A Survey of Feature Selection Strategies for DNA Microarray Classification

Artee Abudayor* Özkan Ufuk Nalbantoğlu
Computer Engineering Department, Engineering Faculty, Erciyes University, Talas, Kayseri, Turkey
* E-mail of the corresponding author: jamendat@gmail.com

Abstract

Classification tasks are difficult and challenging in the bioinformatics field, that used to predict or diagnose patients at an early stage of disease by utilizing DNA microarray technology. However, crucial characteristics of DNA microarray technology are a large number of features and small sample sizes, which means the technology confronts a "dimensional curse" in its classification tasks because of the high computational execution needed and the discovery of biomarkers difficult. To reduce the dimensionality of features to find the significant features that can employ feature selection algorithms and not affect the performance of classification tasks. Feature selection helps decrease computational time by removing irrelevant and redundant features from the data. The study aims to briefly survey popular feature selection methods for classifying DNA microarray technology, such as filters, wrappers, embedded, and hybrid approaches. Furthermore, this study describes the steps of the feature selection process used to accomplish classification tasks and their relationships to other components such as datasets, cross-validation, and classifier algorithms. In the case study, we chose four different methods of feature selection on two-DNA microarray datasets to evaluate and discuss their performances, namely classification accuracy, stability, and the subset size of selected features.

Keywords: Brief survey; DNA microarray data; feature selection; filter methods; wrapper methods; embedded methods; and hybrid methods.
DOI: 10.7176/CEIS/14-2-01
Publication date: March 31st 2023

1. Introduction

In recent years, growing engineering science has been applied in the medical field as well-known bioinformatics, which aids in analysis or diagnosis of the patient's illness to expedite treatment and reduce the risk of diseases such as cancer, leukemia, etc. To diagnose and predict any disease that utilizes medical data (Remeseiro & Bolon-Canedo, 2019), which can be divided into two categories: structured data and unstructured data, The structured data is stored in two dimensions, such that each row corresponds to a sample, and more columns that include the value as genes or features (in the dimensions) that can be used to predict, as with DNA microarray. Unstructured data is usually used in medical imaging and biomedical signal processing fields, which represent waves, sounds, text, or images.

DNA microarray technology provides a profile of thousands of genes to assist many researchers in predicting or diagnosing diseases associated with a patient's illness. Crucial characteristics of the technology (Bolón-Canedo et al., 2014) are as follows: a few samples of data and more than a thousand features (genes); imbalance-class problems; the complexity of classification tasks; the curse of dimensionality; and the dataset shift that occurs when separated into testing and training datasets. According to the technology, this means it confronts a "dimensional curse" in its classification tasks due to the high computational execution needed and the difficulty of discovering biomarkers. To improve the efficiency of the classification task to discover disease-associated biomarkers by employing DNA microarrays, dimensions must be reduced to find a subset of significant features.

Feature selection and feature extraction methods are preprocessing steps of datasets that act as dimensionality reduction techniques. In addition, feature selection algorithms are used to reduce the dimensionality of DNA microarray data by removing redundant and irrelevant features. Feature selection algorithms can decrease computational time, avoid overfitting of the classifier algorithm, increase the classification accuracy, and aid in understanding the association between gene expression and disease [3]-[4]. The original feature can be divided into four categories (Jović et al., 2015) as follows: 1) strongly relevant features 2) weakly relevant but non-redundant features 3) irrelevant features and 4) redundant features. An optimal feature selection algorithm should select deeply relevant and non-redundant features and remove irrelevant and redundant features. The aims of feature selection are to maximize relevance and minimize redundancy in feature subsets. In addition, feature selection methods can be broadly divided into four categories: filter, wrapper, embedded, and hybrid methods (Liu & Wang, 2019), as shown in Figure 1.

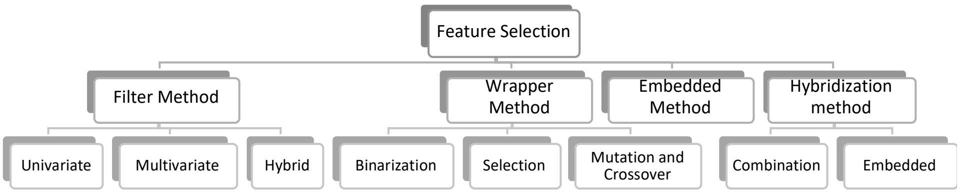


Figure 1. The Overview of Feature Selection Methods

This study briefly surveys feature selection strategies applied via DNA microarray technology for classification tasks. In addition, we provide a summary of the experimental study of the most significant algorithms and evaluation techniques used in the literature, such as microarray datasets, feature selection algorithms, and the case study. In the case study for the classification task, we chose the different feature selection algorithms on two-DNA microarray datasets, such as colon and leukemia datasets. To evaluate the results, we emphasized the achieved classification accuracy, stability, and the number of selected features.

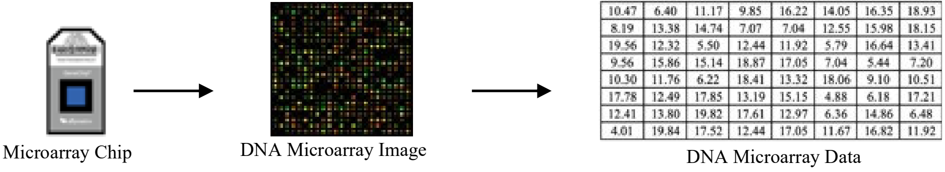
The remainder of the paper is organized as follows: Section 2 introduces the background of DNA microarray datasets. The literature review of filters, wrappers, embedded, and hybrid methods based on feature selection algorithms to be employed in the classification of DNA microarray datasets are presented in Sections 3, 4, 5, and 6, respectively. Section 7 represents the case study of different feature selection methods, with a deep analysis of the findings of this study also provided. Finally, the conclusion is represented in Section 8.

2. DNA Microarray data

DNA-microarray data is medical data that provides the profile of thousands of genes simultaneously in a high-throughput manner, which assists researchers in predicting or diagnosing diseases of the patient's illness to accelerate treatment, reduce sickness risk, and understand the normal as a non-patient. In addition, the data is stored in two dimensions, so that each row corresponds to a sample, and a value in one column represents genes or features (in a dimension) that can be used to predict any disease. Figure 2 represents the general process of acquiring the gene expression data as a feature that corresponds to messenger ribonucleic acid (mRNA) transcribed from DNA. Moreover, DNA inside a nucleus in every cell encodes programs for future organisms.

DNA-microarray data is medical data that provides the profile of thousands of genes simultaneously in a high-throughput manner, which assists researchers in predicting or diagnosing diseases of the patient's illness to accelerate treatment, reduce sickness risk, and understand the normal as a non-patient. In addition, the structure of the data is stored in two dimensions, so that each row corresponds to a sample, and a value in one column represents genes or features (in a dimension) that can be used to predict any disease. Fig 2 represents the general process of acquiring the gene expression data as a feature that corresponds to messenger ribonucleic acid (mRNA) transcribed from DNA. In addition, DNA inside a nucleus in every cell encodes programs for future organisms.

