



## Research Article

# Pulmonary Nodule Recognition Based on Multiple Kernel Learning Support Vector Machine-PSO

Yang Li <sup>1,2</sup>, Zhichuan Zhu,<sup>1,3</sup> Alin Hou <sup>2</sup>, Qingdong Zhao,<sup>2</sup>  
Liwei Liu,<sup>2</sup> and Lijuan Zhang<sup>2</sup>

<sup>1</sup>*School of Mathematics and Statistics, Northeast Normal University, Changchun, Jilin 130024, China*

<sup>2</sup>*School of Computer Science and Engineering, Changchun University of Technology, Jilin 130012, China*

<sup>3</sup>*Faculty of Statistics, Jilin University of Finance and Economics, Changchun, Jilin 130117, China*

Correspondence should be addressed to Alin Hou; [alinhou@163.com](mailto:alinhou@163.com)

Received 28 December 2017; Accepted 12 March 2018; Published 29 April 2018

Academic Editor: Nadia A. Chuzhanova

Copyright © 2018 Yang Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pulmonary nodule recognition is the core module of lung CAD. The Support Vector Machine (SVM) algorithm has been widely used in pulmonary nodule recognition, and the algorithm of Multiple Kernel Learning Support Vector Machine (MKL-SVM) has achieved good results therein. Based on grid search, however, the MKL-SVM algorithm needs long optimization time in course of parameter optimization; also its identification accuracy depends on the fineness of grid. In the paper, swarm intelligence is introduced and the Particle Swarm Optimization (PSO) is combined with MKL-SVM algorithm to be MKL-SVM-PSO algorithm so as to realize global optimization of parameters rapidly. In order to obtain the global optimal solution, different inertia weights such as constant inertia weight, linear inertia weight, and nonlinear inertia weight are applied to pulmonary nodules recognition. The experimental results show that the model training time of the proposed MKL-SVM-PSO algorithm is only 1/7 of the training time of the MKL-SVM grid search algorithm, achieving better recognition effect. Moreover, Euclidean norm of normalized error vector is proposed to measure the proximity between the average fitness curve and the optimal fitness curve after convergence. Through statistical analysis of the average of 20 times operation results with different inertial weights, it can be seen that the dynamic inertial weight is superior to the constant inertia weight in the MKL-SVM-PSO algorithm. In the dynamic inertial weight algorithm, the parameter optimization time of nonlinear inertia weight is shorter; the average fitness value after convergence is much closer to the optimal fitness value, which is better than the linear inertial weight. Besides, a better nonlinear inertial weight is verified.

## 1. Introduction

The number of deaths from lung cancer is as high as 137 million annually around the world, accounting for 18% of cancer mortality ratio. Early surgical treatment is the most effective treatment for lung cancer, but most patients are diagnosed in the late stage of the disease. In 2015, the European Academy of Imaging and the European Respiratory Society published the latest white paper on lung cancer screening in European Respiratory Journal (ERJ) to guide clinical lung cancer screening for early detection and early treatment of lung cancer.

As early representation form of lung cancer in the lung CT image, a pulmonary nodule is defined as a nearly spherical opacity with a diameter smaller than 3 cm. Computed

Tomography (CT) technology is an important means of early detection of pulmonary nodules disease. According to the CT characterization, pulmonary nodules can be divided into solid nodules (such as solitary pulmonary nodules, pulmonary wall adhesion nodules, and vascular adhesion nodules), ground glass nodules, and cavitory nodules.

