

Article

DA-Based Parameter Optimization of Combined Kernel Support Vector Machine for Cancer Diagnosis

Tao Xie, Jun Yao * and Zhiwei Zhou

School of Mechatronic Engineering and Automation, Shanghai University, 99 ShangDa Road, Shanghai 200444, China; xietao1280@shu.edu.cn (T.X.); zhouzwcs@163.com (Z.Z.)

* Correspondence: grandone0529@shu.edu.cn

Received: 31 March 2019; Accepted: 30 April 2019; Published: 6 May 2019



Abstract: As is well known, the correct diagnosis for cancer is critical to save patients' lives. Support vector machine (SVM) has already made an important contribution to the field of cancer classification. However, different kernel function configurations and their parameters will significantly affect the performance of SVM classifier. To improve the classification accuracy of SVM classifier for cancer diagnosis, this paper proposed a novel cancer classification algorithm based on the dragonfly algorithm and SVM with a combined kernel function (DA-CKSVM) which was constructed from a radial basis function (RBF) kernel and a polynomial kernel. Experiments were performed on six cancer data sets from University of California, Irvine (UCI) machine learning repository and two cancer data sets from Cancer Program Legacy Publication Resources to evaluate the validity of the proposed algorithm. Compared with four well-known algorithms: dragonfly algorithm-SVM (DA-SVM), particle swarm optimization-SVM (PSO-SVM), bat algorithm-SVM (BA-SVM), and genetic algorithm-SVM (GA-SVM), the proposed algorithm was able to find the optimal parameters of SVM classifier and achieved better classification accuracy on cancer datasets.

Keywords: cancer; combined kernel function; dragonfly algorithm; support vector machine (SVM); parameter optimization; classification

1. Introduction

In the 21st century, cancer is expected to be the major cause of death all over the world. The GLOBOCAN 2018 cancer morbidity and mortality estimates published by the International Agency for Research on Cancer showed that there were 18.1 million new cancer cases and 9.6 million cancer deaths in 2018 [1]. The correct diagnosis of cancer is essential for patients to receive timely and correct treatment. Machine learning plays a unique and important role in the field of cancer treatment. For example, some researchers applied neural networks to the classification of breast cancer [2,3], and Dongmei Ai et al. [4] identified intestinal microorganisms associated with colorectal cancer by means of decision tree aggregation with a random forest model.

Support vector machine (SVM) is a supervised machine learning method used to solve classification and regression problems, firstly proposed by Vapnik on the basis of statistical learning theory [5]. SVM was applied in many fields, such as economics [6], electric [7], and medical science [8]. Especially in the field of cancer diagnosis, many studies have already proven the excellent performance of SVM classifier [9–11]. SVM uses the principle of structural risk minimization instead of empirical minimization and it can obtain a better generalization ability from limited samples. Moreover, when facing the problem that data cannot be linearly separated, SVM can not only use the idea of kernel function to map nonlinear features to a high-dimensional space, but it can also avoid the problem of “dimensional disaster”.

The performance of SVM classifier depends on three aspects, these being the penalty parameter C of SVM classifier, the type of the kernel function, and its parameters. To improve classification accuracy of SVM classifier, some approaches were presented to search for the optimal parameters, such as grid search [12] and gradient descent [13]. Although these methods have proven the effectiveness in the corresponding literature experiments, they are likely to fall into the local optimum point easily and have the defect of low efficiency. In recent years, some meta-heuristic algorithms, such as the dragonfly algorithm (DA) [14], particle swarm optimization (PSO) [15], bat algorithm (BA), [16] and genetic algorithm (GA) [17–19] have achieved competitive results when they were used to tune SVM classifier's parameters. However, most of this research only focused on the SVM classifier with a single kernel function. Though some literature [20,21] indicates that combining multiple kernel functions can obtain better performance than a single kernel function, little research has provided an in-depth analysis of the performance of SVM classifier with a combined kernel function. There would therefore seem to be a definite need to systematically study the complex optimization problem in the SVM classifier with a combined kernel.

In 2015, Mirjalili proposed a new meta-heuristic algorithm called the dragonfly algorithm (DA) [22], which has already been used to solve different optimization problems, such as feature selection [23,24], the knapsack problem [25], and image processing [26]. Considering that DA has an excellent global search ability and there are few studies on SVM classifier with combined kernels in the field of cancer classification, this paper proposed a novel classification algorithm based on DA and SVM classifier with a combined kernel function (DA-CKSVM) to improve the classification ability for cancer diagnosis. The objective of this research was to construct an SVM classifier with two different kernel functions and use DA to optimize all the parameters in this SVM classifier, such as the parameters in both kernel functions, the weight coefficient of the combined kernel, and SVM's penalty parameter C .

The overall structure of the study takes the form of eight chapters, including this introductory chapter. The remaining part of the paper proceeds as follows: The related work on cancer classification is described briefly in Section 2. Section 3 introduces the basic idea of SVM. Section 4 deals with the construction of the combined kernel. Then, DA is introduced in Section 5. Section 6 is devoted to present the proposed algorithm. Section 7 focuses on the experimental results and discussions. Finally, conclusions and future work are provided in Section 8.

2. Related Work

Since SVM was proposed, many researchers have used it to conduct research on cancer classification. For example, Guangru Xu et al. [27] used SVM analysis to predict the recurrence risk and prognosis for patients with colon cancer. Youlin Tuo et al. [28] constructed an SVM classifier to evaluate the possibility of breast cancer metastasis and obtained high classification accuracy in several independent data sets. Yuanpeng Li et al. [29] used both SVM models and partial-least-square discriminant analysis to diagnose early gastric cancer. The experimental results showed the diagnostic model obtained from SVM was evidently better than the partial-least-square discriminant analysis.

It should be noted that a great deal of research has already recognized that the parameters of SVM classifier play a vital role in improving the effects of classification. The kernel function and its parameters define a nonlinear mapping from the input space to the high-dimensional space, while the trade-off between minimizing the training error and maximizing the classification margin was determined by the penalty parameter C of SVM. Different parameter configurations will lead to different classification results. Therefore, how to select appropriate parameters becomes the main challenge on improving the classification ability of SVM classifier.

A number of researchers have focused on optimizing the parameters of kernel functions to obtain better classification accuracy in cancer diagnosis. M. Prabukumar et al. [30] used SVM classifier to identify the lung cancer and more than 98% accuracy was achieved. The grid search method was employed to search for the optimal parameters in this study. Himar Fabelo et al. [31] used radial basis function (RBF) kernel function, linear kernel function, polynomial kernel function, and

sigmoid kernel function to construct SVM classifiers to recognize brain tumors, respectively, and the cross-validation method was utilized to find the optimum parameters of SVM classifiers. By comparing the classification effects of four different models, the results proved that polynomial kernel had advantages in a few evaluation metrics, but in general RBF kernel function got the best classification results. The literature [17] applied linear, quadratic, RBF, and third-order polynomial SVMs for the screening and classification of prostate cancer, in which the parameter of RBF kernel and parameter C were optimized by the exhaustive method. However, the four models did not perform very well. Then, they combined SVM with the principal component analysis (PCA), the successive projections algorithm (SPA), and GA to automatically optimize penalty parameter C and the parameters in RBF kernel. The experimental results showed that the combination of meta-heuristic algorithm GA and SVM achieved the best performance. Overall, there seems to be some evidence to indicate that polynomial kernel function and RBF kernel function have better performance than other kernel functions, and the meta-heuristic algorithms have a better optimization ability than conventional optimization methods such as grid search and gradient descent.

