RESEARCH ARTICLE

A hybrid feature selection algorithm combining information gain and grouping particle swarm optimization for cancer diagnosis

Fangyuan Yang[1☯]**, Zhaozhao Xu**[2☯]**, Hong Wang** [1]*****, Lisha Sun**[1]*****, Mengjiao Zhai**[1]**, Juan Zhang**[1]

1 Department of Gynecologic Oncology, The First Affiliated Hospital of Henan Polytechnic University, Jiaozuo, Henan, China, **2** School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan, China

☯ These authors contributed equally to this work.

* hongwang197408@outlook.com (HW); 3928077@163.com (LS)

Abstract

Background

Cancer diagnosis based on machine learning has become a popular application direction. Support vector machine (SVM), as a classical machine learning algorithm, has been widely used in cancer diagnosis because of its advantages in high-dimensional and small sample data. However, due to the high-dimensional feature space and high feature redundancy of gene expression data, SVM faces the problem of poor classification effect when dealing with such data.

Methods

Based on this, this paper proposes a hybrid feature selection algorithm combining information gain and grouping particle swarm optimization (IG-GPSO). The algorithm firstly calculates the information gain values of the features and ranks them in descending order according to the value. Then, ranked features are grouped according to the information index, so that the features in the group are close, and the features outside the group are sparse. Finally, grouped features are searched using grouping PSO and evaluated according to in-group and out-group.

Results

Experimental results show that the average accuracy (ACC) of the SVM on the feature subset selected by the IG-GPSO is 98.50%, which is significantly better than the traditional feature selection algorithm. Compared with KNN, the classification effect of the feature subset selected by the IG-GPSO is still optimal. In addition, the results of multiple comparison tests show that the feature selection effect of the IG-GPSO is significantly better than that of traditional feature selection algorithms.

Conclusion

The feature subset selected by IG-GPSO not only has the best classification effect, but also has the least feature scale (*FS*). More importantly, the IG-GPSO significantly improves the *ACC* of SVM in cancer diagnostic.

Introduction

With the rapid increase of cancer incidence and mortality, cancer research has attracted more and more attention [1, 2]. As a genetic mutation-like disease, cancer can be diagnosed and treated by analyzing the genes that are mutated at the molecular level [3, 4]. The traditional diagnostic methods usually use cell morphology and histopathology [5, 6]. However, such methods can not reach the clinical requirements of early diagnosis and treatment [7, 8]. Machine learning, as a new technology, is widely used for cancer diagnosis because of its fast operation and high efficiency [9, 10]. Machine learning based cancer diagnosis refers to the process of using machine learning algorithms to model and train gene expression data, and predicting unknown gene expression data according to the trained model [11, 12].

Common machine learning algorithms include support vector machines, decision trees, naive Bayes and ensemble learning [13, 14]. Among them, support vector machine, as a new machine learning algorithm based on statistical theory learning, has been widely used in cancer diagnosis because of its advantages in high-dimensional small sample data. For example, Huang et al. proposed to optimize support vector machine by using fruit fly optimization algorithm, and applied the optimized support vector machine to breast cancer diagnosis [15]. Similarly, Wang et al. [16] proposed to use the improved whale optimization algorithm for feature selection while optimizing support vector machine parameters. Experimental results show that the proposed algorithm is effective in cancer diagnosis.However, due to the large number of cancer-related genes in gene expression data, and the complex interactions and contradictions among these genes [17], SVM is faced with problems such as high time complexity and poor classification effect when processing such data [18, 19].

For high-dimensional feature space and highly feature correlation of gene expression data, feature selection is required before processing these data [20, 21]. Feature selection [22] can simplify the machine learning model, reduce the training time, and improve the diagnostic effect of the model. According to whether it is related to the classification algorithm, it can be divided into filter [23], wrapper [24] and hybrid [25] algorithms. Filter algorithm uses metric to enhance the correlation between features and class, and reduce the correlation between features. Among them, information gain (IG) [26], Chi-square (Chis) [27] and pearson correlation coefficient (pearson) [28] are common metrics in filter algorithms. Unfortunately, the feature subset selected by filter algorithm have poor classification effect and need to set thresholds artificially, which has great blindness [29].

Wrapper [24] algorithm uses search algorithm to search the original features, and takes ACC index of classification algorithm as the metric of selected feature subset. Common search algorithms include particle swarm optimization (PSO) [30], genetic algorithm (GA) [31] and ant colony algorithm (ant colony algorithm). ACA) [32]. However, the time complexity of such algorithms is extremely high [33], and the search algorithm is easy to fall into local optimal solutions during the search process, resulting in that the selected feature subset is not globally optimal [34, 35]. In recent years, some scholars put forward the strategy of combining filter and wrapper. For example, Got et al. [25] proposed a hybrid algorithm combining mutual

information and whale optimization algorithm. The features below the threshold are filtered by calculating the mutual information of the features. The filtered features are then searched using the whale optimization algorithm, and the searched subset of features is evaluated based on *ACC*. Similarly, Liu et al. [36] proposed a hybrid algorithm combining gain ratio and multi-objective genetic algorithm. Invalid features are filtered by calculating the gain ratio of features. Then multi-objective genetic algorithm is used to search the filtered features, and the selected feature subset is evaluated according to the *ACC* and the feature scale (*FS*). However, the hybrid algorithm does not solve the problem of blindness in threshold setting and the tendency of search algorithm to fall into local optimal solution [37].

Based on the above description, this paper proposes a hybrid algorithm combining information gain and grouped particle swarm (IG-GPSO). The algorithm first computs the information gain value of each feature and sorts it in descending order according to the value size. Then the sorted features are grouped according to the information gain index, and the features with similar information gain values are divided into a group. Finally, the grouped features are searched using the grouped particle swarm optimization algorithm, and the selected feature subset are evaluated according to two methods: in-group and out-group. The experiment is divided into three parts. The firstly is the feature selection process experiment to observe the working principle of the IG-GPSO algorithm. Then, a comparison experiment with other feature selection algorithms is conducted to verify the effectiveness of the IG-GPSO algorithm. Finally, a comparison experiment with other classification algorithms is done to verify the applicability of the IG-GPSO algorithm. In addition, we also select statistical experiments to verify whether the IG-GPSO algorithm is significantly different from other feature selection algorithms.

1. For the blind threshold setting of filter, this paper proposes a feature ranking strategy. The IG value of features is calculated and ranked according to the size of the value. Unlike traditional filter algorithms, at this stage we only rank the features and do not filter any features.

2. For the high time complexity of wrapper, this paper proposes a feature grouping strategy. By calculating the number of groups and grouping the ranked features according to the information index strategy, the number of features in each group is significantly less than the original number of features.

3. For the PSO is easy to fall into local optimal solutions, this paper proposes a group search strategy. In the in-group evaluation, we use *ACC* as a fitness function. In the out-group evaluation, we use *ACC* and *FS* as fitness function.

4. Experimental results show that the *ACC* and number of *FS* selected by the IG-GPSO algorithm are significantly better than the traditional feature selection algorithm. In addition, statistical experiments show that the IG-GPSO algorithm is significantly different from the traditional feature selection algorithm.

The rest of this paper is organized as: the Datasets and methods section presents the gene expression datasets and the IG-GPSO. The Experimental result section presents the comparative and statistical experiments. The Experiment discussion and analysis section is dedicated to the experimental discussion and analysis, and the Conclusion and future work section is the conclusion and future work.

