НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
ноябрь–декабрь 2025 Том 25 № 6 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
November–December 2025 Vol. 25 No 6 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2025-25-6-1185-1196

# DeFs-CBDE: Clustering-guided binary mutation in multi-objective differential evolution for microarray gene selection

**Mohamed Djellal Serandi[1]✉, Fatma Boufera[2], Amina Houari[3], Farid Flitti[4]**

[1,2,3] University Mustapha Stambouli, LISYS laboratory, Mascara, 29000, Algeria
[4] Higher Colleges of Technology in Dubai, Dubai, 500001, United Arab Emirates

[1] mohamed.djellalserandi@univ-mascara.dz✉, https://orcid.org/0009-0009-5775-7956
[2] fboufera@univ-mascara.dz, https://orcid.org/0000-0002-5733-586X
[3] amina.houari@univ-mascara.dz, https://orcid.org/0000-0002-3628-7483
[4] fflitti@hct.ac.ae, https://orcid.org/0000-0002-2480-2580

**Abstract**
DNA microarray technology produces high-dimensional gene expression data, where many genes are irrelevant to disease. Effective feature selection is thus essential to mitigate the curse of dimensionality and enhance classification performance. This study introduces a multi-objective feature selection approach employing a Clustering-Based Binary Differential Evolution (CBDE) mutation to identify a compact set of disease-relevant genes. The proposed DeFs-CBDE algorithm was assessed on four gene expression datasets, i.e. brain, breast, lung, and central nervos system cancer by selecting informative feature subsets and evaluating them using five state-of-the-art classifiers, i.e., Support Vector Machine (SVM), Naive Bayes, K-Nearest Neighbors, Decision Tree (DT), and Random Forest. The DeFs-CBDE method achieved of 100 % accuracy on the brain dataset with three classifiers. On the lung dataset, DeFs-CBDE reached 97.56 % accuracy with SVM and DT. For the breast dataset, DeFs-CBDE attained 93.33 % accuracy very close to the highest score of 93.81 % accuracy. The CNS dataset proved the most challenging, where it achieved 91.67 % accuracy with SVM. Across all datasets, DeFs-CBDE consistently achieved high classification performance.

**Keywords**
microarray data, feature selection, optimization, differential evolution, mutation, multi-objective

**For citation:** Djellal Serandi M., Boufera F., Houari A., Flitti F. DeFs-CBDE: Clustering-guided binary mutation in multi-objective differential evolution for microarray gene selection. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 6, pp. 1185–1196. doi: 10.17586/2226-1494-2025-25-6-1185-1196

УДК 004.93'14

# DeFs-CBDE: бинарная мутация, управляемая кластеризацией, в многокритериальной дифференциальной эволюции для отбора генов с помощью микрочипов

**Мохамед Джеллаль Серанди[1]✉, Фатма Буфера[2], Амина Хуари[3], Фарид Флитти[4]**

[1,2,3] Университет Мустафы Стамбули, Маскара, 29000, Алжир
[4] Высший колледж технологии, колледж в Дубае, Дубай, 500001, Объединенные Арабские Эмираты

[1] mohamed.djellalserandi@univ-mascara.dz✉, https://orcid.org/0009-0009-5775-7956
[2] fboufera@univ-mascara.dz, https://orcid.org/0000-0002-5733-586X
[3] amina.houari@univ-mascara.dz, https://orcid.org/0000-0002-3628-7483
[4] fflitti@hct.ac.ae, https://orcid.org/0000-0002-2480-2580

**Аннотация**
Технология ДНК-микрочипов позволяет получать высокоразмерные данные об экспрессии генов, многие из которых не имеют отношения к заболеванию. Эффективный отбор признаков, таким образом, необходим для смягчения «проклятия размерности» и повышения эффективности классификации. В данном исследовании

представлен многокритериальный подход к отбору признаков с использованием мутации на основе кластеризации для идентификации компактного набора генов, связанных с заболеванием. Предложенный алгоритм DeFs-CBDE был оценен на четырех наборах данных об экспрессии генов: рак головного мозга, молочной железы, легких и центральной нервной системы, путем отбора информативных подмножеств признаков и их оценки с использованием пяти современных классификаторов, а именно: метода опорных векторов, наивного байесовского алгоритма, метода K-ближайших соседей, дерева решений и случайного леса. Метод DeFs-CBDE достиг 100 % точности на наборе данных о мозге с тремя классификаторами. В наборе данных по легким DeFs-CBDE достиг точности 97,56 % с использованием метода опорных векторов и дерева решений. В наборе данных по молочной железе DeFs-CBDE достиг точности 93,33 %, что очень близко к максимальному результату в 93,81 %. Набор данных по центральной нервной системе оказался самым сложным, где точность составила 91,67 % с использованием. Во всех наборах данных DeFs-CBDE стабильно демонстрировал высокую эффективность классификации.

**Ключевые слова**
данные микрочипов, выбор признаков, оптимизация, дифференциальная эволюция, мутация, многокритериальный анализ

**Ссылка для цитирования:** Джеллаль Серанди М., Буфера Ф., Хуари А., Флитти Ф. DeFs-CBDE: бинарная мутация, управляемая кластеризацией, в многокритериальной дифференциальной эволюции для отбора генов с помощью микрочипов // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 6. С. 1185–1196 (на англ. яз.). doi: 10.17586/2226-1494-2025-25-6-1185-1196

## Introduction

A challenge faced in the explosion of microarray gene expression data is its high dimensionality (thousands of genes) combined with a limited sample number (tens or hundreds), leading to the well-known "curse of dimensionality" [1]: As the dimensionality increases, the number of samples required for inference grows rapidly. This problem is addressed by Feature Selection (FS) which aims at selecting a pertinent subset of genes to enhance model performance and interpretability. The development of high-throughput genomics, such as microarrays, has indeed transformed our understanding of diseases such as breast cancer. However, although microarray data provide a wealth of information, appropriate feature selection is necessary to identify genes that significantly influence disease classification.

Choosing the optimal subset of the original features is NP-Hard and requires a lot of time and resources. As a result, selecting an appropriate search algorithm is crucial for feature selection methods. A feature selection algorithm should be assessed from the perspective of effectiveness and efficiency. The effectiveness of a feature selection method is determined by the time it takes to find the final set of [2].

In general, it is possible to represent feature selection as a multi-objective [3] combinatorial optimization problem with two key objectives: minimizing the number of selected features and the Mean Squared Residual (MSR) [4]. Multi-objective evolutionary algorithms enable the search for multiple solutions located on the Pareto front in a single run. These methods are frequently used to solve feature selection problems [5].

One of the most widely used evolutionary algorithms for FS problems is Differential Evolution (DE) which is renowned for its ease of use and effectiveness. The application of DE to the multi-objective scenario has not received much attention, and most current research focuses on maximizing classification accuracy, or the single-objective case [5]. A multi-objective FS approach was initially developed using multi-objective DE and applied to the feature selection task. The results highlight the effectiveness of DE in addressing feature selection challenges.

This research suggests a novel multi-objective feature selection method based on DE for microarray data to improve the model performance and interpretability. The MSR is optimized simultaneously with limiting the number of selected genes. Additionally, a novel mutation operator, named Clustering-Based Binary Differential Evolution (CBDE) is introduced to adapt the algorithm for better optimization in feature selection.

The highlights of the DeFs-CBDE method are:

— A new FS approach for gene expression datasets is proposed.
— A novel CBDE mutation operator is proposed.
— The DeFs-CBDE is extensively validated on four datasets (Brain, Breast, Lung, and Central Nervous System (CNS)).
— The performance obtained with the proposed method is compared to the performance of the FS technique in [6] and with other existing works [7–9].