Computer-Aided Detection (CAD) system of lung is one of the applications of machine vision; it can reduce overload visual fatigue of the radiologist and decrease the possibility of the resulting miscarriage or omission and also provide auxiliary diagnosis results for the doctor as “third party.” Usually, the lung CAD system includes the following modules: acquisition of the lung CT image data, preprocessing of CT image, lung parenchyma segmentation, detection of VOI (Volume of Interest) or ROI (Region of

Interest) in candidate nodules images (mainly refers to the extraction or segmentation), calculation and selection of ROI or VOI features, and recognition of pulmonary nodules, where pulmonary nodules recognition is the core module of the CAD system. The algorithm of Support Vector Machine (SVM) has been widely used in the detection and recognition of pulmonary nodules (see, e.g., [1–15]). Among them, Li et al. [1] applied mixed kernel SVM algorithm to distinguish benign and malignant lung nodules, making the recognition accuracy (ACC) reach 92% and the sensitivity index reach 92.59%; Wang et al. [2] detected lung lesions by use of three-dimensional SVM with Latent Variable algorithm. Furthermore, Demir and Çamurcu [16] and Chang et al. [17] imported the algorithm of Particle Swarm Optimization (PSO) into SVM and selected the optimal parameter group by PSO and then used SVM for identification. In addition, Ma et al. [18] adopted the method of multiple classifiers fusion for lung disease identification.

The Multiple Kernel Learning Support Vector Machine (MKL-SVM) algorithm has achieved good recognition accuracy not just in recognition of lung nodules in [1] but also in other application fields (see [19, 20]). In [21], the Multiple Kernel Learning (MKL) method was elaborated and the latest research progresses were presented in this field. However, the MKL-SVM algorithm involves a large number of parameters, and the selection of parameters will have an important impact on the actual results. In [1], the selection of the optimal parameters is obtained by the grid search algorithm. The advantage of the grid search algorithm is the easiness to get the global optimal solution in the case of dense mesh division, but the disadvantage of the method is that it has a large amount of computation, a long time to search, and a poor real-time performance, which is not easy to form online identification algorithm. The selection of the relevant parameters is an urgent problem to be solved in the MKL-SVM, and the Particle Swarm Optimization (PSO) algorithm based on swarm intelligence algorithm provides an idea to solve the problem.

In this paper, the PSO algorithm and MKL-SVM algorithm are combined to realize the parameter optimization of the MKL-SVM. On this basis, the PSO algorithm with different inertia weights was compared and analyzed in order to obtain the parametric array similar or superior to that of the grid search algorithm aiming at quickly searching the optimal parametric array and the reasonable inertia weight and then precise identification of the pulmonary nodules.

## 2. Multiple Kernel Learning Support Vector Machine (MKL-SVM)

*2.1. Support Vector Machine.* SVM is a learning method using small amount of samples, which can be applied to predict or classify unknown samples by structural risk minimization. The training sample is represented as follows:

$$T = \{(x_i, y_i)\}, \quad i = 1, 2, \dots, l, \quad (1)$$

where  $l$  is the number of training samples,  $x_i$  denotes the input vector of SVM, corresponding to the feature of the

above  $N$ -dimension region of interest (ROI),  $x_i \in R^N$ ;  $y_i \in \{-1, +1\}$  indicates category label; here,  $y_i = 1$  corresponds to nodules and  $y_i = -1$  corresponds to nonnodules.

When SVM is used in the two classification problems, the original model can be written as the following nonlinear optimization problem:

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \\ \text{s.t.} \quad & y_i ((w \cdot \Phi(x_i)) + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, l \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, l, \end{aligned} \quad (2)$$

where  $w$  is the weight vector and  $b$  is the threshold, and the aim of SVM is to maximize the classification interval  $2/\|w\|$ , that is, minimization of  $\|w\|^2$ .  $C$  is the regularization coefficient or penalty parameter, which describes the degree of penalty for misclassification samples. The greater  $C$  is, the more obvious the penalty for misclassification is. When the data cannot be completely separated, the maximum interval will be negative, thus introducing slack variables  $\xi_i$  which can measure the distance between the actual output  $y_i$  and the Support Vector Machine output.