Datasets and methods

Gene expression datasets

In this paper, we selected 4 publicly available gene expression datasets [38, 39], namely Prostate-GE, TOX-171, GLIOMA and Lung-discrete. Prostate-GE contains 5966 genes and 102

samples, including a total of 2 classes of information. TOX-171 contains 5748 genes and 171 samples, including a total of 4 classes of information. GLIOMA contains 4434 genes and 50 samples, including a total of 4 classes of information. Lung-discrete contains 325 genes, 73 samples, and a total of 4 classes of information. All in all, these gene expression data are characterized by a large number of features and a small number of samples.

Support vector machine

SVM map low-dimensional samples into high-dimensional Spaces and obtain a hyperplane to maximize the margin between the two types of samples [40]. This strategy effectively avoids the phenomenon of "overfitting", especially when classifying high-dimensional small sample data [41]. Suppose that the training sample set is a binary classification problem, based on statistical theory, the classification model of SVM can be constructed as follows:

$$\begin{cases} \min \frac{1}{2}(w^T w) + C \sum_{i=1}^{l} \xi_i \\ s.t \, y_i((w^T x_i) + b) \geq 1 - \xi_i, i = 1, 2, \cdots, l \end{cases} \tag{1}$$

Where, $C > 0$ is the regularization parameter, ξ_i is the slack variable, $w \in n$ is the normal vector of the classification hyperplane, and b is the threshold. Using the KKT [42] condition and duality theory in optimization theory, the optimized model of the dual function can be obtained as follows:

$$\begin{cases} \max \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j (x_i^T x_j) \\ s.t \sum_{i=1}^{l} y_i \alpha_i = 0 \\ 0 \leq \alpha_i \leq C; i = 1, 2, \cdots, l; j = 1, 2, \cdots, l \end{cases} \tag{2}$$

Where, α_i is the Lagrange multiplier. The optimization model is a convex quadratic programming problem, so the local optimal solution is the global optimal solution. If $\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_l^*)^T$ is the global optimal solution of the model, then:

$$w^* = \sum_{i=1}^{l} y_i \alpha_i^* x_i \tag{3}$$

According to the KKT complementarity condition given in optimization theory, the optimal solution must satisfy:

$$\alpha_i((y_i(w^T x_i) + b) - 1 + \xi_i) = 0, i = 1, 2, \cdots, l \tag{4}$$

$$(C - \alpha_i)\xi_i = 0, i = 1, 2, \cdots, l \tag{5}$$

According to Eqs (3), (4) and (5), the samples corresponding to the Lagrange multiplier lose their effect on the classification problem, while only the samples corresponding to the Lagrange multiplier $\alpha_i > 0$ play a role, thus deciding the result of the classification.

Support vectors are usually only a small subset of the total sample. After solving the above problem [43], the optimal linear classifier can be obtained as follows:

$$f(x) = \text{sgn}\{(w^{*T} x) + b^*\} = \text{sgn}\left\{\sum_{i=1}^{l} \alpha_i^* y_i (x_i^T x) + b^*\right\} \tag{6}$$

Where, sgn() is the sign function, and b^* is the threshold for classification, which can be obtained from any support vector.

For the nonlinear separable case, SVM constructs the optimal classification in the higher dimensional feature space by mapping the input vector to the higher dimensional feature space. Applying the transformation Φ of x from the input space R^n to the feature space H, get:

$$x \rightarrow \Phi(x) = (\Phi_1(x), \Phi_2(x), \cdots, \Phi_l(x))^T \tag{7}$$

Replacing the input vector x with the eigenvector $\Phi(x)$ gives the optimal classification:

$$f(x) = \text{sgn}\{(w^{*T}\Phi(x)) + b^*\} = \text{sgn}\left\{\sum_{i=1}^{l} \alpha_i^* y_i \left(\Phi_i(x)^T \Phi(x)\right) + b^*\right\} \tag{8}$$

In the above duality problem, the objective function and the classification function are only concerned with the inner product calculation of the training samples. This strategy effectively avoids the complicated calculation of high dimensional space and only needs to calculate the inner product [43].

Methodology

High-dimensional gene expression data greatly affect the diagnostic *ACC* of machine learning algorithms. Therefore, a hybrid feature selection algorithm combining information gain and grouping particle swarm optimization is proposed in this paper. The first is the ranking and grouping stage, which uses IG to rank the features and groups them according to the information index. Then there is the grouping and search stage, which uses the group PSO to search the features after the group, and evaluates them according to the out–group and in-group. Fig 1 shows the flowchart of the IG-GPSO.

A information gain-based ranking and grouping algorithm. Filter algorithm based on information gain calculates the IG value of features, and selects the features whose IG value is higher than the threshold as the selected feature subset [44]. The larger the IG value of the feature, the greater the amount of information contained in the feature. In general, the higher the IG value of a feature, the more distinguishing ability the feature has. Then, the IG of feature *f* is defined as follows:

$$IG(f) = H(f) - H(f|S) \tag{9}$$

In Eq (9), $H(f)$ is the information entropy of the feature. The larger the value of $H(f)$, the more information it carries. Then, the information entropy [45] of feature *f* is defined as follows:

$$H(f) = -\sum_{i=1}^{m} P(C_i)\log_2 P(C_i) \tag{10}$$

Where, $P(C_i)$ is the probability that any sample belongs to class C. $P(C_i) = S_j/S$, m is the number of sample classes. S_j is the number of samples belonging to class C_i and S is the total number of samples.
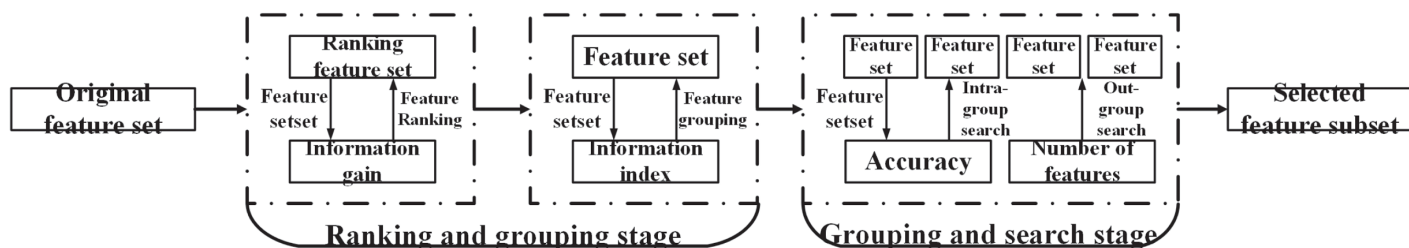

Fig 1. The flowchart of the IG-GPSO.

In Eq (9), $H(f|S)$ is the conditional entropy of feature f, which represents the amount of information that feature f contains given S. Then, the conditional entropy of feature f is defined as follows:

$$H(f|S) = -\sum_{i=1}^{m} \sum_{j=1}^{n} P(C_i, S_j) \log_2 P(C_i|S_j) \tag{11}$$

$H(f, S)$ is the joint entropy of sample S and feature f, which represents the amount of information contained in feature f given S and f. Then, the joint entropy of sample S and feature f is defined as follows:

$$H(f, S) = -\sum_{i=1}^{m} \sum_{j=1}^{n} P(C_i, S_j) \log_2 P(C_i, S_j) \tag{12}$$

In Eq (12), $P(C_i, S_j) = S_{ij}/S_j$ is the probability that the sample in S_j belongs to class C_i. in Eq (9) is the IG of this partition obtained from f, which represents that the feature with the highest IG is the discriminative feature in a given record set.

