Knowledge-Based Systems 260 (2023) 110111

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A hybrid Artificial Immune optimization for high-dimensional feature selection

Yongbin Zhu [a,b], Wenshan Li [a], Tao Li [a,*]

[a] School of Cyber Science and Engineering, Sichuan University, Chengdu, 610065, PR China
[b] Engineering College, Honghe University, Mengzi, 661199, PR China

ARTICLE INFO

Article history:
Received 6 June 2022
Received in revised form 21 October 2022
Accepted 10 November 2022
Available online 19 November 2022

Keywords:
Feature selection
Artificial immune algorithm
Hybrid search
High-dimensional data
Classification

ABSTRACT

For high-dimensional data, the traditional feature selection method is slightly inadequate. At present, most of the existing hybrid search methods have problems of high computational cost and unsatisfactory feature reduction rate. In this paper, a hybrid feature selection method based on artificial immune algorithm optimization (HFSIA) is proposed to solve the feature reduction problem of high-dimensional data. This method combines the filter method with the metaheuristic-based search strategy more effectively. Inspired by biological research results, the method introduces a lethal mutation mechanism and a Cauchy mutation operator with adaptive adjustment factors to improve the search performance of the algorithm. In addition, this method introduces an adaptive adjustment factor in the population update stage to improve the problem of insufficient diversity of the original algorithm. The effective combination of these mechanisms enables the algorithm to obtain better search capability at a lower computational cost. Experimental comparisons with 23 state-of-the-art feature selection methods are conducted on 22 high-dimensional benchmark datasets. The results show that the computational cost of HFSIA is comparable to 5 classical feature selection methods known for their speed. Moreover, it achieves a higher average classification accuracy than 18 hybrid feature selection methods reported in the latest literature with the best feature reduction rate.

1. Introduction

With the continuous updating and development of data collection techniques, high-dimensional data becomes more and more common. However, a large amount of redundant data acts as noise in the pattern recognition models. They not only increase the computational cost of the model, but also lead to the complexity of the model. Moreover, in some cases, they can steer the learning process in the model toward the generation of a weak model, which can lead to erroneous results. This may cause "dimensionality catastrophe" and "overfitting" problems [1]. Reducing the dimensionality of the data is a proven way to solve the classification problem of high-dimensional small sample data [2]. Feature selection techniques are widely used to deal with such problems. The goal of feature selection is to select as few features as possible to effectively describe the entire feature space [3]. It can reduce the time of the model in the training phase while maintaining or improving the classification accuracy.

With the deepening of research and the expansion of application fields, feature selection has been considered to be an important data pre-processing step in the fields of pattern recognition and machine learning [4]. However, the traditional feature selection methods can hardly meet the application requirements in the current large-scale and high-dimensional feature data environment. According to the correlation between the selection basis of features and the learning algorithm, they can be divided into Filter and Wrapper method. The former uses feature correlation criteria to select feature subsets, which has lower computational cost and better universality. However, since the evaluation criteria of the algorithm are independent of the specific learning algorithm, the selected feature subset performs poorly in classification accuracy. Moreover, the evaluation criteria of these methods are mostly based on some statistical measure, so their optimization effect on feature redundancy is general. This is unacceptable for some research fields (such as genome-wide association studies) that require high requirements for feature redundancy optimization. The latter uses heuristics to search for the best possible feature combination and a classifier to evaluate the quality of the selected features. So that a higher quality feature subset can be obtained. But its computational complexity is high because the classifier needs to be trained and tested for each evaluation of the subset. Especially for large-scale datasets, the execution time of the algorithm will double. This is particularly

* Corresponding author.
E-mail address: litao@scu.edu.cn (T. Li).

https://doi.org/10.1016/j.knosys.2022.110111
0950-7051/

serious for some research and application fields that require high-quality feature subsets and are sensitive to computing costs, such as stock forecasting.

To improve the recognition ability and processing efficiency of learning models in high-dimensional feature spaces, scholars have proposed hybrid feature selection methods that use a combination of multiple search strategies. These hybrid methods can overcome the shortcomings of traditional methods in high-dimensional feature spaces to some extent. They enable the algorithms to reduce redundant features and improve the performance of the learned models in an acceptable time. However, from the experimental results reported in the literature, they have the following shortcomings. Some algorithms also suffer from poor classifier performance due to the high redundancy of optimal subsets. In addition, some algorithms also have the problem of low computational efficiency because the algorithm itself is too complex or not optimized enough.

In order to solve the above-mentioned feature selection problems in the high-dimensional feature space, this paper proposes a hybrid feature selection method based on artificial immune algorithm optimization, which is more excellent in computing cost control, namely HFSIA. The method combines the advantages of Filter and Wrapper methods, incorporating the Fisher score algorithm and an improved clone selection algorithm to explore the search space of the optimal subset. The Fisher algorithm is an effective filtering feature selection method. It identifies the importance of features by calculating the mean and variance of the distance between and within classes. The artificial immune algorithm is an efficient heuristic optimization algorithm which simulates the function of natural immune system. It is an intelligent algorithm inspired by the principle, function, and model of biological immunity. Based on the traditional evolutionary algorithm, it introduces the mechanism of affinity maturity, cloning, and memory. It has the characteristics of fast convergence speed and strong global optimization ability. It is widely used to solve problems related to optimization and pattern recognition.

Aiming at the target demand of feature selection, this paper improves the clonal selection algorithm. These improvements include the population initialization, mutation strategy, and population update mechanism of antibodies. Combined with the research results of biology, this paper introduces the lethal mutation mechanism and the Cauchy operator to improve the search performance of the algorithm. In addition, different adaptive adjustment factors are introduced in the mutation and update stages of the algorithm. They are used to improve the search speed of the algorithm and enhance the diversity of the population, respectively. The effective combination of these strategies enables the algorithm to obtain better searchability and lower computational costs. The evaluation results on 23 high-dimensional datasets with features ranging from 1024 to 22,283 demonstrate the effectiveness of the proposed method. Comparative experiments were conducted with five classical feature selection methods and 18 feature selection methods for high-dimensional data reported in the current state-of-the-art literature. The results show that the computational cost of this method is comparable to classical feature selection methods known for their speed. Moreover, it achieves better average classification accuracy than feature selection methods reported in the latest literature with the smallest number of optimal feature subsets. The comparative experimental results fully demonstrate the progressiveness of the proposed method in this paper.

The main contributions of this paper are as follows. (1) In this paper, a hybrid feature selection method is proposed to solve the problem of the high computational cost of the wrapper feature selection method based on a metaheuristic algorithm on high-dimensional datasets. The method incorporates a filter algorithm and an improved clone selection algorithm. It effectively improves its efficiency in obtaining the optimal feature subset on high-dimensional data. (2) To solve the problem of insufficient diversity of the basic artificial immunization algorithm, this paper adopts the Cauchy operator with adaptive adjustment function in the mutation stage to enhance the genetic diversity of the optimal antibody. (3) The conditional lethal mutation is used to guide the search direction of the algorithm in this paper. The improved algorithm obtains better search ability and also improves its efficiency in obtaining the optimal feature subset. (4) In addition, different adaptive adjustment factors are introduced in the mutation and update phases of the algorithm. It not only improves the search speed of the algorithm but also reduces the computational cost.

The rest of this paper is organized as follows. Firstly, the feature selection methods and their related domain knowledge are summarized in Section 2. Secondly, the implementation details of the proposed method are described in Section 3. Then, Section 4 provides a detailed description of the experimental datasets, the evaluation metrics, and the algorithm parameter settings. The experimental results are analyzed and discussed in Section 5. Finally, Section 6 gives the conclusion.

2. Preliminaries and related work

In this section, firstly, the search strategy in feature selection technology is briefly reviewed. And the problems of this kind of method in feature selection in high-dimensional data space are proposed. Secondly, the related work of current hybrid feature selection methods is summarized. Finally, the principle of the immune clonal selection algorithm is introduced.

2.1. Search strategy in feature selection

The first problem that the machine learning model needs to solve is to remove the noise and redundant features in the original data, and reducing the data dimension is a common and effective technology [5]. Currently, there are two main ways to solve this challenge, which are feature extraction and feature selection [6]. Feature extraction transforms the original features into a set of features with distinct physical or statistical significance or kernels [7]. It involves a linear or nonlinear transformation from the original feature space to a new feature space. Feature Selection is also known as Feature Subset Selection (FSS). The purpose of applying feature selection is to reduce or minimize the number of features in the original data in order to obtain some subset that can effectively characterize the class of data. Reducing the number of features means reducing the number of dimensions of the data. This not only helps to reduce the overfitting problem of the classification algorithm, but also enhances the generalization ability of the learning model. Moreover, it facilitates the understanding between features and eigenvalues and enables the model to gain better interpretation. Also, it improves the training efficiency of the model. In most cases, it also leads to better performance of the learning algorithm. Feature selection has become an important data preprocessing step in the process of pattern recognition and has been widely used to solve dimensional reduction problems for datasets in different fields. For example, optimal gene screening in biomedicine [8], hot topic identification in text mining [9], optimal visual content pixel and color selection problems in image analysis [10], etc.

In the process of feature selection, the search strategy and the evaluation of features are critical and they are also often used as a basis for the classification of feature selection algorithms. According to the search strategy used by the algorithm, the feature selection methods can be classified into three types, which are

based on the global optimal search strategy, based on the random search strategy, and based on the heuristic search strategy [11]. The exhaustive method and branch-and-bound method are the main methods used for global optimal search. However, these algorithms are not very practical due to their inefficient execution and certain restrictions on their use. The feature selection algorithm based on a random search strategy takes the classification performance as the criterion for feature evaluation, and then selects the features whose weights exceed the threshold as the optimal feature output. Therefore, they have better application results. However, they also have more obvious disadvantages. That is, the time complexity is high, the subset obtained each time varies greatly, and they are easy to fall into local optimum. The feature selection algorithm based on a heuristic search strategy is designed based on reasonable heuristic rules, and then iterative operations are repeated to produce the optimal feature subset. Heuristic search has low complexity and high execution efficiency, and is widely used in practical applications. However, heuristic algorithms are problem-dependent techniques. Therefore, they are usually well adapted to a particular problem and try to take full advantage of the specificity of that problem. As a result, they usually fall into local optima due to overly greedy search strategies, making it difficult to obtain a global optimal solution.

Feature selection methods are usually implemented by searching the solution space with the aim of maximizing the relevance to the target class and minimizing the redundancy of the selected features [12]. This objective seems to fit best with the global optimal search strategy. However, its computational cost is unacceptable. This is even more unrealistic for large-scale datasets. Since the evaluation of all possible feature subsets is too costly, there is a need to find a suitable feature subset using a method that is acceptable in terms of computational complexity [13]. The metaheuristic algorithm provides an effective solution to such problems. It can find satisfactory near-ideal solutions in an acceptable time, although not the only best solution is found [14]. In recent years, metaheuristic algorithms have been widely used in the problem of searching for the optimal subset of the feature space [15]. Metaheuristics are improvements of heuristics. It overcomes some shortcomings of the heuristic algorithm and is the product of the combination of the random algorithm and the local search algorithm. Metaheuristic-based wrapper feature selection methods have always been the focus of scholars due to their good global search ability [16]. These algorithms mainly include particle swarm optimization (PSO), grey wolf optimization algorithm (GWO), butterfly optimization algorithm (BOA), genetic algorithm (GA) and artificial immune algorithm (IA), etc. They have good global search capabilities and do not require domain knowledge or prior assumptions about the search space. Moreover, they are not limited by the nature of the problem and have better generality. They can effectively handle complex problems that are difficult to solve by traditional optimization algorithms. They can guide the direction of the search, can eliminate combinatorial explosion, and can obtain satisfactory approximate solutions. Because of these advantages, metaheuristic-based search algorithms have received extensive attention from scholars.

Most metaheuristics start by generating several random initial solutions and then evaluating the resulting set of solutions using a fitness function [17]. The approximate optimal solution is searched by continuous loop iteration until one of the termination conditions is satisfied. In addition, people always want to get better results from machine learning models. The strategy of adding different optimization objectives to the fitness function of the feature selection problem comes into being. By using a multi-objective optimization strategy to model the feature selection problem, a set of non-dominated feature subsets can be obtained. In the case where the number of features is not too large, this scheme is able to meet all kinds of application requirements in practical applications. It is undeniable that the wrapper feature selection method based on metaheuristics has an excellent performance in this case. However, as the search space expands, especially when the number of features reaches thousands, its computational cost will increase exponentially [18]. Therefore, there is an urgent need to find a more efficient solution to the feature selection problem in high-dimensional data spaces.