## Related work

Many researchers have focused on FS to improve classification performance on gene expression data. Among the most commonly used approaches are Particle Swarm Optimization (PSO), DE, and hybrid methods, which have shown promising results across various cancer-related datasets. In the following, we present related studies based on these techniques.

Zhang et al. [5] proposed a Multi-Objective Feature Selection method called the Binary Differential Evolution (BDE) with self-learning (MOFS-BDE). The method introduces novel operators, i.e., the probability difference based binary mutation, the One-bit Purifying Search and the efficient non-dominated sorting with crowding distance. Experimental results show that the MOFS-BDE achieves balance between global exploration and local exploitation, and is highly competitive compared to state-of-the-art genetic algorithm, particle swarm, differential evolution, and artificial bee colony-based feature selection algorithms.

Ali et al. [6] proposed a feature selection method based on a hybrid filter-genetic algorithm to enhance the classification of cancer in high dimensional microarray datasets. The method first discards irrelevant features using filter-based techniques including Information Gain (IG), Information Gain Ratio (IGR), and Chi-Squared (CS). Then, a Genetic Algorithm (GA) is used to further optimize the selected features for cancer classification. The approach was tested on four microarray datasets (Brain, Breast, Lung, and CNS) and compared to six other FS methods. The experimental results showed that the GA in the proposed method was able to reduce further 50 % of irrelevant features from the subset selected by filtering.

Ahadzadeh et al. [7] introduced an algorithm called Simple, Fast, and Efficient FS to search a vast area of high-dimensional data, eliminating irrelevant characteristics from the search space in the first phase. Next, the smaller search space is explored using the PSO method to identify significant and pertinent features.

Prajapati et al. [8] proposed a feature selection approach based on DE to reduce the dimensionality of microarray datasets. The method improves classification accuracy by combining DE with Random Forest (RF), Decision Tree (DT), and LR (Logistic Regression), and it outperforms models without feature selection.

Hamla and Ghanem [9] proposed a hybrid feature selection method that combines Fisher score filtering and Support Vector Machine Recursive Feature Elimination (SVM-RFE) to improve gene selection in microarray data.

Paul et al. [10] introduced a multi-objective PSO-based feature selection method designed for multi-label classification with online arrival of features. The approach automatically identifies the most relevant subset of features through a structured three-phase filtering process. In the first phase, a multi-objective PSO selects candidate features from each incoming group. The second phase eliminates redundancy by comparing newly selected features with those previously chosen. In the final phase, the method prunes previously selected features that become insignificant due to the arrival of new data. This adaptive and evolutionary framework effectively maintains a compact and relevant feature set, improving multi-label classification.

Han et al. [11] suggested a Multi-Objective PSO with Adaptive Strategies (MOPSO-ASFS) for FS. The approach uses an adaptive penalty of Penalty Boundary Interaction mechanism to enhance the selection pressures of the archive. An adaptive leading particle selection based on feature information combines the opposite mutation and the feature frequencies to improve the selection pressure of each particle. The experimental results from 14 UCI datasets and 6 gene expression datasets demonstrated the effectiveness of MOPSO-ASFS in high-dimensional space compared to classical multi-objective FS algorithms, the multi-dimensional regions.

Dashtban et al. [12] proposed a bioinspired multi-objective algorithm that improves the Bat Algorithm with refined formulations, effective multi-objective operators, and social learning-based random walks. Fisher criterion is used as initial filtering to reduce irrelevant genes. Applied to three cancer microarray datasets, results revealed new combinations of important biomarkers associated with the development of three challenging cancers. Also, the algorithm showed strong classification performance.

Ghosh et al. [13] introduced a two-phase feature selection framework designed for microarray data analysis. In the first phase, they applied an ensemble filtering strategy that combined the top-ranked features identified by symmetrical uncertainty, chi-square, and ReliefF methods using union and intersection operations. The refined feature subsets were then further optimized with a genetic algorithm to select the most informative features.

According to this state-of-the-art, GA, PSO, and DE have demonstrated strong potential in addressing the challenges of feature selection, particularly in high-dimensional and noisy data environments. Among them, GA and DE are widely recognized for their effectiveness. DE, in particular, is distinguished by its simplicity, adaptability, and its ability to incorporate advanced mutation strategies such as CBDE.

## Proposed method

In this paper, an improved multi-objective differential evolution approach for FS (DeFs-CBDE) method, based on a CBDE mutation operator, is presented. This novel method uses a Non-dominated Sorting Genetic Algorithm (NSGA-II).

**Differential evolution**

Differential evolution is a technique that has been effectively used in many scientific and technical domains in recent years to solve continuous optimization challenges. DE was first proposed by [14] as a straightforward but effective heuristic for global optimization in continuous spaces. The DE algorithm consists of five main steps: initialization, mutation, crossover, selection, and termination [15].

**Architecture of the proposed method**

As shown in Fig. 1, the data set is first preprocessed and then divided into training and testing sets. The training set is then used to apply the proposed DeFs-CBDE algorithm resulting in a selected feature subset (New subset training with FS). The same features, without applying the DeFs-CBDE algorithm, are extracted to form the corresponding test set (New subset testing with FS). These new generated training and testing subsets are then fed into classifiers to evaluate the performance of selected features.

**DeFs-CBDE algorithm description**

By incorporating DE into the feature selection procedure, the suggested method described in Algorithm 1 guides the search for the best feature subsets.

**Algorithm 1.** The pseudocode of DeFs-CBDE approach.

**Input:** A partial training matrix of size $n \times m$ ($n$ genes and $m$ samples), Initial Population size $p$, Initial crossover probability $C$, Initial factor $F$;

**Output:** A matrix of size $s \times m$ (the dataset with $s$ optimally selected features);

**Begin**

Initialize population $P$ by generating $p$ solutions, each gene random in [0, 1];

Compute fitness value ($f1, f2$) for each solution in $P$;