Filter algorithm selects the features whose IG value is higher than the threshold as the selected feature subset. However, the artificial setting of the threshold has great blindness. Therefore, this paper proposes a information gain-based ranking and grouping algorithm. First, we give the required number of groups:

$$k = \sqrt[\gamma]{\sum_{i=1}^{M} |f_i|} \tag{13}$$

Where, γ is the feature index of the dataset. $||$ is the *FS* of the statistics. The number of features in each group and the sum of the IG can be calculated by the number of groups. Then, the information index of each grouping can be defined as:

$$I\_index = \sum_{i=1}^{M} |IG(f_i)|/k \tag{14}$$

The purpose of feature grouping is to combine the features with close IG values, so that the IG values of the features in the group are close, and the IG values of the features outside the group is sparse. Then grouped features are re-ranked according to the IG value. **Algorithm 1** presents the steps of the IG.

Algorithm 1: A information gain-based ranking and grouping algorithm (IG)

```
Require: Feature set: F = {f₁, f₂, ···, fₘ}; Number of feature groups: k
Ensure: Feature groups: F = {F₁, F₂, ···, Fₖ}
   for i = 1, 2, ···, m do
 1: Calculate the information entropy, conditional entropy and joint
    entropy of feature f according to Eqs (10), (11) and (12).
 2: Calculate the IG value of the feature according to Eq (9).
 3: Ranked feature set: IG(F) = {IG(f₁), IG(f₂), ···, IG(fₘ)}
   for i = 1, 2, ···, m do
 4: Calculate the number k of required groups according to Eq (13).
 5: Calculate the information index of each group according to Eq
    (14).
 6: The ranked feature sets are grouped: F = {F₁, F₂, ···, Fₖ}
   return Feature groups: F = {F₁, F₂, ···, Fₖ}
```

A grouping particle swarm optimization-based wrapper algorithm. *Particle parameter setting*. PSO is a meta-heuristic search algorithm that simulates the foraging behavior of birds. In the search process, the swarm treats particles as points in space, and these particles search at a certain speed, adjusting their flight speed and direction based on their own flight experience and the flight experience of other particles.

Suppose the i-th particle has position $X_i = (x_{i1}, x_{i2}, \cdots, x_{iN})$ and velocity $V_i = (v_{i1}, v_{i2}, \cdots, v_{iN})$ [46]. The best position through which particle passes is denoted as $P_i = (p_{i1}, p_{i2}, \cdots, p_{iN})$.

The best position that all particles passes through is denoted as $P_{gbest} = (p_{g1}, p_{g2}, \cdots, p_{gN})$. Then, update the position and velocity of each particle according to Eqs (15) and (16) in the iterative search:

$$v_{in}^{k+1} = w \times v_{in}^k + c_1 r_1 (p_{in}^k - x_{in}^k) + c_2 r_2 (p_{gn}^k - x_{in}^k) \tag{15}$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \tag{16}$$

Where, w is the inertia weight; k is the number of iterations; r_1 and r_2 are random numbers between 0 and 1. c_1 and c_2 are used to accelerate the particle. Moreover, v_{in} is limited by the maximum velocity v_{max}, and if the velocity of a particle in a dimension exceeds the maximum velocity v_{max}, then the velocity of the particle in that dimension is limited below the maximum velocity [47].

In the search process, c_1 is used to control the convergence speed of the particle, and c_2 is used to control the search speed of the particle. When $c_2 = 0$, the velocity of the particle will no longer change, resulting in the failure to find the optimal position, and then fall into the local optimal solution [48]. Based on this, this paper proposes a grouping particle swarm optimization-based wrapper algorithm.

In this algorithm, the speed and position of the particles are updated by adding the previously grouped feature groups to the search space when the particles fall into the local optimal solution. To this end, we also propose in-group and out-group particle fitness functions.

Particle initialization. Grouping PSO-based Wrapper algorithm refers to the process of selecting M optimal features from N features ($M \leq N$). Each particle in a particle swarm during the search process represents a potentially optimal subset of features. For each particle, if the i-th bit is 1, it means the feature is selected, otherwise it means the feature is not selected. Thus, each particle represents a potentially optimal subset of a subset of features.

In the feature search process, PSO is easy to be affected by a single particle, and then fall into the local optimal solution, so it can not find the global optimal solution. Therefore, we initialize the feature set using a grouping strategy. Suppose that the number of features corresponding to each particle is P. In each search round, P is N of the total number of features. When the PSO falls into the local optimal solution, the diversity of particles is improved by adding a new feature group.

Fitness function. Wrapper algorithm based on PSO generally use the ACC of the classification algorithm as the fitness function to evaluate the feature subset. Instead, this paper proposes two fitness functions, in-group and out-group, to evaluate the selected feature subset. In in-group evaluation, we use ACC as a fitness function for the selected feature subset. Specifically defined as follows:

$$Fintness(F_j) = ACC(F_j) \tag{17}$$

Where, ACC is the performance index of the classification algorithm. In the out-group evaluation, we use the ACC and the FS as the fitness function for the selected feature subset. Specifically defined as follows:

$$Fintness\left(F_j\right) = \lambda \times ACC\left(F_j\right) - (1 - \lambda) \times \frac{Size(F_j)}{Size(F)} \tag{18}$$

Where, $Size()$ is used to count the FS. λ is the weight of the FS, which is used to balance the ACC and the FS. According to Eq (18), the feature subset selected based on the out-group fitness function has the least FS and the best ACC.

Algorithm description and analysis. For the problem that PSO is easy to fall into local optimal solution, this paper proposes a grouping particle swarm optimization-based wrapper algorithm. The algorithm searches for grouped features and evaluates them using in-group and out-group fitness functions. In the in-group evaluation, we use *ACC* as the evaluation function. In out-group evaluation, we use *ACC* and *FS* as fitness functions.

Step 1 is the feature input process, which is used to sequentially output the feature group F_j. Step 2 is the particle initialization process, which is used to initialize the feature group F_j. Steps 3-6 are the in-group feature evaluation process, which is used to search and evaluate the feature F_j, and output the optimal feature group F_j' in this round. Steps 7-8 are the out-group feature evaluation process, which is used to evaluate the searched F_j' and output the globally optimal feature group.

Algorithm 2: A grouping particle swarm optimization-based Wrapper algorithm (GPSO)

```
Require: Feature groups: F = {F₁, F₂, ···, Fₖ}
Ensure: Optimal feature groups: F' = {F'₁, F'₂, ···, F'ₖ}
    for j = 1, 2, ···, k do
  1: Input the corresponding feature groups in turn.
    for i = 1, 2, ···, Size(Fⱼ) do
    2: Initialize the position and velocity of each particle i, and
       set the current optimal particle as pbestᵢ and gbestᵢ.
    for i = 1, 2, ···, Size(Fⱼ) do
    3: Calculate the fitness of each particle i in the group accord-
       ing to Eq (17).
    4: For each particle i, compare Fitnessᵢ with the local size
       pbestᵢ. If Fitnessᵢ is better than pbestᵢ, then pbestᵢ =
       Fitnessᵢ.
    5: For each particle i, compare Fitnessᵢ with the global size
       gbestᵢ. If Fitnessᵢ is better than gbestᵢ, then gbestᵢ =
       Fitnessᵢ.
    6: Update the position and velocity of particle i according to
       Eqs (15) and (16).
  7: Calculate the fitness of the feature group outside the group
     according to Eq (18).
  8: For feature groups Fⱼ, compare Fitnessᵢ with the size of the
     global Fitnessⱼ. If Fitnessᵢ is better than Fitnessⱼ, then Fit-
     nessⱼ = Fitnessᵢ
    return Optimal feature groups: F' = {F'₁, F'₂, ···, F'ₖ}
```