2.2. Hybrid feature selection approaches

High-dimensional data analysis is a major component of current research in machine learning and data mining. Feature selection approximately reduces the dimensionality of the original data by finding the optimal subspace of the data representation. It improves the performance of learning models while reducing the computational cost, thus becoming an effective means of data preprocessing in data analysis. However, among the current feature selection methods, the traditional single feature selection method can hardly provide a satisfactory solution when dealing with high-dimensional data. In recent years, the outstanding performance of hybrid algorithms in solving optimization problems has received wide attention from scholars. Hybrid algorithms are a class of algorithms that effectively combine different algorithms to solve more complex problems.

In the feature selection problem, several hybrid algorithms have been used to solve the problem of finding the optimal subset of features for high-dimensional data. Hybrid methods combine the advantages of different feature selection methods. The combination of multiple algorithms increases the probability of achieving an efficient and fast way to find the optimal solution. Moreover, there are some algorithms that combine the best features of different algorithms to develop new algorithms. Therefore, hybrid algorithms are able to reduce the search space of optimal feature subsets to a greater extent. In this case, hybrid algorithms based on metaheuristics reduce the possibility of falling into local optimal solutions. Because of the removal of a large amount of noisy data, they are better able to avoid premature convergence and explore the entire data space more efficiently. In most cases, the improved hybrid algorithms are able to obtain better quality global optimal solutions. Moreover, they are able to make a better trade-off between the search quality of the algorithm and the development quality. Therefore, the hybrid methods have a better application value compared to the single feature selection methods.

So far, a variety of hybrid feature selection methods have been proposed. Combining Information Gain and Butterfly Optimization Algorithm (IG-bBOA), Zohre et al. [19] propose a three-stage hybrid feature selection method. The method first uses the minimum redundancy-maximum new classification information (MR-MNCI) method to filter the features, then uses the IG-bBOA to search for candidate feature subsets, and finally uses the similarity-based ranking method to select the final feature subsets. Abhilasha et al. [20] proposed a method combining filter and wrapper for feature selection of high-dimensional data. In this method, the multi-attribute decision-making method is used as a filter to extract the feature information, and then the binary Jaya algorithm with a time-varying transfer function is used to search for the optimal feature subset. Combining global search and local search methods, Liu et al. [21] proposed a hybrid feature selection method based on a genetic algorithm and embedded regularization. Lu et al. [22] proposed a hybrid feature selection algorithm that combines mutual information maximization and an adaptive genetic algorithm. The algorithm first removes a large

number of redundant features by maximizing mutual information, and then searches for the optimal feature subset by adaptive genetic algorithm. Ma et al. [23] proposed a two-stage hybrid ant colony algorithm for high-dimensional feature selection. It uses an interval strategy to determine the optimal subset size of features searched by the additional stage. Huang et al. [24] designed a two-stage hybrid feature selection algorithm. The method combines binary state transition algorithm and ReliefF algorithm, and realizes fast search of feature space with the help of new operators. Additionally, Yan et al. [25] proposed a hybrid feature selection algorithm that combines simulated annealing and an improved coral reef optimization algorithm. And the algorithm is used to solve the feature selection problem of high-dimensional biomedical datasets. Hussain et al. [26] proposed a hybrid optimization method integrating the sine–cosine algorithm into Harris hawks. And the method is used to solve numerical optimization and feature selection problems. Song et al. [27] proposed a hybrid feature selection method based on correlation-guided clustering and particle swarm optimization, and used it to solve the feature selection problem for high-dimensional data. Chen et al. [28] proposed a two-stage hybrid feature selection algorithm. The method first evaluates the importance of features through the ReliefF algorithm and establishes two related tasks on the target concept. Then the PSO algorithm is used to search for the optimal feature subset by combining the information sharing between the two related tasks. Salesi et al. [29] proposed a hybrid feature selection method that combines filters and wrappers. The method first evaluates the features by Fisher algorithm, and then uses a wrapper method combining tabu search and genetic algorithm to obtain the optimal feature subset.

All these hybrid methods are able to overcome the shortcomings of traditional methods in solving feature selection for high-dimensional data to some extent. Compared with the classical filter feature selection methods and other intelligent algorithm-based feature selection methods, they can reduce feature redundancy and computational cost to a greater extent. Moreover, they have better performance in improving the classification performance of classifiers. However, they still have the problem of unsatisfactory classifier performance due to excessive noise in the filtered feature subsets to varying degrees. In addition, most of the algorithms have the problem of high computational cost because the algorithms themselves are too complex or not optimized enough.

2.3. Artificial immune optimization algorithm

The artificial immune algorithm is an intelligent computing method designed to solve complex optimization problems. It is also a meta-heuristic algorithm that simulates the operation mechanism of the biological immune system. Since Burnet [30] fully elaborated the principle of clone selection in 1959, the algorithm has been generally recognized by the immunology community. This theory states that B cells with a high affinity for antigens in organisms are retained by the immune system and have the characteristics of clonal proliferation. And, these proliferating cells will differentiate into two types of cells with different functions. Some are memory cells that function as antigen markers. Others are plasma cells that destroy antigens, known as antibodies. The theory of clonal selection is used to explain the characteristics of immune responses to antigenic stimulation. The core idea is to select only those cells that can recognize antigens for cloning and proliferation. It describes the properties of the acquired immunity of the biological immune system. The clonal selection mechanism corresponds to the process of affinity maturation of immune cells. That is, under the action of this mechanism, immune cells with lower affinity to antigens undergo a process of "maturing" by gradually increasing their affinity after undergoing clonal proliferation and mutation. During this process, mutations in cloned individuals are inversely proportional to antigen affinity. The production of antibodies is the learning process of the immune system.

Based on clonal selection theory, de Castro et al. proposed a famous clonal selection algorithm (CSA) in 2000 [31]. It is pointed out in the literature that the algorithm is mainly composed of population initialization, clone selection, clone proliferation, high-frequency mutation, and population renewal. Among the important features of the clonal selection algorithm, the high-frequency mutation is an important part. It is the basic guarantee to realize the diversity of algorithms, but the choice is the premise.

The basic clonal selection algorithm consists of the following steps [32]. Each time these steps are performed, a new generation of immune cells will be generated.

- Antibody initialization: Generate a set P of candidate solutions.
- Affinity evaluation: Calculate the affinity of each antibody in the antibody pool.
- Selection and cloning: Select n antibodies with the highest affinity, and clone these n antibodies in proportion to their affinity with the antigen to form a clone group C.
- High-frequency mutation: The clone population is subjected to high-frequency mutation, and a mature antibody population D is generated.
- Population update: D is reselected to form a memory cell set M. Generate d new antibodies to replace the lower affinity antibodies in P.
- Repeat steps 2–5 until the termination conditions are met.

In the last decade, the CSA has attracted the attention of a large number of researchers due to its good global optimization capability and convergence performance. Correspondingly, many different variants of the CSA have evolved and been applied to different research fields. Shang et al. [33] improved CSA in four aspects, which made it achieve better convergence than other algorithms when solving multi-objective optimization problems. Dai et al. [34] improved the performance of the algorithm by combining two-way quantum crossover with the basic CSA algorithm. Thus, the algorithm achieves better convergence speed and robustness than other CSA algorithms and heuristics in solving path planning problems. The computational speed of CSA was improved by Xu et al. [35] using the degenerate identification (DR) method. In addition, Yan et al. [36] successfully employed CSA to solve the nonlinear optimization problem in oil exploration and obtained higher detection accuracy. The basic CSA was improved and successfully used for the solution of dynamic multimodal optimization problems by Luo et al. [37].

3. Proposed method

In this paper, a hybrid feature selection method (HFSIA) combining the filter feature selection method with a multi-objective artificial immune algorithm is proposed. This method effectively combines the advantages of the Fisher filtering algorithm and improved clonal selection algorithm. According to the target requirements of feature selection in high-dimensional data, this method greatly improves the initialization and mutation strategy of antibody population of the CSA. In this section, the implementation details of the HFSIA algorithm and several improvements to the basic CSA will be described in detail. These descriptions include the following aspects: antibody coding mode, affinity evaluation operator, population initialization mode, mutation strategy, and renewal mechanism of the antibody in the population. Finally, the pseudo code of the algorithm is given at the end of this section, and the symbols used in the algorithm are also explained.

3.1. Solution encoding

For feature selection methods based on metaheuristics, binary coding strategy is mostly used to represent feature space. This is because binary vectors can not only easily represent subsets of features, but also simplify the operation of the algorithm. Therefore, for feature selection problems, binary encoding is usually adopted to represent the individuals in the solution. In this paper, the antibody representing the subset of features also adopts a binary coding strategy. Moreover, this can also make better use of the advantages of the algorithm itself.

In this paper, the antibody gene is composed of a binary vector of length n. Where each locus corresponds to a feature, and n is the dimension of the feature space. Each antibody represents a subset of candidate features. The coding indicates whether the feature of the position is included in the feature subset by the values '1' and '0'. A value of '1' indicates that the feature at this location is selected, and a value of '0' indicates that it is not selected.

3.2. Evaluating the fitness

In this paper, a multi-objective optimization strategy is used to model the feature selection problem. The aim is to obtain a better balance between the classification performance and the number of feature subsets. That is, it seeks to obtain higher classification accuracy with a smaller number of feature subsets. In this paper, for the feature selection problem, there are two optimization objectives, which are the classification accuracy and the number of feature subsets. To this end, in combination with the multi-objective optimization decision model, Eq. (1) shown below is used as the fitness function in this paper. The construction method of the fitness function is also widely used in other literature [38–41] to evaluate the quality of feature subsets.

$$fitness = \omega \times E_r + (1-\omega) \times \frac{p}{q} \tag{1}$$

Among them, $\omega \in (0, 1)$ is a given real number. E_r is the classification error rate. It is obtained by evaluating this subset of features by an evaluator (usually a classifier). q is the total number of features in the dataset, while p is the number of selected features in the feature subset. In this paper, K-nearest Neighbor (kNN) is used as the classification estimator for feature subsets, where $k = 5$.

3.3. Initial population generation

In the basic clonal selection algorithm, a random distribution is used to implement the population initialization. In fact, many biological phenomena appear in the form of a probability distribution of continuous random variables. In addition, the probability distribution of many random variables takes the normal distribution as its limit distribution under certain conditions. The normal distribution is also known as the Gaussian distribution. The Gaussian distribution is a very important probability distribution in many fields such as mathematics, physics, and engineering. In evolutionary computing, Gaussian distribution is often used in the population mutation link of evolutionary algorithms. However, extensive literature studies have shown that populations are unlikely to experience large mutations when using Gaussian distributions. This may cause the problem of insufficient algorithm diversity. This increases the risk of the search algorithm falling into a local optimum and reduces the convergence speed of the algorithm. Cauchy distribution is another continuous probability distribution function. Compared with Gaussian distribution, its attenuation speed is slower and allows a larger mutation step. This greatly increases the possibility of the algorithm jumping out of the local optimum. Moreover, it has been reported that even in terms of the diversity of methods, the Cauchy distribution is better than the Gaussian distribution in the search process of the evolutionary Algorithm [42–44]. The probability density function of the one-dimensional Cauchy distribution is shown in the following formula.

$$f(x; x_0, \gamma) = \frac{1}{\gamma\pi[1 + (\frac{x-x_0}{y})^2]} = \frac{1}{\pi}[\frac{\gamma}{(x-x_0)^2 + \gamma^2}] \tag{2}$$

The Cauchy distribution has two parameters, x_0 and γ. x_0 is the position parameter and γ is the scale parameter. They determine the shape of the Cauchy distribution. If the value of γ is larger, the peak height of the probability density function will be smaller and the width will be larger. Conversely, if the value of γ is small, the peak height of the probability density function will be higher, and the peak width will be smaller. When $\gamma = 1$ and $x_0 = 0$, it is called the standard Cauchy distribution. Its probability density function is shown in the following formula.

$$f(x; 0, 1) = \frac{1}{\pi(1 + x^2)} \tag{3}$$

Its corresponding cumulative distribution function is shown in the following formula.

$$F(x; x_0, \gamma) = \frac{1}{\pi} arctan(\frac{x - x_0}{\gamma}) + 0.5 \tag{4}$$

When the parameters are the same, the probability density functions of the Cauchy and Gaussian distributions are shown in Fig. 1 below. The following conclusions can be drawn intuitively from the figure. Compared with the Gaussian distribution, the Cauchy distribution has a slower decay rate and a larger range of values.

Therefore, in order to obtain the optimal feature subset more quickly, in this paper, the Cauchy distribution will be applied to the initialization, mutation, and update stages of the population. At the same time, this is also to reduce the computational cost of the algorithm. The initial population space is generated using the standard Cauchy distribution, which is then transformed into a feature code for antibodies. The standard Cauchy distribution function is shown in Eq. (3). The outline of the population initialization algorithm is given in Algorithm 1. Firstly, the algorithm generates the real initial population space through the standard Cauchy distribution function. Then, according to the threshold value η, the real number bits in the initial population space are converted into binary bits that can represent the feature code. In this paper, the conversion domain value η is taken as -0.2 when the antibody population is initialized. That is, when the gene locus of the antibody takes a value greater than η, the gene locus is assigned a value of '1', otherwise, it is '0'.