In the feature space, SVM is used to map the input data ( $X_i$ ) into a high-dimensional feature space  $Z$  by nonlinear transformation  $\Phi(X)$ , and then the optimal classification hyperplane is constructed in the high-dimensional feature space  $Z$  to realize the SVM. In the process of constructing the hyperplane in the feature space  $Z$ , the training algorithm uses the dot product and the kernel function  $K(x_i, x_j)$  to represent the inner products  $\Phi(x_i)$  and  $\Phi(x_j)$ ; that is, a function  $K$  can be found to form the next formula:

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j). \quad (3)$$

Thus, it is not necessary to construct and solve the convex quadratic programming problem for a given training sample, and the problem is transformed into the following optimization problem by using Lagrange multiplier:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j K(x_i, x_j) \alpha_i \alpha_j - \sum_{j=1}^l \alpha_j \\ \text{s.t.} \quad & \sum_{i=1}^l y_i \alpha_i = 0, \\ & 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l. \end{aligned} \quad (4)$$

The offset  $b$  in (2) can be solved by means of the following equation:

$$b = y_j - \sum_{i=1}^l y_i \alpha_i K(x_i, x_j). \quad (5)$$

The decision function is constructed as follows:

$$f(x) = \text{sgn}(g(x)), \quad (6)$$

where

$$g(x) = \sum_{i=1}^l \alpha_i y_i K(x_i, x_j) + b. \quad (7)$$

2.2. *Multiple Kernel Learning SVM (MKL-SVM)*. Various kernel functions have diverse advantages. One of the keys to improve the performance of SVM is to design an appropriate kernel function for a given problem. The common basic kernel functions are polynomial kernel function and radial basis function (RBF), which are presented, respectively, as follows:

$$\begin{aligned} K_{\text{poly}}(x, y) &= (x^t y + 1)^d \\ K_{\text{rbf}}(x, y) &= \exp\left(-\frac{\|x - y\|^2}{2g^2}\right), \end{aligned} \quad (8)$$

where the parameter  $d$  represents the order of the polynomial kernel, the parameter  $g$  denotes the width of the RBF kernel, and  $d$  and  $g$  need to be given in advance.

**Proposition 1.** *The convex combination form of the kernel function is still a kernel function:*

$$K_{\text{mix}}(x, x') = \sum_{p=1}^U m_p K_p(x, x'), \quad (9)$$

where

$$\sum_{p=1}^U m_p = 1, \quad 0 < m_p < 1, \quad p = 1, \dots, U, \quad (10)$$

and  $K_p$  is the  $p$ th species of basic kernel function and  $m_p$  corresponds to the weights of the  $p$ th species of basic kernel function in the total multiple kernel function.  $U$  species basic kernel functions are used in the multiple kernel function, and the weight sum of various basic kernel functions is one so as to limit the weight proportion of various basic kernel functions in the multiple kernel functions in proportion.

*Proof.* Let  $\{x_1, x_2, \dots, x_l\}$  be a set of  $l$  points in any given  $R^N$ ; we just need to prove that the Gram matrix in (9) is positive semidefinite matrix.

Let  $K_1, K_2, \dots, K_p$  be the Gram matrix of  $K_1(x, x')$ ,  $K_2(x, x')$ ,  $\dots$ ,  $K_p(x, x')$  for  $\{x_1, x_2, \dots, x_l\}$ ; for any  $\alpha \in R^l$ , we obtain

$$\begin{aligned} \alpha^T K_{\text{mix}}(x, x') \alpha &= \alpha^T \left( \sum_{p=1}^U m_p K_p(x, x') \right) \alpha \\ &= \alpha^T (m_1 K_1(x, x') + m_2 K_2(x, x') + \dots \\ &\quad + m_U K_U(x, x')) \alpha = m_1 \alpha^T K_1(x, x') \alpha \\ &\quad + m_2 \alpha^T K_2(x, x') \alpha + \dots + m_U \alpha^T K_U(x, x') \alpha \geq 0. \end{aligned} \quad (11)$$

So  $K_{\text{mix}}(x, x') = \sum_{p=1}^U m_p K_p(x, x')$  is positive semidefinite matrix; that is,  $K_{\text{mix}}(x, x')$  is a kernel function, and the evidence is proven.  $\square$

It is proven that the kernel function expressed by (9) satisfies the Mercer condition and can be used for the training

and classification of SVM. By using the above MKL-SVM, we can use nonlinear transformation of the sample points to get the corresponding kernel matrix so as to obtain the classification results during the training of SVM classifier.