Experimental result

Experimental setting

Feature selection algorithms usually use the classification index of classification algorithms to evaluate the goodness of the selected feature subset. In addition, we also select the *FS* as an evaluation index for the algorithm. In this section, 11 feature selection algorithms are selected for comparison experiments. Among them, the information gain (IG) [26], chi-square (Chis) [27] and pearson correlation coefficient (Pearson) [28] are filter algorithms. Particle swarm optimization (PSO) [30], genetic algorithm (GA) [31] and ant colony algorithm (ACA) [32] are wrapper algorithms. The combination of information gain and particle swarm optimization (IG-PSO) [26, 30], information gain and genetic algorithm (IG-GA) [26, 31], information gain and ant colony algorithm (IG-ACA) [26, 32], chi-square and particle swarm optimization (Chis-PSO) [27, 30], and pearson correlation coefficient and particle swarm optimization (Pearson-PSO) [28, 30] are hybrid algorithms.

The comparison experiment is divided into three parts: firstly, the experiments on the feature selection process, in which the feature ranking, feature grouping and feature selection process of the IG-GPSO is observed by setting the interrupt procedure. Then the comparison experiment with other algorithms is carried out, and the IG-GPSO is compared with 11 traditional feature selection algorithms to verify the effectiveness. Finally, comparison experiment with other classification algorithms, SVM is compared with KNN to verify the applicability. In addition, the multiple comparison tests between the IG-GPSO and 11 traditional feature selection algorithms is performed to verify whether the IG-GPSO and 11 traditional feature selection algorithms have significant differences.

Experiments on the feature selection process

This section presents experiments on the feature selection process, in order to observe the influence of the threshold on the IG-GPSO, we use the IG to rank the features, and the ranked features were grouped according to the information index. The threshold of the filter stage is set to 1, 2, $\cdots$, k. In addition, we use SVM to test the selected feature subset. Fig 2 shows the feature ranking grouping process of the IG-GPSO.

Observing Fig 2, the feature selection results of IG on different datasets show different trends, respectively. On the lung-discrete dataset, when the threshold is 2, the *ACC* index of SVM reaches 91.8%, and the overall trend is increasing. On the Prostate-GE dataset, with the increase of the threshold, the classification effect of SVM shows a downward trend, and the *ACC* index reaches 91.2% when the threshold is 20. On the TOX-171 dataset, with the increase of the threshold, the SVM reaches the maximum value when the threshold is 8, and the *ACC* index is 98.8%. In summary, the threshold may vary greatly depending on the dataset. Fig 3 shows the feature selection process of the IG-GPSO.

Observing Fig 3, the classification effects of SVM on different datasets also show different trends. At the beginning of the out-group search, the classification effect of SVM is relatively poor, especially on the TOX-171 dataset, where the *ACC* index is only 96.0%. This is mainly because the number of features in the group is very small, which makes the PSO fall into a
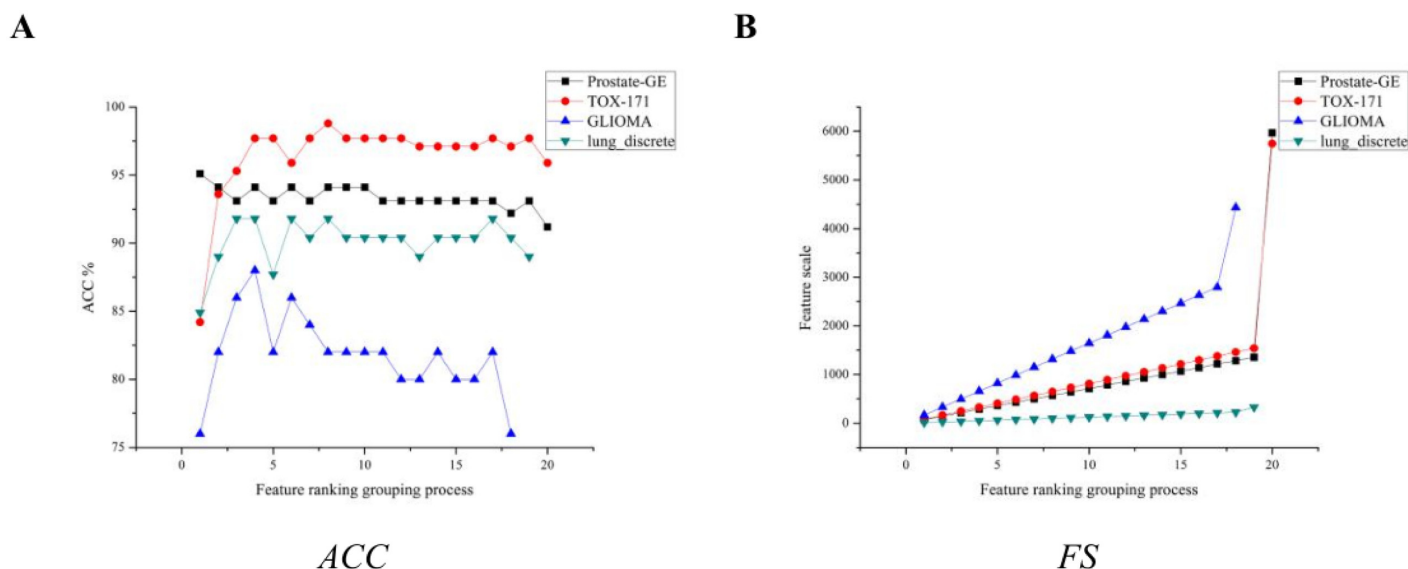

Fig 2. The feature ranking grouping process of the IG-GPSO. (A)Accuracy of the feature ranking grouping process. (B)The number of feature subsets of the feature ranking grouping process.
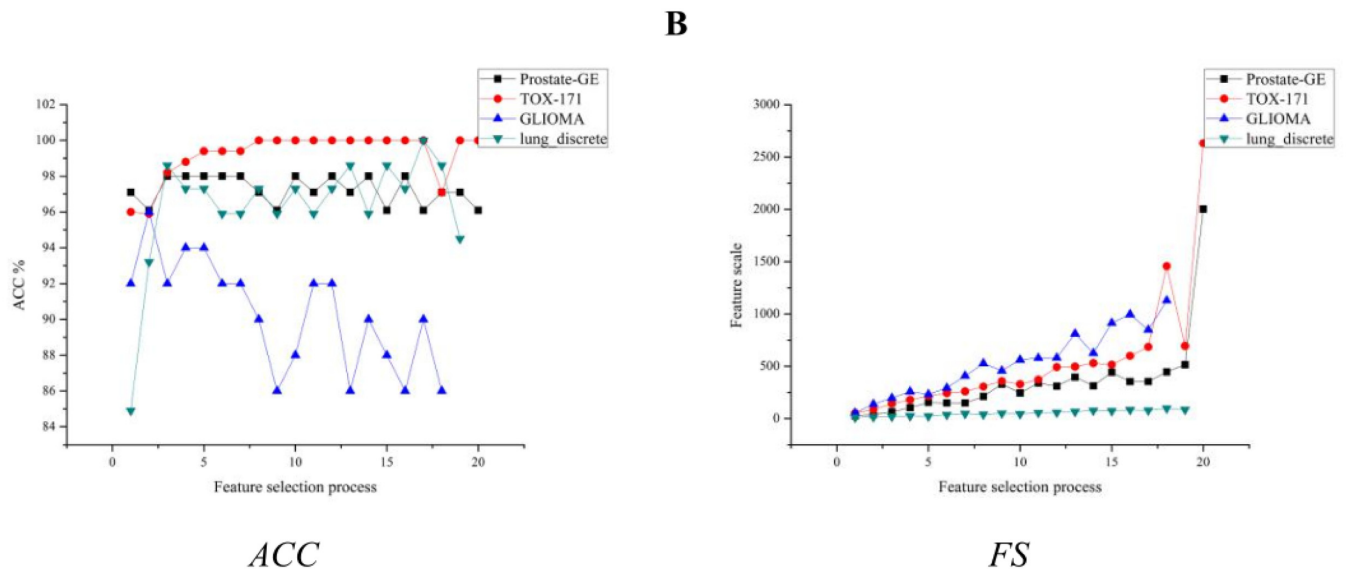
Fig 3. The feature selection process of the IG-GPSO. (A)Accuracy of the feature selection process. (B)The number of feature subsets of the feature selection process.