3.4. Mutation and update strategy

Hypermutation is an important mechanism for the biological immune system to recognize external invasion. It obtains a higher affinity with antigen through the mutation mechanism of the antibody gene. The clonal selection algorithm introduces the mutation theory of biological cells to promote the proliferation and evolution of individuals in the population. The mutation mechanism plays an important role in the operation steps of the clonal selection algorithm. It gives the algorithm the ability of local random search. At the same time, it also maintains the diversity of the population and suppresses the premature convergence of the algorithm. Mutation of individuals in a population is carried out at randomly selected genetic loci. The fundamental

Algorithm 1 Pseudo-code of population initialization

```
procedure GeneratePop
    Initialize population P as per Eq. (3). P = {Ab_1, Ab_2, ..., Ab_N}, Ab_i = [x_1, x_2, ..., x_k], x_k is the antibody locus, and k is feature dimension.
    /* Transform real code of Ab_i to binary */ .
    For each Ab in P Do
        If x_k > η Then x_k = 1 Else x_k = 0 EndIf.
    End for loop
    Return P
end procedure
```

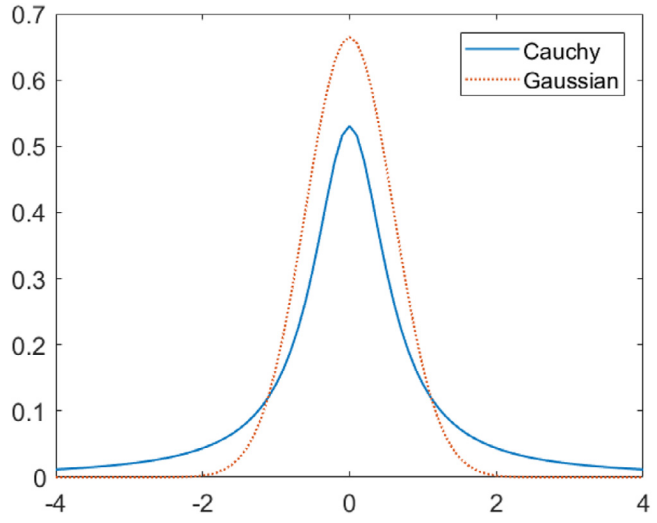


Fig. 1. Comparison of probability density functions between Cauchy distribution and Gaussian distribution.

purpose is to make the population more diverse. Various mutation strategies have been proposed to improve the algorithm to enhance its search performance. These mutation strategies, however, usually differ for different problem domains. Moreover, some improvements can obtain performance gains but at the expense of computational cost. Therefore, a better balance between the performance of the algorithm and the computational cost is needed.

The goal of feature selection is to find a smaller number of high-quality subsets of features to represent the entire feature space. That is, to find a subset with less feature redundancy and higher classification accuracy. Based on this, this paper is inspired by the phenomenon of lethal mutation in gene mutation theory and performs lethal mutation operations on elite antibodies in the population. In this way, the algorithm is accelerated to search in the direction of a smaller number of feature subsets. From a biological point of view, although lethal mutations are detrimental to lethal individuals, they are beneficial for maintaining the heterozygous state of the population. The experimental results show that it can make the algorithm obtain better one-way search ability in solving the feature selection problem of high-dimensional data. Therefore, a feature subset with less feature redundancy can be obtained while ensuring classification accuracy. Moreover, it can also reduce the computational cost of the algorithm to a greater extent.

Furthermore, to steer the mutation process in the direction required by the problem domain, we added an adaptive linear acceleration factor $\delta \in (-0.5, 0.5]$ to the mutation process of the elite antibodies. Its role is to accelerate the rate of lethal decay of elite antibody genes. It works on the premise that the affinity of the local optimal antibody to the antigen can be consistently increased. That is, under the condition of ensuring the classification performance of the optimal feature subset, it is accelerated

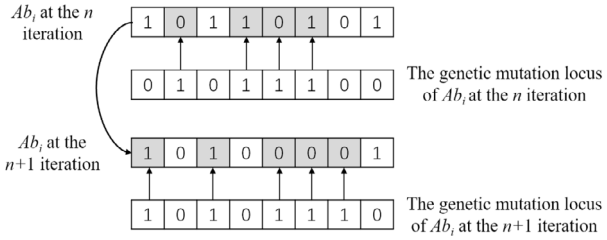


Fig. 2. Schematic diagram of conditional lethal mutation of an antibody.

to search in the direction of a smaller number of features. The calculation equation of the acceleration factor is shown in the following formula.

$$\delta = 0.5 - \frac{t}{T_{max}} \tag{5}$$

Among them, t is the current number of iterations, and T_{max} is the total number of iterations.

During each iteration, each antibody involved in the mutation operation needs to first determine the gene locus at which it performs the mutation. In this paper, the location and number of genes involved in mutation for each antibody are determined jointly by the Cauchy distribution function and the transformation domain value δ. First, a random real number sequence with the same length as the feature space is generated from the Cauchy distribution function. And then it is converted into a binary sequence representing the mutation position of the antibody gene by the domain value. In the binary sequence, '1' indicates that the bit will be mutated, and '0' does not. Therefore, the domain value δ has an important impact on the position and number of genes involved in the variation in an antibody. As the number of iterations increases, the value of δ will gradually tend to -0.5 from 0.5. It can be seen from the schematic diagram of the probability density function of the Cauchy distribution in Fig. 1 that when the value of the transformation threshold is smaller, there will be more '1' in the generated binary sequence. That is, the number of genes involved in the mutation will increase. Therefore, when conditional lethal mutations are used, the genes in the antibody will gradually die out. That is, the number of selected features (i.e. loci that are '1') in the antibody will be less and less. Therefore, the mutation strategy will guide the algorithm to speed up the search in the direction of fewer features. Fig. 2 depicts the gene change when the antibody undergoes lethal mutation when the change of the fitness value of the locally optimal antibody satisfies the mutation condition. Algorithm 2 lists the main steps to perform a lethal mutation operation on the elite antibodies selected from the population in each iteration.

In the classical clonal selection algorithm, the update of the population is carried out on the premise of maintaining the population number unchanged. In the HFSIA, the number of genes in the antibody decays rapidly during the mutation process due to the use of lethal mutation strategies that lead to gene extinction. The purpose is to guide the algorithm to search in the direction of

Algorithm 2 Pseudo-code of Mutation based on gene lethal mutation mechanism

procedure MUTATION
  $C^* \leftarrow CloneCopy(C, n)$, Clone these n elite antibodies according to the cloning size q_i. $n = c_r \times N$, $q_i = n - j, j \in [1, n]$, j is the fitness ranking.
  $flag \leftarrow Check(bestAb_i, k)$, Check the fitness changes of $bestAb_i$, $flag = 1$ when $f(bestAb_i)_{t_0+k} < f(bestAb_i)_{t_0}$, $k = 3$.
  $\delta \leftarrow$ Compute δ as per Eq. (5) when $flag = 1$.
  $C' \leftarrow$ Generating gene mutation markers for each antibody in C^* by Algorithm 1 and the domain value as δ.
  $C \leftarrow C^*\{C'\} = 0$, Perform genetic lethal mutation operation.
  Return C, Return the mutation antibody set.
end procedure

fewer features, so as to obtain high-quality feature subsets with fewer features. Compared with other intelligent algorithms based on metaheuristics, the search of the algorithm is not completely random. The advantage of this strategy is that the algorithm's search is better guided. On the other hand, this also helps to reduce the complexity and computational cost of the algorithm itself. But correspondingly, it is also easier to cause the algorithm to fall into a locally optimal solution. In order to eliminate the risk of falling into a local optimum due to excessively rapid fitness decay, it is necessary to enhance the diversity of the population during the iterative process.

Therefore, two strategies are adopted to compensate. One is to scale up the population when it is updated, i.e., to adopt an incremental update strategy. That is, the number of updates in HFSIA is set to N, while the number of updates in the basic algorithm is $d(d < N)$. This does not mean that the population size will keep increasing during the iteration. Selection is used to maintain the overall size of the population constant. The other is to add the linear incremental adjustment factor $\theta \in (0, 1]$ in the population update stage. Its purpose is to dynamically adjust the number of antibody genes newly added to the population according to iterative changes. That is, the mutation probability of individuals in the population is enhanced, so as to achieve the purpose of improving the diversity of the population. The calculation equation of the adjustment factor is shown in the following Eq. (6).

$$\theta = \frac{t}{T_{max}} \tag{6}$$

Among them, t is the current number of iterations, and T_{max} is the total number of iterations. For its specific implementation and application, please refer to the algorithm framework code part in Section 3.5.

3.5. The proposed algorithm framework and notation

In theory, the higher the dimension of the data, the more detailed the description of things. This plays an important role in some fields of research. But for classification problems, too much redundant feature data will cause a serious decline in the performance of the classifier, and even lead to the problem of dimensional disaster. For the feature selection problem in high-dimensional data space, the main idea is to use a hybrid feature selection method. But how to combine different algorithms more effectively is worthy of further study by scholars. Through many experiments, this paper finds a more efficient hybrid feature selection method than the current literature reports to solve the feature selection problem of high-dimensional data. The method solves the problem of selecting optimal feature subsets for high-dimensional data through a two-stage screening operation. The framework diagram of HFSIA algorithm is shown in Fig. 3.

The algorithm evaluates and ranks all features in the data space by the Fisher scoring function in the first stage. The Fisher score algorithm calculates the mean and variance of the distances between different categories of features and within the same category. It identifies the importance of features through the calculated mean and variance. It is an efficient univariate filtered feature selection method and has the advantage of fast computation. The calculation method will be briefly introduced next [45].

Given a set of labeled data samples, $\{Ab_i, Ab_j\}, Ab_j \in \{1, \ldots, c\}, i = 1, \ldots n$, where c is the number of categories, and n_k represents the number of data samples in the kth category. u_k^i represents the mean of all data samples on the ith feature. μ and σ are the mean and variance of the category k corresponding to the ith feature respectively. The Fisher Score of the ith feature can be calculated by the following formula.

$$F_k = \frac{\sum_{k=1}^{C} n_k (u_k^i - u^i)^2}{\sum_{k=1}^{C} n_k (\sigma_k^i)^2} \tag{7}$$

where $\sum_{k=1}^{C} n_k (u_k^i - u^i)^2$ is the variance of the ith feature between different categories, and $\sum_{k=1}^{C} n_k (\sigma_k^i)^2$ is the variance of the kth feature within the same category.

After the classification importance scores of the features are obtained, the features can be filtered to obtain a reduced subset of candidate data. Compared with the full feature set, the feature subset obtained through the first stage has been greatly reduced in the number of features. Theoretically, for any filtering-type feature selection method, as long as an optimal threshold value is selected, the desired feature subset can be obtained. Although the selection of this optimal domain value can be achieved by a simple exhaustive method. However, there is no guarantee that the feature combined with a high score is the one with the best quality. Because the feature score obtained by any univariate evaluation rule does not guarantee that the combination with a higher score is the optimal feature subset [46]. Therefore, experimental analysis was conducted for the quality of the feature subset screened by the Fisher Score algorithm and the selection of domain values. The filtering and classification experiments of the feature subsets were performed on different datasets with increasing domain values. The experimental results show that the classification accuracy of the features scored and sorted by Fisher Score always oscillates within a certain range after being screened by different thresholds. Moreover, the peak value of its oscillation does not have a linear proportional relationship with the selected threshold value. The following Fig. 4 is the relationship between the increase of the threshold value of the Fisher Score of *GLI-85* and the classification accuracy. The experimental results on other datasets are similar to this figure. It will not be repeated here.

Therefore, this paper adopts the artificial immune algorithm with good global search performance to perform a secondary search on the feature subset after the initial screening. A hybrid feature selection method based on the Fisher filtering method combined with the wrapper method optimized by the artificial immune algorithm is constructed. After experimental analysis, considering both the quality of the optimal feature subset and the computational cost of the algorithm itself, this paper chooses the filter domain value of Fisher Score to be 200. The structural framework and main steps of the HFSIA algorithm are shown in the following algorithm 3.

In addition, Table 1 lists the symbols used in this section.