Further understanding of the nature of kernel functions will help to build a more powerful SVM classifier. In 2001, Scholkopf et al. first divided the kernel function into local kernel function and global kernel function [32]. Surveys conducted by Simts and Jordaan [33] showed that local kernels have a good interpolation ability, while global kernels have a good extrapolation ability. They linearly combined a local kernel function with a global kernel function to obtain a novel kernel function that exploited the advantages of both kernel functions to make SVM classifiers achieve better performance. Another important finding was that the performance of SVM classifier with the combined kernel function is influenced by the weight coefficient of the kernel function. To stratify and predict clinical outcomes in patients with ovarian cancer, Jaya Thomas and Lee Sael [34] combined a linear kernel function and a RBF kernel function in a weighted linear combination. They achieved satisfactory results after optimizing parameters and the weight coefficient. However, they only used a separate dataset that accounted for 25% of the total sample size to determine the parameters of kernels, which may have led to less precise parameters. In terms of the weight coefficient optimization, they used the optimization method proposed by Zien et al. [35]. Therefore, optimizing the weight coefficient and parameters of kernel functions separately may cause uncoordinated results. Nguyen et al. [36] combined three kernel functions which included inverse multi-quadric kernel, RBF kernel, and sigmoid kernel, and optimized parameters with an evolutionary algorithm to construct an SVM classifier. The classification results on cancer datasets indicated the methodology was superior to a single kernel function. However, the weight coefficients of kernel functions were not taken into account.

The traditional optimization methodologies such as grid search and gradient descent are not only time-consuming, but also insufficient to synchronously optimize all parameters in the SVM classifier with multiple kernel functions. Meta-heuristic algorithms have already been applied to search for the optimal parameters of SVM classifier with a single kernel function and have been proven to have high classification accuracy and stability. However, up until now, far too little attention has been paid to the use of the meta-heuristic algorithm to optimize parameters in an SVM classifier with multiple kernel functions by solving the complex optimization problem.

3. The Basic Idea of SVM and the Construction of the Combined Kernel Function

As a powerful supervised learning method, SVM is suitable to deal with classification problems and regression problems. On the one hand, SVM has a high accuracy rate for linearly inseparable data. On the other hand, linearly inseparable data can be mapped into a high-dimensional space through kernel functions in SVM. This chapter will briefly introduce the basic principles of SVM with the classical binary classification as an example.

3.1. Linear SVM

Suppose that training data $U = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ is linearly separable, where $x_i \in R^n$ represents the i -th training sample that is represented by m features, and $y_i \in \{-1, 1\}$ represents the corresponding class label. The hyperplane $\omega^T x + b = 0$ is the decision boundary of two types of data, where a weight vector ω is a normal vector of the hyperplane, b is the deviation, and x is the training sample. The goal of SVM is to determine the appropriate ω and b to make the hyperplane as far as possible from the nearest samples. Therefore, the training samples can be correctly classified by:

$$\omega^T x_i + b \geq +1 \quad \text{for } y_i = +1 \quad (1)$$

$$\omega^T x_i + b \leq -1 \quad \text{for } y_i = -1 \quad (2)$$

One inequality can be obtained for the two formulas above:

$$y_i(\omega^T x_i + b) \geq 1, \forall i = 1, 2, \dots, n \quad (3)$$

In order to get the optimal hyperplane $(\omega \cdot x) + b = 0$, SVM needs to deal with the optimization problem which is shown below:

$$\begin{cases} \min \Phi(\omega) = \frac{1}{2} \|\omega\|^2 \\ \text{s.t. } y_i(\omega^T x_i + b) \geq 1, i = 1, 2, \dots, n \end{cases} \quad (4)$$

where $(\min \Phi(\omega) = \frac{1}{2} \|\omega\|^2)$ is the objective function and $y_i(\omega^T x_i + b) \geq 1$ is the constraint. The optimal solution of Equation (4) is the saddle point of the following Lagrange function:

$$\begin{aligned} L(\omega, b, \alpha) &= \frac{1}{2} \|\omega\|^2 - \sum_{i=1}^n \alpha_i [y_i(\omega^T x_i + b) - 1] \\ \alpha_i &\geq 0, i = 1, 2, \dots, n \end{aligned} \quad (5)$$

where α is the Lagrange multiplier. Since the gradients of ω and b at the saddle point are zero, therefore:

$$\frac{\partial L}{\partial \omega} = \omega - \sum_{i=1}^n \alpha_i y_i x_i = 0 \Rightarrow \omega = \sum_{i=1}^n \alpha_i y_i x_i \quad (6)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^n \alpha_i y_i = 0 \Rightarrow \sum_{i=1}^n \alpha_i y_i = 0 \quad (7)$$

Substituting Equations (6) and (7) into Equation (5), the problem of constructing an optimal hyperplane is translated into a dual quadratic programming problem:

$$\begin{cases} \max W(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j \\ \text{s.t. } \sum_{i=1}^n \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, 2, \dots, n \end{cases} \quad (8)$$

ω , b and α can be obtained by solving Equations (6)–(8). Only a small part of α_i are greater than zero, and these corresponding samples which are closest to the hyperplane are called support vectors (SVs).

For an unknown sample x , the following formula can be used to determine its class:

$$y = \text{sgn}(\omega^T x + b) \quad (9)$$

However, in practical applications, data are usually not linearly separable, which will result in a large number of misclassified samples. Hence, a relaxation term $\xi_i \geq 0$ needs to be added into linear SVM to relax the constraints as follows:

$$y_i(\omega^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, 2, \dots, n \quad (10)$$

The corresponding objective function is:

$$\Phi(\omega, \xi) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i \quad (11)$$

where penalty parameter C represents the degree of punishment for the error. The larger the C is, the heavier the penalty will be.

3.2. Nonlinear SVM

For the non-linearly separable data, a kernel function is used to map them into a high-dimensional space to make them linearly separable. The kernel function is defined as follows:

$$K(x_1, x_2) = \phi(x_1)^T \phi(x_2) \quad (12)$$

The optimization problem of SVM is shown in Equation (13):

$$\begin{cases} \min \Phi(\omega) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t. } y_i(\omega^T \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, 2, \dots, n \end{cases} \quad (13)$$

4. The Construction of the Combined Kernel Function

Various types of kernel functions are usually applied in SVM classifier, such as the RBF kernel function, linear kernel function, polynomial kernel function, and sigmoid kernel function. Each type of kernel function has its own advantages and disadvantages. Studies conducted by Simts and Jordaan [33] showed that in local kernels, only the data close to each other will affect the value of kernels, while in global kernels, the data far from each other will also affect the value of kernels. This means that the learning ability of local kernels is stronger than that of global kernels, but the generalization ability of local kernels is weaker than that of global kernels. In order to improve the classification accuracy of SVM classifier, this paper linearly combined a local kernel function with a global kernel function to take advantages of these two kernel functions.