local optimal solution when searching. With the increase of out-group search, except for GLIOMA dataset, the classification effect of SVM showed an increasing trend. On lung-discrete and TOX-171 datasets, the *ACC* index of SVM achieves 100.0%. This result shows that the group search strategy effectively avoids the situation of the PSO falling into the local optimal solution.

Comparison experiments with other feature selection algorithms

This section presents the comparison experiments with other feature selection algorithms, and we select 11 traditional feature selection algorithms for experiments. Among them, the threshold of the filter algorithm is set to half of the total number of informative features. The evaluation algorithm of the wrapper algorithm is the SVM. The participation of the hybrid algorithm is the same as those for the filter and wrapper algorithms. Tables 1 and 2 shows the *ACC* and *FS* results of SVM on the selected feature subset.

From Tables 1 and 2, the average *ACC* index of SVM on the original gene expression datasets is only 88.03%, especially the *ACC* index on the GLIOMA dataset is only 76.0%. This shows that the high-dimensional feature space and high feature redundancy of data greatly damage the classification effect of SVM. Compared with the filter algorithm, IG has the best feature selection effect on the Prostate-GE dataset, while Chis has the best feature selection effect on the Lung-discrete dataset. Overall, Chis has the best feature selection effect in the

Table 1. *ACC* index of SVM on the selected feature subset.

Dataset	ALL	IG	Chis	Pearson	PSO	GA	ACA	IG-PSO	IG-GA	IG-ACA	Chis-PSO	Pearson-PSO	IG-GPSO
Prostate-GE	91.2	94.1	93.1	91.2	96.1	92.2	93.1	**98.0**	97.1	**98.0**	**98.0**	95.1	**98.0**
TOX-171	95.9	97.7	97.7	98.8	99.4	**100.0**	98.8	**100.0**	98.8	**100.0**	**100.0**	**100.0**	**100.0**
GLIOMA	76.0	82.0	82.0	82.0	84.0	84.0	82.0	90.0	88.0	90.0	95.0	84.0	**96.0**
Lung-discrete	89.0	89.0	91.8	89.0	94.5	93.2	95.9	97.3	94.5	98.6	94.5	94.5	**100.0**
Average	88.03	90.70	91.15	90.25	93.50	92.35	92.45	96.33	94.60	96.65	96.88	93.40	**98.50**

Table 2. *FS* index of SVM on the selected feature subset.

Dataset	ALL	IG	Chis	Pearson	PSO	GA	ACA	IG-PSO	IG-GA	IG-ACA	Chis-PSO	Pearson-PSO	IG-GPSO
Prostate-GE	5966	677	677	2983	2525	2927	2380	283	331	285	130	1141	**63**
TOX-171	5748	769	769	2874	1732	2325	2314	338	388	364	363	1198	**305**
GLIOMA	4434	1398	1398	2217	1561	1856	2010	426	590	530	317	1076	**138**
Lung-discrete	325	111	111	163	142	153	139	52	**48**	53	50	59	77
Average	4118.25	738.75	738.75	2059.25	1490.00	1815.25	1710.75	274.75	339.25	308.00	215.00	869.50	**145.75**

https://doi.org/10.1371/journal.pone.0290332.t002

filter algorithms, and the average *ACC* index of SVM is 91.15%. Compared with the wrapper algorithm, PSO has the best feature selection effect, and the average *ACC* index of SVM is 93.50%.

Compared with the hybrid algorithm, Chis-PSO has the best feature selection effect on the TOX-171 dataset, and the *ACC* index of SVM is 98.0%, and the *FS* index is 130. IG-ACA has the best feature selection effect on the Lung-discrete dataset, the *ACC* index of SVM is 98.6%, and the *FS* index is 53. Overall, Chis-PSO has the best feature selection effect in the hybrid algorithms. Moreover, the feature selection effect of the hybrid algorithm is significantly better than the filter and wrapper algorithms. Compared with the traditional feature selection algorithms, the feature selection effects of the IG-GPSO are optimal, and the average *ACC* index of the SVM is 98.50%. In addition, the *FS* index of the IG-GPSO is also the least. This shows that the IG-GPSO effectively avoids the blindness of threshold setting and the PSO is easy to fall into a local optimal solution.

Comparison experiments with other classification algorithms

This section presents the comparison experiments with other classification algorithms, in order to avoid the limitations brought by using a single evaluation algorithm, we select KNN as the evaluation algorithm for the wrapper and hybrid algorithms. Similarly, we also selected 11 traditional feature selection algorithms for comparative experiments, and used KNN to test the datasets after feature selection. Tables 3 and 4 shows the *ACC* and *FS* results of KNN on the selected feature subset.

From Tables 3 and 4, KNN has extremely poor classification effect on the original gene expression datasets, and the average *ACC* index is 81.58%. This shows again that the high-dimensional feature space and high feature redundancy of the data greatly damage the classification effect. Similarly, the average *ACC* index of KNN on the feature subset selected by IG is 87.25%, and the *FS* index is 738.75. PSO has the best feature selection effect in the wrapper algorithms, and the average *ACC* index of KNN is 91.18%. Overall, the *FS* selected by the filter algorithm is less than the wrapper algorithm, but the classification effect of the feature subset selected by the filter algorithm is worse than the wrapper algorithm.

Table 3. *ACC* index of KNN on the selected feature subset.

Dataset	ALL	IG	Chis	Pearson	PSO	GA	ACA	IG-PSO	IG-GA	IG-ACA	Chis-PSO	Pearson-PSO	IG-GPSO
Prostate-GE	85.3	91.2	90.2	87.3	90.2	88.2	91.2	94.1	92.2	95.1	94.1	91.2	**96.1**
TOX-171	85.4	89.5	90.6	89.5	95.3	91.2	93.6	97.1	94.7	95.9	**97.7**	95.9	**97.7**
GLIOMA	72.0	82.0	82.0	78.0	86.0	76.0	84.0	82.0	90.0	92.0	86.0	86.0	**94.0**
Lung-discrete	83.6	86.3	84.9	83.6	93.2	90.4	91.8	94.5	94.5	**97.3**	**97.3**	94.5	**97.3**
Average	81.58	87.25	86.93	84.60	91.18	86.45	90.15	91.93	92.85	95.08	93.78	91.90	**96.28**

https://doi.org/10.1371/journal.pone.0290332.t003

Table 4. *FS* index of KNN on the selected feature subset.