Interestingly, DNA can be divided into two segments: coding and non-coding, where the coding segments are known as genes in the structure of proteins (Bolón-Canedo et al., 2014). Additionally, DNA microarray data can be divided into two types of problems, namely, two-class and multiclass. The two-class problems dedicate the data to separating healthy patients from tumor patients, whereas the multiclass problems dedicate the data to differentiating between distinct types of cancer (Cilia et al., 2019). Table 1 shows an example of DNA microarray data as colon, ovarian, leukemia, lung, and breast with a list of characteristics of the datasets in terms of the total number of samples, number of genes, number of classes, and description of the number of normal and tumors.



10.47	6.40	11.17	9.85	16.22	14.05	16.35	18.93
8.19	13.38	14.74	7.07	7.04	12.55	15.98	18.15
19.56	12.32	5.50	12.44	11.92	5.79	16.64	13.41
9.56	15.86	15.14	18.87	17.05	7.04	5.44	7.20
10.30	11.76	6.22	18.41	13.32	18.06	9.10	10.51
17.78	12.49	17.85	13.19	15.15	4.88	6.18	17.21
12.41	13.80	19.82	17.61	12.97	6.36	14.86	6.48
4.01	19.84	17.52	12.44	17.05	11.67	16.82	11.92

Microarray Chip
DNA Microarray Image
DNA Microarray Data

Figure 2. The general process of acquiring the gene expression data (Bolón-Canedo et al., 2014)

The crucial characteristics of DNA microarray data are as follows: Firstly, small sample sizes and more than a thousand features (genes). Generally, the number of samples is often under a hundred, and the number of genes is more than a thousand, as shown in Table 1. The table presents the number of sample breast datasets that refer to less than 100 patients; the number of genes is more than 24000. Secondly, there is an imbalance-class problem because the number of cancer patients tends to be lower than healthy patients. Thirdly, the complexity of classification tasks is cursed by dimensionality. Finally, dataset shift occurs when dividing data into testing and training sets, in which testing sets tend to be unseen data with a phenomenal change in distribution that depends on any features or class boundaries.

Table 1. An example of DNA-Microarray Datasets

Dataset	No. of Samples	No. of Genes	No. of Classes	Description
Colon	62	2,000	2	cancer: 40 and normal:22
Ovarian	253	15,154	2	cancer: 162 and normal:91
Leukemia	72	7,129	2	AML: 25 and ALL: 47
Lung	203	12,600	5	1: 139, 2:17, 3:6, 4:21, 5:20
Breast	97	24,481	2	non-relapse: 51, relapse:46

3. Filter-based Feature selection

The filter method is a technique that gives scores to each feature or feature subset to find the most significant features by using relevant and non-redundancy features, which is an independent learning algorithm in only one iteration process, as shown in Figure 3. Moreover, the filter methods can be divided into three approaches: univariate, multivariate, and others.

Figure 3. The Overview Step of Filter-based Feature Selection for Classification

Table 2 represents the list of the literature review of filter-based feature selection on DNA-microarray with the type of filter algorithm, the classifier, the number of selected features, cross-validation approaches, and datasets of their experimentation, respectively. According to the table, we can divide the filter method into three strategies: univariate, multivariate, and other algorithms, as follows:

Table 2. The literature review of filter-based feature selection for cancer classification

Algorithm	Ref	Filter-FS Methods			NO. of Selected Feature	Classification Algorithm	Cross-Validation	No. of Datasets
		Uni.	Multi.	Other				
IG	(Cilia et al., 2019)	/	-	-	5-10,000	DT, RF, 1NN and NN	10 k-folds	Six
MI	(Roffo et al., 2017)	-	/	-	10-200	SVM (linear)	5 k-fold with 2/3 training set and 1/3 testing set	Eight
F-Score	(Roffo et al., 2017)	-	/	-	10-200	SVM (linear)	5 k-fold with 2/3 training set and 1/3 testing set	Eight
L-Score	(Tabakhi et al., 2015)	/	-	-	10-100	NB, DT and SVM	5 kfolds	Five
ReliefF	(Lee et al., 2020)	/	-	-	10-130	SVM and KNN	LOOCV	Six
mRMR	(Paul et al., 2017)	-	/	-	-	NB, SVM, and C4.5	Holdout (0.4)	Nine
PCC	(Tabakhi et al., 2015)	-	/	-	0-100	NB, DT and SVM	5 kfolds	Five
ACO	(Tabakhi et al., 2015)	-	/	-	0-100	NB, DT and SVM	5 kfolds	Five
DE	(Hancer, 2020) 2020	-	/	-	-	5NN and DT	10kfolds	Ten
Hybrid F-Score and other	(Das, 2016)	-	-	/	-	NB, MLP, SVM	N/A	Six
Hybrid MRMR and CFC	(Das, 2016)	-	-	/	-	NB, MLP, SVM	N/A	Six
Parallel	(Venkataramana et al., 2019)	-	-	/	10-52	DT, RF	9 kfold with 70% training set and 30% testing set	Two

Note: Uni: Univariate Approach **Multi:** Multivariate Approach

3.1 Univariate approaches

Univariate approaches usually rank each feature by scoring it, which can be divided into four categories: information, statistics, similarity, and distance, as follows:

- Information Measure is an entropy-based feature evaluation method and is usually defined to evaluate the effectiveness of a given feature in predicting the class of unknown or known samples, such as Information Gain (IG), which was proposed by Yang and Pederson in 1997 (Yang & Pedersen, 1997). Mutual Information (MI) estimates a measure of the mutual dependence between a random variable Y (a feature or class label) and another variable X (Gao et al., 2018).
- Statistic Measure: Fisher score (F-Score) is a kind of supervised learning technique to find a subset of significant features by reducing the dimension of the variables (Gu et al., 2012).
- Similarity Measure: Laplacian Score (L-Score) was proposed by Xiaofei He, etc.(He et al., 2006), which is essentially based on Laplacian Eigenmaps (Belkin & Niyogi, 2001) and Locality Preserving Projection (He & Niyogi, 2004). L-Score begins with constructing a nearest neighbor graph to model the local structure and then selects those features with their scores that have the minimal L-Score value.
- Distance Measure: ReliefF measures to search nearest neighbors: one to find the same class, called the nearest hit H, whereas the other different class, which is called nearest miss M (Kira & Rendell, 1992). ReliefF is enhanced from the Relief for dealing with the multi-class problem by taking Euclidean distance as a correlation index to weigh features based on the difference's instances of different classes. The advantage of ReliefF is more robust than Relief and can manipulate incomplete and noisy data.

3.2 Multivariate

For multivariate filter-based feature selection that is evaluated the rank of the entire feature subset and considered feature dependency can be divided into four categories, as follows:

- **Statistic Measure**: Pearson correlation coefficient (PCC) has measured the linear correlation between two variables feature X and class \Upsilon by giving a value between +1 and -1. If the a value is as 1,0,-1 are total positive correlation, no correlation, and total negative correlation, respectively (Lee et al., 2020).
- **Information Measure**: Minimum Redundancy Maximum Relevance (mRMR) was proposed by Ding and Peng in 2005 (Peng et al., 2005). The main idea of mRMR is to find feature subsets to maximize the relevance with the target class and minimize the redundancy amongst the selected features. Interesting, significant features are a top rank feature among the other features and closest relevant to the class variable and less redundancy within the feature.
- **Meta-Heuristic Algorithms (MAs**) are stochastic population-based algorithms, which successfully solve real optimization problems such as Feature selection. A differential evolution (DE) algorithm with the fuzzy rough set theory is called DEFRS (Hancer, 2020). To evaluate the performance of DEFRS that had utilized a fitness function, as shown in Eq. (1).

  $$fitness = \beta \times \frac{\left|\mu_{POS_R(P)}(X)\right|}{|U|} + (1+\beta) \times \frac{1-|SF|}{|NF|} \quad (1)$$

  Where the parameter β ∈ (0,1] Controls the importance of the significant subset of feature sizes. Moreover, $|SF|$is represented by the subset of selected features., and $|NF|$ is represented by the whole number of features.