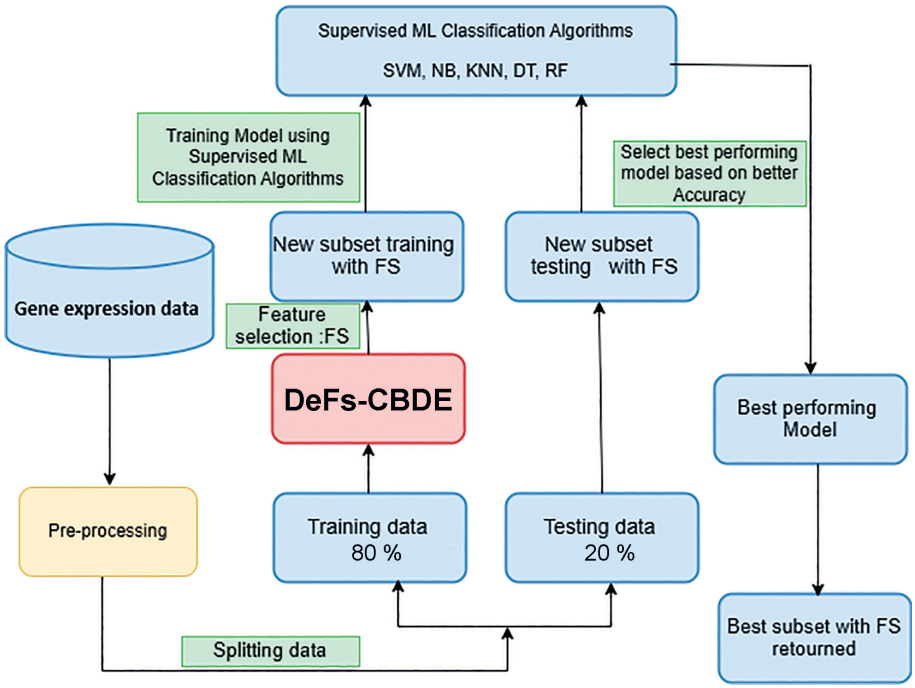


*Fig. 1*. Illustration of DeFs-CBDE scheme

Initialize array $C$ and array $F$ for all individuals in $P$ using initial $C$ and $F$;
**While** termination criteria not met **do**
**For** each solution $i$ in $P$ **do**
Select 3 randomly chosen solutions ($x$, $y$, $z$) from $P$ ranked from best to worst based on fitness, ensuring $x \neq y \neq z \neq i$;
Generate a mutant vector using **CBDE** operator: **CBDE** ($x$, $y$, $z$, $Fi$);
Create a trial vector by performing binomial crossover between the mutant vector and the current solution $i$ using crossover probability $C$;
Evaluate the fitness of the trial vector using $f1$ and $f2$;
**if** the trial vector is better than the current solution then Replace the current solution $i$ with the trial vector;
**else** $Ci$ = normrnd (mean($C$), 0.1); $Fi$ = normrnd (mean ($F$), 0.1);
**End for**
Merge current population $P$ with offspring population;
Perform Non-Dominated Sorting using NSGA-II based on Pareto dominance;
Calculate crowding distance for solutions within each rank and select top $p$ solutions for the next generation based on rank and crowding distance;
**end while**
**Return the final set of selected features;**
**Function CBDE** ($x$, $y$, $z$, $F$) $w$ = **GENEINSERTION** ($x$, **GENEREMOVAL** ($y$, $z$, $F$), $F$);
**Function GENEINSERTION** ($x$, $y$, $F$) **if** $x_i$ = 1 or ($y_i$ = 1 and $randi < F$) then return 1; **else** return $x_i$;
**Function GENEREMOVAL** ($x$, $y$, $F$) **if** $x_i$ = 0 or ($y_i$ = 1 and $randi < F$) then return 0; **else** return $x_i$.

**Binary representation of features**

The method uses BDE to perform efficient FS in a binary search space where each gene is represented by a bit. A variation of DE called BDE was created to address optimization issues with binary variables [15]. The aim of BDE is to effectively search discrete solution spaces where decision variables are binary. In BDE, the crossover and mutation operators have been adapted to work with binary variables. For every potential solution, a binary string of size $n$ (genes) is used. Every bit in the string corresponds to a feature: a value of 1 means that the feature has been selected, whereas a value of 0 means that it has not. As illustrated in Fig. 2.

**Population initialization**

A parent population is initialized by randomly generating binary strings, each representing a potential subset of features.

**Multi-Objective approach**

A multi-objective approach uses many objective functions, as the name suggests. Many objectives or various criteria are prevalent in the majority of practical decision-making difficulties in many real-world or practical scenarios.

— Pareto-based evaluation: the Pareto-based rank assignment is used to evaluate individuals [16].
— NSGA-II: the fitness-sharing concept is applied within NSGA-II to promote solution diversity [17].

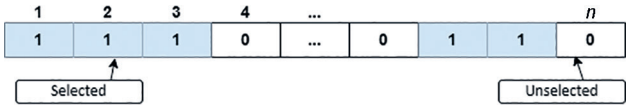


*Fig. 2*. Representation of binary string encoding for FS solution

**CBDE mutation operator**

The proposed CBDE operator is inspired by the Biclustering Binary Differential Evolution (BBDE) [15] strategy but introduces key improvements to better suit gene clustering for feature selection. Unlike BBDE, which simultaneously handles row and column structures, CBDE eliminates biclustering constraints and focuses exclusively on clustering genes into compact and discriminative subsets, a requirement more relevant to genomic feature selection.

Furthermore, CBDE operator is designed to enhance the mutation process by prioritizing solutions with better fitness scores, ensuring more effective exploration of the search space. Specifically, three vectors ($\mathbf{x_{r1}}$, $\mathbf{x_{r2}}$, and $\mathbf{x_{r3}}$) are randomly selected and ranked by fitness, with the best ($\mathbf{x_{r1}}$) strongly influencing mutation and the worst ($\mathbf{x_{r3}}$) being suppressed, thus biasing the mutation toward high-quality solutions and increasing the chance of producing better offspring. After ordering, the mutation operator combines the vectors based on the chosen strategy such as difference-based perturbation. A factor $F$ (initially 0.65) controls gene insertion or removal by comparing a random value $randi \in [0, 1]$, where $i$ specifies the position of the current bit in the individual. This mechanism prevents solutions from degenerating into trivial cases (e.g., all 0s or all 1s), preserving population diversity and enabling efficient gene clustering. The CBDE method performs gene insertion or removal if the generated number *randi* is less than $F$; otherwise, the gene remains unchanged.

Empirically, our results originally presented in Tables 3 to 7 (provided in link[1]) demonstrate that this mechanism improves both classification accuracy and stability. Although these tables are no longer included directly in the manuscript, the provided link ensures access to the full experimental details. The consistent superiority of CBDE over other state-of-the-art methods further confirms its practical relevance.

The proposed CBDE operator introduces several innovations, including a clustering-oriented design for compact and non-redundant gene subsets, a fitness-guided ranking mechanism for knowledge-driven exploration, an adaptive dual insertion-removal strategy to preserve diversity, a perturbation tailored to gene clustering for biological interpretability and empirical validation, showing statistically significant improvements over conventional binary DE approaches.

$$CBDE = \text{GENEINSERTION}(\mathbf{x_{r1}}, \text{GENEREMOVAL}(\mathbf{x_{r2}}, \mathbf{x_{r3}}, F), F). \quad (1)$$

The new individual is evaluated using the fitness function.

$$Fi = \text{normrnd}(\text{mean}(F), 0.1), \quad (2)$$

$$CRi = \text{normrnd}(\text{mean}(CR), 0.1). \quad (3)$$

$Fi$ and $CRi$ represent the adapted versions of the mutation factor ($F$) and crossover probability ($CR$) for individual $i$, adjusting the mutation intensity and crossover rate respectively only when the individual shows no improvement, thereby ensuring effective self-adaptation and enhancing the algorithm's exploration capability.

If its score improves, the values of $F$ and $CR$ remain unchanged; otherwise, they are slightly adjusted using predefined Eqs. (2) and (3).

A summary of the CBDE mutation operation, using the adapted gene insertion and removal mechanisms tailored for clustering tasks is as follows:

$$\text{GENEINSERTION}(\mathbf{x}, \mathbf{y}, F) = \begin{cases} 1, \text{ if } x_i = 1 \text{ or } (y_i = 1 \text{ and } randi < F) \\ x_i, \text{ otherwise} \end{cases}, \quad (4)$$

$$\text{GENEREMOVAL}(\mathbf{x}, \mathbf{y}, F) = \begin{cases} 0, \text{ if } x_i = 0 \text{ or } (y_i = 1 \text{ and } randi < F) \\ x_i, \text{ otherwise} \end{cases}. \quad (5)$$

**Illustrative example of the CBDE mutation operator**

Consider the gene expression matrix $\mathbf{M} \in R^{5\times5}$, where rows represent genes and columns represent samples, and the initial population $\mathbf{P}$, consisting of four individuals represented by binary vectors (5 bits, one per gene):

$$\mathbf{M} = \begin{pmatrix} 42 & 15\,36 & 71\,24 \\ 33 & 58\,17 & 39\,65 \\ 13 & 45\,92 & 12\,26 \\ 57 & 81\,40 & 59\,31 \\ 21 & 14\,29 & 76\,54 \end{pmatrix}; \mathbf{P} = \begin{pmatrix} \mathbf{a:}\ 1\,0 & 1\,0 & 1 \\ \mathbf{b:}\ 0\,1 & 1\,0 & 0 \\ \mathbf{c:}\ 1\,1 & 0\,1 & 0 \\ \mathbf{d:}\ 0\,0 & 1\,1 & 1 \end{pmatrix}.$$

We apply the CBDE ($\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$) mutation using Eq. (1) with scaling factor $F = 0.65$ and a random vector: $\mathbf{rand} = [0.8, 0.9, 0.2, 0.5, 0.6]$.