Dataset	ALL	IG	Chis	Pearson	PSO	GA	ACA	IG-PSO	IG-GA	IG-ACA	Chis-PSO	Pearson-PSO	IG-GPSO
Prostate-GE	5966	677	677	2983	2630	2506	2685	269	299	267	224	1283	**27**
TOX-171	5748	769	769	2874	1800	2144	1890	287	376	336	**282**	1038	305
GLIOMA	4434	1398	1398	2217	1305	968	1319	1398	486	480	726	982	**60**
Lung-discrete	325	111	111	163	112	113	126	59	52	49	57	61	**27**
Average	4118.25	738.75	738.75	2059.25	1461.25	1432.75	1505.00	503.25	303.25	283.00	322.25	841.00	**104.75**

https://doi.org/10.1371/journal.pone.0290332.t004

Compared with the hybrid algorithm, IG-ACA has the best feature selection effect on the Prostate-GE dataset, and the *ACC* index of KNN is 95.1%. The Chis-PSO has the best feature selection effect on the TOX-171 dataset, and the *ACC* index of KNN is 97.7%. Overall, IG-ACA has the best feature selection effect in the hybrid algorithms, and the average *ACC* index of KNN is 95.08%. The filter algorithm has the worst feature selection effect, while the wrapper algorithm has the largest *FS* selected. However, the feature selection effect of the hybrid algorithm is significantly better than the filter and wrapper algorithms. More importantly, KNN has the best classification effect on the dataset after the IG-GPSO feature selection, the average *ACC* index is 96.28%, and the *FS* index is also the smallest.

Statistical experiments with other feature selection algorithms

This section presents statistical experiments, in order to compare whether there are significant differences algorithms, we choose Friedman test for statistical experiments. The ranking values of all algorithms on each dataset are counted. For all algorithms, all ranking values are obtained as comparison values. According to the Friedman test, the following results are obtained:

$$\chi_F^2 = \frac{12 \times N}{k \times (k+1)} \left[\sum_{j=1}^{k} R_j^2 - \frac{k \times (k+1)^2}{4} \right] \tag{19}$$

Where, *N* is the number of datasets, *k* is the number of feature selection algorithms, and R_j is the average of the ranking values of each feature selection algorithm. For computational convenience, we transform the χ_F^2 distribution into a distribution obeying F_F. Specifically:

$$F_F = \frac{(N-1) \times \chi_F^2}{N \times (k-1) - \chi_F^2} \tag{20}$$

Where, the F_F distribution has $k-1$ and $(N-1) \times (k-1)$ degrees of freedom. Then, the *ACC* and *FS* index results of SVM and KNN on 4 datasets are tested. When the significance level $\alpha = 0.05$, the null hypothesis is that there is no significant difference between all algorithms. According to Eqs (19) and (20), when $N = 8$, the Friedman test result is as follows:

When the significance level $\alpha = 0.05$, $F(12, 84) = 1.869$. The results based on the SVM indexes are: $\chi_F^2 = 75.43$, $F_F = 25.67$. However, F_F is significantly greater than 1.869, which rejects the null hypothesis. Similarly, the results based on the KNN indexes are: $\chi_F^2 = 80.19$, $F_F = 35.95$. However, F_F is significantly greater than 1.869, which again rejects the null hypothesis. In summary, the IG-GPSO is significantly different from traditional feature selection algorithms.

Experiment discussion and analysis

In experiments on feature selection process, with the increase of threshold, the filter stage of the IG-GPSO shows different trends on different datasets, such as an increasing trend on

GLIOMA dataset and a decreasing trend on Prostate-GE dataset. These results show that different thresholds may have different effects for different datasets. Therefore, there is a great blindness in setting the threshold artificially. The filter and wrapper stage of the IG-GPSO also shows different trends on different datasets. Overall, with the increase of out-group search, the *ACC* index of SVM on all datasets shows an upward trend. These results show that the feature grouping stage of the IG-GPSO reduces the time complexity of the search algorithm and effectively avoids the situation that the PSO is easy to solve locally. Therefore, the IG-GPSO takes into account both time complexity and *ACC*.

Comparison experiments with other feature selection algorithms, in the filter algorithms, Chis has the best feature selection effect, and the average *ACC* index of SVM is 91.15%. Unfortunately, Pearson has the worst feature selection effect, and the average *ACC* index of SVM is 90.25%. In the wrapper algorithms, PSO has the best feature selection effect, and the average *ACC* index of SVM is 93.50%. Similarly, GA has the worst feature selection effect, and the average *ACC* index of SVM is 92.35%. In the hybrid algorithms, Chis-PSO has the best feature selection effect, and the average *ACC* index of SVM is 96.88%. Pearson-PSO has the worst feature selection effect, and the average *ACC* index of SVM is 93.40%. The difference is that hybrid algorithm has a better feature selection effect than the filter and wrapper algorithm. Therefore, the filter algorithm only reduces the feature space of the data, but does not solve the problem of high feature redundancy.

In addition, the *FS* index by the filter algorithm is a fixed number, and it is half of the total number of informative features. In the wrapper algorithm, the *FS* index by PSO is the least, and the average *FS* index of SVM is 1490.00. The *FS* index by GA is the largest, and the average *FS* index of SVM is 1815.25. Overall, the *FS* index by the wrapper algorithm is not fixed, and it is more than the filter algorithm. The wrapper algorithm effectively avoids the blindness of threshold setting by evaluating the feature subset. The wrapper algorithm effectively avoids the blindness of threshold setting by evaluating the feature subset. However, the time complexity of the search algorithm is very high, which significantly limits the efficiency of the wrapper. In the hybrid algorithm, the *FS* index by Chis-PSO is the least, and the average *FS* index of SVM is 215.00. The *FS* index by Pearson-PSO is the largest, and the average *FS* index of SVM is 869.50. In general, the *FS* index by the hybrid algorithm is much less than the filter and wrapper algorithms.

Comparison experiments with other classification algorithms, in the filter algorithms, KNN still has poor classification effect on the original gene expression datasets, and the average *ACC* index is 81.58%. IG has the best feature selection effect, and the average *ACC* index of KNN is 87.25%. In the wrapper algorithms, PSO has the best feature selection effect, and the average *ACC* index of KNN is 91.18%. The difference is that IG-ACA has the best feature selection effect in the hybrid algorithm, and the average *ACC* index of KNN is 95.08%. Pearson-PSO has the worst feature selection effect, and the average *ACC* index of SVM is 91.90%. In addition, the *FS* index by the hybrid algorithm is less than that of the filter and the wrapper algorithm. Compared with the traditional feature selection algorithm, the feature subset selected by the IG-GPSO has the best classification effect, and the average *ACC* index of KNN is 96.28%. The result shows that no matter which classification algorithm is used as the evaluation algorithm, the feature selection effect of the IG-GPSO is optimal.

To further verify the effectiveness of the IG-GPSO, we use the Friedman test for multiple comparisons. We selected the *ACC* and *FS* indexes of SVM and KNN on the selected feature subset as data values for testing. In multiple comparisons, the assumption is that there is no significant difference between all algorithms. When the significance level $\alpha = 0.05$, the multiple comparison tests based on Friedman are all able to reject the null hypothesis. It can be concluded that no matter which classification algorithm is used as the evaluation algorithm, the

feature selection effect of the IG-GPSO is significantly better than the traditional feature selection algorithm. The result shows that the difference of evaluation algorithms does not affect the feature selection effect of the IG-GPSO, and the selected feature subset has certain applicability.