  An unsupervised Gene Selection ant colony optimization method is called MGSACO (Yu & Liu, 2003). Moreover, to evaluate the fitness function had expressed in Eq. (2).

  $$fitness = \alpha \times prediction_{accuracy} + \beta \frac{1}{|SF|} \quad (2)$$

  Where $prediction_{accuracy}$ is the classification accuracy score that is given by the classifier algorithm. The parameters α and β were weighted values between classification accuracy and the set of selected features, which were set at 1 and 0.001, respectively.

3.3 Others

The other methods are in (Das, 2016) proposed hybridization between univariate and multivariate approaches for optimal gene selection. The hybridization can be divided into two strategies: one based on Fisher score and correlation coefficient features, while the other is based on mRMR and FCBF, respectively. The parallelized hybrid feature selection method is called HFS (Venkataramana et al., 2019). The HFS method includes parallelized correlation feature subset selection followed by rank-based feature selection methods.

4. Wrapper-based Feature Selection

Wrapper methods estimate the significance of selected feature subsets with classification accuracies that depend on learning algorithms. In addition, wrapper methods iteratively produce different candidate feature subsets in some strategies and utilize a classifier algorithm to calculate the corresponding classification accuracy. Generally, feature subsets are retained until a new feature subset with a higher classification accuracy replaces them, as shown in Figure 4.

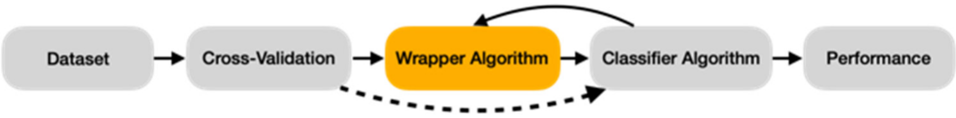


Figure 4. The overview step of Wrapper-based Feature Selection for Classification

Table 3 lists the literature review of wrapper-based feature selection on DNA-Microarray data and other components, such as the strategy of selecting the significant features, fitness functions, the classifier, cross-validation approaches, and datasets of their experimentation, respectively. According to the literature reviews of the wrapper method, we can divide strategies for the selection of the significant feature subsets into three categories as follows:

4.1 Binarization vector

In the search space of the FS problem, it is essential to convert the continuous search space of the solution to a binary version as 0 or 1. Thus, each n-dimensional of the solution represents a feature that takes two values, either "1" if the feature is selected, or "0" if the feature is not selected. Therefore, many researchers attempt to employ two transfer functions, S-shaped and V-shaped, to help choose the feature subsets, as follows:

4.1.1 S-shaped transfer function

In Binary Cuckoo Search [24] and Improved Binary PSO (Chuang et al., 2008), Altruistic Whale Optimization Algorithm (AltWOA) (Kundu et al., 2022), improved Salp Swarm Algorithm (iSSA) (Balakrishnan et al., 2021), improved WOA (iWOA) (Khaire & Dhanalakshmi, 2022), the algorithms utilized search spaces and converted them into a binary vector in each dimension by a sigmoid transfer function, as shown in Eq. (3). Excited binary grey wolf optimizer (EBGWO) (Segera et al., 2020) utilized S-shaped transfer function in Eq. (4) and Binary Moth Flame Optimization Algorithm (BMFO) (Khurma et al., 2020), which utilized S-shaped transfer function in Eq. (5), respectively. In addition, Chimp optimization algorithm (ChOA) (Pashaei & Pashaei, 2022) employed S-shaped transfer function in Eq. (6) and V-shaped transfer function in Eq. (7), respectively.

$$S(x_{i,j}(t)) = \frac{1}{1+e^{-x_{i,j}(t)}} \quad (3)$$

$$S(x_{i,j}(t)) = \frac{1}{1+e^{-10(x_{i,j}(t)-0.5)}} \quad (4)$$

$$S(x_{i,j}(t)) = \frac{1}{e^{x_{i,j}(t)}} \quad (5)$$

$$S(x_{i,j}(t)) = \frac{1}{1+e^{-2x_{i,j}(t)}} \quad (6)$$

Table 3. The literature review of wrapper-based feature selection for classification

Algorithm	Ref	Filter-FS Methods				Fitness Function	Classification Algorithm	Cross-Validation	No. of Datasets
		B	S	MnC	O				
CSA	(De Souza et al., 2018)	/	-	-	-	classification accuracy	NB	10 K-folds	Six
GA	(Too & Abdullah, 2021)	-	-	/	-	Eq. (10)	KNN	10 K-folds	Three
GWO	(Segera et al., 2020)	/	-	-	-	Eq. (10)	KNN (k=5) (Emary et al., 2016)	10 Kfolds (Friedman et al., 2001)	Seven
AltWOA	(Kundu et al., 2022)	/	-	-	-	Eq. (12)	SVM	5 Kfolds	Eight
ChOA	(Pashaei & Pashaei, 2022)	/	-	-	-	Eq. (12)	KNN (k=5)	10 K-folds	Five
SSA	(Balakrishnan et al., 2021)	/	-	-	-	Binary cross-entropy	KNN (k=5), NB (Gaussian kernel)	70% training sets and 30% testing sets	Five

Algorithm	Ref	Filter-FS Methods				Fitness Function	Classification Algorithm	Cross-Validation	No. of Datasets
		B	S	MnC	O				
WOA	(Mafarja & Mirjalili, 2018)	-	/	-	-	Eq. (11)	KNN (k=5) (Emary et al., 2016)	10 Kfolds (Friedman et al., 2001)	Two
	(Mafarja & Mirjalili, 2018)	-	/	-	-	Eq. (11)	KNN (k=5) (Emary et al., 2016)	10 Kfolds (Salesi & Cosma, 2017)	Two
	(Mafarja & Mirjalili, 2018)	-	-	/	-	Eq. (11)	KNN (k=5)	10 Kfolds	Two
CS	(Salesi & Cosma, 2017)	/	-	/	-	classification accuracy and Eq (11)	All kernel of SVM	2 Kfolds < 1000 features 3 Kfolds >1000 features	Five
	(Pereira et al., 2014)	/	-	-	-	classification accuracy	OPF	10 Kfolds	Four
PSO	(Chuang et al., 2008)	/	-	-	-	classification accuracy	KNN (k= 1)	LOOCV	Eleven
ACO	(Fahrudin et al., 2016)	-	-	-	-	classification accuracy	KNN, NB, DT, Jrip	-	Twelve
MFO	(Khurma et al., 2020)	/	-	-	-	Eq. (11)	KNN (k=5) (Emary et al., 2016)	80% training sets and 20 % testing sets	Twenty-three
CS	(Alzaqebah et al., 2021)	-	-	-	/	classification accuracy	SVM	80% training sets and 20 % testing sets	nine
IWOA	(Khaire & Dhanalakshmi, 2022)	/	-	-	/	Cross enrtropy loss function	SVM (polynomial)	10 Kfolds	Six
ACTFRO	(Meenachi & Ramakrishnan, 2021)	-	-	-	/	Fuzzy rough degree of dependency	Rule based classifier , NN, FNN, Random Forest, SVM, FRNN	10 Kfolds	Five
GATFRO	(Meenachi & Ramakrishnan, 2021)	-	-	-	/	Fuzzy rough degree of dependency	Rule based classifier , NN, FNN, Random Forest, SVM,FRNN	10 Kfolds	Five
GLEO	(Too & Mirjalili, 2021)	-	-	-	/	Eq. (10)	KNN (k=5)	80% training sets and 20 % testing sets	Sixteen