**Step 1:** Gene Removal. We compute GENEREMOVAL ($\mathbf{c}$, $\mathbf{d}$, $F$) using Eq. (5). Applying it to $\mathbf{c} = [1, 1, 0, 1, 0]$ and $\mathbf{d} = [0, 0, 1, 1, 1]$, we obtain $\mathbf{x'} = [1, 1, 0, 0, 0]$.

If the resulting fitness does not improve, Eqs. (2) and (3) are applied to adjust $F$ and $CR$.

**Step 2:** Gene Insertion. We then compute GENEINSERTION ($\mathbf{b}$, $\mathbf{x'}$, $F$) using Eq. (4). Applying it to $\mathbf{b} = [0, 1, 1, 0, 0]$ and $\mathbf{x'} = [1, 1, 0, 0, 0]$, we obtain CBDE ($\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$) = [1, 1, 1, 0, 0].

**Interpretation:** The resulting solution selects genes 1, 2, and 3. Therefore, the gene cluster constructed is:

$$\begin{pmatrix} 42 & 15\,36 & 71\,24 \\ 33 & 58\,17 & 39\,65 \\ 13 & 45\,92 & 12\,26 \end{pmatrix}.$$

This cluster is then evaluated using the fitness function which combines the number of selected genes and the MSR.

**Fitness evaluation**

The algorithm evaluates the quality of each solution using two objective functions: the size of the selected features and the MSR, aiming to guide the search toward solutions that balance minimal feature size with high classification accuracy.

In this study, the MSR [18] is used as the second objective function to assess the coherence of gene clusters

[1]Available at: https://doi.org/10.5281/zenodo.17201993 (accessed: 16.11.2025).

by measuring the similarity of gene expression patterns across all samples; a lower MSR indicates more consistent and biologically relevant gene groups. The MSR for a cluster $C$ that consists of $I$ rows and $J$ columns is defined as follows:

$$MSR(C) = \frac{1}{|I| \times |J|} \sum_{i=1}^{|I|} \sum_{j=1}^{|J|} (c_{ij} - c_{iJ} - c_{Ij} + c_{IJ})^2, \quad (6)$$

where $c_{ij}$ is the expression value of gene $i$ under condition $j$; $c_{iJ} = \frac{1}{|J|} \sum_{j \in J} c_{ij}$ is the average expression of gene $i$ across all conditions; $c_{Ij} = \frac{1}{|I|} \sum_{i \in I} c_{ij}$ is the average expression of all genes under condition $j$; $c_{IJ} = \frac{1}{|I| \times |J|} \sum_{i \in I} \sum_{j \in J} c_{ij}$ is the overall average expression within the cluster.

**Example**

Consider a gene cluster $C$ composed of 2 genes and their expression values across 3 samples (conditions) as follows:

$$C = \begin{pmatrix} 10 & 15 & 30 \\ 25 & 20 & 40 \end{pmatrix}.$$

We compute the MSR as defined in Eq. (6). For example, when $i = 1, j = 1$, we have:

$$c_{11} = 10,\ c_{IJ} \approx 23.333,$$
$$c_{iJ} \approx 18.333,\ c_{Ij} = 17.5.$$

We then substitute these into the MSR equation and iterate for all values in the matrix:

$$MSR(C) = \frac{1}{6} \sum_{i=1}^{2} \sum_{j=1}^{3} (c_{ij} - c_{iJ} - c_{Ij} + c_{IJ})^2.$$

After plugging in all values and performing the calculations: $MSR(C) \approx 22.22$.

**Selection process**

After evaluating fitness, the algorithm chooses the best individuals to create the next generation based on dominance (NSGA-II), ensures diversity and convergence.

**Iteration and convergence**

The process is repeated (mutation, crossover, fitness evaluation, and selection) the predetermined number of iterations or until the convergence criteria are completed.

## Experiments and Evaluation

**Experimental Settings**

To get the best results, the study suggested DeFs-CBDE feature selection method optimal parameters were chosen by trial and error. The parameter settings for the suggested DeFs-CBDE approach on each experimental dataset are displayed in Table 1.

**Description of considered datasets**

In this work, we evaluated the effectiveness of the suggested DeFs-CBDE feature selection strategy using four high-dimensional carcinogenic microarray datasets. These four datasets include brain cancer [19], breast cancer [20], lung cancer [21], and CNS [21], as illustrated in Table 2. We focused on these four widely adopted benchmark datasets because they are well-curated, publicly available, and have been extensively used in previous studies on feature selection and classification of gene expression data. This choice allows us to ensure fair comparison with prior work and to validate our method under standard experimental conditions. However, we acknowledge that relying only on these datasets may limit the generalizability of the conclusions. Therefore, as discussed in the Conclusion and Future Work, we plan to extend the evaluation to larger and more diverse datasets, to confirm the robustness of DeFs-CBDE across different cancer types and experimental platforms.

*Table 1.* Hyperparameters used in the proposed DeFs-CBDE

| Hyper Parameter | Value |
|---|---|
| Initial $F$ (Factor) | 0.65 |
| Crossover rate | 0.6 |
| Number of generations | 50 |
| Number of Individuals | 20 |

The 5-fold cross-validation method is employed to divide the dataset into training and testing sets. Each algorithm is executed independently 20 times to ensure statistical reliability. All experiments were conducted on a Windows 10 PC with a 2.40 GHz Xeon E5620 8-core processor and 40 GB of RAM, using the parameter settings shown in Table 1.

## Experimental results and discussion

**Evaluation of the suggested DeFs-CBDE Performance**

The performance of machine learning models using features selected by the proposed DeFs-CBDE method was compared to the FS techniques presented in [6]. Fig. 3 shows the classification results using all features, features selected by individual filter methods (Chi-squared (CS), Information Gain (IG), and Gain Ratio (IGR)), hybrid filter-GA methods (CS-GA, IG-GA, and IGR-GA) from [6], and those selected by DeFs-CBDE.

For the brain dataset, as shown in Fig. 3, *a*, the DeFs-CBDE method shows superior performance across all classifiers, achieving 100 % in accuracy, recall, precision, and F-measure with SVM, K-Nearest Neighbors (KNN), and RF. It remains competitive with NB (85.71 %) and matches or exceeds other methods with DT. These results confirm its robustness and effectiveness in FS.