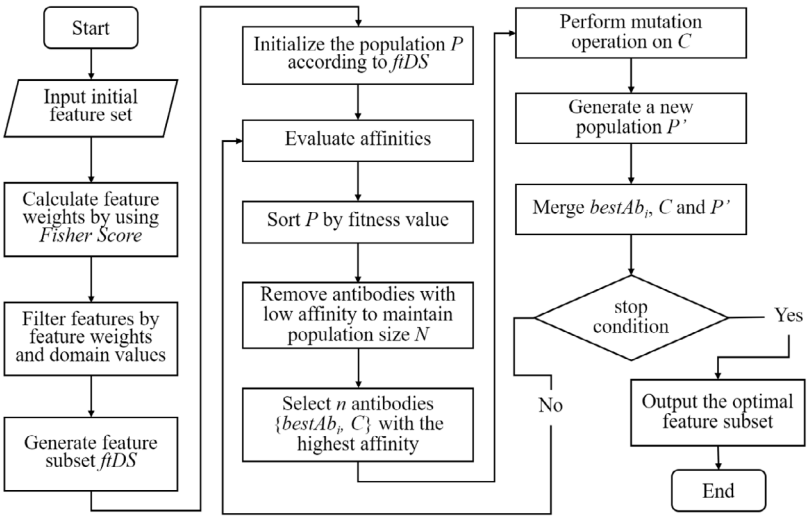


Fig. 3. The framework diagram of HFSIA.

Algorithm 3 HFSIA for feature selection

procedure
input:
  Training sample *orData*.
  The antibody group size N.
  Maximum number of iterations T_{max}.
  Select rate c_r. Proportion of antibodies selected for cloning.
output: The optimal feature subset R_S.
  $/*$ *Initialization* $*/$, Set the initial parameters of the algorithm.
Apply filtering algorithm (Fisher score) and generate feature subsets *ftD*.
$P \leftarrow GeneratePop(N, ftD)$, Generate initial population P based on feature subset *ftD* and Algorithm 1.
for $i < T_{max}$ **do**
    1). $Fitness \leftarrow FitnessFun(P)$, Evaluate affinities for antibodies in population P as per Eq. (1).
    2). $sort(P)$, Sort P by fitness value.
    3). $Select(P, N)$, Remove antibodies with low affinity to maintain population size N.
    4). $bestAb_i, C \leftarrow Select(P, c_r)$, Select n antibodies with the highest fitness and $bestAb_i$.
    5). $Mutation(C)$, Submit C to Algorithm 2 for mutation operation.
    6). $P' \leftarrow$ Generate a new population based on Algorithm 1 and domain value θ as per Eq. (6).
    7). $M \leftarrow \{bestAb_i \cup C \cup P'\}$
    8). $P \leftarrow M$, update P by M.
end for loop
$R_S = best(P)$, Obtain the optimal feature subset.
Return R_S, Return the optimal feature subset.
end procedure

Table 1
The description of the important symbols.

Symbols	Description	Symbols	Description
orData	The training sample data.	δ	The adaptive linear acceleration factor.
ftD	Feature subset by Fisher score.	θ	The Linear Increment Factor.
P	The set of antibodies.	c_r	Select rate.
bestAb	The local optimal antibody.	T_{max}	The maximum number of iterations.
R_S	The optimal subset of features returned.	C	The antibody selection set.
N	The population size.	P'	The antibody mutation set.
n	$n = N \times c_r$, The selection pool size.	M	Population update set.
fitness	The fitness of antibodies in population P.	η	Cauchy transform threshold for initial population.

4. Experiment methodology

In this section, datasets used in the experiment are first introduced, then the performance evaluation criteria of the classification test are explained, and finally, the parameters setting of the HFSIA algorithm in the experiment are described.

4.1. Datasets

In the experiments in this paper, a total of 22 real datasets are used to verify the performance of the proposed feature selection algorithm. These datasets cover varying numbers of features from 1024 to 22,283. They are datasets from UCI Repository [47], feature selection datasets from Arizona State University [48], and

Table 2
Summary of the experimental datasets.

ID	Datasets	Instances	Features	Classes	ID	Datasets	Instances	Features	Classes
1	Yale	1024	165	15	12	NCI9	9712	60	9
2	Colon	2000	62	2	13	Pixraw10P	10000	100	10
3	SRBCT	2308	83	4	14	Orlraws10P	10304	100	10
4	WrapAR10P	2400	130	10	15	Brain Tumor2	10367	50	4
5	Leukemia1	5327	72	3	16	Prostate Tumor	10509	102	2
6	DLBCL	5469	77	2	17	Leukemia3	11225	72	3
7	9Tumor	5726	60	9	18	11Tumor	12533	174	11
8	TOX-171	5748	171	4	19	Lung	12600	203	5
9	Brain Tumor1	5920	90	5	20	Ovarian	15154	253	2
10	Leukemia2	7129	72	3	21	SMK-CAN-187	19993	187	2
11	CNS	7129	60	2	22	GLI-85	22283	85	2

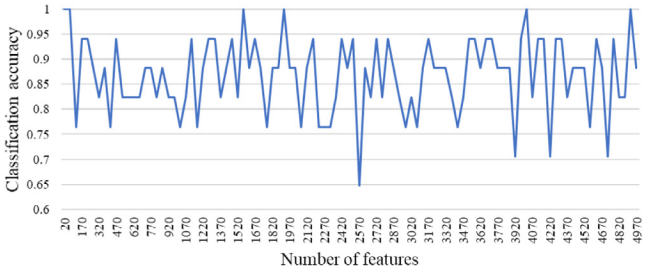


Fig. 4. Relationship between Fisher Score screening threshold and classification accuracy of GLI-85.

gene expression datasets [28], respectively. The UCI dataset is used in the evaluation of feature selection algorithms in many pieces of literature. In addition, the ASU feature selection dataset and gene expression dataset are specially selected to examine the performance of the algorithm on high-dimensional datasets. Table 2 shows the details of these datasets.

4.2. Performance evaluation criteria

In this paper, the *cross-validation* [49] is used to evaluate the accuracy of the classification algorithm. It is a commonly used validation technique and is widely used to evaluate the performance of machine learning models. In this paper, the average classification accuracy of kNN is used to evaluate the quality of the selected optimal feature subset. The classification accuracy, the number of features of the optimal feature subset, and the average and deviation of the computational cost obtained from the experimental results are all statistical results after the algorithm runs 20 times independently on each dataset. And based on these statistical results to evaluate the performance of the algorithm. For other parameters in the experiment, the setting values described in the corresponding literature were used.

4.3. Parameter settings

All experiments are done on a PC with an Intel Core i5 and 8 GB of RAM. Also, all algorithms are performed on different datasets using the same settings. In all experiments, the parameter configuration of the HFSIA algorithm is as follows. The maximum number of consecutive iterations of the algorithm is $T = 50$, the population size $N = 10$, the select rate $c_r = 0.5$, and the initial transformation domain value of Cauchy random numbers $\eta = -0.2$. The parameter w of the fitness function is set to 0.99. The filter domain value of Fisher score is 200.

5. Experiments and discussion

In this section, the proposed feature selection method HFSIA is comprehensively evaluated and analyzed through experiments. Firstly, the performance of HFSIA on all 22 datasets involved in the experiment is analyzed. The experimental results are compared with the results using all features. These comparisons include the reduction degree of redundant features and the improvement of classifier performance. Secondly, the HFSIA algorithm is compared with several feature selection methods reported in other literature. These feature selection methods include 5 classical filtering feature selection algorithms and 18 metaheuristic-based feature selection methods reported in the latest literature. The experimental analysis and comparison will be carried out in the following three aspects. They are the classification accuracy and the number of features of the optimal feature subset, as well as the computational cost of the algorithm.

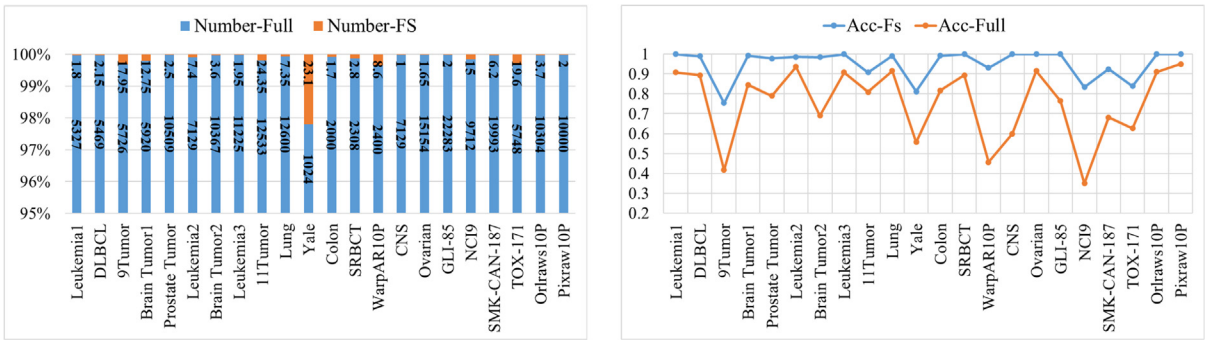
5.1. Performance evaluation

In this section, the effectiveness of the HFSIA feature selection method in improving classifier performance is verified by experiments. Fig. 5(a) shows the quantitative comparison between the optimal feature subset obtained using HFSIA and the full features of the dataset. Fig. 5(b) depicts a comparison between the classification accuracy obtained by the kNN classifier with or without HFSIA feature selection method. It should be noted that the results in Fig. 5 are the average number and average classification accuracy of the optimal feature subsets of each dataset after repeated execution 20 times.

As can be seen from Fig. 5(a), HFSIA obtained very good performance in eliminating feature redundancy on all experimental datasets. The elimination rates of feature redundancy for all datasets are above 97.7%. Statistically, the optimal number of features in the subset of features filtered by HFSIA is within 2.26% of the total number of features in all the participating experimental datasets. In them, the largest percentage of the average number of features in the optimal feature subset is *Yale*, with a ratio of 2.26%. The smallest percentage is obtained by *GLI-85*, with a ratio of only 0.00898%.

From the perspective of improving the classification performance of the classification algorithm, in all datasets, the HFSIA method improves the classifier performance by 5%-48.333%. On 52% of the datasets, HFSIA improves classification performance by more than 10%. On 36% of the datasets, it improves classification performance by more than 20%. On 16% of the datasets, it improves classification performance by more than 30%. The datasets with the highest classification performance improvement are *NCI9*, *WarpAR10P* and *CNS*. On these datasets, the performance of the classifier is improved by more than 40%.

From the comparative analysis of the above two aspects, it can be concluded that HFSIA achieves better classification accuracy than the entire feature space with a very small number of



(a) Feature number comparison.

(b) Accuracy of kNN with FS and full.

Fig. 5. The Classification performance with HFSIA and full.

features. This fully demonstrates the effectiveness of HFSIA in eliminating redundant features.

In addition, in order to observe the convergence speed of the algorithm, 9 datasets were selected in stages according to the number of features of the experimental data to plot the variation curve of their fitness values. The variation of the fitness value with the number of iterations is shown in Fig. 6. In the figure, the horizontal axis is the number of iterations and the vertical axis is the fitness value. The fitness value records the evaluation value of the subset of locally optimal features selected during each iteration. It shows the category identification performance of the subset. According to the definition of evaluation metrics in Section 3.2, the fitness value is the classification error rate resulting from the classification evaluation of this feature subset. That is, the lower the fitness value of the subset, the better its class identification, and the more useful it is for the classifier. The shape of the fitness curve shows the searchability of the algorithm. The faster the fitness curve declines, the better the searchability of the algorithm. From the figure, it can be seen that HFSIA achieves within 20 iterations from the initial value to the optimal value on different datasets with different numbers of features (2308–22,283). And it can achieve fast convergence to the optimal value within 10 iterations on more than 80% of the datasets. This demonstrates the powerful global search capability of HFSIA. It also implies that the improvement of the basic clone selection algorithm in this paper is successful and effective.

5.2. Comparative analysis

In this paper, HFSIA is analyzed in comparison with many advanced feature selection algorithms reported in the recent literature. To verify the computational efficiency of HFSIA, it is compared with five classical filtered feature selection algorithms known for their speed. In addition, to verify the advancedness of the performance of the proposed algorithm, it is compared with 18 feature selection methods reported in the recent literature. These comparative analyses include the following three aspects. They are the feature reduction rate, the classification accuracy, and the computational cost of the algorithm. In all tabular data, the best results for each criterion are marked in bold.

5.2.1. Comparison with classical feature selection methods

This paper conducts comparative experiments with 5 classical feature selection methods on 10 benchmark datasets. The five methods are CFS (statistical-based) [50], FCBF (information theoretical-based) [51], ReliefF (similarity-based) [52], SBMLR (sparsity-based) [53] and SPEC (graph theory-based) [54]. The experimental results are shown in Tables 3–5. Table 3 presents a comparison of the classification accuracies of the optimal feature subsets obtained by different feature selection methods. Table 4 describes the comparison of the number of feature subsets for different algorithms to achieve optimal accuracy. Table 5 describes the computational cost of all algorithms to achieve optimal accuracy on these datasets. The classification accuracy data in the table is the best value obtained after 20 runs on each dataset. In this comparative experiment, all the results are obtained with the same classification algorithm and experimental parameter settings. It should be noted that the experimental data of the five classical feature selection methods in the table are all from the literature [28].