Note: B: Binarization Approach **S:** Selection Approach **MnC**: Mutation and Crossover **O**: Others

4.1.2 V-shaped transfer function

In (De Souza et al., 2018) proposed a new wrapper based on a "V-shaped" binarization of crow search algorithm (CSA), as follows in Eq. (7). As the binary version of the Cuckoo Search, namely BCS in (Pereira et al., 2014) using a sigmoid in Eq. (3) and hyperbolic tangent transfer function as shown in Eq. (8), to map the continuous values to binary ones, respectively:

$$V(x_{i,j}(t)) = \frac{x_{i,j}(t)}{1+\left(x_{i,j}(t)\right)^2} \quad (7)$$

$$V(x_{i,j}(t)) = \left|tanh\, x_{i,j}(t)\right| \quad (8)$$

Then, the output is obtained by both S-shaped and V-shaped transfer functions, which are still in a continuous manner, and it should be converted to the binary-valued one by threshold or rand ∈[0,1], which can be calculated by Eq. (9).

$$x_{i,j}(t) = \begin{cases} 0, & if\ r \le V(x_{i,j}(t)), S(x_{i,j}(t)) \\ 1, & otherwise \end{cases} \quad (9)$$

where $x_{i,j}(t)$ indicates the i-th search agent and j-th dimension at the current iteration t. if $x_{i,j}(t)$ = 1 represents the value element being selected as a relevant attribute, while where $x_{i,j}(t)$= 0 indicates the j-th corresponding element being ignored.

4.2 Selection Approach

Whale Optimization Algorithm (WOA) with tournament and Roulette Wheel selection mechanisms in the search (Mafarja & Mirjalili, 2018) are called WOA-T and WOA-R, respectively. The WOA-T and WOA-R algorithms were tested on the two-DNA microarray.

4.3 Mutation and Crossover operators

A WOA algorithm with a mutation operator is called WOA-CM (Mafarja & Mirjalili, 2018). The mutation operation is employed to simulate changing the position of a specific solution around a randomly selected solution (search for prey) or around the best-found solution (encircling prey movement), and the mutation rate is linearly decremented from 0.9 to 0 depending on the iteration number. The genetic algorithm (GA) as a fundamental optimization tool has been widely used in feature selection tasks. However, GA suffers from the hyperparameter setting, high computational complexity, and the randomness of selection operation. Therefore, we propose a new rival genetic algorithm, as well as a fast version of rival genetic algorithm, to enhance the performance of GA in feature selection. The proposed approaches utilize the competition strategy that combines the new selection and crossover schemes, which aim to improve the global search capability. Moreover, a dynamic mutation rate is proposed to enhance the search behavior of the algorithm in the mutation process (Too & Abdullah, 2021). CS with mutation neighborhood search (Salesi & Cosma, 2017) in which a procedure is embedded in the binary CS with an s-shaped transfer function (in Eq. 3) for converting them into a binary vector as 0 or 1, and then the mutated operation begins, as follows:

- The selected variable value was 0 in the binary vector and should be inverted to 1 in the mutated vector.
- Because the selected variable value was 1, the mutated operator inverts it to 0.

4.4 Others

As a strategy for selecting and unselecting feature subsets, the Cuckoo Search Algorithm (CA) with its memory-based mechanism and Levy Flight, was proposed in (Alzaqebah et al., 2021). Then, randomly locating the solution was used to select the feature subsets. ACTFRO, based on ACO and Tabu Search (TS), and GATFRO, based on GA and TS, were presented (Meenachi & Ramakrishnan, 2021) by employing a fuzzy rough set to optimally select feature subsets. A general learning equilibrium optimizer (GLEO), which was suggested in (Too & Mirjalili, 2021). It employs this strategy to select and deselect significant features by converting them to binary-valued ones through the application of a threshold.

4.5 Fitness function

A fitness function is utilized to evaluate the performance of the solution in the FS problem, which can be calculated in two ways, such as classification accuracy or utilizing Eq. (10) – (12). Moreover, the equation of FS aims to maximize the performance of the classification and to minimize the number of the selected features to achieve a balance between two objectives, which are combined into one by setting a weight factor, as shown in Eq. (10)-(12) (Mafarja & Mirjalili, 2018).

$$fitness = \alpha.\gamma_R(D) + \beta\frac{|SF|}{|NF|} \quad (10)$$

$$fitness = \left((1-\alpha)\times Z_{acc}\right) + \alpha\frac{|SF|}{|NF|} \quad (11)$$

$$fitness = \left((1-\alpha)\times Z_{acc}\right) + \alpha\frac{|NF|-|SF|}{|NF|} \quad (12)$$

Where $\gamma_R(D)$ is the classification error rate given by a classifier algorithm.

5. Embedded-based Feature Selection

As shown in Figure 5, the embedded methods utilize a learning algorithm that embeds into the process of feature selection. In addition, the learning algorithm has trained features with feature coefficients simultaneously by minimizing the fitting errors and then obtained the subset of the selected features (Liu & Wang, 2019). However, the embedded method did not need to evaluate the classification accuracy of the subset like the wrapper method. So, these methods have less computational cost.

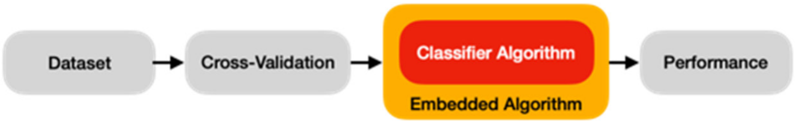


Figure 5. The overview step of Embedded-based Feature Selection for Classification

The literature review of the embedded method is shown in Table 4. The table lists the literature review of the embedded-based feature selection on DNA-Microarray data. In (Hameed et al., 2018) proposed Least Absolute Shrinkage and Selection Operator is called LASSO for both high and low-dimensional datasets. A combination between Mean Decrease Accuracy (MDA) and Mean Decrease Gini (MDG) is called MDA-MDG (Shafi et al., 2020). In (Maldonado & López, 2018) proposed embedded method with two support vector machines (SVM) that extend the ideas of KP-SVM to Cost-Sensitive SVM (CS-SVM) and Support Vector Data Description (SVDD) for the class-imbalanced problem, which are called KP-CSSVM and KP-SVDD, respectively. Improving SVM-RFE has utilized an optimized extreme learning machine (OELM) model instead of the SVM classifier algorithm, which is called OELM-RFE(Ding et al., 2021). In (Guyon et al., 2002) proposed SVM method based on Recursive Feature Elimination (RFE) is called SVM-RFE.