In general, the high-dimensional feature space and high feature redundancy of data greatly harm the classification effect of the classification algorithms. SVM and KNN have a significant improvement in each classification index on the datasets after feature selection. Specifically, Pearson has the worst feature selection effect in the single feature selection algorithms. PSO has the best feature selection effect in the single feature selection algorithms. Chis-PSO and IG-ACA have the best feature selection effect in the traditional feature selection algorithms, and the *FS* index is also the least. Pearson-PSO is even worse than some single algorithms in feature selection, which is mainly due to the fact that Pearson removes some important features in the filter stage. Compared with the traditional feature selection algorithm, the *ACC* and the *FS* indexes by the IG-GPSO are optimal. In addition, SVM has the best classification results in all datasets. Therefore, we use SVM as the applied algorithm for cancer diagnosis.

Conclusion and future work

Machine learning is widely used in cancer diagnosis. However, due to the inherent high-dimensional feature space and high feature redundancy of gene expression data, the application effect of existing machine learning algorithms is poor. Based on this, this paper proposes a hybrid feature selection algorithm combining information gain and grouping particle swarm optimization. Different from the traditional filter algorithm, we use the information gain to calculate the IG value of each feature, and rank the value in descending order. Furthermore, this paper proposes a information gain-based ranking and grouping algorithm. By grouping the features, the IG of the features in the group is close. Finally, we use the grouping PSO algorithm to search for the grouped features and evaluate them according to both in-group and out-group. Experimental results show that the IG-GPSO has the best feature selection effect, and the *ACC* indexes of SVM and KNN on 4 gene expression datasets is 98.50% and 96.28%, respectively. In addition, multiple comparison tests show that the IG-GPSO is significantly better than traditional feature selection algorithms. SVM has the best classification effect, and we selected SVM as the applied algorithm for the cancer diagnosis. However, on some gene expression datasets, the number of feature subset selected by the IG-GPSO is not the smallest. This may be due to the fact that the feature grouping does not consider the correlation between features, which results in too many feature subsets. Therefore, the future work is to consider introducing mutual information into feature groups in order to screen the feature groups with very low correlation.

Acknowledgments

Author Contributions

Conceptualization: Fangyuan Yang.

Data curation: Fangyuan Yang, Lisha Sun, Mengjiao Zhai.

Formal analysis: Hong Wang, Lisha Sun.

Funding acquisition: Hong Wang.

Investigation: Fangyuan Yang.

Methodology: Fangyuan Yang, Zhaozhao Xu.

Project administration: Hong Wang, Juan Zhang.

Resources: Hong Wang, Juan Zhang.

Software: Zhaozhao Xu, Hong Wang.

Supervision: Zhaozhao Xu, Hong Wang.

Validation: Zhaozhao Xu, Hong Wang, Mengjiao Zhai, Juan Zhang.

Visualization: Hong Wang, Mengjiao Zhai, Juan Zhang.

Writing – original draft: Fangyuan Yang.

Writing – review & editing: Zhaozhao Xu, Hong Wang, Lisha Sun, Mengjiao Zhai, Juan Zhang.

References

1. Kim Y, Zhu L, Zhu H, et al. Characterizing cancer and COVID-19 outcomes using electronic health records. Plos one. 2022; 17(5): e0267584. https://doi.org/10.1371/journal.pone.0267584 PMID: 35507598

2. Xu Z, Shen D, Nie T, et al. A hybrid sampling algorithm combining M-SMOTE and ENN based on Random forest for medical imbalanced data. Journal of Biomedical Informatics. 2020; 107: 103465. https://doi.org/10.1016/j.jbi.2020.103465 PMID: 32512209

3. Nassif A B, Talib M A, Nasir Q, et al. Breast cancer detection using artificial intelligence techniques: A systematic literature review. Artificial Intelligence in Medicine. 2022: 102276. https://doi.org/10.1016/j.artmed.2022.102276 PMID: 35430037

4. Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC medical informatics and decision making. 2020; 20(1): 1–16. https://doi.org/10.1186/s12911-020-1023-5 PMID: 32013925

5. He Z, Chen Z, Tan M, et al. A review on methods for diagnosis of breast cancer cells and tissues. Cell proliferation. 2020; 53(7): e12822. https://doi.org/10.1111/cpr.12822 PMID: 32530560

6. Veta M, Pluim J P W, Van Diest P J, et al. Breast cancer histopathology image analysis: A review. IEEE transactions on biomedical engineering. 2014; 61(5): 1400–1411. https://doi.org/10.1109/TBME.2014.2303852 PMID: 24759275

7. Chen R J, Lu M Y, Wang J, et al. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE Transactions on Medical Imaging. 2020; 41(4): 757–770. https://doi.org/10.1109/TMI.2020.3021387

8. Pu B, Zhu N, Li K, et al. Fetal cardiac cycle detection in multi-resource echocardiograms using hybrid classification framework. Future Generation Computer Systems. 2021; 115: 825–836. https://doi.org/10.1016/j.future.2020.09.014

9. Mohamed E A, Rashed E A, Gaber T, et al. Deep learning model for fully automated breast cancer detection system from thermograms. PLoS one. 2022; 17(1): e0262349. https://doi.org/10.1371/journal.pone.0262349 PMID: 35030211

10. Li H, Pu B, Kang Y, et al. Research on massive ECG data in XGBoost. Journal of Intelligent & Fuzzy Systems. 2019; 36(2): 1161–1169. https://doi.org/10.3233/JIFS-169890

11. Xu Z, Shen D, Kou Y, et al. A synthetic minority oversampling technique based on Gaussian mixture model filtering for imbalanced data classification. IEEE Transactions on Neural Networks and Learning Systems. 2022. https://doi.org/10.1109/TNNLS.2022.3197156 PMID: 35984792

12. Xu Z, Shen D, Nie T, et al. A cluster-based oversampling algorithm combining SMOTE and k-means for imbalanced medical data. Information Sciences. 2021; 572: 574–589. https://doi.org/10.1016/j.ins.2021.02.056

13. Lee S J, Tseng C H, Yang H Y, et al. RRandom rotboost: An ensemble classification method based on rotation forest and adaboost in random subsets and its application to clinical decision support. Entropy. 2022; 24(5): 617. https://doi.org/10.3390/e24050617 PMID: 35626502

14. Irwan D, Ali M, Ahmed A N, et al. Predicting Water Quality with Artificial Intelligence: A Review of Methods and Applications. Archives of Computational Methods in Engineering. 2023: 1–20.