RBF kernel has a strong ability to learn, and polynomial kernel has strong generalization ability; thus the combination of the two can take into account the ability of both learning and generalization. If we use only two kinds of basic kernel functions of both RBF kernel and polynomial kernel, that is,  $U = 2$ ,  $K_1 = K_{\text{poly}}$ , and  $K_2 = K_{\text{rbf}}$ , the multiple kernel function of (12) is able to be formed. Compared with the single kernel function, we need to estimate a set of kernel parameters and a weight coefficient  $m$ . The weight coefficient  $m$  can regulate freely the proportion of different kernel functions mixed in multiple kernels, adjust flexibly the ability of learning and generalization, and make the results unbiased towards the promotion of a particular target.

$$K(x, x') = m K_{\text{poly}}(x, x') + (1 - m) K_{\text{rbf}}(x, x'). \quad (12)$$

In [1], the grid search algorithm in the sense of CV is used to find the optimal regularization coefficient  $C$ , the order  $d$  of the polynomial kernel, the kernel width  $g$  of the RBF kernel, and the weight coefficient  $m$  of the multiple kernels. The optimal parameter group can be obtained by the grid search algorithm during the CV process corresponding to the highest classification accuracy. Lots of parameters and short step length of mesh induce large amount of calculation and long running time. The global optimal solution could be found by the heuristic algorithm, not needing to traverse all the parameter points in the grid.

2.3. *MKL-SVM Based on Modified Particle Swarm Optimization Algorithm*. Particle Swarm Optimization (PSO) is a typical heuristic algorithm on the basis of swarm intelligence optimization theory. In 1955, PSO was first proposed by Kennedy and Eberhart in [22], whose basic principle was originated from the research on the predation behavior of artificial life and birds. When birds prey upon food, the simplest and most effective method of finding food is to search the current area around the food nearest to birds. Compared with Generic Algorithm (GA), PSO searches the optimal particles by tracking the particles in the solution space without selection, crossover, and mutation.

It is assumed that the population  $X = (X_1, X_2, \dots, X_n)$  consists of  $n$  particles in a  $D$ -dimensional search space, where  $X_i$  represents the position of the  $i$ th particle in  $D$ -dimensional search space and also is a candidate solution of problem denoted by a vector of  $D$  dimensions as  $X_i = (X_{i1}, X_{i2}, \dots, X_{iD})^T$ . According to the objective function, we can calculate the fitness value of each particle position  $X_i$ . The speed of the  $i$ th particle is  $V_i = (V_{i1}, V_{i2}, \dots, V_{iD})^T$ , and its individual extreme value and group extreme value are  $P_i = (P_{i1}, P_{i2}, \dots, P_{iD})^T$  and  $P_g = (P_{g1}, P_{g2}, \dots, P_{gD})^T$ , respectively. During each iteration, the particle updates its velocity and

position by the individual extrema and the group extrema, which are given, respectively, as follows:

$$\begin{aligned} V_{id}^{k+1} &= \omega V_{id}^k + c_1 r_1 (P_{id}^k - X_{id}^k) + c_2 r_2 (P_{gd}^k - X_{id}^k), \\ X_{id}^{k+1} &= X_{id}^k + V_{id}^{k+1}, \end{aligned} \quad (13)$$

where  $\omega$  is an inertia weight;  $d = 1, 2, \dots, D$ ;  $D$  represents the number of parameters to be searched;  $k$  is the number of the present iterations;  $V_{id}$  is the velocity of particles,  $c_1$  and  $c_2$  are acceleration factors, which are nonnegative constants, and  $r_1$  and  $r_2$  are random numbers distributed within the interval  $[0, 1]$ . In order to prevent the blind search of particles, the position and velocity are usually limited to the range of  $[-X_{\max}, X_{\max}]$  and  $[-V_{\max}, V_{\max}]$ .