In terms of improving the performance of the classifier, the following results can be obtained from the observation and comparison of the data in Table 3. In the 10 datasets participating in the experiment, the classification accuracy of HFSIA on all datasets is higher than that of the other 5 classical feature selection methods. According to the statistics in Table 3, on these datasets, the classification accuracy obtained by HFSIA is 4.11%–32% higher than the maximum value of the other five algorithms. On 40% of the dataset, HFSIA outperforms the maximum classification accuracy obtained by other methods by more than 10%. The highest proportion of classification accuracy is 9Tumor and Brain Tumor2. On these datasets, the performance gains of the classifiers are more than 30% higher than the maximum value of other methods. This fully shows that compared with these five classical feature selection methods, the HFSIA method is the best in improving the performance of the classifier.

In terms of eliminating redundant features, the following results can be drawn from the data in Table 4. In these 10 datasets, the optimal feature subset obtained by HFSIA has a lower number of features than other methods. It is only 2.94%–23.08% of the minimum value of other methods. On 80% of the datasets, the optimal subset obtained by HFSIA has less than 15% of the minimum features of other methods. On 60% of the datasets, the number of features is below 10% of the minimum of other methods. The smallest proportion of features is *DLBCL*, *Prostate Tumor*, *Leukemia3*, and *Lung*. On these datasets, the number of features of the optimal feature subset obtained by HFSIA is all below 7% of the minimum value of other methods. This fully shows that the HFSIA method has the best effect in eliminating redundant features compared with these five classical feature selection methods.

In terms of the computational cost of the algorithm, the following results can be drawn from the data in Table 5. Among the five classical feature selection methods involved in the experiment, the SPEC method has the lowest computational cost. It is the fastest on all datasets. This is followed by the ReliefF and FCBF methods, which are close in computational cost on 70% of the dataset and outperform the rest of the methods. Again, the SBMLR method, which outperforms the CFS method on all

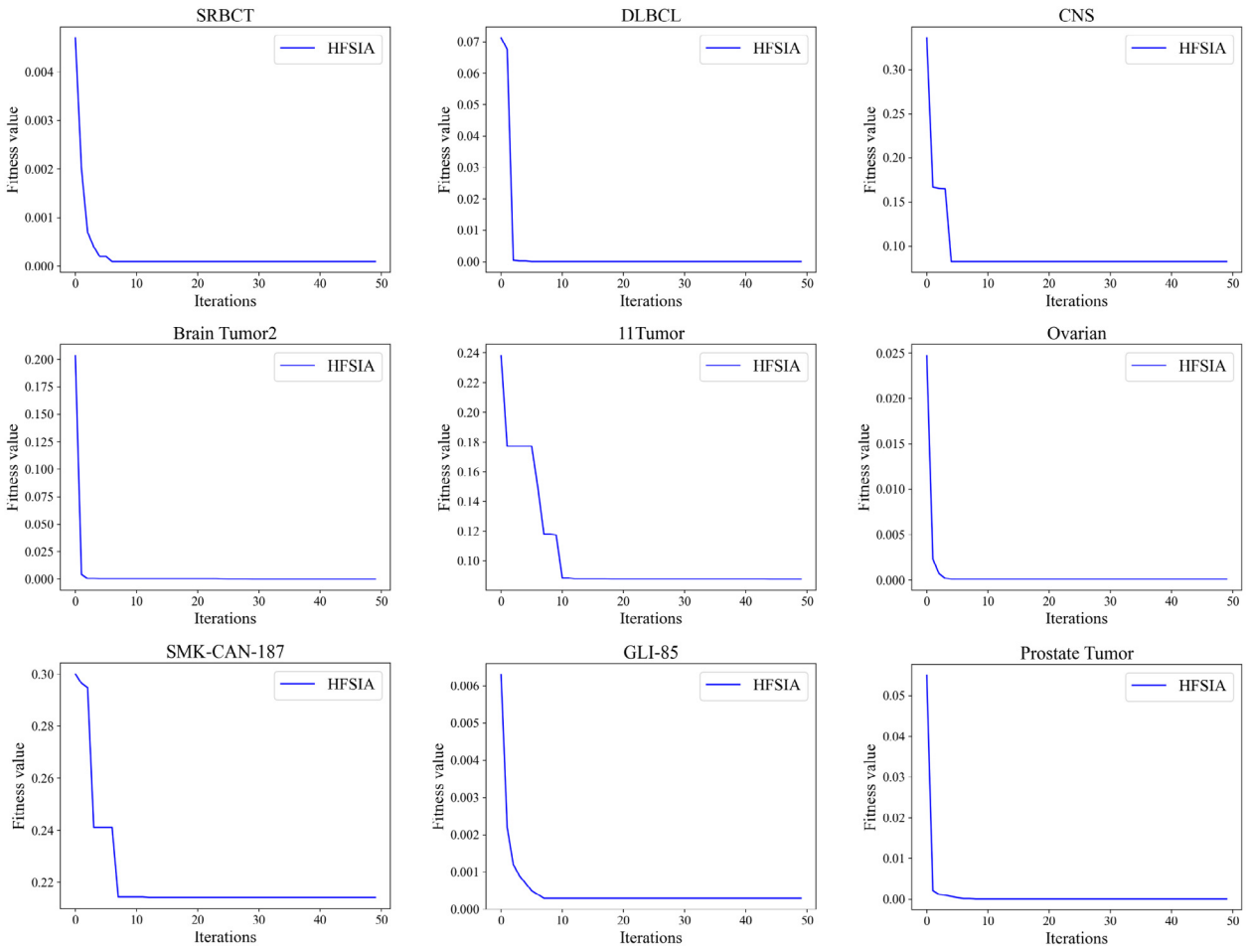


Fig. 6. Change curve of fitness value on 8 datasets.

Table 3
Comparison of classification accuracy of optimal feature subsets obtained by different methods.

Dataset	CFS	FCBF	ReliefF	SBMLR	SPEC	HFSIA
Leukemia1	95.89	91.61	94.46	93.04	94.29	**100**
DLBCL	91.96	93.21	92.82	92.75	74.11	**100**
9Tumor	56.67	41.67	61.67	48.33	38.33	**91.667**
Brain Tumor1	86.67	80	85.78	82.22	84.44	**100**
Prostate Tumor	94.09	91.12	92.09	95.09	85.36	**100**
Leukemia2	88.57	85.89	91.29	84.64	90	**100**
Brain Tumor2	68	64	68	64	48	**100**
Leukemia3	94.01	93.14	94.29	93.04	84.29	**100**
11Tumor	83.91	82.94	84.91	70.13	83.3	**97.059**
Lung	93.31	92.06	90.17	92.62	81.29	**100**

Table 4
Comparison of the number of features of the optimal subsets obtained by different methods.

Dataset	CFS	FCBF	ReliefF	SBMLR	SPEC	HFSIA
Leukemia1	97	49	603	14	2489	**1**
DLBCL	88	66	343	34	2371	**1**
9Tumor	47	32	544	26	1049	**6**
Brain Tumor1	142	106	771	27	5388	**4**
Prostate Tumor	59	49	433	32	405	**1**
Leukemia2	119	71	645	18	3544	**2**
Brain Tumor2	117	75	1171	21	4420	**2**
Leukemia3	138	80	978	17	4917	**1**
11Tumor	379	394	1114	15	6158	**13**
Lung	550	453	1440	30	2378	**2**

Table 5
Computational cost comparison of different feature selection methods (s).

Dataset	CFS	FCBF	ReliefF	SBMLR	SPEC	HFSIA
Leukemia1	741.57	1.49	1.3	0.92	**0.19**	4.8145
DLBCL	685.15	1.52	1.37	2.01	**0.34**	4.194
9Tumor	652.7	1.53	1.11	7.14	**0.41**	5.4977
Brain Tumor1	1248.17	2.69	2.04	9.2	**0.66**	5.8385
Prostate Tumor	572.82	3.57	2.21	9.84	**0.83**	4.9926
Leukemia2	810.32	1.68	1.63	10.79	**0.99**	5.255
Brain Tumor2	765.5	2.84	1.54	11.5	**1.16**	5.3947
Leukemia3	677.51	4.04	2.84	13.46	**1.4**	5.341
11Tumor	4681.4	25.06	15.54	13.87	**2.28**	7.0041
Lung	10029	37.95	16.69	28.13	**3.11**	6.2198

datasets. Undoubtedly, the CFS method is the most computationally expensive among these 5 classical feature selection methods. By comparing the data in the table, it can be concluded that the HFSIA method proposed in this paper outperforms the SBMLR method on 80% of the datasets. It outperforms FCBF and ReliefF methods on 20% of the dataset. It outperforms the CFS method on all datasets. It can be concluded that the computational cost of HFSIA is between SBMLR and FCBF. This fully demonstrates that HFSIA is very competitive in terms of computational cost control, even compared with classical filtering feature selection methods known for their speed.

After summarizing the above analysis, the following conclusions can be drawn. In terms of computational cost alone, the HFSIA algorithm proposed in this paper is comparable to the classical feature selection method known for its speed. Moreover, the obtained feature subset is much better than these 5 classical feature selection methods in terms of quantity and performance improvement of the classifier. To sum up, compared with the five classical feature selection methods, the HFSIA method proposed in this paper can obtain a higher-quality feature subset while taking into account the computational cost.

5.2.2. Comparison with other hybrid feature selection methods

In this paper, 22 benchmark datasets are compared with 18 other feature selection methods for high-dimensional data reported in the latest literature. It should be noted that based on the consideration of many uncertainties such as the experimental environment, algorithm source code, and algorithm parameters, the experimental data of the comparative algorithm in the following table are all the experimental results disclosed in the corresponding comparative literature.

Tables 6–8 describe the comparison of experimental results between HFSIA and the feature selection method mentioned in [28]. In the literature, the authors propose an evolutionary multitask-based feature selection method (PSO-EMT) and use it to solve the classification problem of high-dimensional data. It performs well in improving classification accuracy and computational cost. This paper compares this method with 4 other feature selection algorithms on 10 gene expression datasets. These datasets all have high dimensionality, the number of features varies from 5327 to 12,600, and the number of samples is small. These 4 feature selection methods are PSO, CSO, AMSO, and VLPSO. In the experiment, the classification results of the 10-fold cross-validation of the kNN algorithm are used as the basis for the performance evaluation of the algorithm. The experimental results are shown in Tables 6–8. Table 6 describes the comparison of the average classification accuracy between the HFSIA algorithm and the five methods after 20 repetitions on all experimental datasets. Table 7 describes the comparison between the average numbers of optimal feature subsets obtained by different methods. Table 8 describes the average computational cost of all algorithms on these 10 datasets.

In terms of improving the performance of the classifier, the following results are obtained by observing and comparing the data in Table 6. Among the 10 datasets participating in the experiment, the classification accuracy of HFSIA is higher than that of the other 5 feature selection methods on all datasets. After statistical analysis of the data in the table, the following results are obtained. On these datasets, the classification accuracy obtained by HFSIA is 4.4%–17.77% higher than the maximum value of the other five algorithms. On 80% of the datasets, HFSIA outperforms the maximum classification accuracy obtained by other methods by more than 5%. On 30% of the datasets, HFSIA outperforms the maximum values of other methods by more than 10%. The highest proportion of classification accuracy is *9Tumor* and *Brain Tumor2*. On these datasets, the classification accuracy obtained by HFSIA is more than 17% higher than the maximum value of other methods. This fully shows that the HFSIA method is the best in improving the performance of the classifier compared with these five feature selection methods.

In terms of eliminating redundant features, the following results can be drawn from the data in Table 7. In all datasets participating in the experiment, the number of features of the optimal feature subset obtained by the HFSIA method is less than 50% of the minimum value of other methods. It is only 3.5%–47.52% of the minimum value of other methods. On 80% of the datasets, the number of features of the optimal feature subset obtained by the HFSIA method is below 14% of the minimum value of the other methods. On 70% of the dataset, it is below 10% of the minimum of other methods. The optimal feature subsets with the smallest proportion of features are *Leukemia1*, *Brain Tumor2*, *Leukemia3*, and *Lung*. On these datasets, the number of features of the optimal feature subset obtained by HFSIA is less than 6% of the minimum value of other methods. This fully shows that the HFSIA method has the best performance in eliminating redundant features compared with these five methods.