15. Huang H, Feng X, Zhou S, et al. A new fruit fly optimization algorithm enhanced support vector machine for diagnosis of breast cancer based on high-level features. BMC bioinformatics. 2019; 20: 1–14. https://doi.org/10.1186/s12859-019-2771-z PMID: 31182028

16. Wang M, Chen H. Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis. Applied Soft Computing. 2020; 88: 105946. https://doi.org/10.1016/j.asoc.2019.105946

17. Radovic M, Ghalwash M, Filipovic N, et al. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. BMC bioinformatics. 2017; 18(1): 1–14. https://doi.org/10.1186/s12859-016-1423-9 PMID: 28049413

18. Polat H, Danaei Mehr H, Cetin A. Diagnosis of chronic kidney disease based on support vector machine by feature selection methods. Journal of medical systems. 2017; 41: 1–11. https://doi.org/10.1007/s10916-017-0703-x PMID: 28243816

19. Ghaddar B, Naoum-Sawaya J. High dimensional data classification and feature selection using support vector machines. European Journal of Operational Research. 2018; 265(3): 993–1004. https://doi.org/10.1016/j.ejor.2017.08.040

20. Li J, Zhou Z, Dong J, et al. Predicting breast cancer 5-year survival using machine learning: A systematic review. PloS one. 2021; 16(4): e0250370. https://doi.org/10.1371/journal.pone.0250370 PMID: 33861809

21. Lee S J, Xu Z, Li T, et al. A novel bagging C4. 5 algorithm based on wrapper feature selection for supporting wise clinical decision making. Journal of biomedical informatics. 2018; 78: 144–155. https://doi.org/10.1016/j.jbi.2017.11.005 PMID: 29137965

22. Remeseiro B, Bolon-Canedo V. A review of feature selection methods in medical applications. Computers in biology and medicine. 2019; 112: 103375. https://doi.org/10.1016/j.compbiomed.2019.103375 PMID: 31382212

23. Hancer E, Xue B, Zhang M. Differential evolution for filter feature selection based on information theory and feature ranking. Knowledge-Based Systems. 2018; 140: 103–119. https://doi.org/10.1016/j.knosys.2017.10.028

24. Chen G, Chen J. A novel wrapper method for feature selection and its applications. Neurocomputing. 2015; 159: 219–226. https://doi.org/10.1016/j.neucom.2015.01.070

25. Got A, Moussaoui A, Zouache D. Hybrid filter-wrapper feature selection using whale optimization algorithm: A multi-objective approach. Expert Systems with Applications. 2021; 183: 115312. https://doi.org/10.1016/j.eswa.2021.115312

26. Azhagusundari B, Thanamani A S. Feature selection based on information gain. International Journal of Innovative Technology and Exploring Engineering (IJITEE). 2013; 2(2): 18–21.

27. Alshaer H N, Otair M A, Abualigah L, et al. Feature selection method using improved CHI Square on Arabic text classifiers: analysis and application. Multimedia Tools and Applications. 2021; 80: 10373–10390. https://doi.org/10.1007/s11042-020-10074-6

28. Liu Y, Mu Y, Chen K, et al. Daily activity feature selection in smart homes based on pearson correlation coefficient. Neural Processing Letters. 2020; 51: 1771–1787. https://doi.org/10.1007/s11063-019-10185-8

29. Lu H, Chen J, Yan K, et al. A hybrid feature selection algorithm for gene expression data classification. Neurocomputing. 2017; 256: 56–62. https://doi.org/10.1016/j.neucom.2016.07.080

30. Sharkawy R M, Ibrahim K, Salama M M A, et al. Particle swarm optimization feature selection for the classification of conducting particles in transformer oil. IEEE Transactions on Dielectrics and Electrical Insulation. 2011; 18(6): 1897–1907. https://doi.org/10.1109/TDEI.2011.6118628

31. Li S, Wu H, Wan D, et al. An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine. Knowledge-Based Systems. 2011; 24(1): 40–48. https://doi.org/10.1016/j.knosys.2010.07.003

32. Kabir M M, Shahjahan M, Murase K. A new hybrid ant colony optimization algorithm for feature selection. Expert Systems with Applications. 2012; 39(3): 3747–3763. https://doi.org/10.1016/j.eswa.2011.09.073

33. Wang A, An N, Chen G, et al. Accelerating wrapper-based feature selection with K-nearest-neighbor. Knowledge-Based Systems. 2015; 83: 81–91. https://doi.org/10.1016/j.knosys.2015.09.030

34. Lu Y, Liang M, Ye Z, et al. Improved particle swarm optimization algorithm and its application in text feature selection. Applied Soft Computing. 2015; 35: 629–636. https://doi.org/10.1016/j.asoc.2015.07.005

35. Tran B, Xue B, Zhang M. A new representation in PSO for discretization-based feature selection. IEEE Transactions on Cybernetics. 2017; 48(6): 1733–1746. https://doi.org/10.1109/TCYB.2017.2714145 PMID: 28650835

36. Liu Z, Chang B, Cheng F. An interactive filter-wrapper multi-objective evolutionary algorithm for feature selection. Swarm and Evolutionary Computation. 2021; 65: 100925. https://doi.org/10.1016/j.swevo.2021.100925

37. Song X F, Zhang Y, Gong D W, et al. A fast hybrid feature selection based on correlation-guided clustering and particle swarm optimization for high-dimensional data. IEEE Transactions on Cybernetics. 2021; 52(9): 9573–9586. https://doi.org/10.1109/TCYB.2021.3061152

38. Zhu P, Zuo W, Zhang L, et al. Unsupervised feature selection by regularized self-representation. Pattern Recognition. 2015; 48(2): 438–446. https://doi.org/10.1016/j.patcog.2014.08.006

39. Cooper L A D, Kong J, Gutman D A, et al. An integrative approach for in silico glioma research. IEEE Transactions on Biomedical Engineering. 2010; 57(10): 2617–2621. https://doi.org/10.1109/TBME.2010.2060338 PMID: 20656651

40. Tan Z H, Tan P, Jiang Y, et al. Multi-label optimal margin distribution machine. Machine Learning. 2020; 109: 623–642. https://doi.org/10.1007/s10994-019-05837-8

41. Maldonado S, Weber R, Famili F. Feature selection for high-dimensional class-imbalanced data sets using support vector machines. Information sciences. 2014; 286: 228–246. https://doi.org/10.1016/j.ins.2014.07.015

42. Ghosh D, Singh A, Shukla K K, et al. Extended Karush-Kuhn-Tucker condition for constrained interval optimization problems and its application in support vector machines. Information Sciences. 2019; 504: 276–292. https://doi.org/10.1016/j.ins.2019.07.017

43. Shawe-Taylor J, Sun S. A review of optimization methodologies in support vector machines. Neurocomputing. 2011; 74(17): 3609–3618. https://doi.org/10.1016/j.neucom.2011.06.026

44. Omuya E O, Okeyo G O, Kimwele M W. Feature selection for classification using principal component analysis and information gain. Expert Systems with Applications. 2021; 174: 114765. https://doi.org/10.1016/j.eswa.2021.114765

45. Białynicki-Birula I, Mycielski J. Uncertainty relations for information entropy in wave mechanics. Communications in Mathematical Physics. 1975; 44: 129–132. https://doi.org/10.1007/BF01608825

46. Xu X, Lin P. Parameter identification of sound absorption model of porous materials based on modified particle swarm optimization algorithm. PloS One. 2021; 16(5): e0250950. https://doi.org/10.1371/journal.pone.0250950 PMID: 33945538

47. Xu W, Ismail M M, Liu Y, et al. PParameter optimization of adaptive flux-weakening strategy for permanent-magnet synchronous motor drives based on particle swarm algorithm. IEEE Transactions on Power Electronics. 2019; 34(12): 12128–12140. https://doi.org/10.1109/TPEL.2019.2908380

48. Zhang Q, Wang P, Chen Z. An improved particle filter for mobile robot localization based on particle swarm optimization. Expert Systems with Applications. 2019; 135: 181–193. https://doi.org/10.1016/j.eswa.2019.06.006