Four alternative kernel functions are listed below, which are usually used in SVM classifier:

- (1) Linear kernel function

$$K_{lin}(x, x_i) = x^T * x_i \quad (14)$$

- (2) Polynomial kernel function

$$K_{poly}(x, x_i) = (x^T * x_i + 1)^d \quad (15)$$

- (3) RBF kernel function

$$K_{RBF}(x, x_i) = \exp(-\text{gamma} * |x - x_i|^2) \quad (16)$$

- (4) Sigmoid kernel function

$$K_{sig}(x, x_i) = \tanh(\text{gamma} * x^T * x_i + r) \quad (17)$$

RBF kernel function is a local kernel function, while the other three are all global kernel functions. According to the composition conditions of the kernel function, if a combined kernel function is linearly

constructed from existing kernels, it is still a kernel function [37] and can be used in SVM classifier. In view of local kernel function's predominant learning ability and global kernel function's outstanding generalization ability, this paper combined a RBF kernel with a polynomial kernel to obtain a novel combined kernel function which is shown in Equation (18):

$$K_c = \lambda K_{RBF} + (1 - \lambda) K_{poly} \quad (18)$$

where λ is the weight coefficient in the range from 0–1. In this SVM classifier, the parameter tuning problem is to search for the optimal combination of the parameters set (C , γ , d , λ). The dragonfly algorithm is applied in this paper to solve this complex optimization problem.

5. Dragonfly Algorithm (DA)

The dragonfly algorithm (DA) is a novel swarm intelligence algorithm presented by Mirjalili in 2015 [22]. The exclusive cluster behaviors of dragonflies, namely hunting and migration, are the main source of inspiration for the algorithm. The hunting swarm is called the static swarm where dragonflies gather into small groups and fly back and forth in a small area to hunt prey. On the other hand, in the dynamic swarm, a big swarm of dragonflies can fly a long distance in one direction. These two states are very similar to the exploratory phase and the exploitation phase of meta-heuristic algorithms. The flight of the static swarm in a small area is similar to the exploring stage of the optimization algorithm, while the flying of the dynamic group along one direction is beneficial to exploitation.

The behavior of any swarms follows the principles given by Reynolds [38]:

- Separation, whose aim is to avoid the collision between individuals and their neighbors in the static swarm.
- Alignment, whose purpose is to match the individual velocity with others in the same group.
- Cohesion, which is used to indicate the tendency of individuals to move towards the center of the group.

In order to survive, individuals in the group should be attracted to food and distracted by outward enemies. Considering these behaviors, five primary factors are used to update individuals' positions in a swarm. The mathematical models for these behaviors are as follows:

(1) Separation

$$S_i = - \sum_{j=1}^N X - X_j \quad (19)$$

where X represents the position of current individual, X_j shows the position of j -th adjacent individual, and N is the amount of neighboring individuals.

(2) Alignment

$$A_j = \frac{\sum_{j=1}^N V_j}{N} \quad (20)$$

where V_j is the velocity of the j -th neighboring individual.

(3) Cohesion

$$C_i = \frac{\sum_{j=1}^N X_j}{N} - X \quad (21)$$

(4) Attraction

$$F_i = X^+ - X \quad (22)$$

where X^+ represents the position of the food source.

(5) Distraction

$$E_i = X^- + X \quad (23)$$

where X^- represents the position of the enemy.

The behavior of dragonflies is represented by the combination of the behaviors above. Similar to the velocity vector in PSO, the step vector ΔX is used in DA to update the position of dragonflies, which is defined as follows:

$$\Delta X_{t+1} = (sS_i + aA_i + cC_i + fF_i + eE) + w\Delta X_t \quad (24)$$

where s and S_i represent the separation weight and the separation of i -th individual, respectively; a and A_i are the alignment weight and the alignment of the i -th individual, respectively; c and C_i indicate the cohesion weight and the cohesion of the i -th individual, respectively; f and F_i show the food factor and the food source of the i -th individual, respectively; e and E_i are the enemy factor and the position of the enemy of the i -th individual, respectively; w represents the inertia weight, and t is the iteration counter. The position vector is represented by the following formula:

$$X_{t+1} = X_t + \Delta X_{t+1} \quad (25)$$

where t is the current iteration.

In the absence of adjacent solutions, DA uses random walk or *Levy flight* [39] to improve the randomness, stochastic behavior, and exploration of dragonflies. Therefore, the dragonflies update their position as follows:

$$X_{t+1} = X_t + Levy(d) \times X_t \quad (26)$$

where d is the dimension of the position vector. The *Levy flight* can be calculated according to Equation (27):

$$Levy(x) = 0.01 \times \frac{r_1 \times \sigma}{|r_2|^{\frac{1}{\beta}}} \quad (27)$$

where r_1, r_2 are two random numbers in $[0, 1]$, β is a constant, and σ is calculated by Equation (28):

$$\sigma = \left(\frac{\Gamma(1 + \beta) \times \sin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right) \times \beta \times 2^{\left(\frac{\beta-1}{2}\right)}} \right)^{1/\beta} \quad (28)$$

where $\Gamma(x) = (x-1)!$.

6. Proposed Algorithm: DA-CKSVMS

This section will elaborate the proposed algorithm DA-CKSVMS, which uses DA to optimize parameters of SVM with a combined kernel.

6.1. The Basic Process of DA-CKSVMS

In the DA-CKSVMS algorithm, each dragonfly represents a solution, that is, a combination of the parameters set (C, γ, d, λ) which defines a four-dimensional search space for the optimization problem. The main process of the proposed algorithm is given below:

Algorithm 1: The main process of DA-CKSVMS

Step 1: Set the maximum iteration times, the number of dragonflies, and the upper and lower bounds of each parameter in the parameters set (C, γ, d, λ) .

Step 2: Initialize the step vectors, the values of s, a, c, f, e and w in Equation (24), and the position of each individual.

Step 3: Train the SVM classifier with the training set and test it with the testing set.

Step 4: Evaluate the fitness value of each individual and update the enemy and food source.

Step 5: Update the values of s, a, c, f, e and w .

Step 6: Calculate S, A, C, F and E according to Equations (19)–(23).

Step 7: Update the neighboring radius.

Step 8: If the dragonfly has at least one neighbor, the step vector and the position vector of the dragonfly will be calculated according to Equations (24) and (25). If not, the position vector will be updated by Equation (26).

Step 9: Adjust the new position based on boundaries of the parameters.

Step 10: If the maximum iteration times is reached, go to the Step 11. Otherwise, loop to Step 3.

Step 11: Output the final SVM classifier with optimal parameters.

The schematic diagram in Figure 1 shows the whole process of the DA-CKSVM algorithm. As can be seen from the diagram, DA-CKSVM algorithm first initializes the relevant parameters. After that, the DA-CKSVM algorithm uses the normalized data set to train and test the SVM classifier with a combined kernel function by means of cross-validation. In the optimization process, the DA algorithm is employed to search for the best solution of the parameters set (C, γ, d, λ). When the maximum number of iterations is reached, the algorithm is terminated and the final SVM classifier with the optimal parameters is achieved.