*Table 2.* Description of the high-dimensional microarray datasets used in this study

| Dataset | #Features | #Samples | #Classes |
|---|---|---|---|
| Brain | 5,597 | 42 | 5 |
| Breast | 24,481 | 97 | 2 |
| Lung | 12,600 | 203 | 5 |
| CNS | 7,129 | 60 | 2 |

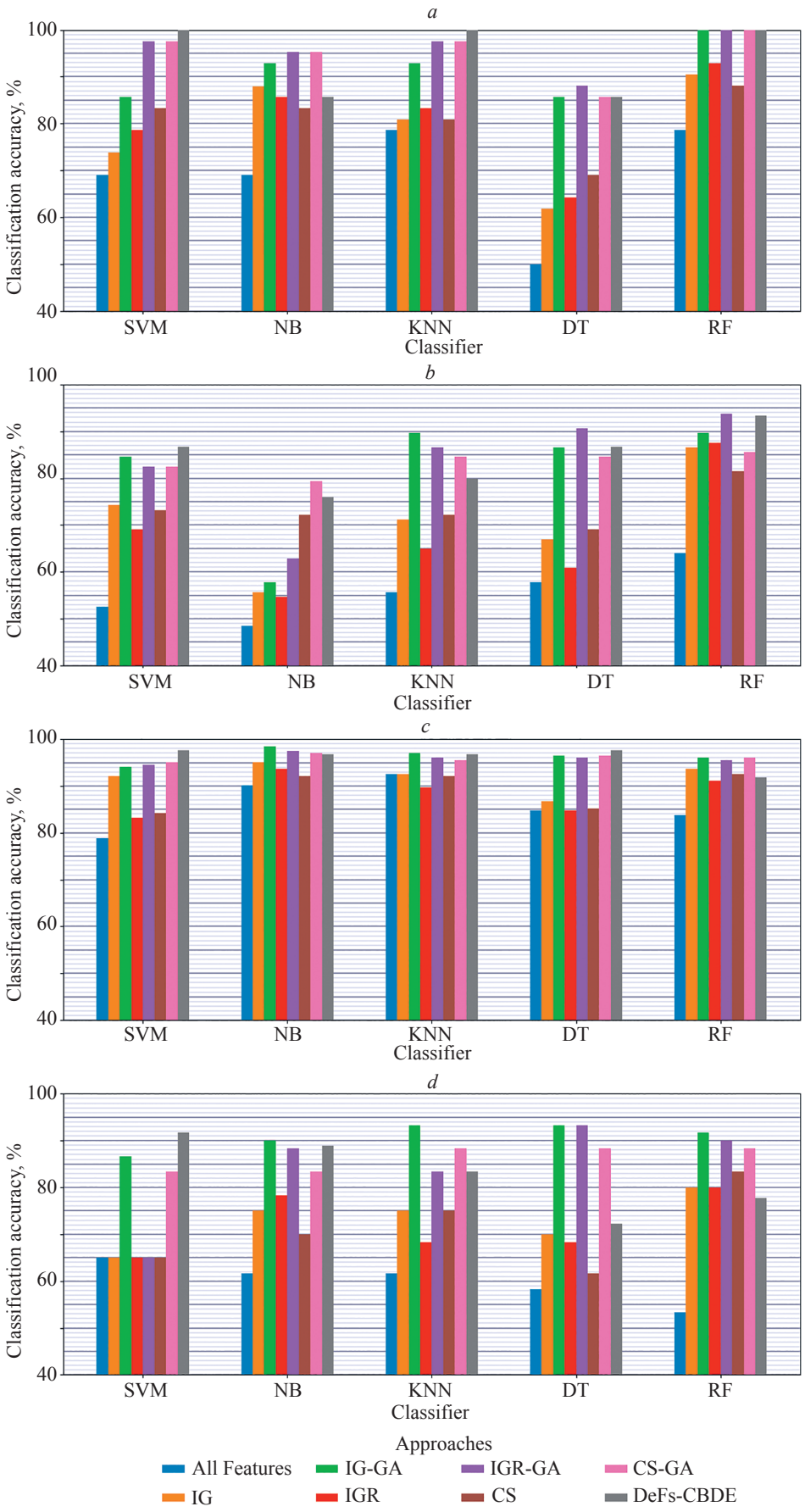


*Fig. 3.* Comparison of classifier accuracies on the data sets: Brain (*a*); Breast (*b*); Lung (*c*); CNS (*d*) for DeFs-CBDE method

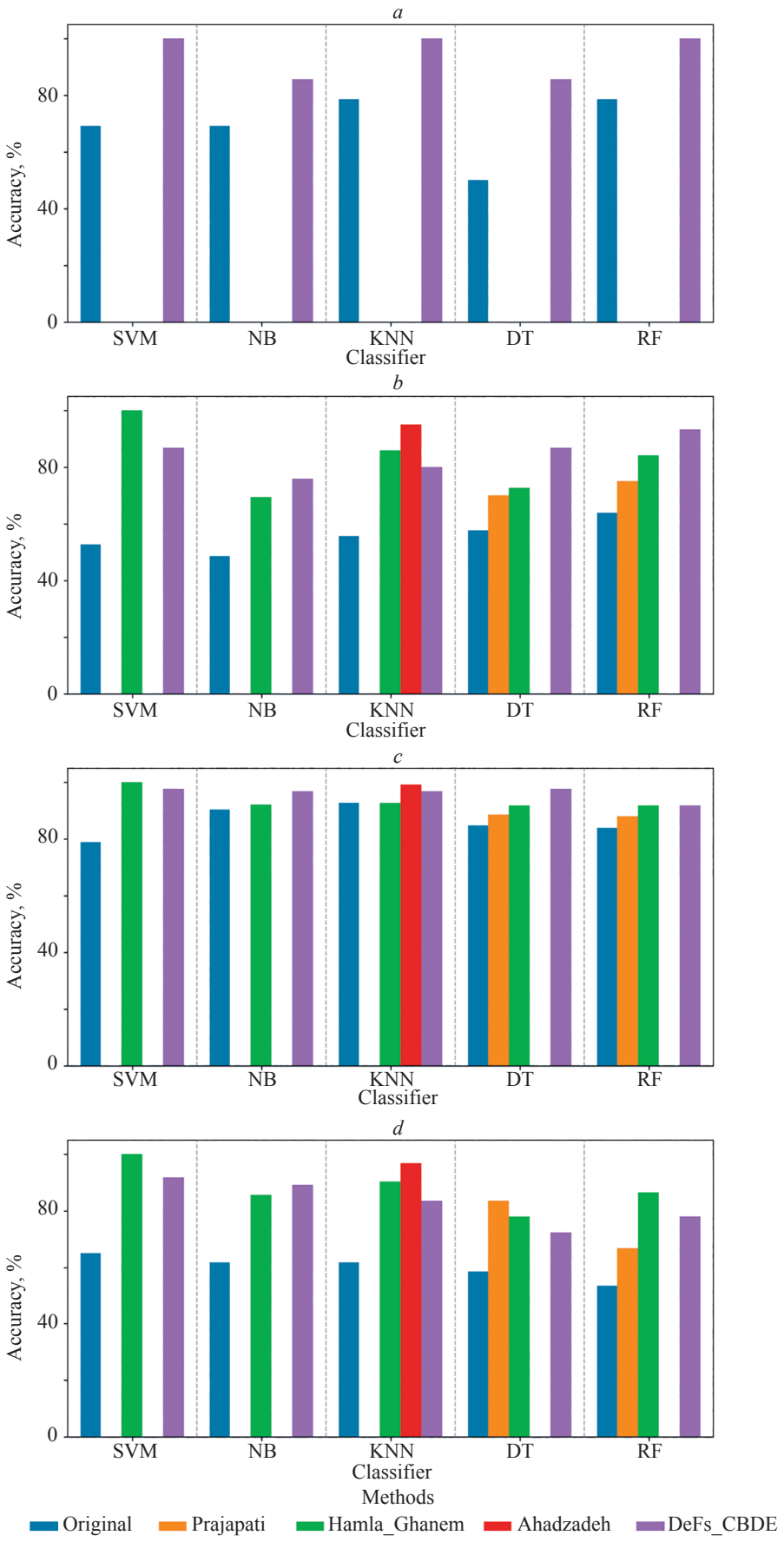


*Fig. 4*. Comparison of classifier accuracies on the data sets: Brain (*a*); Breast (*b*); Lung (*c*); CNS (*d*) with existing works

For the Breast dataset, as shown in Fig. 3, *b*, the DeFs-CBDE method shows consistently strong performance across classifiers. It outperforms other methods with SVM (86.67 %) and RF (93.33 %), and delivers stable results with NB (76 %) and DT (86.67 %). Although slightly below IGR-GA with KNN, it remains competitive, confirming its robustness in FS.

