the empirical evaluation is restricted to five cancer microarray datasets, and the performance of AEMO-DEFS on other omics modalities remains to be evaluated. Second, the computational overhead incurred by repeated nondominated sorting and distance calculations may become prohibitive when scaling to ultra-high-dimensional datasets. Third, the evaluation focused on a specific classifier (k-NN) and cross-validation procedure.

These limitations suggest several promising directions for future research, including investigating the performance of AEMO-DEFS on other types of high-dimensional data, addressing scalability concerns through parallelization or surrogate modeling, and conducting ablation studies to quantify the individual contributions of each adaptive mechanism.

6 Conclusion

This paper introduced the Adaptive Ensemble Multi-Objective Differential Evolution for Feature Selection (AEMO-DEFS), a multi-objective metaheuristic framework tailored to high-dimensional feature selection under the concurrent goals of maximizing classification performance and minimizing subset cardinality. The distinctive contribution of AEMO-DEFS lies in two coupled adaptive mechanisms: first, an ensemble of differential evolution mutation–crossover strategies whose sampling probabilities are updated online according to their observed contribution to generating competitive offspring during the evolutionary process; second, a memory-driven parameter control scheme that adaptively samples and updates the scaling factor F and crossover rate CR based on historically successful parameter realizations. By operationalizing operator choice and parameterization as data-driven control variables, the method aims to sustain an effective exploration–exploitation balance across heterogeneous landscapes and across different stages of the search.

Empirical studies on five cancer microarray datasets provided consistent evidence of improved search quality and solution utility. Across all datasets, AEMO-DEFS attained the lowest Inverted Generational Distance and favorable ϵ-indicator values, supporting the claim that the algorithm yields Pareto sets that are both close to the reference front and well distributed. Importantly, these gains translated into application-level outcomes: AEMO-DEFS simultaneously achieved the highest classification accuracy while selecting the fewest features, suggesting that performance improvements arise from more effective multi-objective optimization rather than from privileging one objective at the expense of the other. Future work will investigate richer credit-assignment signals for operator adaptation (e.g., hypervolume contribution), scalability strategies for larger feature spaces (including surrogate-assisted evaluation and parallelization), and stronger domain validation through stability analysis and biological interpretation of selected biomarkers, complemented by cross-study generalization tests and additional classifiers to assess robustness.

References

[1] Behrouz Ahadzadeh, Moloud Abdar, Fatemeh Safara, Abbas Khosravi, M. B. Menhaj, and P. N. Suganthan. 2023. SFE: A Simple, Fast, and Efficient Feature Selection Algorithm for High-Dimensional Data. *IEEE Transactions on Evolutionary Computation* 27, 6 (2023), 1896–1911.

[2] Qasem Al-Tashi, S. J. Abdulkadir, H. M. Rais, Seyedali Mirjalili, and H. Alhussian. 2020. Approaches to Multi-Objective Feature Selection: A Systematic Literature Review. *IEEE Access* 8 (2020), 125076–125096.

[3] X. Bing, M. Zhang, W. N. Browne, and Y. Xin. 2016. A Survey on Evolutionary Computation Approaches to Feature Selection. *IEEE Transactions on Evolutionary Computation* 20, 4 (2016), 606–626.

[4] C. A. C. Coello Coello, G. T. Pulido, and M. S. Lechuga. 2004. Handling Multiple Objectives with Particle Swarm Optimization. *IEEE Transactions on Evolutionary Computation* 8, 3 (2004), 256–279.

[5] Kalyanmoy Deb and Himanshu Jain. 2014. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems with Box Constraints. *IEEE Transactions on Evolutionary Computation* 18, 4 (2014), 577–601.

[6] P. Dhal and C. Azad. 2022. A Comprehensive Survey on Feature Selection in the Various Fields of Machine Learning. *Applied Intelligence* 52, 4 (2022), 4543–4581.

[7] F. Han, W. T. Chen, Q. H. Ling, and H. Han. 2021. Multi-Objective Particle Swarm Optimization with Adaptive Strategies for Feature Selection. *Swarm and Evolutionary Computation* 62 (2021), 100847.

[8] S. S. Khan, M. S. Kawoosa, B. Bannerjee, S. C. Chauhan, and S. Khan. 2024. Revolutionizing Feature Selection: A Breakthrough Approach for Enhanced Accuracy

and Reduced Dimensions, with Implications for Early Medical Diagnostics. *Journal of Medical Systems* 48, 2 (2024), 1–15.

[9] X. Li. 2003. A Non-Dominated Sorting Particle Swarm Optimizer for Multiobjective Optimization. In *Genetic and Evolutionary Computation Conference (GECCO)*. Springer, 37–48.

[10] M. Nssibi, G. Manita, and O. Korbaa. 2023. Advances in Nature-Inspired Metaheuristic Optimization for Feature Selection Problem: A Comprehensive Survey. *Computer Science Review* 49 (2023), 100559.

[11] D. Savic. 2002. Single-Objective vs. Multiobjective Optimisation for Integrated Decision Support. Unspecified venue in the source paper.

[12] B. Venkatesh and J. Anuradha. 2019. A Review of Feature Selection and Its Methods. *Cybernetics and Information Technologies* 19, 1 (2019), 3–26.

[13] Q. Zhang and H. Li. 2007. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Transactions on Evolutionary Computation* 11, 6 (2007), 712–731.

[14] A. Zhou, B. Y. Qu, H. Li, S. Z. Zhao, P. N. Suganthan, and Q. Zhang. 2011. Multiobjective Evolutionary Algorithms: A Survey of the State of the Art. *Swarm and Evolutionary Computation* 1, 1 (2011), 32–49.

[15] Eckart Zitzler, Lothar Thiele, Marco Laumanns, Carlos M. Fonseca, and V. Grunert da Fonseca. 2003. Performance Assessment of Multiobjective Optimizers: An Analysis and Review. *IEEE Transactions on Evolutionary Computation* 7, 2 (2003), 117–132.