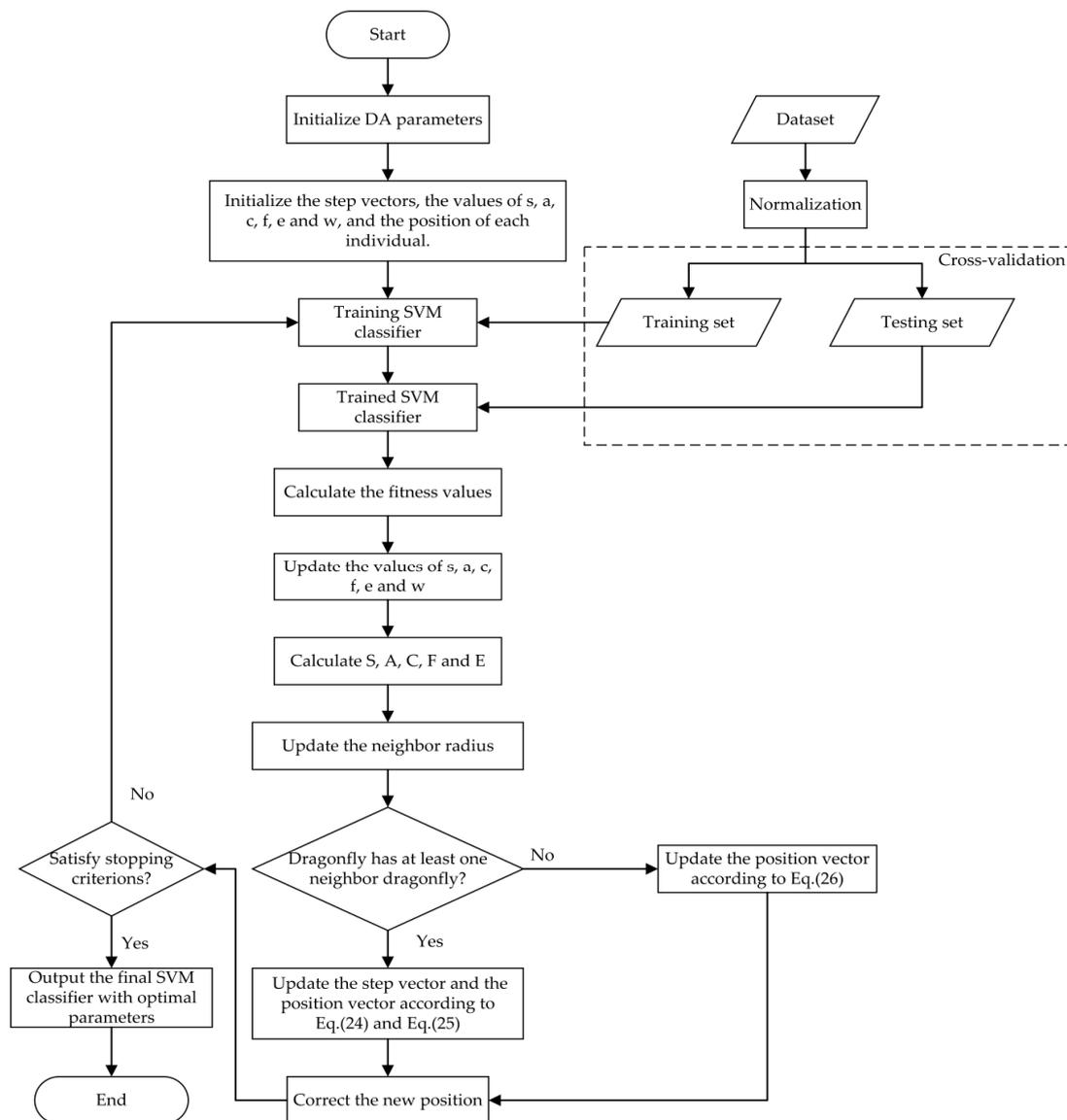


Figure 1. Flow chart of the proposed DA-CKSVM algorithm.

6.2. Fitness Function

The fitness function is applied to evaluate the performance of the parameter searching. The fitness function in this paper is defined as the classification error rate on test sets, as shown in Equation (29):

$$Fitness(I_i^t) = 100 - \frac{1}{K} \sum_{k=1}^K \frac{1}{N} \sum_{j=1}^N \delta(c(x_j), y_j) \quad (29)$$

The performance of the individual at the iteration t is evaluated by above equation. K and N denote the number of folds of cross-validation and the number of samples in the test set, respectively. $c(x_j)$ represents the classification result of the j -th sample in the test set, y_j indicates the label of each sample, and δ shows the relationship between $c(x_j)$ and y_j , as shown in the following formula:

$$\delta(c(x_j), y_j) = \begin{cases} 1 & \text{if } c(x_j) = y_j \\ 0 & \text{otherwise} \end{cases} \quad (30)$$

7. Experimental Results and Discussion

7.1. Data Sets and Experimental Platform

In order to validate the proposed DA-CKSVM algorithm's performance in cancer classification, experiments were carried out on six cancer data sets from University of California, Irvine (UCI) machine learning repository and two cancer data sets from Cancer Program Legacy Publication Resources. Breast Cancer Coimbra (BCC), Haberman's Survival (HS), Hepatocellular Carcinoma (HCC), Thoracic Surgery (TS), Breast Cancer Wisconsin Diagnostic (BCWD), and Breast Cancer Wisconsin Prognostic (BCWP) are from UCI machine learning repository; Diffuse Large B-cell Lymphoma (DLBCL_D) and Breast_A (B_A) come from Cancer Program Legacy Publication Resources. Table 1 lists the descriptions of all data sets.

Table 1. Data sets descriptions.

Data Sets	Instances	Features	Classes
Breast Cancer Coimbra (BCC)	116	10	2
Haberman's Survival (HS)	306	3	2
Hepatocellular Carcinoma (HCC)	165	49	2
Thoracic Surgery (TS)	470	17	2
Breast Cancer Wisconsin Diagnostic (BCWD)	569	30	2
Breast Cancer Wisconsin Prognostic (BCWP)	198	33	2
Diffuse Large B-cell Lymphoma (DLBCL_D)	129	3795	4
Breast_A (B_A)	98	1213	3

All the experiments in this paper were implemented on Matlab 2014(a) and SVM classifier was trained with a library for support vector machines (LIBSVM) [40]. The details of the experimental platform are given in Table 2.

Table 2. Experimental Platform.

NAME	Detailed Settings
Hardware	
Central Processing Unit (CPU)	Advanced Micro Devices (AMD) Ryzen 7 2700X
Frequency	3.70GHz
Random Access Memory (RAM)	16G
Hard Drive	250G
Software	
Operating System	Windows 10
Programming Language	MATLAB R2014a
Tool for support vector machine (SVM)	LIBSVM

7.2. Data Preprocessing

Given that the characteristics in larger numerical ranges may dominate those in smaller numerical ranges [27], each feature is normalized and scaled in the range of [0, 1] in this paper:

$$f' = \frac{f - \min}{\max - \min} \quad (31)$$

where f is the original value, f' is the scaled value, “min” represents the minimum value of the characteristic, “max” represents the maximum value of the characteristic.

7.3. Cross-Validation

All experiments used a k -fold cross-validation in which the original data set was randomly divided into k subsets of (approximately) equal size. Each time, $k-1$ subsets were selected as the training set and the remaining subset was used as the test set. This process was repeated k times. Finally, the average of the classification accuracy on the testing set was used as the evaluation value. In this paper, k was set to 10.