For the Lung dataset, as shown in Fig. 3, *c*, DeFs-CBDE shows strong and stable performance across classifiers, achieving top accuracy with SVM and DT (97.56 %) and solid results with NB and KNN (96.72 %). While RF reaches 91.80 %, slightly below other methods, the approach remains competitive overall, confirming its reliability and versatility in FS.

For the CNS dataset, as illustrated in Fig. 3, *d*, DeFs-CBDE demonstrates competitive performance across all classifiers. It achieves 91.67 % accuracy with SVM and 88.89 % with NB, outperforming or matching most existing methods. While slightly lower with KNN, DT, and RF, the method maintains balanced and reliable metrics, confirming its robustness in FS.

**Comparison of DeFs-CBDE with existing works**

In addition to the main comparison with several state of the art FS algorithms in [6] reported in the previous section, we have compared the proposed method to other existing works [7–9]. Across all four datasets, the proposed DeFs-CBDE approach achieves notably superior accuracies. On the Brain dataset, it reaches 100 % with SVM, KNN, and RF, exceeding by far the baseline of 78.57 %. For the Breast dataset, it achieves 93.33 % with RF, outperforming the best competing result (84.09 % by Hamla and Ghanem). On the Lung dataset, DeFs-CBDE obtains 97.56 % with SVM and DT, very close to the top scores while maintaining consistency. Finally, on the CNS dataset, it reaches 91.67 % with SVM, compared to 65 % without FS, confirming its effectiveness across diverse data. A comprehensive comparison of classification accuracies between DeFs-CBDE and the methods reported in [7–9] is presented in Fig. 4.

The initial data for the diagrams (Fig. 4) are provided in link[1].

**Complexity of computation**

The computational cost of the DeFs-CBDE method can be approximated based on its main parameters, as summarized below. $T$: Number of iterations, $p$: Population size, $n$: Number of genes (features), $m$: Number of samples, $s$: Number of selected features, $Tfitness$: Time complexity of computing the fitness function ($f1$,$f2$).

The overall time complexity is:

$$O(T \times (p \times (n + Tfitness)) + p^2).$$

According to this analysis, the main factors influencing computational effort are the population size ($p$), the number of iterations ($T$), and the number of genes ($n$). By striking an effective balance among these variables, our method can handle high-dimensional data at a computationally affordable cost.

**Biological Relevance of Selected Features (FS) — Lung Dataset**

Fig. 4, *a* shows the functional enrichment results obtained from the genes selected by the FS method applied to the Lung dataset. The analysis, performed with the g:Profiler tool [22], highlights a strong biological significance of the retained genes. Indeed, among the selected features, about 70 % are mapped to significantly enriched pathways (p-value < 0.05 after multiple testing correction). The most enriched categories include cell cycle regulation, cell proliferation, and immune response which are all closely related to lung cancer mechanisms. Additionally, several enriched terms from Gene Ontology (GO) were identified, particularly within *Biological Process* and *Molecular Function*. For instance, genes involved in apoptosis regulation and intracellular signaling pathways are significantly represented (adjusted p-value $\approx 10^{-5}$). These findings demonstrate that the FS-selected genes are not random but rather capture key biological mechanisms of lung cancer, thus providing strong biological validation for the feature selection process.

The initial data for the diagrams (Fig. 5) are provided in link[2].

[1] Available at: https://doi.org/10.5281/zenodo.17201993 (accessed: 16.11.2025).

[2] Available at: https://doi.org/10.5281/zenodo.17201993 (accessed: 16.11.2025).

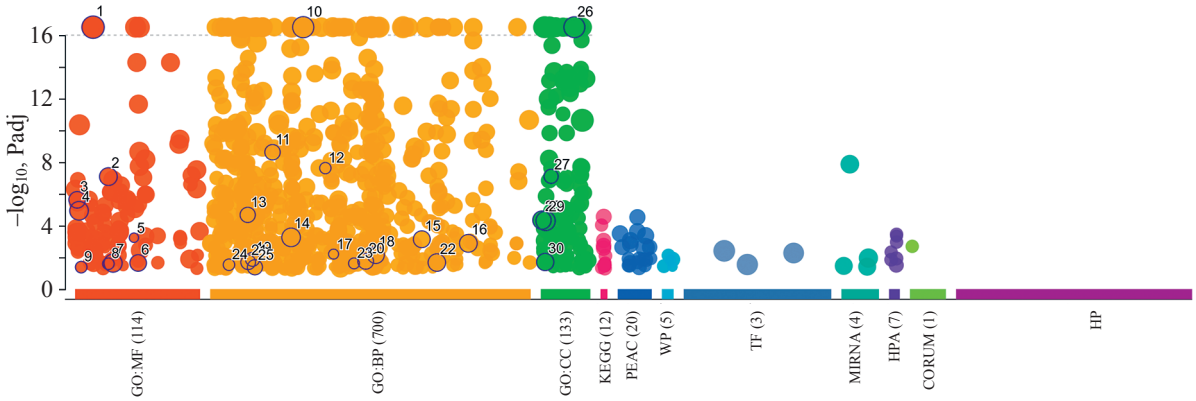


*Fig. 5*. Functional enrichment analysis of FS-selected genes on the Lung dataset using, g:Profiler

We now provide concrete examples of top-ranked genes selected by CBDE and highlight their known involvement in lung cancer. For example:

— Epidermal Growth Factor Receptor (EGFR): frequently mutated in non-small cell lung cancer (NSCLC) and a target of tyrosine kinase inhibitors [23].
— Kirsten Rat Sarcoma Viral Oncogene: one of the most common driver mutations in lung adenocarcinoma, associated with poor prognosis [24].
— Tumor Protein p53: the most frequently mutated tumor suppressor gene in lung cancer, playing a central role in cell cycle regulation and apoptosis [24].
— Anaplastic Lymphoma Kinase (ALK): gene rearrangements involving ALK are important biomarkers guiding targeted therapy in lung adenocarcinoma [23].
— Mucin 1: overexpressed in lung cancer and associated with tumor progression and metastasis [25].

Furthermore, the functional enrichment analysis performed on the broader set of genes selected by CBDE (e.g., pathways such as 'p53 signaling pathway' and 'EGFR tyrosine kinase inhibitor resistance') provides additional statistical support that our method captures not only individual key players but also biologically coherent pathways and processes critically involved in lung carcinogenesis.

## Conclusion

This paper presents a new and effective technique for using Multi-Objective Differential Evolution as a feature selection strategy to find high-quality features. A novel mutation operator, called Clustering-Based Binary Differential Evolution (CBDE), which blends gene insertion and removal, is introduced. This allows for better solutions. Throughout the experimentation phase, datasets on the brain cancer, breast cancer, lung cancer, and the Central Nervous System were used to illustrate the benefits of our strategy over current approaches. The outcomes were remarkable, and the chosen features showed a great deal of variability.

We used five classifiers Support Vector Machine, Naive Bayes, K-Nearest Neighbors, Decision Tree and Random Forest across several datasets to evaluate the effect of DeFs-CBDE. We compared the proposed method with three feature selection algorithms: Chi-Squared (CS), Information Gain (IG), and Gain Ratio (IGR), as well as three hybrid filter-GA feature selection techniques: IG-GA, IGR-GA, and CS-GA.