The PSO algorithm is applied into MKL-SVM algorithm of (12). Because the corresponding order of polynomial kernel is defined as positive integer for  $d \geq 2$  and with the increase of  $D$ , generalization ability of polynomial kernel decreases gradually, so only the two values  $d = 2$  and  $d = 3$  were calculated, and there is no need to search other parameters. Here the dimension of the search space of the particle is set to  $D = 3$ ;  $X_i = (X_{i1}, X_{i2}, X_{i3})^T$  represents the solution of the  $i$ th particles, where  $X_{i1}$ ,  $X_{i2}$ , and  $X_{i3}$  of each dimension are corresponding to the regularization coefficient  $C$ , the kernel width  $g$  of RBF, and the multiple kernel weight  $m$  to be searched, respectively.

### 3. Application of MKL-SVM-PSO Algorithm in Pulmonary Nodule Recognition

After introducing the classic PSO algorithm, the recognition accuracy rate (ACC) of pulmonary nodules in the sense of CV is regarded as the final target and determined as the fitness function value of PSO, and then ACC is defined as follows:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (14)$$

where TP denotes the detected true positive nodule; FP is the detected false positive nodule; FN indicates the undetected false negative nodule; TN is the detected true negative nodule, that is, nonnodule. ACC measures total recognition accuracy to measure the actual detection rate of pulmonary nodules; the SEN is defined as follows:

$$\text{SEN} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (15)$$

The parameter optimization algorithm of MKL-SVM-PSO algorithm is described in Figure 1.

## 4. Experimental Results and Analysis

*4.1. The Experimental Data and the Results Analysis of MKL-SVM-PSO Algorithm.* The experimental data were collected from 20 groups from a third-grade class A hospital in Jilin province with a total of about 700 images, and each group was diagnosed with the diagnostic criteria of doctor. The size of each CT image was  $512 \times 512$ , and the slice thickness is

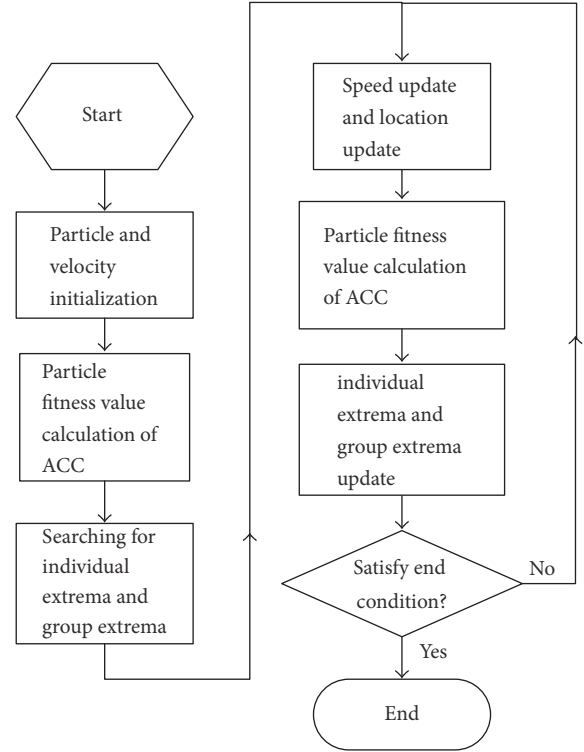


FIGURE 1: The diagram of MKL-SVM-PSO algorithm.

5.0 mm. 270 ROI are extracted: 80 pulmonary nodules and 190 false positive nodules. After feature selection, the data samples were randomly divided into two groups: the training group including 170 samples and the testing group including 100 samples.