7.4. Experimental Results

To validate the performance of the proposed DA-CKSVM algorithm, experiments in this section were carried out to compare the DA-CKSVM algorithm with dragonfly algorithm-SVM (DA-SVM) [14], particle swarm optimization-SVM (PSO-SVM) [15], bat algorithm-SVM (BA-SVM) [16], and genetic algorithm-SVM (GA-SVM) [19]. According to the literature, all the algorithms for comparison employed a single RBF kernel to implement SVM classifier. Table 3 shows the initial parameters of each algorithm. The parameters of the proposed algorithm were consistent with those of DA algorithm. The iteration times and the population size in all algorithms were respectively set to 300 and 30 to obtain fair and reliable experiment results. Furthermore, 10 experiment trials were carried out in each algorithm to evaluate the final results by average classification accuracy and standard deviation, which may minimize the influence of randomness.

Table 3. Initial parameters of algorithms.

Algorithm	Parameter	Value
Dragonfly algorithm (DA)	Number of dragonflies	30
	Generations	300
Particle swarm optimization (PSO)	c_1	2
	c_2	2
	Inertia w	1
	Number of particles	30
	Generations	300
Bat algorithm (BA)	Minimum frequency	0
	Maximum frequency	2
	Loudness	0.5
	Pulse rate	0.5
	Number of bats	30
	Generations	300
Genetic algorithm (GA)	Crossover ratio	0.6
	Mutation ratio	0.1
	Selection mechanism	Roulette wheel
	Population size	30
	Generations	300

Table 4 lists each algorithm's final classification accuracy over 10 trials in the form of average \pm standard deviation. It is shown in Table 4 that in comparison with the PSO-SVM algorithm and GA-SVM algorithm, the DA-CKSVM algorithm achieves the highest classification accuracy on all datasets. The DA-CKSVM algorithm outperforms the BA-SVM algorithm and obtains the best performance on seven data sets out of eight data sets. The DA-CKSVM algorithm achieves better accuracy than the DA-SVM algorithm on six data sets out of eight data sets. The best accuracy over 10 trials for each algorithm is shown in Table 5. As shown, the DA-CKSVM algorithm achieves the best result on seven data sets out of eight data sets. Tables 4 and 5 reveal that better overall classification accuracy on cancer data sets can be obtained by means of constructing a combined kernel function for SVM classifier and searching for optimal kernel parameters.

Table 4. Classification accuracy and standard deviation for all data sets.

Data Set	Average Classification Accuracy and Standard Deviation (%)				
	DA-CKSVM	DA-SVM	PSO-SVM	BA-SVM	GA-SVM
BCC	<u>84.00 \pm 1.21</u>	82.80 \pm 0.00	82.81 \pm 0.02	80.47 \pm 3.44	75.67 \pm 2.94
HS	<u>77.94 \pm 0.56</u>	76.95 \pm 0.28	75.05 \pm 2.76	76.78 \pm 1.01	73.70 \pm 2.91
HCC	75.88 \pm 1.15	<u>77.43 \pm 0.00</u>	75.26 \pm 5.07	76.47 \pm 3.02	67.82 \pm 6.59
TS	<u>85.19 \pm 0.15</u>	85.11 \pm 0.00	82.04 \pm 0.11	85.11 \pm 0.00	81.83 \pm 0.11
BCWD	98.07 \pm 0.00	<u>98.30 \pm 0.08</u>	97.89 \pm 0.18	96.71 \pm 0.78	97.15 \pm 0.07
BCWP	<u>82.06 \pm 0.71</u>	81.39 \pm 0.16	81.34 \pm 0.00	78.13 \pm 0.87	77.46 \pm 0.35
DLBCL_D	<u>80.53 \pm 0.03</u>	75.77 \pm 0.00	75.77 \pm 0.00	75.77 \pm 0.00	40.01 \pm 6.10
B_A	<u>95.00 \pm 0.00</u>	92.00 \pm 0.00	92.00 \pm 0.00	92.00 \pm 0.00	56.61 \pm 11.91

The bold values with underline represent that the corresponding algorithm obtains the highest results.

One noteworthy problem can be found by analyzing Tables 4 and 5. The DA-CKSVM algorithm shows excellent performance on most cancer datasets. But for two datasets, namely HCC and BCWD, the DA-CKSVM algorithm has no advantage over the comparison algorithms. In particular, Table 5 indicates the best result in the DA-CKSVM algorithm on dataset HCC equals to that in the comparison algorithms, but the DA-CKSVM algorithm does not perform better than others when comparing the average results over 10 trials. This implies that though the complementary characteristics of two different kernels in the combined kernel function may improve the SVM classifier's classification ability

on most of data sets, for certain data sets, the best result can be obtained only when optimizing the SVM classifier with a single RBF kernel. Even if the optimal weight coefficient λ is very close to 1, the existence of the polynomial kernel in the combined kernel function may reduce the classification accuracy slightly in several trials.

Table 5. The highest result of each algorithm.

Data Set	The Best Results (%)				
	DA-CKSV	DA-SVM	PSO-SVM	BA-SVM	GA-SVM
BCC	86.21	82.80	82.88	82.88	81.97
HS	78.77	77.47	77.16	77.47	76.83
HCC	77.43	77.43	77.43	77.43	77.43
TS	85.53	85.11	82.13	85.11	81.91
BCWD	98.07	98.42	98.24	97.19	97.18
BCWP	83.37	81.84	81.34	79.39	77.82
DLBCL_D	80.58	75.77	75.77	75.77	57.37
B_A	95.00	92.00	92.00	92.00	89.89

The bold values with underline represent that the corresponding algorithm obtains the highest results.

Figure 2a–h shows the fitness curves for each algorithm. The curves are drawn from the average fitness values over 10 trials.