In terms of the computational cost of the algorithm, the following results can be drawn from the data in Table 8. Among the five feature selection methods participating in the comparative experiments, the VLPSO method has the lowest computational cost. Its computational cost is lower than the other 4 methods on 70% of the datasets. This is followed by AMSO and PSO-EMT methods, which are close in computational cost on 60% of the dataset and outperform the rest of the methods. Next is PSO, which outperforms CSO methods on all datasets. Undoubtedly, the CSO method is the most computationally expensive of them all. By comparing the data in the table, it can be concluded that the HFSIA method proposed in this paper outperforms the five methods on 90% of the datasets. Only 8.6%–87.84% of the minimum value of other methods. Computational cost on 80% of the datasets is below 70% of the lowest value of other methods. The computational cost on 30% of the datasets is below 40% of the lowest value of other methods. Among them, *11Tumor* and *Lung* have the lowest computational cost, which is less than 12% of the lowest value of other methods. This fully shows that, compared with these five feature selection methods, the HFSIA has significant advantages in controlling the computational cost.

Table 6
Comparison of classification accuracy of optimal feature subsets obtained by different methods.

Dataset	PSO	CSO	AMSO	VLPSO	PSO-EMT	HFSIA
Leukemia1	80.60/2.55	90.79/2.88	94.01/1.58	93.31/2.34	91.11/2.79	**100/0**
DLBCL	83.67/1.52	94.60/3.26	94.10/1.95	86.51/2.88	93.76/2.80	**99/2.44**
9Tumor	42.72/1.42	59.78/3.55	50.11/3.61	54.94/4.80	58.00/4.02	**75.42/7.39**
Brain Tumor1	73.73/2.21	80.41/3.93	72.67/3.79	71.19/3.52	87.37/1.50	**99.17/2.04**
Prostate Tumor	84.50/1.64	79.95/3.18	89.58/1.35	88.74/2.23	89.65/1.82	**97.75/2.55**
Leukemia2	78.61/2.02	80.83/2.28	87.52/2.00	85.82/2.96	90.07/2.47	**98.57/3.01**
Brain Tumor2	61.99/2.91	80.73/5.62	74.96/3.48	66.78/4.10	72.27/4.09	**98.5/3.66**
Leukemia3	89.83/1.00	91.49/3.84	94.45/1.04	91.56/1.67	94.51/1.50	**100/0**
11Tumor	71.81/1.75	83.50/1.70	83.10/1.31	80.92/2.39	86.15/1.45	**90.74/4.19**
Lung	78.77/1.53	88.94/1.75	89.97/1.80	89.55/1.68	91.09/0.94	**99.13/1.6**

Table 7
Comparison of the number of features of the optimal subsets obtained by different methods.

Dataset	PSO	CSO	AMSO	VLPSO	PSO-EMT	HFSIA
Leukemia1	2615.5	170.12	51.49	54.7	198.4	**1.8/0.41**
DLBCL	2681	30.08	50.56	48.14	83.55	**2.15/2.11**
9Tumor	2811.9	220.34	52.16	47.05	263.09	**17.95/4.637**
Brain Tumor1	2917.2	207.61	93.54	26.83	351.21	**12.75/3.117**
Prostate Tumor	2926.6	207.98	44.36	35.97	149.86	**2.5/1.147**
Leukemia2	3513.8	389.4	71.54	53.39	224.44	**7.4/5.336**
Brain Tumor2	5117.2	90.43	62.08	81.46	499.69	**3.6/1.903**
Leukemia3	5535.7	88.64	57.19	35.23	268.08	**1.95/0.605**
11Tumor	6205	589.36	319	249.3	541.45	**24.35/3.104**
Lung	6234.7	230.41	193.47	176	617.61	**7.35/4.848**

Table 8
Computational cost comparison of different feature selection methods (*s*).

Dataset	PSO	CSO	AMSO	VLPSO	PSO-EMT	HFSIA
Leukemia1	41.2	247.29	6.8	6.09	9.28	**5.3494/0.252**
DLBCL	47.59	389.67	8.34	7.18	7.02	**4.7939/0.34**
9Tumor	39.18	370.4	**5.52**	5.65	8.09	6.4389/0.396
Brain Tumor1	66.65	457.24	11.65	9.55	15.43	**6.4799/0.303**
Prostate Tumor	78.77	410.83	14.31	11.05	16.88	**5.5481/0.441**
Leukemia2	66.09	445.99	9.66	8.96	12.19	**6.0048/0.485**
Brain Tumor2	80.5	945.7	12.06	11.76	11.51	**5.7796/0.334**
Leukemia3	120.64	1837.91	15.64	15.94	14.72	**5.7177/0.223**
11Tumor	418.54	6278.54	91.22	67.41	106.53	**7.5514/0.651**
Lung	574.17	5419.71	255.32	78	134.59	**6.7417/0.313**

Through the above comparative analysis, the following conclusions can be drawn. The optimal feature subset obtained by HFSIA is better than the other five methods in average classification accuracy and average number. Moreover, its computational cost on 90% of the datasets is better than these five methods. Therefore, the following conclusions can be further drawn. Compared with these five feature selection methods, the HFSIA algorithm proposed in this paper has strong competitive advantages in both the quality of the optimal feature subset and the computing speed of the algorithm. This fully proves the progressiveness of the HFSIA algorithm in solving the feature selection problem of high-dimensional data.

Tables 9–11 describe the comparison of experimental results between HFSIA and the feature selection method mentioned in [27]. In this paper, a hybrid feature selection method (HFS-C-P) based on correlation-guided clustering and particle swarm optimization is proposed to solve the classification problem of high dimensional microarray data. This method performs well in improving classification accuracy and calculation cost. This paper compares this method with other 7 advanced feature selection algorithms on 9 gene expression datasets. The seven feature selection methods are BPSO, BBPSO, Rc-BBFA, SaPSO, HPSO-SSM, BPSO+fiter, and H-BQPSO, respectively. The experimental results are shown in Tables 9–11. Table 9 describes the comparison of the average classification accuracy of the HFSIA algorithm with the 7 feature selection methods on these datasets. Table 10 describes the comparison between the average numbers of optimal feature subsets obtained by these methods. Table 11 describes the average computational cost of all algorithms on these datasets.

In terms of improving the performance of the classifier, the following results are obtained by comparing the data in Table 9. Among the eight feature selection methods involved in the comparative experiments, HFS-C-P obtained the highest classification accuracy and outperformed other methods on all datasets. The average classification accuracy of the proposed HFSIA method on 66.7% of the datasets is greater than or equal to the optimal value of other methods. This fully shows that compared with these 8 feature selection methods, the HFSIA method has strong competitiveness in improving the performance of the classifier.

In terms of feature number control for the optimal feature subset, the following results can be obtained after comparing the data in Table 10. The HFSIA method outperforms the other methods on all datasets participating in the experiment. The number of optimal feature subsets is only 3.42%–72.19% of the minimum values of other methods. Moreover, in 77.78% of the datasets, the number of features in the optimal feature subset obtained by the HFSIA method is less than 30% of the minimum of the other methods. On 55.56% of the datasets, it is below 17% of the minimum value of other methods. The features with the smallest proportion of the optimal feature subset are *CNS*, *Colon*, and *SRBCT*. On these datasets, the number of optimal feature subsets obtained by HFSIA is less than 15% of the minimum of the other methods. This is a good indication that the HFSIA method performs well in eliminating redundant features compared to the feature selection methods in this table.

Table 9
Comparison of classification accuracy of optimal feature subsets obtained by different methods.

Datasets	BPSO	BBPSO	Rc-BBFA	SaPSO	HPSO-SSM	BPSO+fiter	H-BQPSO	HFS-C-P	HFSIA
Yale	77.49(+)	78.05(+)	78.10(+)	75.07(+)	75.24(+)	67.06(+)	71.73(+)	79.52	**81.21/0.047**
Colon	90.52(+)	90.79(+)	90.70(+)	83.04(+)	87.40(+)	88.11(+)	89.73(+)	92.47	**99.17/0.026**
SRBCT	94.94(+)	96.03(+)	96.18(+)	98.93(+)	99.01(=)	97.73(+)	98.78(+)	**100**	**100/0**
WarpAR10P	86.43(+)	88.24(+)	86.90(+)	93.05(+)	95.56(+)	89.31(+)	92.33(+)	**100**	93.08/0.04
DLBCL	82.46(+)	83.16(+)	82.57(+)	91.54(+)	95.85(+)	97.78(+)	99.29(+)	**100**	99/0.024
leukemia2	74.40(+)	75.02(+)	74.15(+)	94.89(+)	97.69(+)	100.00(=)	100.00(=)	**100**	98.57/0.03
CNS	76.23(+)	74.77(+)	74.55(+)	71.87(+)	72.48(+)	83.69(+)	80.88(+)	85.91	**100/0**
Lung	53.28(+)	60.55(+)	55.36(+)	78.93(+)	93.71(+)	87.02(+)	96.67(+)	98.01	**99.13/0.016**
Ovarian	89.70(+)	94.97(+)	90.09(+)	97.91(+)	99.43(=)	99.55(=)	100.00(=)	**100**	**100/0**

Table 10
Comparison of the number of features of the optimal subsets obtained by different methods.

Datasets	BPSO	BBPSO	Rc-BBFA	SaPSO	HPSO-SSM	BPSO+fiter	H-BQPSO	HFS-C-P	HFSIA
Yale	509.6	507	507.2	505.8	450	54	32	357.4	**23.1/5.384**
Colon	977	805.2	1002	796.4	591.6	30.6	29.4	155	**1.7/0.823**
SRBCT	112.2	1 099.0	1130.8	915.8	357.6	19.6	26.4	99.4	**2.8/0.696**
WarpAR10P	1174.6	1171.4	1172.4	949.4	186.6	82.4	29	204.6	**8.6/5.892**
DLBCL	2717.2	2732.4	2725	2196	589.8	36.2	12.8	227	**2.15/2.11**
leukemia2	3588.5	2924.3	3547.5	2831.8	655.2	48.2	59	152.7	**7.4/5.336**
CNS	3564.7	3542.3	3613	2843.4	1941.8	622	29.2	367.6	**1/0**
Lung	6295.4	5155.2	6295.8	5054.4	504.2	275	19.6	522.4	**7.35/4.848**
Ovarian	7577	6114.6	7495.4	6066	409.8	277	7	42.5	**1.65/0.489**

Table 11
Computational cost comparison of different feature selection methods (m).

Datasets	BPSO	BBPSO	Rc-BBFA	SaPSO	HPSO-SSM	BPSO+fiter	H-BQPSO	HFS-C-P	HFSIA
Yale	36.384	28.882	35.705	26.08	27.265	7.808	9.414	22.693	**0.3521/0.04**
Colon	14.76	10.184	14.766	12.231	11.238	3.311	2.936	5.974	**0.0627/0.009**
SRBCT	18.049	15.014	24.462	15.801	11.948	6.993	6.817	6.67	**0.1006/0.004**
WarpAR10P	40.483	36.485	46.341	27.048	29.617	9.066	5.574	12.494	**0.3002/0.03**
DLBCL	56.455	43.554	55.872	34.877	43.075	4.461	2.95	6.766	**0.0799/0.006**
leukemia2	67.845	30.059	57.388	38.984	42.902	9.512	8.753	6.278	**0.1001/0.008**
CNS	46.169	24.043	48.535	34.408	34.679	4.416	3.147	7.874	**0.1487/0.008**
Lung	564.316	405.192	564.05	343.779	357.687	48.521	4.969	16.595	**0.1124/0.005**
Ovarian	1205.176	961.59	996.491	502.948	669.449	48.609	9.359	7.859	**0.2033/0.013**

In terms of the computational cost of the algorithms, a comparison of the data in Table 11 shows the following results. Among the 8 feature selection methods involved in the comparison experiments, the HFSIA method outperforms the other methods on all datasets. Moreover, its computational cost is only 1.51%–5.39% of the minimum value of other methods. Its computational cost on 88.89% of the datasets is below 5% of the minimum of the other methods. On 66.67% of the datasets, the computational cost is less than 3% of the minimum of other methods. Among them, the lowest percentage is *Colon* and *leukemia2*, whose computational cost is less than 1.6% of the lowest value of other methods. This fully shows that the HFSIA method has the highest computational efficiency compared to these eight feature selection methods.

The following conclusions can be drawn from the above analysis. Compared with the 8 feature selection methods involved in the experiment, the HFSIA algorithm has strong advantages in improving the classifier performance and reducing redundant features and computational cost. This shows that the HFSIA algorithm can obtain higher quality feature subsets with less computational cost. This also fully demonstrates the superiority of the HFSIA algorithm proposed in this paper.