The proposed algorithm was used on high-dimensional datasets described in the section 'Description of Considered Datasets' to improve the performance of Machine Learning methods by eliminating irrelevant features and retaining only the most relevant subsets. Experiments showed that reducing the dimensionality up to 50 % has improved Machine Learning performance. Despite its effectiveness demonstrated in the results section, the proposed method has some limitations. First, the CBDE increases operator exploration to escape local optima. However, this increases the calculation burden and the search time specifically for large datasets with high dimensionality. Second, the algorithm parameters require careful tuning that may not generalize well across all types of datasets. Additionally, adding new unseen data may require re-running the selection algorithm.

Future work will focus on optimizing the computational efficiency and robustness of DeFs-CBDE by exploring hybrid and adaptive approaches. In particular, we plan to incorporate recent metaheuristic optimization techniques, such as the Whale Optimization Algorithm, Shark Smell Optimization, and Grey Wolf Optimizer, which have shown strong performance in high-dimensional and nonlinear optimization problems. In addition, we intend to evaluate DeFs-CBDE on a wider range of datasets, to further demonstrate its generalizability across different cancer types and experimental platforms. Finally, we plan to extend this line of research with a formal ablation analysis to better quantify the contribution of each component of the CBDE operator.

## References


1. Trunk G.V. A problem of dimensionality: A simple example. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1979, vol. PAMI-1, no. 3, pp. 306–307. https://doi.org/10.1109/tpami.1979.4766926
2. Saberi-Movahed F., Rostami M., Berahmand K., Karami S., Tiwari P., Oussalah M., Band S.S. Dual regularized unsupervised feature selection based on matrix factorization and minimum redundancy with application in gene selection. *Knowledge-Based Systems*, 2022, vol. 256 pp. 109884. https://doi.org/10.1016/j.knosys.2022.109884
3. Deb K. Multi-objective optimisation using evolutionary algorithms: an introduction. *Multi-objective Evolutionary Optimisation for Product Design and Manufacturing*, 2011, pp. 3–34.
4. Cheng Y., Church G.M. Biclustering of expression data. *Proc. of the 8th International Conference on Intelligent Systems for Molecular Biology*, 2000, pp. 93–103.
5. Zhang Y., Gong D., Gao X., Tian T., Sun X. Binary differential evolution with self-learning for multi-objective feature selection. *Information Sciences*, 2020, vol. 507, pp. 67–85. https://doi.org/10.1016/j.ins.2019.08.040
6. Ali W., Saeed F. Hybrid filter and genetic algorithm-based feature selection for improving cancer classification in high-dimensional


## Литература


1. Trunk G.V. A problem of dimensionality: A simple example // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1979. V. PAMI-1. N 3. P. 306–307. https://doi.org/10.1109/tpami.1979.4766926
2. Saberi-Movahed F., Rostami M., Berahmand K., Karami S., Tiwari P., Oussalah M., Band S.S. Dual regularized unsupervised feature selection based on matrix factorization and minimum redundancy with application in gene selection // Knowledge-Based Systems. 2022. V. 256. P. 109884. https://doi.org/10.1016/j.knosys.2022.109884
3. Deb K. Multi-objective optimisation using evolutionary algorithms: an introduction // Multi-objective Evolutionary Optimisation for Product Design and Manufacturing. 2011. P. 3–34.
4. Cheng Y., Church G.M. Biclustering of expression data // Proc. of the 8th International Conference on Intelligent Systems for Molecular Biology. 2000. P. 93–103.
5. Zhang Y., Gong D., Gao X., Tian T., Sun X. Binary differential evolution with self-learning for multi-objective feature selection // Information Sciences. 2020. V. 507. P. 67–85. https://doi.org/10.1016/j.ins.2019.08.040
6. Ali W., Saeed F. Hybrid filter and genetic algorithm-based feature selection for improving cancer classification in high-dimensional

microarray data. *Processes*, 2023, vol. 11, no. 2, p. 562. https://doi.org/10.3390/pr11020562
7. Ahadzadeh B., Abdar M., Safara F., Khosravi A., Menhaj M.B., Suganthan P.N. SFE: A simple, fast, and efficient feature selection algorithm for high-dimensional data. *IEEE Transactions on Evolutionary Computation*, 2023, vol. 27, no. 6, pp. 1896–1911. https://doi.org/10.1109/tevc.2023.3238420
8. Prajapati S., Das H., Gourisaria M.K. Feature selection using differential evolution for microarray data classification. *Discover Internet of Things*, 2023. vol. 3, no. 1, p. 12. https://doi.org/10.1007/s43926-023-00042-5
9. Hamla H., Ghanem K. A hybrid feature selection based on fisher score and svm-rfe for microarray data. *Informatica*, 2024, vol. 48, no. 1, pp. 57–68. https://doi.org/10.31449/inf.v48i1.4759
10. Paul D., Jain A., Saha S., Mathew J. Multi-objective pso based online feature selection for multi-label classification. *Knowledge-Based Systems*, 2021, vol. 222, pp. 106966, https://doi.org/10.1016/j.knosys.2021.106966
11. Han F., Chen W.-T., Ling Q.-H., Han H. Multi-objective particle swarm optimization with adaptive strategies for feature selection. *Swarm and Evolutionary Computation*, 2021, vol. 62, pp. 100847. https://doi.org/10.1016/j.swevo.2021.100847
12. Dashtban M., Balafar M., Suravajhala P. Gene selection for tumor classification using a novel bio-inspired multi-objective approach. *Genomics*, 2018, vol. 110, no. 1, pp. 10–17. https://doi.org/10.1016/j.ygeno.2017.07.010
13. Ghosh M., Adhikary S., Ghosh K.K., Sardar A., Begum S., Sarkar R. Genetic algorithm based cancerous gene identification from microarray data using ensemble of filter methods. *Medical and Biological Engineering and Computing*, 2019, vol. 57, no. 1, pp. 159–176. https://doi.org/10.1007/s11517-018-1874-4
14. Storn R., Price K. Differential evolution: a simple and efficient adaptive scheme for global optimization over continuous spaces. *International Computer Science Institute*, 1995, vol. 95, no. 12, pp. 1–12.
15. Charfaoui Y., Houari A., Boufera F. AMoDeBic: An adaptive multi-objective differential evolution biclustering algorithm of microarray data using a biclustering binary mutation operator. *Expert Systems with Applications*, 2024, vol. 238, part B, pp. 121863, https://doi.org/10.1016/j.eswa.2023.121863
16. Goldberg D.E. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, 1989, 412 p.
17. Deb K., Pratap A., Agarwal S., Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 2002, vol. 6, no. 2, pp. 182–197. https://doi.org/10.1109/4235.996017
18. Noronha M.D., Henriques R., Madeira S.C., Zárate L.E. Impact of metrics on biclustering solution and quality: a review. *Pattern Recognition*, 2022, vol. 127, pp. 108612. https://doi.org/10.1016/j.patcog.2022.108612
19. Pomeroy S.L., Tamayo P., Gaasenbeek M., Sturla L.M., Angelo M., McLaughlin M.E., et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature*, 2002, vol. 415, no. 6870, pp. 436–442. https://doi.org/10.1038/415436a
20. van't Veer L.J., Dai H., van de Vijver M.J., He Y.D., Hart A.A.M., Mao M., et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 2002, vol. 415, no. 6871, pp. 530–536. https://doi.org/10.1038/415530a
21. Zhao G., Wu Y. Feature subset selection for cancer classification using weight local modularity. *Scientific Reports*, 2016, vol. 6, pp. 34759. https://doi.org/10.1038/srep34759
22. Kolberg L., Raudvere U., Kuzmin I., Adler P., Vilo J., Peterson H. g: Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Research*, 2023, vol. 51, no. W1, pp. W207–W212. https://doi.org/10.1093/nar/gkad347
23. Herbst R.S., Morgensztern D., Boshoff C. The biology and management of non-small cell lung cancer. *Nature*, 2018, vol. 553, no. 7689, pp. 446–454. https://doi.org/10.1038/nature25183
24. Skoulidis F., Heymach J.V. Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nature Reviews Cancer*, 2019, vol. 19, no. 9, pp. 495–509. https://doi.org/10.1038/s41568-019-0179-8
25. Nath S., Mukherjee P. Muc1: a multifaceted oncoprotein with a key role in cancer progression. *Trends in Molecular Medicine*, 2014, vol. 20, no. 6, pp. 332–342. https://doi.org/10.1016/j.molmed.2014.02.007