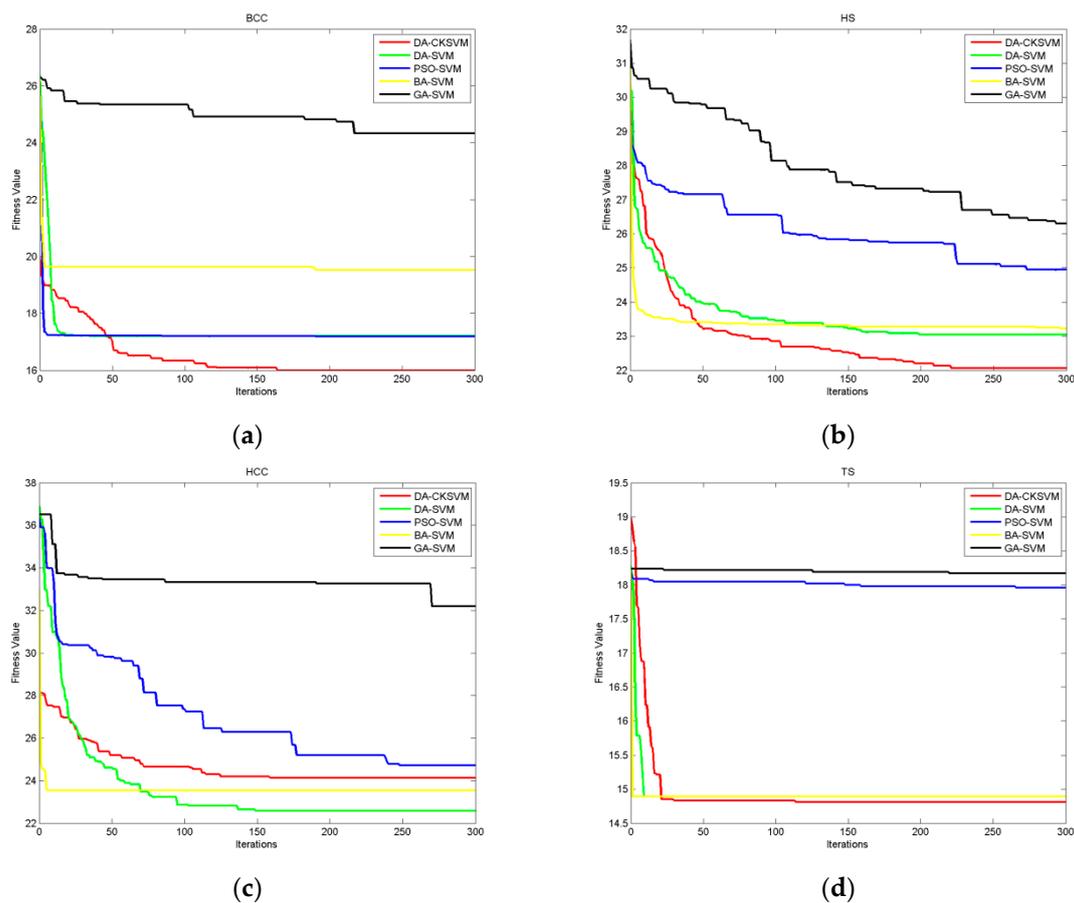


Figure 2. Cont.

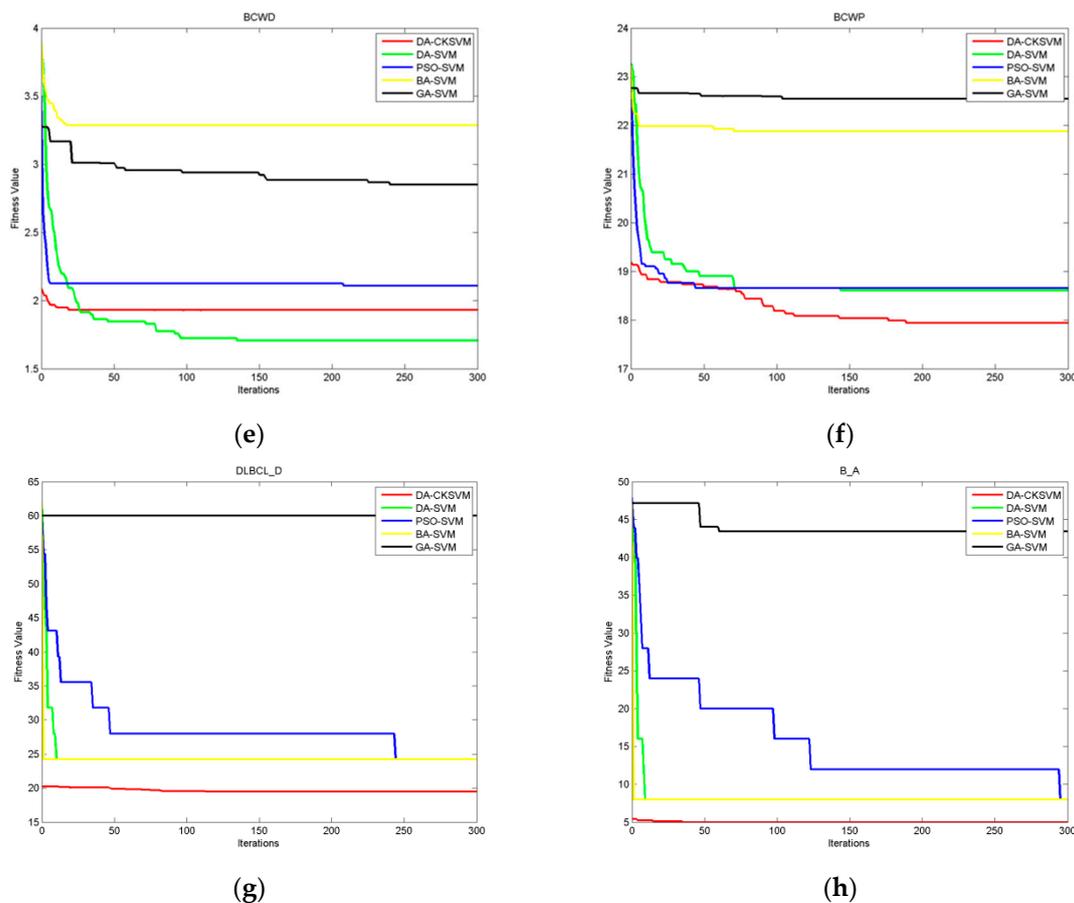


Figure 2. Fitness curves of all algorithms on six data sets listed in Table 1. (a) BCC data set; (b) HS data set; (c) HCC data set; (d) TS data set; (e) BCWD data set; (f) BCWP data set; (g) DLBCL_D data set; (h) B_A data set.

Figure 2a displays the fitness values of the BCC data set. As shown, the DA-CK SVM algorithm achieves the best result and the result of the GA-SVM algorithm was the worst.

Figure 2b shows the optimization process of the HS data set, where the DA-CK SVM algorithm achieves the best result and the similar result is obtained by the DA-SVM algorithm and the BA-SVM algorithm. As can be seen from the figure, the DA-CK SVM algorithm gains advantages in early iterations and further improved classification accuracy in the subsequent optimization process.

The fitness curves of HCC are displayed in Figure 2c. The DA-SVM algorithm is the best of all the algorithms, followed by the BA-SVM algorithm, and the DA-CK SVM algorithm achieves the third best result.

Figure 2d illustrates the fitness value of TS. In this data set, the DA-CK SVM algorithm gains the best result. The DA-SVM algorithm and the BA-SVM algorithm have the same result.

Figure 2e shows the fitness curves of BCWD. As can be seen, the DA-SVM algorithm obtains the best accuracy for this data set and the DA-CK SVM algorithm ranks the second.

Figure 2f represents the fitness values of BCWP. The DA-CK SVM algorithm gains advantages in earlier iterations and performs the best in the end. The result of the DA-SVM algorithm is similar to the PSO-SVM algorithm.

Figure 2g demonstrates the fitness values of DLBCL_D. The DA-CK SVM algorithm gets the best fitness values, while other algorithms except GA-SVM obtain the same result.

Figure 2h represents the fitness values of B_A. The conclusion is consistent with that in Figure 2g.