Tables 12–13 describe the comparison of the experimental results between HFSIA and the feature selection methods mentioned in [29]. In the literature, a two-stage hybrid feature selection method (TAGA) that integrates filters and wrappers was proposed. The method uses an evolutionary algorithm-based filter feature selection algorithm to solve the classification problem of high-dimensional data and achieves a better performance in improving the classification accuracy. This paper compares this method and 4 other algorithms on 8 benchmark datasets. The four feature selection methods are mRMR-mid, QPFS, SPECCMI, and CGA, respectively. The experimental results are shown in Tables 12 and 13. Table 12 describes the comparison of the average classification accuracy of the HFSIA algorithm with the five feature selection methods. Table 13 describes the comparison between the average numbers of optimal feature subsets obtained by these methods.

In terms of the classification accuracy of the selected optimal feature subset, the following results are obtained by comparing the data in Table 12. Among the six feature selection methods involved in the experiment, the HFSIA method proposed in this paper outperforms the other five methods on 87.5% of the datasets. On these datasets, the classification accuracy obtained by HFSIA is 0.03%–18.13% higher than the optimal values of the other five algorithms. Among them, the average classification accuracy obtained on 50% of the dataset is more than 3% higher than the best value of other methods. This fully shows that, compared with these five feature selection methods, the optimal feature subset obtained by HFSIA has a strong competitive advantage in classification performance.

In terms of eliminating redundant features, the following results can be drawn from the comparison of the data in Table 13. In all datasets participating in the experiment, the HFSIA method outperformed the other 5 methods on 75% of the datasets, only 28.57%–88.57% of the minimum values of the other methods. And, on 62.5% of the datasets, the number of optimal feature subsets obtained by the HFSIA method is below 42% of the minimum values of other methods. The smallest proportion of features in the optimal feature subset is Lymphoma and Pixraw10P. On these

Table 12
Comparison of classification accuracy of optimal feature subsets obtained by different methods.

Algorithm	CLN	GLI	NCI	SMK	TOX	LYM	ORP	PIW
TAGA	94.03/0.8	98.8/0.0	79.5/0.4	75.2/0.5	77.7/0.8	94.3/0.5	99.8/0.4	97.0/0.0
mRMR-mid	98.4	95.3	65	68.4	72.5	**97.9**	97	96
QPFS	91.9	94.1	83.3	74.3	72.5	**97.9**	98	97
SPECCMI	93.5	96.5	80	71.1	77.2	93.8	94	95
CGA	95.2/1.3	95.9/0.6	78.8/2.4	72.7/1.5	74.3/2.6	94.1/1.0	98.5/1.0	97.2/0.4
HFSIA	**99.17/2.64**	**100/0**	**83.33/5.56**	**92.43/3.32**	**83.82/3.73**	93.16/2.69	**100/0**	**100/0**

* Note: CLN, GLI, NCI, SMK, TOX, LYM, ORP, PIW in the table represent the datasets Colon, GLI-85, NCI9, SMK-CAN-187, TOX-171, Lymphoma, Orlraws10P, Pixraw10P, respectively.

Table 13
Comparison of the number of features of the optimal subsets obtained by different methods.

Algorithm	CLN	GLI	NCI	SMK	TOX	LYM	ORP	PIW
AGA	10.6/8.1	14.8/4.3	40.5/3.5	13.1/4.4	23.6/6.5	20.3/3.7	13.4/4.7	8.1/0.7
mRMR-mid	**1**	5	43	7	28	22	20	7
QPFS	3	12	39	17	**15**	23	9	20
SPECCMI	32	16	38	13	24	36	21	11
CGA	7.9/3.9	16.2/7.0	39.0/5.8	8.8/6.3	28.8/3.9	30.3/7.0	13.6/3.6	10.2/5.1
HFSIA	1.7/0.82	**2/0.67**	**15/11.19**	**6.2/2.94**	19.6/17.87	**6.95/3.34**	**3.7/0.95**	**2/0**

Table 14
Computational cost comparison of different feature selection methods (s).

Algorithm	CLN	GLI	NCI	SMK	TOX	LYM	ORP	PIW
Algorithm	CLN	GLI	NCI	SMK	TOX	LYM	ORP	PIW
TAGA	123	122	127	139	132	117	129	148
SFS	34	48	33	140	123	59	62	63
BE	183	223	162	649	587	294	288	301
CGA	158	181	168	183	174	161	199	183
HFSIA	**3.76/0.56**	**4.46/0.2**	**5.79/0.06**	**9.73/0.88**	**5.29/0.19**	**14.05/0.52**	**5.69/0.27**	**4.66/0.06**

datasets, the optimal number of optimal feature subsets obtained by HFSIA is less than 32% of the minimum value of other methods. This fully shows that, compared with these five feature selection methods, the redundancy of the optimal feature subset selected by the HFSIA method is very advantageous.

The following conclusions can be drawn from the above analysis. Compared with the five feature selection methods involved in the experiment, the HFSIA algorithm has strong advantages in improving the performance of the classifier and reducing redundant features. This shows that the optimal feature subset obtained by the HFSIA algorithm has higher quality. This also fully proves the superiority of the HFSIA algorithm proposed in this paper.

Table 14 depicts the average computational cost of HFSIA with the other four algorithms (TAGA, SFS, BE and CGA) on these datasets. A comparison of the data in the table yields the following results. Among the five feature selection methods involved in the experiment, the HFSIA method outperformed the other methods on all datasets. Moreover, its computational cost is only 4.3%–23.83% of the minimum of the other methods. Its computational cost on 75% of the datasets is below 15% of the lowest value of other methods. The computational cost on 62.5% of the datasets is below 10% of the lowest value of other methods. Among them, *SMK-CAN-187*, *TOX-171*, and *Pixraw10P* have the lowest computational cost, which is less than 7.5% of the lowest value of other methods. This fully demonstrates that the HFSIA method has higher computational efficiency compared with these five feature selection methods.

Through the comparative analysis of the above experimental results, it can be concluded that, combined with the evaluation results of the two indicators of the quality of the optimal feature subset and the computational cost, the HFSIA method has an outstanding competitive advantage in feature selection of high-dimensional data compared with the 18 feature selection methods reported in the latest literature.

6. Ablation experiments

Ablation experiments were also performed to verify the necessity of the functional modules of each part of the proposed model. In this paper, ablation tests are performed on all datasets involved in the experiments. For the sake of simplicity, only the classification results of the kNN algorithm are used as the analysis metrics for the experiments. As shown in Fig. 7, the experimental results under each metric are presented. Among them, "Full" represents the method of removing all feature subset evaluation modules. "Fisher (200)" represents a module that removes the search part of the artificial immune algorithm. HFSIA means the fusion of all functional modules. Through the observation and analysis of the experimental results, the following results can be obtained. From the performance of each module on the test dataset, the results obtained by HFSIA are significantly higher than the other two schemes and have very obvious advantages. From the performance of the kNN algorithm in classification accuracy on all datasets, the fusion method of HFSIA has a better performance advantage than any other individual method. Through the above analysis, the following conclusions can be drawn. For the model proposed in this paper, the fusion scheme of HFSIA is effective, which is very helpful for the performance improvement of the classifier.

7. Conclusions

In this paper, an efficient hybrid feature selection method (HFSIA) based on an artificial immune algorithm is proposed. The algorithm combines the Fisher filter algorithm and improved artificial immune algorithm to optimize the search process of the optimal feature subset for high-dimensional data. According to the target requirements of feature selection, the method improves the population initialization and mutation strategy of the antibody in the algorithm, as well as the population update method.

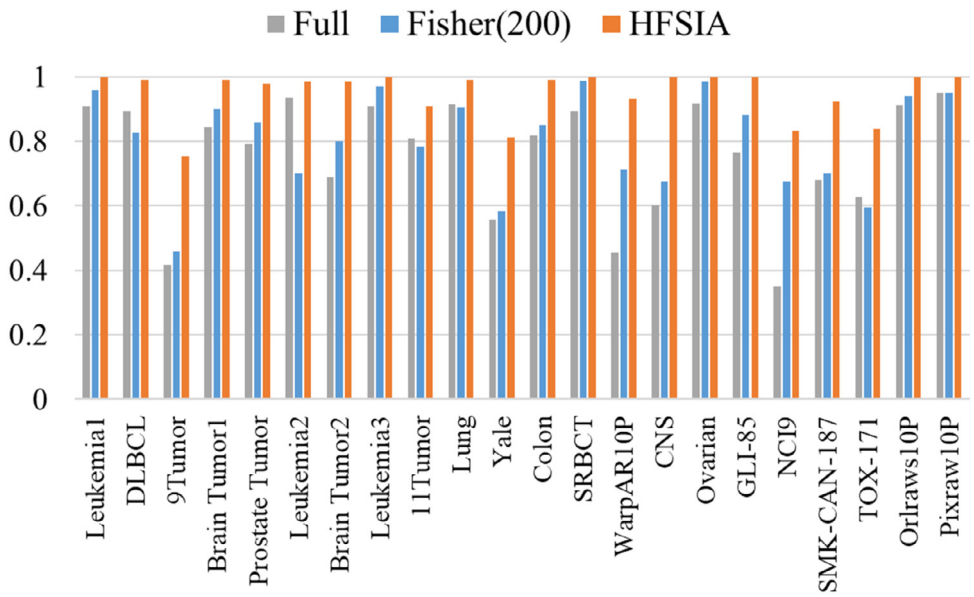


Fig. 7. Relationship between Fisher Score screening threshold and classification accuracy of GLI-85.

In order to verify the effectiveness of the HFSIA algorithm, extensive experimental validation and analysis are conducted on 22 high-dimensional datasets with a number of features ranging from 1024 to 22,283. These experimental analyses cover the following three aspects. (1) The algorithm improves the classification performance of the classifier. The experimental results obtained by HFSIA are analyzed in comparison with that without feature selection. These analyses include the reduction of feature redundancy and the improvement of classification accuracy. In addition, the convergence performance of the algorithm is experimentally verified. (2) Comparative analysis with other feature selection methods. The experimental results are compared with the results of 23 feature selection methods mentioned in other literature. These feature selection methods include 5 classical feature selection algorithms and 18 metaheuristic feature selection methods reported in the latest literature. These comparative analyses include classification accuracy, the number of optimal feature subsets, and the computational cost of the algorithm. (3) The structural validity analysis of the algorithm itself. Ablation experiments are performed to verify the necessity of the functional modules of each part of the proposed model.

Based on the results of the above analysis, the following conclusions can be drawn. The optimal feature subset obtained by HFSIA is effective in improving the performance of the classifier. Compared with the five classical Filter algorithms, the quality advantage of the optimal feature subset obtained by HFSIA is obvious and it is competitive in controlling the computational cost. The search efficiency performs better than the SBMLR algorithm on 80% of the datasets and significantly outperforms the CFS algorithm on all data. Compared with the 18 feature selection methods proposed in the recent literature, the feature subsets obtained by HFSIA have lower redundancy and higher average classification accuracy. Moreover, its advantage in computational efficiency is obvious, significantly outperforming the comparison algorithms on 9/10, 9/9, and 8/8 datasets, respectively. Therefore, it can fully illustrate the effectiveness and progressiveness of this method in solving the problem of feature selection of high-dimensional data. In the future work, we will deeply study the performance of the HFSIA algorithm on ultra-high dimensional datasets, and further improve the performance of the algorithm. In addition, we hope to further improve its efficiency.

CRediT authorship contribution statement

Yongbin Zhu: Conceptualization, Methodology, Software, Writing – original draft, Validation, Formal analysis, Investigation, Visualization, Writing – review & editing. **Wenshan Li:** Review & editing. **Tao Li:** Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared my data/code in the attached file named 'HFSIA'.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China (No. 2020YFB1805400); in part by the National Natural Science Foundation of China (No. U1736212, No. U19A2068, No. 62032002, and No. 62002248); in part by the China Postdoctoral Science Foundation (No. 2019TQ-0217, No. 2020M673277 and 2020M683345); in part by the Provincial Key Research and Development Program of Sichuan, PR China (No. 20ZDYF3145); in part by the Fundamental Research Funds for the Central Universities, PR China (No. YJ201933); in part by the China International Postdoctoral Exchange Fellowship Program (Talent-Introduction).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.knosys.2022.110111.

References

[1] B.H. Nguyen, B. Xue, M. Zhang, A survey on swarm intelligence approaches to feature selection in data mining, Swarm Evol. Comput. 54 (2020) 100663, http://dx.doi.org/10.1016/j.swevo.2020.100663.

[2] J. Hua, W.D. Tembe, E.R. Dougherty, Performance of feature-selection methods in the classification of high-dimension data, Pattern Recognit. 42 (3) (2009) 409–424, http://dx.doi.org/10.1016/j.patcog.2008.08.001.

[3] M. Dash, H. Liu, Feature selection for classification, Intell. Data Anal. 1 (3) (1997) 131–156, http://dx.doi.org/10.3233/IDA-1997-1302.