Table 4. The literature review of embedded-based feature selection for classification

Algorithm	Ref	Cross-Validation	Classification Algorithm	Comparison with	No. of Datasets
LASSO	(Hameed et al., 2018)	10k-fold	-	NB, Bayes Net, SVM	Seven
MDA-MDG	(Shafi et al., 2020)	10k-fold	Random Forest	FCBFS + SVM	One
KP-SVDD	(Maldonado & López, 2018)	LOOCV	SVM	Fisher+CSSVM etc.	Four
OELM-RFE	(Ding et al., 2021)	Separated training and Testing sets	OELM, SVM and SVMB	SVM-RFE with (OELM, SVM and SVMB)	Four
SVM-RFE	(Guyon et al., 2002)	LOOCV	SVM	SVM and Baseline Two	Two

6. Hybrid Method Feature Selections

The hybrid method combines two approaches: filter with wrapper methods or wrapper with embedded methods, etc. Many studies, however, concentrated on the hybridization of filter and wrapper methods for feature selection algorithms. Essentially, the filter methods are fast and dependently learn algorithms, but any variable may be lost in the process. On the other hand, the wrapper methods have good performance but are slow. So, finding a way to hybridize the benefits of these methods can improve the performance of classifiers.

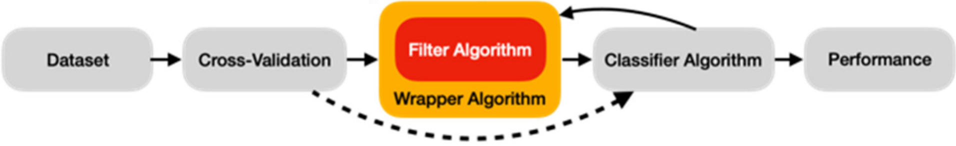


Figure 6. The combination of Hybrid-based Feature Selection for Classification.

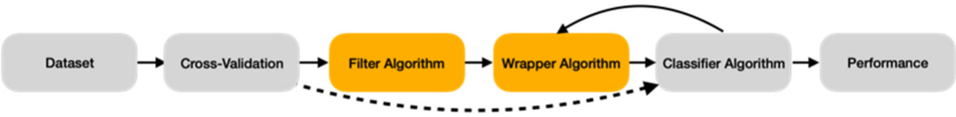


Figure 7. The embedded of Hybrid-based Feature Selection for Classification.

This study divided the steps of the hybrid method into two directions. Firstly, it has utilized a filter algorithm for finding the significant selected features by ordering them by ranking or score. Then, the selected feature subsets from the filter methods are put into the wrapper method for finding the relationship between the features, as shown in Figure 6. Secondly, the filter algorithm embeds the wrapper methods to evaluate the feature subsets, as shown in Figure 7. Both hybrid algorithms are evaluated by the classifier algorithm and employ cross-validation to separate between training and testing sets. Table 5 shows the list of review papers of hybrid method feature selection for classification with the criterion of the selected feature, both filter and wrapper methods, fitness function, cross-validation, classifier algorithm, and the datasets used in their experimentation, as follows:

Table 5. The literature review of hybrid-based feature selection for classification

Algorithm	Ref	Criterion for filter FS		Filter algorithm and the number of selected features in filter-based FS method	Criterion of feature subsets on wrapper FS			Fitness functions	Classification Algorithm	Cross-Validation	No. of Datasets
		Comb.	Emb.		B	M	O				
GA	(Jansi Rani & Devaraj, 2019)	/	-	MI (50 top features)	-	-	/	Classification Accuracy and	All kernel of SVM	-	Three
	(Shukla et al., 2019)	/	-	MI (threshold=0.5)	-	-	/	Eq. (16)	NB, DT, SVM and kNN	10 kfold	Five
	(Salem et al., 2017)	/	-	IG (threshold between 0.4-0.7)	-	-	/	Classification Accuracy	GP	10 kfold	Seven
GWO	(Loey et al., 2020)	/	-	IG (threshold > 0.29 for Colon and > 0.129 for Breast)	-	-	/	Eq. (15)	SVM	5kfold	Two
ACO	(Ghosh et al., 2019)	/	-	Similarly (10 features)	-	-	/	Classification Accuracy	-	LOOCV	Ten
	(Sun et al., 2019)	/	-	ReliefF (score > 0.02)	-	-	/	Eq. (15)	KNN, MLP=20 neurals , SVM (linear), ELM	10 kfold	Six
ECWSA	(Guha et al., 2020)	-	/	mRMR for local search	-	-	/	Classification Accuracy	KNN (k=5)	10 kfold	Seven
PSO	(Jain et al., 2018)	/	-	CFS	/	-	-	Classification Accuracy	NB	10 kfold	Eleven
CS	(Gunavathi & Premalatha, 2015)	/	-	T-statistic (10 features), SNR (50 features) and F-statistic (100 features)	-	-	-	Classification Accuracy	KNN (k=3)	5 kfold	Ten
DFA	(Singh & Singh, 2019)	/	-	ReliefF (k =10) with cup point 30%	/		-	Classification Accuracy	SVM	10 kfold	Six
FOA	(Nouri-Moghaddam et al., 2020)	/	-	F-score (500 features)	/	/	-	Classification Accuracy	SVM (RBF)	10 kfold	Six
FA	(Almugren & Alshamlan, 2019)	/	-	F-score (100, 200, 300, 400 and 500	-	-	/	Classification Accuracy	SVM	LOOCV	Five
ALO	(Singh & Singh, 2019)	/	-	FAST (2, 3,4,5 and 6 features)	/	-	-	Eq. (17)	KNN (k=3)	10 kfold	Twelve
BA	(Alomari et al., 2017)	/	-	mRMR (Top 50features)	/	-	-	Eq. (15)	SVM (RBF)	10 kfold	Ten
MFO	(Dabba et al., 2020)	/	-	mRMR (Top 100 features)	/	-	-	Eq. (15)	SVM (RBF)	LOOCV	twelve
VNLHHO	(Qu et al., 2021)	/	-	F-score	/	-	-	Eq. (18)	Discriminant analysis (discr), DT, KNN, SVM	10 kfold	Eight

Algorithm	**Ref**	**Criterion for filter FS**		**Filter algorithm and the number of selected features in filter-based FS method**	**Criterion of feature subsets on wrapper FS**			**Fitness functions**	**Classification Algorithm**	**Cross-Validation**	**No. of Datasets**
		Comb.	Emb.		**B**	**M**	**O**				
									and NB		
TOPSIS-Jaya	(Chaudhuri & Sahu, 2021)	/	-	IG, GR, ReliefF Correlation, ANOVA	/	-	-	Classification Accuracy	NB (Gaussian kernel)	10 kfold	Ten

Note: Comb: Combination Approach **Emb.:** Embedded Approach **B**: Binarization Approach **M**: Multi-objective approach **O**: Others

- Filter phase: this phase is divided into two parts by the hybrid method: Firstly, the filter and wrapper methods work with a sequential filter algorithm that can help reduce the initial dimensionality of gene expression datasets and obtain the candidate gene subset shown in Figure 6. On the other hand, the filter algorithm embeds wrapper methods for the local search mechanism, as shown in Figure 7.
- Wrapper phase: In hybrid method wrapper methods, the criterion for determining significant feature subsets can be divided into three approaches: binarization methods, multi-objective mechanisms, and others.