The simulation experiments are carried out using the platform MATLAB with libsvm toolbox. In the process of model parameter optimization, 5-fold cross-validation is used to obtain the optimal parameter set corresponding to the highest ACC. Let the number  $n$  of cluster particles be 20, that is,  $n = 20$  in the MKL-SVM-PSO algorithm, and the dimension  $D$  of each particle be 3, that is,  $D = 3$ , and then the maximum number of iterations  $\text{maxgen}$  of the algorithm is set to be 200, and the inertia weight is constant, that is,  $\omega = 1$ . The simulation experiment will show the convergence performance through the optimal fitness curve and the average fitness curve as in Figure 2.

The optimal individual fitness curve of MKL-SVM-PSO algorithm is obtained as shown in Figure 2. The most optimal individual fitness value, namely, the nodule recognition accuracy, is 94.1176% through 5-fold cross-validation in training set; and the correspondent optimal particle position is  $C = 29.7267$ ,  $g = 19.0653$ , and  $d = 2$ ,  $m = 0.8$ , respectively. In this case, the running time of the proposed MKL-SVM-PSO algorithm is 363.4640 s, less than 2815.0786 s compared with 3178.5426 s of the grid search algorithm [1], and it accounts for just 11.43% of the grid search algorithm running time. Applying the proposed method to the test set, the test results of ACC reached 91% and those of SEN reached 88.89%.



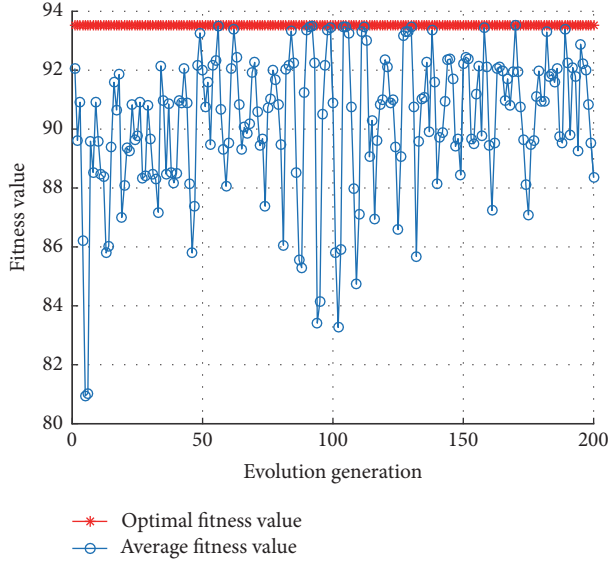


FIGURE 2: The fitness curve of the MKL-SVM-PSO algorithm with constant inertia weight.

**4.2. The Influence and Comparison of Different Inertia Weight on Lung Nodule Recognition.** Compared with the grid search algorithm, the computational time of MKL-SVM-PSO algorithm requires shorter time, but as the iteration times gradually increase, the shock amplitude of average fitness value in each generation is more severe, and a certain gap exists with the optimal fitness value, which can be found in Figure 2.

In [23], the inertia weight  $\omega$  was first introduced into PSO algorithm, and the larger inertia weight value that was conducive to the global search was pointed out, and the smaller inertia weight that was more conducive to the local search was also presented, and, moreover, the inertia weight that can reflect the ability of the particle to inherit the previous speed was also discussed. In order to balance the global search ability and local search ability of the algorithm, the linear decreasing inertia weight (LDIW) method proposed by Shi and Eberhart [23] will be used to reassign  $\omega$  as follows:

$$\omega(k) = \frac{\omega_{\text{start}}(\omega_{\text{start}} - \omega_{\text{end}})(T_{\text{max}} - k)}{T_{\text{max}}}, \quad (16)$$

where  $\omega_{\text{start}}$  is the initial inertia weight;  $\omega_{\text{end}}$  is the inertia weight of the iteration to the maximum number of times;  $k$  is the number of the current iteration generations;  $T_{\text{max}}$  is the maximal iteration number. In order to ensure that the above iteration algorithm not only has better global search ability in initial phase but also has strong local search ability to obtain the optimal solution in later iterations, commonly let  $\omega_{\text{start}} = 0.9$  and  $\omega_{\text{end}} = 0.4$ ; the inertia weight decreases linearly from initial 0.9 to 0.4, and this is also an empirical approach. The fitness curve of the optimal parameters obtained by the MKL-SVM-PSO algorithm is shown in Figure 3.