[4] Z. Sun, Parallel Feature Selection Based on MapReduce, Springer, 2014, pp. 299–306, http://dx.doi.org/10.1007/978-3-319-01766-2_35.

[5] Z.M. Hira, D.F. Gillies, A review of feature selection and feature extraction methods applied on microarray data, Adv. Bioinform. 2015 (2015) 1–13, http://dx.doi.org/10.1155/2015/198363.

[6] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, J. Saeed, A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction, J. Appl. Sci. Technol. Trends 1 (2) (2020) 56–70, http://dx.doi.org/10.38094/jastt1224.

[7] Y. Zhang, Safety management of civil engineering construction based on artificial intelligence and machine vision technology, in: S.-B. Tsai (Ed.), Adv. Civ. Eng. 2021 (2021) 1–14, http://dx.doi.org/10.1155/2021/3769634.

[8] S.N. Qasem, F. Saeed, Hybrid feature selection and ensemble learning methods for gene selection and cancer classification, Int. J. Adv. Comput. Sci. Appl. 12 (2) (2021) 193–200, http://dx.doi.org/10.14569/IJACSA.2021.0120225.

[9] S. Manochandar, M. Punniyamoorthy, Scaling feature selection method for enhancing the classification performance of Support Vector Machines in text mining, Comput. Ind. Eng. 124 (2018) 139–156, http://dx.doi.org/10.1016/j.cie.2018.07.008.

[10] S. Chidambaram, A. Sumathi, Optimal feature selection for the classification of hyperspectral imagery using adaptive spectral–spatial clustering, Int. J. Parallel Program. 48 (5) (2020) 813–832, http://dx.doi.org/10.1007/s10766-018-0607-5.

[11] Z. Manbari, F. AkhlaghianTab, C. Salavati, Hybrid fast unsupervised feature selection for high-dimensional data, Expert Syst. Appl. 124 (2019) 97–118, http://dx.doi.org/10.1016/j.eswa.2019.01.016.

[12] J. Too, M. Mafarja, S. Mirjalili, Spatial bound whale optimization algorithm: an efficient high-dimensional feature selection approach, Neural Comput. Appl. 33 (23) (2021) 16229–16250, http://dx.doi.org/10.1007/s00521-021-06224-y.

[13] P. Agrawal, H.F. Abutarboush, T. Ganesh, A.W. Mohamed, Metaheuristic algorithms on feature selection: A survey of one decade of research (2009–2019), IEEE Access 9 (2021) 26766–26791, http://dx.doi.org/10.1109/ACCESS.2021.3056407.

[14] G. Jothi, H. Inbarani H., Hybrid tolerance rough set–firefly based supervised feature selection for MRI brain tumor image classification, Appl. Soft Comput. 46 (2016) 639–651, http://dx.doi.org/10.1016/j.asoc.2016.03.014.

[15] V. Santucci, M. Baioletti, A. Milani, An algebraic framework for swarm and evolutionary algorithms in combinatorial optimization, Swarm Evol. Comput. 55 (2020) 100673, http://dx.doi.org/10.1016/j.swevo.2020.100673.

[16] S. Zhang, C. Lee, H. Chan, K. Choy, Z. Wu, Swarm intelligence applied in green logistics:A literature review, Eng. Appl. Artif. Intell. 37 (2015) 154–169, http://dx.doi.org/10.1016/j.engappai.2014.09.007.

[17] S. Barak, J.H. Dahooie, T. Tichý, Wrapper ANFIS-ICA method to do stock market timing and feature selection on the basis of Japanese Candlestick, Expert Syst. Appl. 42 (23) (2015) 9221–9235, http://dx.doi.org/10.1016/j.eswa.2015.08.010.

[18] J. Kim, J. Kang, M. Sohn, Ensemble learning-based filter-centric hybrid feature selection framework for high-dimensional imbalanced data, Knowl.-Based Syst. 220 (2021) 106901, http://dx.doi.org/10.1016/j.knosys.2021.106901.

[19] Z. Sadeghian, E. Akbari, H. Nematzadeh, A hybrid feature selection method based on information theory and binary butterfly optimization algorithm, Eng. Appl. Artif. Intell. 97 (2021) 104079, http://dx.doi.org/10.1016/j.engappai.2020.104079.

[20] A. Chaudhuri, T.P. Sahu, A hybrid feature selection method based on Binary Jaya algorithm for micro-array data classification, Comput. Electr. Eng. 90 (December 2020) (2021) 106963, http://dx.doi.org/10.1016/j.compeleceng.2020.106963.

[21] X.Y. Liu, Y. Liang, S. Wang, Z.Y. Yang, H.S. Ye, A hybrid genetic algorithm with wrapper-embedded approaches for feature selection, IEEE Access: Pract. Innov. Open Solutions 6 (2018) 22863–22874, http://dx.doi.org/10.1109/ACCESS.2018.2818682.

[22] H. Lu, J. Chen, K. Yan, Q. Jin, Y. Xue, Z. Gao, A hybrid feature selection algorithm for gene expression data classification, Neurocomputing 256 (2017) 56–62, http://dx.doi.org/10.1016/j.neucom.2016.07.080.

[23] W. Ma, X. Zhou, H. Zhu, L. Li, L. Jiao, A two-stage hybrid ant colony optimization for high-dimensional feature selection, Pattern Recognit. 116 (2021) 107933, http://dx.doi.org/10.1016/j.patcog.2021.107933.

[24] Z. Huang, C. Yang, X. Zhou, T. Huang, A hybrid feature selection method based on binary state transition algorithm and relieff, IEEE J. Biomed. Health Inf. 23 (5) (2019) 1888–1898, http://dx.doi.org/10.1109/JBHI.2018.2872811.

[25] C. Yan, J. Ma, H. Luo, A. Patel, Hybrid binary coral reefs optimization algorithm with simulated annealing for feature selection in high-dimensional biomedical datasets, Chemometr. Intell. Lab. Syst. 184 (August 2018) (2019) 102–111, http://dx.doi.org/10.1016/j.chemolab.2018.11.010.

[26] K. Hussain, N. Neggaz, W. Zhu, E.H. Houssein, An efficient hybrid sine-cosine Harris hawks optimization for low and high-dimensional feature selection, Expert Syst. Appl. 176 (March) (2021) 114778, http://dx.doi.org/10.1016/j.eswa.2021.114778.

[27] X.F. Song, Y. Zhang, D.W. Gong, X.Z. Gao, A fast hybrid feature selection based on correlation-guided clustering and particle swarm optimization for high-dimensional data, IEEE Trans. Cybern. (2021) 1–14, http://dx.doi.org/10.1109/TCYB.2021.3061152.

[28] K. Chen, B. Xue, M. Zhang, F. Zhou, An evolutionary multitasking-based feature selection method for high-dimensional classification, IEEE Trans. Cybern. (2020) 1–15, http://dx.doi.org/10.1109/TCYB.2020.3042243.

[29] S. Salesi, G. Cosma, M. Mavrovouniotis, TAGA: Tabu Asexual Genetic Algorithm embedded in a filter/filter feature selection approach for high-dimensional data, Inform. Sci. 565 (2021) 105–127, http://dx.doi.org/10.1016/j.ins.2021.01.020.

[30] F.M. Burnet, The Clonal Selection Theory of Acquired Immunity, Vanderbilt University Press, Nashville, 1959, http://dx.doi.org/10.5962/bhl.title.8281.

[31] L.N. De Castro, F.J. Von Zuben, The clonal selection algorithm with engineering applications, in: Proceedings of GECCO, 2000, 2000, pp. 36–39.

[32] B. Haktanirlar Ulutas, S. Kulturel-Konak, A review of clonal selection algorithm and its applications, Artif. Intell. Rev. 36 (2) (2011) 117–138, http://dx.doi.org/10.1007/s10462-011-9206-1.

[33] R. Shang, L. Jiao, F. Liu, W. Ma, A novel immune clonal algorithm for MO problems, IEEE Trans. Evol. Comput. 16 (1) (2012) 35–50, http://dx.doi.org/10.1109/TEVC.2010.2046328.

[34] H. Dai, Y. Yang, H. Li, C. Li, Bi-direction quantum crossover-based clonal selection algorithm and its applications, Expert Syst. Appl. 41 (16) (2014) 7248–7258, http://dx.doi.org/10.1016/j.eswa.2014.05.053.

[35] N. Xu, Y. Ding, L. Ren, K. Hao, Degeneration recognizing clonal selection algorithm for multimodal optimization, IEEE Trans. Cybern. 48 (3) (2018) 848–861, http://dx.doi.org/10.1109/TCYB.2017.2657797.

[36] X. Yan, P. Li, K. Tang, L. Gao, L. Wang, Clonal selection based intelligent parameter inversion algorithm for prestack seismic data, Inform. Sci. 517 (2020) 86–99, http://dx.doi.org/10.1016/j.ins.2019.12.083.

[37] W. Luo, X. Lin, T. Zhu, P. Xu, A clonal selection algorithm for dynamic multimodal function optimization, Swarm Evol. Comput. 50 (February 2018) (2019) 100459, http://dx.doi.org/10.1016/j.swevo.2018.10.010.

[38] M. Aladeemy, S. Tutun, M.T. Khasawneh, A new hybrid approach for feature selection and support vector machine model selection based on self-adaptive cohort intelligence, Expert Syst. Appl. 88 (2017) 118–131, http://dx.doi.org/10.1016/j.eswa.2017.06.030.

[39] E. Emary, H.M. Zawbaa, A.E. Hassanien, Binary ant lion approaches for feature selection, Neurocomputing 213 (2016) 54–65, http://dx.doi.org/10.1016/j.neucom.2016.03.101.

[40] M.M. Mafarja, S. Mirjalili, Hybrid Whale Optimization Algorithm with simulated annealing for feature selection, Neurocomputing 260 (2017) 302–312, http://dx.doi.org/10.1016/j.neucom.2017.04.053.

[41] M.A.E. Aziz, A.E. Hassanien, Modified cuckoo search algorithm with rough sets for feature selection, Neural Comput. Appl. 29 (4) (2018) 925–934, http://dx.doi.org/10.1007/s00521-016-2473-7.

[42] K.-T. Lan, C.-H. Lan, Notes on the distinction of Gaussian and Cauchy mutations, in: 2008 Eighth International Conference on Intelligent Systems Design and Applications, IEEE, 2008, pp. 272–277, http://dx.doi.org/10.1109/ISDA.2008.237.

[43] H. Wang, C. Li, Y. Liu, S. Zeng, A hybrid particle swarm algorithm with Cauchy mutation, in: 2007 IEEE Swarm Intelligence Symposium, IEEE, 2007, pp. 356–360, http://dx.doi.org/10.1109/SIS.2007.367959.

[44] W. chuan Wang, L. Xu, K. wing Chau, D. mei Xu, Yin-Yang firefly algorithm based on dimensionally Cauchy mutation, Expert Syst. Appl. 150 (2020) 113216, http://dx.doi.org/10.1016/j.eswa.2020.113216.

[45] Q. Gu, Z. Li, J. Han, Generalized Fisher score for feature selection, in: Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence, UAI 2011, 2011, pp. 266–273, arXiv:1202.3725.

[46] J. Pirgazi, M. Alimoradi, T. Esmaeili Abharian, M.H. Olyaee, An Efficient hybrid filter-wrapper metaheuristic-based gene selection method for high dimensional datasets, Sci. Rep. 9 (1) (2019) 1–15, http://dx.doi.org/10.1038/s41598-019-54987-1.

[47] D. Dua, C. Graff, UCI machine learning repository, 2017.

[48] Arizona State University, Arizona state university's (ASU) repository, 2018, url https://github.com/chappers/scikit-feature.

[49] S. Arlot, A. Celisse, A survey of cross-validation procedures for model selection, Stat. Surv. 4 (none) (2010) http://dx.doi.org/10.1214/09-SS054.

[50] M. Hall, L. Smith, Practical feature subset selection for machine learning, Comput. Sci. 98 (1998) 181–191.
[51] L. Yu, H. Liu, Feature selection for high-dimensional data: A fast correlation-based filter solution, in: Proceedings, Twentieth International Conference on Machine Learning, Vol. 2, 2003, pp. 856–863.
[52] R.S. Marko, K. Igor, Theoretical and empirical analysis of relief and rrelieff, Mach. Learn. J. 53 (2003) 23–69.
[53] G. Cawley, N. Talbot, M. Girolami, Sparse multinomial logistic regression via bayesian L1 regularisation, Adv. Neural Inf. Process. Syst. 19 (2006).
[54] Z. Zhao, H. Liu, Spectral feature selection for supervised and unsupervised learning, in: Proceedings of the 24th International Conference on Machine Learning - ICML '07, ACM Press, New York, New York, USA, 2007, pp. 1151–1157, http://dx.doi.org/10.1145/1273496.1273641.