The Wilcoxon rank sum test with 5% significance level was applied on the average accuracy results to further evaluate the overall performance of the DA-CK SVM algorithm and comparison

algorithms. The Wilcoxon rank sum test is a nonparametric statistical test to demonstrate that one algorithm is significantly different from others. Table 6 lists the Wilcoxon test p -values between the DA-CKSVM algorithm and the comparison methods. A p -value lower than 0.05 indicates a statistically significant difference between the DA-CKSVM algorithm and other methods. A p -value higher than 0.05 (underlined) indicates no significant difference compared with comparison methods. In Table 6, only in three cases: DA-CKSVM vs. DA-SVM on TS, DA-CKSVM vs. PSO-SVM on HCC, and DA-CKSVM vs. BA-SVM on TS, were the p -values larger than the predicted statistical significance level of 0.05, while the other p -values were smaller than the significance level of 0.05. Such conclusions can be drawn from the fact that the results of the DA-CKSVM algorithm are statistically significant with those of other methods on all eight data sets.

Table 6. The DA-CKSVM algorithm vs. four comparison algorithms in terms of p -values of the Wilcoxon rank sum test over eight cancer data sets ($p \geq 0.05$ are underlined).

	BCC	HS	HCC	TS	BCWD	BCWP	DLBCL_D	B_A
DA-CKSVM vs. DA-SVM	<0.05	<0.05	<0.05	<u>0.08</u>	<0.05	<0.05	<0.05	<0.05
DA-CKSVM vs. PSO-SVM	<0.05	<0.05	<u>0.15</u>	<0.05	<0.05	<0.05	<0.05	<0.05
DA-CKSVM vs. BA-SVM	<0.05	<0.05	<0.05	<u>0.08</u>	<0.05	<0.05	<0.05	<0.05
DA-CKSVM vs. GA-SVM	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05

In summary, compared with the SVM classifiers using a single RBF kernel function, the proposed DA-CKSVM algorithm with a combined kernel function has a better classification ability in cancer diagnosis. Given the complementarity between RBF kernel and the polynomial kernel in the combined kernel function, DA-CKSVM is able to map input data into high-dimensional efficiently. Moreover, DA-CKSVM can effectively optimize the parameters of SVM along with the weight coefficient in the combined kernel function. Due to the excellent searching performance of the dragonfly algorithm, the local optima problem can be avoided when solving a wider range of classification problems.

8. Conclusions and Future Work

The methods to select or construct an appropriate kernel function and tune its parameters for SVM classifier have received considerable critical attention in cancer diagnosis. A novel classification algorithm, DA-CKSVM, based on SVM with a combined kernel function, was proposed in this paper. The combined kernel function in DA-CKSVM was constructed with an RBF kernel function and a polynomial kernel function. Parameters of the combined kernel SVM classifier were optimized by DA. The performance of the proposed DA-CKSVM in cancer classification was compared with four algorithms from the literature, which were the DA-SVM algorithm, PSO-SVM algorithm, BA-SVM algorithm, and GA-SVM algorithm. The experimental results showed that due to its excellent learning ability and generalization ability, the DA-CKSVM algorithm has better classification accuracy than these algorithms with a single kernel function. This reflects the good practical value of the proposed algorithm in the field of cancer classification.

Future research needs to be carried out in several directions. First, it should be noted that, like in the comparison literature, the SVM classifier in this paper was not trained with independent test sets, which may lead to over-fitting. An independent test set will be applied to evaluate the trained model in the future. Second, DA-CKSVM did not perform better than comparison methods on two data sets, which reveals the limitation of the proposed DA-CKSVM algorithm. In some cases, the existence of a polynomial kernel in a combined kernel function may slightly reduce the classification accuracy even if the polynomial kernel plays a minor role in the combined kernel function. In future, a mechanism should be established to evaluate the advantages of a single kernel function and a combined kernel function during the process of parameter optimization and remove the ineffective kernel if necessary. In this way, the classification ability of the proposed algorithm can be further improved. Third, the

proposed algorithm should be extended to more applications of medical diagnostic classification to explore its potential.

Author Contributions: Conceptualization, T.X.; Data curation, T.X. and Z.Z.; Formal analysis, T.X.; Investigation, T.X.; Methodology, T.X. and J.Y.; Project administration, J.Y.; Resources, J.Y.; Software, T.X. and J.Y.; Supervision, J.Y.; Validation, Z.Z.; Visualization, T.X.; Writing—original draft, T.X.; Writing—review & editing, J.Y.

Funding: This research was supported by Shanghai Key Laboratory of Power station Automation Technology.

Acknowledgments: The authors would like to thank Professor Du Dajun and Professor Wang Ling from Shanghai University, China for thoroughly proofreading this paper.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [[CrossRef](#)] [[PubMed](#)]
2. Ting, F.F.; Tan, Y.J.; Sim, K.S. Convolutional neural network improvement for breast cancer classification. *Expert Syst. Appl.* **2019**, *120*, 103–115. [[CrossRef](#)]
3. Alom, M.Z.; Yakopcic, C.; Nasrin, M.S.; Taha, T.M.; Asari, V.K. Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network. *J. Digit. Imaging* **2019**. [[CrossRef](#)] [[PubMed](#)]
4. Ai, D.; Pan, H.; Han, R.; Li, X.; Liu, G.; Xia, L.C. Using Decision Tree Aggregation with Random Forest Model to Identify Gut Microbes Associated with Colorectal Cancer. *Genes* **2019**, *10*, 112. [[CrossRef](#)] [[PubMed](#)]
5. Vapnik, V.N. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **1999**, *10*, 988–999. [[CrossRef](#)] [[PubMed](#)]
6. Dai, S.; Niu, D.; Han, Y. Forecasting of Power Grid Investment in China Based on Support Vector Machine Optimized by Differential Evolution Algorithm and Grey Wolf Optimization Algorithm. *Appl. Sci.* **2018**, *8*, 636. [[CrossRef](#)]
7. Illias, H.A.; Zhao Liang, W. Identification of transformer fault based on dissolved gas analysis using hybrid support vector machine-modified evolutionary particle swarm optimisation. *PLoS ONE* **2018**, *13*, e0191366. [[CrossRef](#)]
8. Zhou, J.; Li, L.; Wang, L.; Li, X.; Xing, H.; Cheng, L. Establishment of a SVM classifier to predict recurrence of ovarian cancer. *Mol. Med. Rep.* **2018**, *18*, 3589–3598. [[CrossRef](#)] [[PubMed](#)]
9. Kavitha, M.S.; Shanthini, J.; Sabitha, R. ECM-CSD: An Efficient Classification Model for Cancer Stage Diagnosis in CT Lung Images Using FCM and SVM Techniques. *J. Med. Syst.* **2019**, *43*, 73. [[CrossRef](#)] [[PubMed](#)]
10. Geeitha, S.; Thangamani, M. Incorporating EBO-HSIC with SVM for Gene Selection Associated with Cervical Cancer Classification. *J. Med. Syst.* **2018**, *42*, 225. [[CrossRef](#)]
11. Zhang, L.; Zhou, W.; Wang, B.; Zhang, Z.; et al. Applying 1-norm SVM with squared loss to gene selection for cancer classification. *Appl. Intell.* **2018**, *48*, 1878–1890. [[CrossRef](#)]
12. Hsu, C.W.; Lin, C.J. A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Netw.* **2002**, *13*, 415–425. [[CrossRef](#)] [[PubMed](#)]
13. Chapelle, O.; Vapnik, V.; Bousquet, O.; Mukherjee, S. Choosing Multiple Parameters for Support Vector Machines. *Mach. Learn.* **2002**, *46*, 131–159. [[CrossRef](#)]
14. Tharwat, A.; Gabel, T.; Hassanien, A.E. Parameter optimization of support vector machine using dragonfly algorithm. Paper Presented at the 3rd International Conference on Advanced Intelligent Systems and Informatics (AISII 2017), Cairo, Egypt, 9–11 September 2017. [[CrossRef](#)]
15. Lin, S.-W.; Ying, K.-C.; Chen, S.-C.; Lee, Z.-J. Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Syst. Appl.* **2008**, *35*, 1817–1824. [[CrossRef](#)]
16. Tharwat, A.; Hassanien, A.E.; Elnaghi, B.E. A BA-based algorithm for parameter optimization of Support Vector Machine. *Pattern Recognit. Lett.* **2017**, *93*, 13–22. [[CrossRef](#)]
17. Siqueira, L.F.S.; Morais, C.L.M.; Araújo Júnior, R.F.; de Araújo, A.A.; Lima, K.M.G. SVM for FT-MIR prostate cancer classification: An alternative to the traditional methods. *J. Chemom.* **2018**, *32*, e3075. [[CrossRef](#)]