It can be seen from Figure 3 that the shock amplitude is reduced by using the inertia weight of (16) rather than using constant inertia weight, close to the optimal solution in the

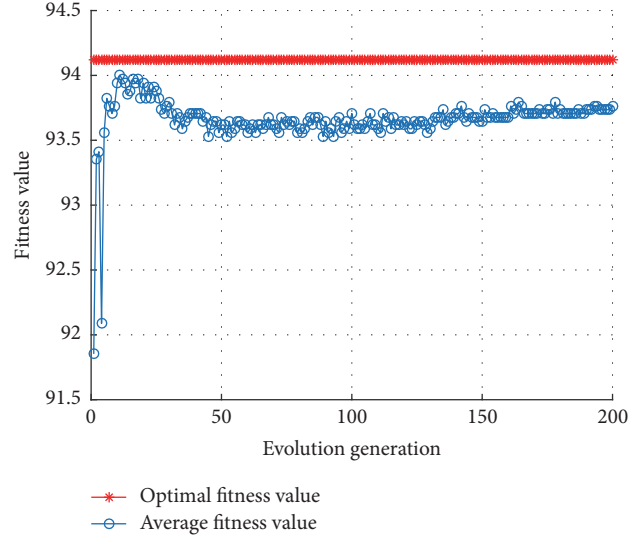


FIGURE 3: The fitness curve of MKL-SVM-PSO algorithm of (16).

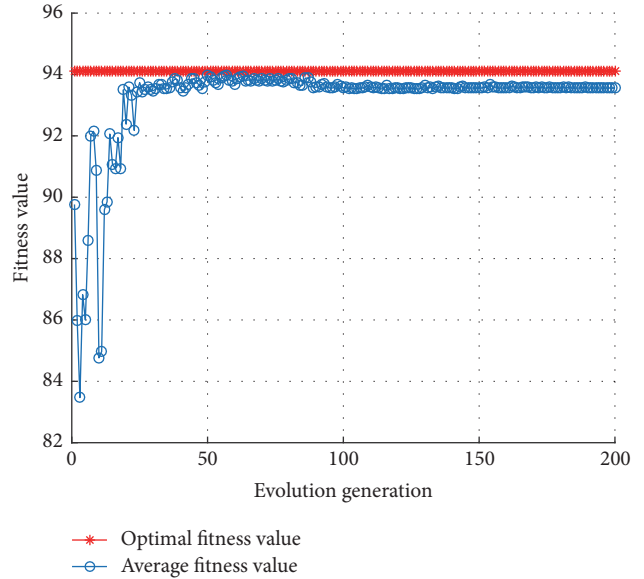


FIGURE 4: The fitness curve using the MKL-SVM-PSO algorithm of (17).

early stage. Besides, commonly used linear inertia weight is shown in the following equation:

$$\omega(k) = \omega_{\text{end}} + (\omega_{\text{start}} - \omega_{\text{end}}) \left( \frac{T_{\text{max}} - k}{T_{\text{max}}} \right). \quad (17)$$

Figure 4 shows the fitness curve corresponding to (17). In Figure 4, the shock amplitude induced by inertia weight of (17) is slightly larger, but it quickly converges to the optimal individual fitness value. The linear inertia weight represented by (16) and (17) can make the average fitness curve smooth, though it is easy to fall into local optimum in the early stage.