microarray data // Processes. 2023. V. 11. N 2. P. 562. https://doi.org/10.3390/pr11020562
7. Ahadzadeh B., Abdar M., Safara F., Khosravi A., Menhaj M.B., Suganthan P.N. SFE: A simple, fast, and efficient feature selection algorithm for high-dimensional data // IEEE Transactions on Evolutionary Computation. 2023. V. 27. N 6. P. 1896–1911. https://doi.org/10.1109/tevc.2023.3238420
8. Prajapati S., Das H., Gourisaria M.K. Feature selection using differential evolution for microarray data classification // Discover Internet of Things. 2023. V. 3. N 1. P. 12. https://doi.org/10.1007/s43926-023-00042-5
9. Hamla H., Ghanem K. A hybrid feature selection based on fisher score and svm-rfe for microarray data // Informatica. 2024. V. 48. N 1. P. 57–68. https://doi.org/10.31449/inf.v48i1.4759
10. Paul D., Jain A., Saha S., Mathew J. Multi-objective pso based online feature selection for multi-label classification // Knowledge-Based Systems. 2021. V. 222. P. 106966, https://doi.org/10.1016/j.knosys.2021.106966
11. Han F., Chen W.-T., Ling Q.-H., Han H. Multi-objective particle swarm optimization with adaptive strategies for feature selection // Swarm and Evolutionary Computation. 2021. V. 62. P. 100847. https://doi.org/10.1016/j.swevo.2021.100847
12. Dashtban M., Balafar M., Suravajhala P. Gene selection for tumor classification using a novel bio-inspired multi-objective approach // Genomics. 2018. V. 110. N 1. P. 10–17. https://doi.org/10.1016/j.ygeno.2017.07.010
13. Ghosh M., Adhikary S., Ghosh K.K., Sardar A., Begum S., Sarkar R. Genetic algorithm based cancerous gene identification from microarray data using ensemble of filter methods // Medical and Biological Engineering and Computing. 2019. V. 57. N 1. P. 159–176. https://doi.org/10.1007/s11517-018-1874-4
14. Storn R., Price K. Differential evolution: a simple and efficient adaptive scheme for global optimization over continuous spaces // International Computer Science Institute. 1995. V. 95. N 12. P. 1–12.
15. Charfaoui Y., Houari A., Boufera F. AMoDeBic: An adaptive multi-objective differential evolution biclustering algorithm of microarray data using a biclustering binary mutation operator // Expert Systems with Applications. 2024. V. 238. Part B. P. 121863, https://doi.org/10.1016/j.eswa.2023.121863
16. Goldberg D.E. Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, 1989. 412 p.
17. Deb K., Pratap A., Agarwal S., Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II // IEEE Transactions on Evolutionary Computation. 2002. V. 6. N 2. P. 182–197. https://doi.org/10.1109/4235.996017
18. Noronha M.D., Henriques R., Madeira S.C., Zárate L.E. Impact of metrics on biclustering solution and quality: a review // Pattern Recognition. 2022. V. 127. P. 108612. https://doi.org/10.1016/j.patcog.2022.108612
19. Pomeroy S.L., Tamayo P., Gaasenbeek M., Sturla L.M., Angelo M., McLaughlin M.E., et al. Prediction of central nervous system embryonal tumour outcome based on gene expression // Nature. 2002. V. 415. N 6870. P. 436–442. https://doi.org/10.1038/415436a
20. van't Veer L.J., Dai H., van de Vijver M.J., He Y.D., Hart A.A.M., Mao M., et al. Gene expression profiling predicts clinical outcome of breast cancer // Nature. 2002. V. 415. N 6871. P. 530–536. https://doi.org/10.1038/415530a
21. Zhao G., Wu Y. Feature subset selection for cancer classification using weight local modularity // Scientific Reports. 2016. V. 6. P. 34759. https://doi.org/10.1038/srep34759
22. Kolberg L., Raudvere U., Kuzmin I., Adler P., Vilo J., Peterson H. g: Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update) // Nucleic Acids Research. 2023. V. 51. N W1. P. W207–W212. https://doi.org/10.1093/nar/gkad347
23. Herbst R.S., Morgensztern D., Boshoff C. The biology and management of non-small cell lung cancer // Nature. 2018. V. 553. N 7689. P. 446–454. https://doi.org/10.1038/nature25183
24. Skoulidis F., Heymach J.V. Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy // Nature Reviews Cancer. 2019. V. 19. N 9. P. 495–509. https://doi.org/10.1038/s41568-019-0179-8
25. Nath S., Mukherjee P. Muc1: a multifaceted oncoprotein with a key role in cancer progression // Trends in Molecular Medicine. 2014. V. 20. N 6. P. 332–342. https://doi.org/10.1016/j.molmed.2014.02.007

**Authors**

**Mohamed Djellal Serandi** — PhD Student, University Mustapha Stambouli, LISYS laboratory, Mascara, 29000, Algeria, https://orcid.org/0009-0009-5775-7956, mohamed.djellalserandi@univ-mascara.dz
**Fatma Boufera** – PhD, Full Professor, University Mustapha Stambouli, LISYS laboratory, Mascara, 29000, Algeria, https://orcid.org/0000-0002-5733-586X, fboufera@univ-mascara.dz

**Amina Houari** — PhD, Associate Professor, University Mustapha Stambouli, LISYS laboratory, Mascara, 29000, Algeria, SC 57021480300, https://orcid.org/0000-0002-3628-7483, amina.houari@univ-mascara.dz
**Farid Flitti** — PhD, Associate Professor, Higher Colleges of Technology in Dubai, Dubai, 500001, United Arab Emirates, SC 24821958500, https://orcid.org/0000-0002-2480-2580, fflitti@hct.ac.ae

**Авторы**

**Джеллаль Серанди Мохамед** — аспирант, Университет Мустафы Стамбули, лаборатория LISYS, Маскара, 29000, Алжир, https://orcid.org/0009-0009-5775-7956, djellalserandi@univ-mascara.dz
**Буфера Фатма** — кандидат технических наук, профессор, профессор, Университет Мустафы Стамбули, лаборатория LISYS, Маскара, 29000, Алжир, https://orcid.org/0000-0002-5733-586X, fboufera@univ-mascara.dz
**Хуари Амина** — PhD, доцент, Университет Мустафы Стамбули, лаборатория LISYS, Маскара, 29000, Алжир, SC 57021480300, https://orcid.org/0000-0002-3628-7483, amina.houari@univ-mascara.dz
**Флитти Фарид** — PhD, доцент, Высший колледж технологии, колледж в Дубае, Дубай, 500001, Объединенные Арабские Эмираты, SC 24821958500, https://orcid.org/0000-0002-2480-2580, fflitti@hct.ac.ae

*Received 16.07.2025*
*Approved after reviewing 30.09.2025*
*Accepted 21.11.2025*

*Статья поступила в редакцию 16.07.2025*
*Одобрена после рецензирования 30.09.2025*
*Принята к печати 21.11.2025*