6.1 Binarization vectors

Many wrapper approaches utilized the sigmoid transfer function in Eq. 3 to update the selected feature's new position at each iteration. The example of the output value of the sigmoid transfer function is greater than 0.5, as the binary value of the gene will be equal to 1; otherwise, it will be 0. Here, 1 means gene selection, and 0 means not selected. As an example of the binary vectors approaches, as improved-Binary Particle Swarm Optimization (iBPSO) with CFS (Jain et al., 2018), Binary Dragonfly Algorithm (BDFA) with ReliefF (Karizaki & Tavassoli, 2019), Multi-Objective Forest Optimization Algorithm (MOFOA) with Fisher score (Nouri-Moghaddam et al., 2020), Ant Lion Optimization (ALO) with FAST (Singh & Singh, 2019), Binary Bat Algorithm (BBA) with mRMR (Alomari et al., 2017), and Quantum moth flame optimization algorithm (QMFOA) with mRMR (Dabba et al., 2020). Variable neighborhood learning Harris Hawks optimizer (VNLHHO) (Qu et al., 2021) utilized the sigmoid transfer function in Eq. 13, whereas TOPSIS-Jaya employed the sigmoid transfer function in Eq. 14, respectively.

$$S(x_{i,j}(t)) = 1 - e^{\frac{\left(-\frac{A(i)}{ite\times ite_{max}}\right)}{2\tau}} + \frac{e^{x_{i,j}(t)} - e^{-x_{i,j}(t)}}{e^{x_{i,j}(t)} + e^{-x_{i,j}(t)}} \quad (13)$$

$$S(x_{i,j}(t)) = \frac{1}{1 + e^{\frac{-x_{i,j}(t)}{2\tau}}} \quad (14)$$

where A(i) denotes the statistical value of iterations, ite_{max} and ite are the maximum and current numbers of iterations, respectively. τ is the time-varying variable.

6.2 Multi-objective approach

Multi-objective flamework for wrapper methods step on wrapper methods feature selection that can simultaneously optimize the conflicting objectives of the multiple fitness functions, thereby generating a set of Pareto-optimal solutions. MOFOA with Fisher score is called GSMOFOA, (Nouri-Moghaddam et al., 2020) in which the criterion of search is the subset features as the repository for separate population, crowding-distance, and binary tournament selection.

6.3 Other

Binary Genetic Algorithm with CMIM (Shukla et al., 2019), Genetic Algorithm with Mutual Information (Jansi Rani & Devaraj, 2019), and Genetic Algorithm with Information gain (Salem et al., 2017) on the wrapper approaches as the initialized population is encoded by a binary string, with the chromosome size representing the number of features. The bit value 1 represents a selected feature, whereas the bit value 0 represents an unselected feature. In (Gunavathi & Premalatha, 2015) Cuckoo Search (CS) with filter methods such as T-statistics, Signal-to-Noise Ratio (SNR), and F-statistics values. CS is binary encoding in the egg representation binary encoding between 1 and 0 that a feature or gene is selected or not and is ignored for cuckoo search, respectively. Firefly algorithm with an f-score filter-based feature selection method and an SVM classifier is called FFF-SVM (Almugren & Alshamlan, 2019). In (Loey et al., 2020) proposed GWO with IG, which was tested on two datasets, the breast and colon datasets. Ant colony optimization algorithm for wrapper algorithms and similarity for filter algorithms (Ghosh et al., 2019) and RFACO-GS (Sun et al., 2019) have utilized the update of the pheromone values of the ACO algorithm to select the significant feature by random selection. ECWSA is an embedded chaotic whale survival algorithm (Guha et al., 2020). The ECWSA has utilized the chaos to select the type of movement to improve search capacity. Moreover, ECWSA using a local search (mRMR) enhanced the algorithms' ability to remove any unnecessary features without affecting the computational time.

6.4 Fitness function

In the hybrid method, at each iteration, the number of features is reduced with the criterion-evaluated feature subsets by the wrapper method. Each iteration of the wrapper algorithm attains better classification accuracy with the utilization of fewer features. As a result, the fitness function is commonly evaluated using two methods: classification accuracy and Eq. (11)–(13).

$$fitness = \alpha.\gamma_R(D) + \beta \frac{|SF|-|SF|}{|NF|} \quad (15)$$

$$fitness = NB_{acc} \times \omega + \beta \left(\frac{SF}{mod(NF - SF)} \right) \quad (16)$$

$$fitness = \frac{NB_{acc}}{|NF|} \quad (17)$$

$$fitness = \alpha \times Acc + \beta \times \left(1 - \frac{|SF|}{|NF|} \right) \quad (18)$$

Where NB_{acc} represents the classification accuracy given by the NB classifier algorithm, ω is a constant value.

7. The Case Study

The case study aims to demonstrate the performance of different methods of feature selection algorithms, such as filters, wrappers, embedded, and hybrid algorithms. We chose two DNA-Microarray datasets to evaluate and compare the performance, namely the colon and leukemia datasets. The detailed distributions of samples, features, classes and descriptions for each dataset are outlined in Table 1.

In this study, we employ a 10 k-fold cross-validation to divide the training and testing sets and choose K-Nearest Neighbor (K = 5) for the classifying algorithm, and the results are averaged over 10 independent runs to achieve statistically average results. The performance of the results is evaluated and compared against well-known FS algorithms in terms of the average classification accuracy (Acc), the number of selected features (SF), precision (PR), recall (RE), F-score (FSc), and CPU computational time in seconds (Time). To compare the performance of the feature selection, we chose four filter methods, two wrapper methods, two embedded methods, and three hybrid methods for comparison against algorithms. Examples of the algorithms are ReliefF, PCC, GWO, WOA, Lasso, GWO-ReliefF, WOA-mRMR. Moreover, the parameter settings and experimental results of WOA-mRMR were obtained from (Guha et al., 2020).

Table 6 outlines the results of the algorithms, and the best results in the table are highlighted in bold. The experiment results of WOA-mRMR perform superior to other algorithms in terms of classification accuracy, precision, recall, and F-score as WOA-mRMR can obtain the highest scores on two datasets having smaller selected feature sizes. In addition, the filter method has the least execution time of the others. According to Table 6, it seems clear that the hybrid method's feature selection outperformed other methods as far as classification accuracy and the set of selected feature sizes, which means the method is stable in manipulating across two different DNA-microarray datasets.