18. Phan, A.V.; Nguyen, M.L.; Bui, L.T. Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems. *Appl. Intell.* **2016**, *46*, 455–469. [[CrossRef](#)]
19. Huang, C.-L.; Wang, C.-J. A GA-based feature selection and parameters optimization for support vector machines. *Expert Syst. Appl.* **2006**, *31*, 231–240. [[CrossRef](#)]
20. Song, H.; Ding, Z.; Guo, C.; Li, Z.; Xia, H. Research on combination kernel function of support vector machine. Paper Presented at the International Conference on Computer Science and Software Engineering (CSSE 2008), Wuhan, Hubei, China, 12–14 December 2008. [[CrossRef](#)]
21. Dash, C.S.K.; Sahoo, P.; Dehuri, S.; Cho, S.-B. An Empirical Analysis of Evolved Radial Basis Function Networks and Support Vector Machines with Mixture of Kernels. *Int. J. Artif. Intell. Tools* **2015**, *24*. [[CrossRef](#)]
22. Mirjalili, S. Dragonfly algorithm: A new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. *Neural Comput. Appl.* **2015**, *27*, 1053–1073. [[CrossRef](#)]
23. Sayed, G.I.; Tharwat, A.; Hassanien, A.E. Chaotic dragonfly algorithm: An improved metaheuristic algorithm for feature selection. *Appl. Intell.* **2018**, *49*, 188–205. [[CrossRef](#)]
24. Mafarja, M.M.; Eleyan, D.; Jaber, I.; Hammouri, A.; Mirjalili, S. Binary Dragonfly Algorithm for Feature Selection. Paper Presented at the 2017 International Conference on New Trends in Computing Sciences (ICTCS 2017), Amman, Jordan, 11–13 October 2017. [[CrossRef](#)]
25. Abdel-Basset, M.; Luo, Q.; Miao, F.; Zhou, Y. Solving 0–1 Knapsack Problems by Binary Dragonfly Algorithm. Paper Presented at the International Conference on Intelligent Computing (ICIC 2017), Liverpool, UK, 7–10 August 2017. [[CrossRef](#)]
26. Díaz-Cortés, M.-A.; Ortega-Sánchez, N.; Hinojosa, S.; Oliva, D.; Cuevas, E.; Rojas, R.; Demin, A. A multi-level thresholding method for breast thermograms analysis using Dragonfly algorithm. *Infrared Phys. Technol.* **2018**, *93*, 346–361. [[CrossRef](#)]
27. Xu, G.; Zhang, M.; Zhu, H.; Xu, J. A 15-gene signature for prediction of colon cancer recurrence and prognosis based on SVM. *Gene* **2017**, *604*, 33–40. [[CrossRef](#)] [[PubMed](#)]
28. Tuo, Y.; An, N.; Zhang, M. Feature genes in metastatic breast cancer identified by MetaDE and SVM classifier methods. *Mol. Med. Rep.* **2018**, *17*, 4281–4290. [[CrossRef](#)] [[PubMed](#)]
29. Li, Y.; Xie, X.; Yang, X.; Guo, L.; Liu, Z.; Zhao, X.; Luo, Y.; Jia, W.; Huang, F.; Zhu, S.; et al. Diagnosis of early gastric cancer based on fluorescence hyperspectral imaging technology combined with partial-least-square discriminant analysis and support vector machine. *J. Biophotonics* **2018**, *12*, e201800324. [[CrossRef](#)]
30. Prabukumar, M.; Agilandeewari, L.; Ganesan, K. An intelligent lung cancer diagnosis system using cuckoo search optimization and support vector machine classifier. *J. Ambient Intell. Hum. Comput.* **2017**, *10*, 267–293. [[CrossRef](#)]
31. Fabelo, H.; Ortega, S.; Casselden, E.; Loh, J.; Bulstrode, H.; Zolnourian, A.; Grundy, P.; Callico, G.M.; Bulters, D.; Sarmiento, R. SVM Optimization for Brain Tumor Identification Using Infrared Spectroscopic Samples. *Sensors* **2018**, *18*, 4487. [[CrossRef](#)]
32. Li, M.; Lu, X.; Wang, X.; Lu, S.; Zhong, N. Biomedical classification application and parameters optimization of mixed kernel SVM based on the information entropy particle swarm optimization. *Comput. Assist. Surg.* **2016**, *21* (Suppl. 1), 132–141. [[CrossRef](#)]
33. Smits, G.F.; Jordaan, E.M. Improved SVM regression using mixtures of kernels. Paper Presented at the 2002 International Joint Conference on Neural Networks (IJCNN'02), Honolulu, HI, USA, 12–17 May 2002. [[CrossRef](#)]
34. Thomas, J.; Sael, L. Multi-kernel ls-svm based bio-clinical data integration: Applications to ovarian cancer. *Int. J. Data Min. Bioinform.* **2017**, *19*, 150–167. [[CrossRef](#)]
35. Zien, A.; Ong, C.S. Multiclass multiple kernel learning. Paper Presented at the 24th International Conference on Machine Learning (ICML 2007), Corvallis, OR, USA, 20–24 June 2007. [[CrossRef](#)]
36. Nguyen, H.-N.; Ohn, S.-Y.; Park, J.; Park, K.-S. Combined kernel function approach in SVM for diagnosis of cancer. Paper Presented at the First International Conference on Natural Computation (ICNC 2005), Changsha, China, 27–29 August 2005. [[CrossRef](#)]
37. Tan, Y.; Wang, J. A support vector machine with a hybrid kernel and minimal vapnik-chervonenkis dimension. *IEEE Trans. Knowl. Data Eng.* **2004**, *16*, 385–395. [[CrossRef](#)]
38. Reynolds, C.W. Flocks, herds, and schools: A distributed behavioral model. Paper Presented at the 14th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1987, Anaheim, CA, USA, 27–31 July 1987. [[CrossRef](#)]

39. Yang, X.S. *Nature-Inspired Metaheuristic Algorithms*, 2nd ed.; Luniver Press: Beckington, UK, 2010; p. 106.
40. Chang, C.-C.; Lin, C.-J. LIBSVM: A Library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).