Table 6. The average of Acc, SF, PR, RE, FSc, Time by different feature selection algorithms on three DNA-Microarray datasets

Algorithms	**Methods**	**The results**											
		Colon						**Leukemia**					
		Acc	PR	RE	FSc	SF	Time	Acc	PR	RE	FSc	SF	Time
ReliefF	Filter-U	0.85	0.84	0.81	0.83	10	1.48	0.93	0.93	0.92	0.93	10	4.94
PCC	Filter-M	0.86	0.85	0.83	0.84	10	0.17	0.96	0.97	0.94	0.95	10	0.18
BGWO	Wrapper	0.87	0.87	0.84	0.85	878.4	131.66	0.96	0.96	0.96	0.96	3412.2	181.84
WOA	Wrapper	0.85	0.85	0.81	0.83	968.1	125.94	0.96	0.96	0.96	0.96	3048.1	168.15
Lasso	Embedded	0.93	0.93	0.92	0.92	2000	66.26	1.00	1.00	1.00	1.00	7129	205.63
GWO+PCC	Hybrid	0.89	0.88	0.88	0.88	89.3	60.84	0.98	0.98	0.98	0.98	93.2	58.79
WOA-mRMR	Hybrid	1.00	1.00	1.00	1.00	30	-	1.00	1.00	1.00	1.00	4	-

Table 7. The advantages and disadvantages of the feature selection methods

Methods	**Examples**	**Advantages**	**Disadvantages**
Filter	Fisher Score, MI	fast, less computation time, and independent	the lowest classification accuracy
Wrapper	BCSA, BGWO	accuracy depend on classifier algorithm	large number of selected features long computation time
Embedded	Lasso	faster than wrapper methods	Classifier depend on selection
Hybrid	MI-GWO, WOA-mRMR	high classification accuracy appropriated deal with high dimensional data	Hybrid need two high-performance algorithms

As shown in Table 7, we summed up the advantages and disadvantages of the whole feature selection method. Filter methods are fast, low-computing, and independent of classifier algorithms but have low classification accuracy. Wrapper methods, which have a slow execution time, use classification accuracy in selecting a new subset of feature subset obtained by a classifier algorithm. Embedded methods employ a learning algorithm that simultaneously trains features with feature coefficients by minimizing fitting errors. The hybrid method combines two approaches, such as the filter and wrapper methods, that employ two benefits of different methods that can improve the performance of classifiers and be easy to deal with very high-dimensional data.

8. Conclusions

This review paper focuses on addressing classification tasks on DNA microarray data, which are known characteristics of DNA microarray data. Feature selection is an approach to reduce computational time and help with classification tasks by finding significant subsets of selected features, which may be generated using a filter, wrapper, embedded, and hybrid methods. For each approach, we provided brief literature reviews about the process of feature selection and listed components for classification, such as classifier algorithms, cross-validation, etc.
Throughout this study, in the literature review and the case study, we presented the current research going on feature selection techniques; that indicated the aim of achieving feature selection, such as stability, classification accuracy, and subset size on DNA-microarray data.

References

Almugren, N., & Alshamlan, H. M. (2019). New bio-marker gene discovery algorithms for cancer gene expression profile. *IEEE Access*, *7*, 136907–136913.

Alomari, O. A., Khader, A. T., Al-Betar, M. A., & Abualigah, L. M. (2017). Gene selection for cancer classification by combining minimum redundancy maximum relevancy and bat-inspired algorithm. *International Journal of Data Mining and Bioinformatics*, *19*(1), 32–51.

Alzaqebah, M., Briki, K., Alrefai, N., Brini, S., Jawarneh, S., Alsmadi, M. K., Mohammad, R. M. A., ALmarashdeh, I., Alghamdi, F. A., & Aldhafferi, N. (2021). Memory based cuckoo search algorithm for feature selection of gene expression dataset. *Informatics in Medicine Unlocked*, *24*, 100572.

Balakrishnan, K., Dhanalakshmi, R., & Khaire, U. M. (2021). Improved salp swarm algorithm based on the levy flight for feature selection. *The Journal of Supercomputing*, *77*(11), 12399–12419.

Belkin, M., & Niyogi, P. (2001). Laplacian eigenmaps and spectral techniques for embedding and clustering. *Nips*, *14*(14), 585–591.

Bolón-Canedo, V., Sánchez-Marono, N., Alonso-Betanzos, A., Benítez, J. M., & Herrera, F. (2014). A review of microarray datasets and applied feature selection methods. *Information Sciences*, *282*, 111–135.

Chaudhuri, A., & Sahu, T. P. (2021). A hybrid feature selection method based on Binary Jaya algorithm for micro-array data classification. *Computers & Electrical Engineering*, *90*, 106963.

Chuang, L.-Y., Chang, H.-W., Tu, C.-J., & Yang, C.-H. (2008). Improved binary PSO for feature selection using gene expression data. *Computational Biology and Chemistry*, *32*(1), 29–38. https://doi.org/10.1016/j.compbiolchem.2007.09.005

Cilia, N. D., De Stefano, C., Fontanella, F., Raimondo, S., & Scotto di Freca, A. (2019). An experimental comparison of feature-selection and classification methods for microarray datasets. *Information*, *10*(3), 109.

Dabba, A., Tari, A., & Meftali, S. (2020). Hybridization of Moth flame optimization algorithm and quantum computing for gene selection in microarray data. *Journal of Ambient Intelligence and Humanized Computing*, 1–20.

Das, K. (2016). *Informative gene selection for effective classification of gene expression datasets* [Department of Computer Science, Shiksha o Anusandhan University]. http://hdl.handle.net/10603/136635

De Souza, R. C. T., dos Santos Coelho, L., De Macedo, C. A., & Pierezan, J. (2018). A V-shaped binary crow search algorithm for feature selection. *2018 IEEE Congress on Evolutionary Computation (CEC)*, 1–8.

Ding, X., Yang, F., Jin, S., & Cao, J. (2021). An Efficient Alpha Seeding Method for Optimized Extreme Learning Machine-based Feature Selection Algorithm. *Computers in Biology and Medicine*, 104505.

Emary, E., Zawbaa, H. M., & Hassanien, A. E. (2016). Binary ant lion approaches for feature selection. *Neurocomputing*, *213*, 54–65.

Fahrudin, T. M., Syarif, I., & Barakbah, A. R. (2016). Ant colony algorithm for feature selection on microarray datasets. *2016 International Electronics Symposium (IES)*, 351–356.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1, Issue 10). Springer series in statistics New York.

Gao, W., Hu, L., & Zhang, P. (2018). Class-specific mutual information variation for feature selection. *Pattern Recognition*, *79*, 328–339. https://doi.org/10.1016/j.patcog.2018.02.020

Ghosh, M., Guha, R., Sarkar, R., & Abraham, A. (2019). A wrapper-filter feature selection technique based on ant colony optimization. *Neural Computing and Applications*, 1–19.

Gu, Q., Li, Z., & Han, J. (2012). Generalized fisher score for feature selection. *ArXiv Preprint ArXiv:1202.3725*.

Guha, R., Ghosh, M., Mutsuddi, S., Sarkar, R., & Mirjalili, S. (2020). Embedded chaotic whale survival algorithm for filter–wrapper feature selection. *Soft Computing*, *24*(17), 12821–12843.

Gunavathi, C., & Premalatha, K. (2015). Cuckoo search optimisation for feature selection in cancer classification: A new approach. *International Journal of Data Mining and Bioinformatics*, *13*(3), 248–265.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1–3), 389–422.

Hameed, S. S., Petinrin, O. O., Osman, A., & Hashi, F. S. (2018). Filter-wrapper combination and embedded feature selection for gene expression data. *Int. J. Advance Soft Compu. Appl*, *10*(1), 90–105.

Hancer, E. (2020). New filter approaches for feature selection using differential evolution and fuzzy rough set theory. *Neural Computing and Applications*, 1–16.

He, X., Cai, D., & Niyogi, P. (2006). Laplacian score for feature selection. *Advances in Neural Information Processing Systems*, 507–514.

He, X., & Niyogi, P. (2004). Locality preserving projections. *Advances in Neural Information Processing Systems*, *16*(16), 153–160.

Jain, I., Jain, V. K., & Jain, R. (2018). Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification. *Applied Soft Computing*, *62*, 203–215.

Jansi Rani, M., & Devaraj, D. (2019). Two-stage hybrid gene selection using mutual information and genetic algorithm for cancer data classification. *Journal of Medical Systems*, *43*, 1–11.

Jović, A., Brkić, K., & Bogunović, N. (2015). A review of feature selection methods with applications. *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1200–1205.

Karizaki, A. A., & Tavassoli, M. (2019). A novel hybrid feature selection based on ReliefF and binary dragonfly for high dimensional datasets. *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*, 300–304.

Khaire, U. M., & Dhanalakshmi, R. (2022). Stability investigation of improved whale optimization algorithm in the process of feature selection. *IETE Technical Review*, *39*(2), 286–300.

Khurma, R. A., Aljarah, I., & Sharieh, A. (2020). An Efficient Moth Flame Optimization Algorithm using Chaotic Maps for Feature Selection in the Medical Applications. *ICPRAM*, 175–182.

Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. In *Machine Learning Proceedings 1992* (pp. 249–256). Elsevier.

Kundu, R., Chattopadhyay, S., Cuevas, E., & Sarkar, R. (2022). AltWOA: Altruistic Whale Optimization Algorithm for feature selection on microarray datasets. *Computers in Biology and Medicine*, *144*, 105349.

Lee, J., Choi, I., & Jun, C.-H. (2020). An efficient multivariate feature ranking method for gene selection in high-dimensional microarray data. *Expert Systems with Applications*, 113971.

Li, Z., Xie, W., & Liu, T. (2018). Efficient feature selection and classification for microarray data. *PloS One*, *13*(8), e0202167.

Liu, W., & Wang, J. (2019). A brief survey on nature-inspired metaheuristics for feature selection in classification in this decade. *2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC)*, 424–429.

Loey, M., Jasim, M. W., El-Bakry, H. M., Taha, M. H. N., & Khalifa, N. E. M. (2020). Breast and Colon Cancer Classification from Gene Expression Profiles Using Data Mining Techniques. *Symmetry*, *12*(3), 408.

Mafarja, M., & Mirjalili, S. (2018). Whale optimization approaches for wrapper feature selection. *Applied Soft Computing*, *62*, 441–453.

Maldonado, S., & López, J. (2018). Dealing with high-dimensional class-imbalanced datasets: Embedded feature selection for SVM classification. *Applied Soft Computing*, *67*, 94–105.

Meenachi, L., & Ramakrishnan, S. (2021). Metaheuristic search based feature selection methods for classification of cancer. *Pattern Recognition*, *119*, 108079.

Nouri-Moghaddam, B., Ghazanfari, M., & Fathian, M. (2020). A novel filter-wrapper hybrid gene selection approach for microarray data based on multi-objective forest optimization algorithm. *Decision Science Letters*, *9*(3), 271–290.

Pashaei, E., & Pashaei, E. (2022). An efficient binary chimp optimization algorithm for feature selection in biomedical data classification. *Neural Computing and Applications*, *34*(8), 6427–6451.

Paul, A., Sil, J., & Mukhopadhyay, C. D. (2017). Gene selection for designing optimal fuzzy rule base classifier by estimating missing value. *Applied Soft Computing*, *55*, 276–288.

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(8), 1226–1238.

Pereira, L. A. M., Rodrigues, D., Almeida, T. N. S., Ramos, C. C. O., Souza, A. N., Yang, X.-S., & Papa, J. P. (2014). A binary cuckoo search and its application for feature selection. In *Cuckoo Search and Firefly Algorithm* (pp. 141–154). Springer.

Qu, C., Zhang, L., Li, J., Deng, F., Tang, Y., Zeng, X., & Peng, X. (2021). Improving feature selection performance for classification of gene expression data using Harris Hawks optimizer with variable neighborhood learning. *Briefings in Bioinformatics*, *22*(5), bbab097.

Remeseiro, B., & Bolon-Canedo, V. (2019). A review of feature selection methods in medical applications. *Computers in Biology and Medicine*, *112*, 103375.

Roffo, G., Melzi, S., Castellani, U., & Vinciarelli, A. (2017). Infinite latent feature selection: A probabilistic latent graph-based ranking approach. *Proceedings of the IEEE International Conference on Computer Vision*, 1398–1406.

Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, *23*(19), 2507–2517.

Salem, H., Attiya, G., & El-Fishawy, N. (2017). Classification of human cancer diseases by gene expression profiles. *Applied Soft Computing*, *50*, 124–134.

Salesi, S., & Cosma, G. (2017). A novel extended binary cuckoo search algorithm for feature selection. *2017 2nd International Conference on Knowledge Engineering and Applications (ICKEA)*, 6–12.

Segera, D., Mbuthia, M., & Nyete, A. (2020). An Excited Binary Grey Wolf Optimizer for Feature Selection in Highly Dimensional Datasets: *Proceedings of the 17th International Conference on Informatics in Control, Automation and Robotics*, 125–133. https://doi.org/10.5220/0009805101250133

Shafi, A. S. M., Molla, M. I., Jui, J. J., & Rahman, M. M. (2020). Detection of colon cancer based on microarray dataset using machine learning as a feature selection and classification techniques. *SN Applied Sciences*, *2*(7), 1–8.

Shukla, A. K., Singh, P., & Vardhan, M. (2019). A new hybrid feature subset selection framework based on binary genetic algorithm and information theory. *International Journal of Computational Intelligence and Applications*, *18*(03), 1950020.

Singh, D., & Singh, B. (2019). Hybridization of feature selection and feature weighting for high dimensional data. *Applied Intelligence*, *49*(4), 1580–1596.

Sun, L., Kong, X., Xu, J., Zhai, R., & Zhang, S. (2019). A hybrid gene selection method based on ReliefF and ant colony optimization algorithm for tumor classification. *Scientific Reports*, *9*(1), 1–14.

Tabakhi, S., Najafi, A., Ranjbar, R., & Moradi, P. (2015). Gene selection for microarray data classification using a novel ant colony optimization. *Neurocomputing*, *168*, 1024–1036.

Too, J., & Abdullah, A. R. (2021). A new and fast rival genetic algorithm for feature selection. *The Journal of Supercomputing*, *77*, 2844–2874.

Too, J., & Mirjalili, S. (2021). General learning equilibrium optimizer: A new feature selection method for biological data classification. *Applied Artificial Intelligence*, *35*(3), 247–263.

Venkataramana, L., Jacob, S. G., Ramadoss, R., Saisuma, D., Haritha, D., & Manoja, K. (2019). Improving classification accuracy of cancer types using parallel hybrid feature selection on microarray gene expression data. *Genes & Genomics*, *41*(11), 1301–1313.

Yang, Y., & Pedersen, J. O. (1997). A comparative study on feature selection in text categorization. *Icml*, *97*(412–420), 35.

Yu, L., & Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, 856–863.