In order to ensure obtaining the global optimal solution, the following three kinds of nonlinear inertia weight are adopted to control convergence precision and convergence

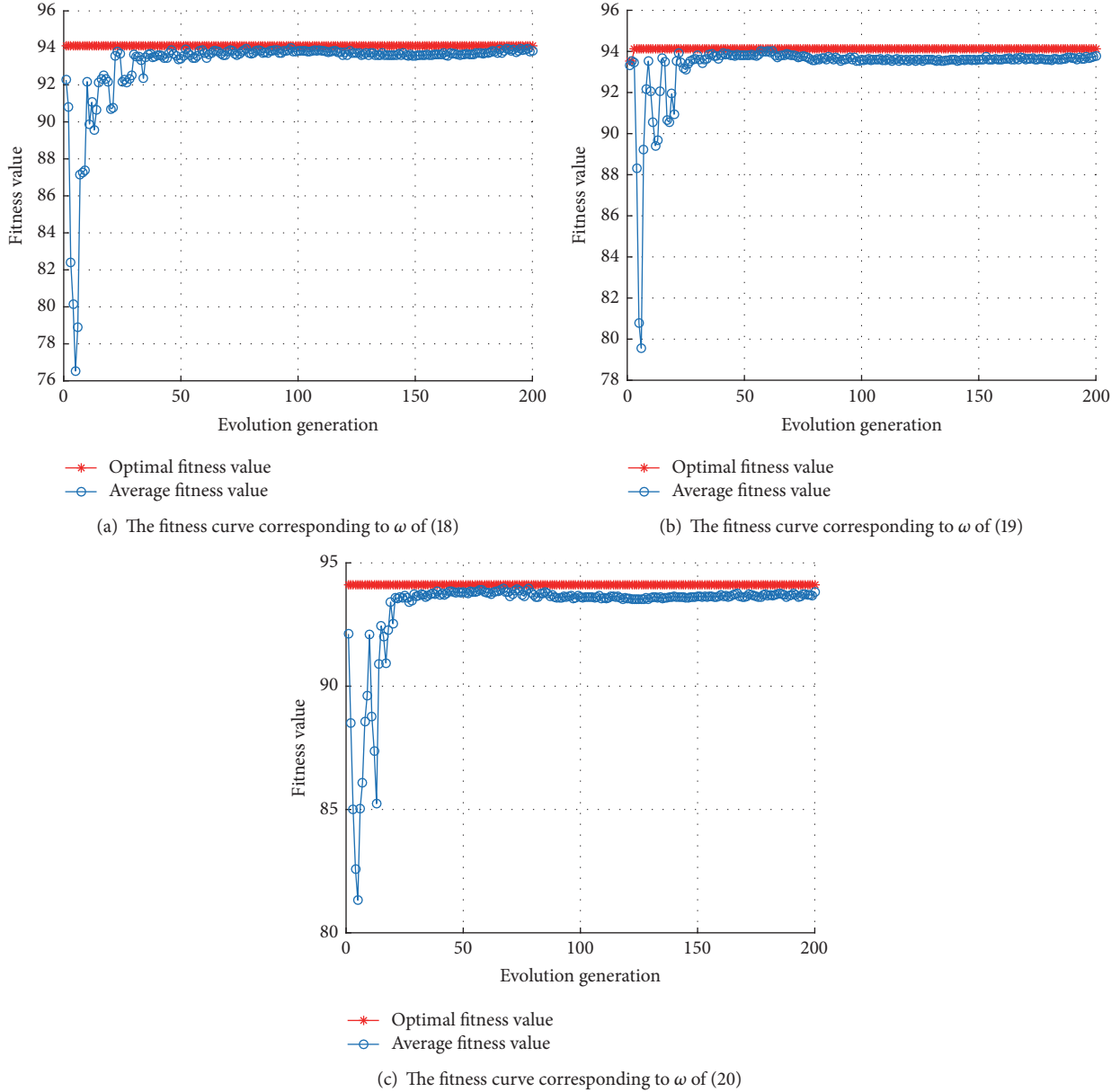


FIGURE 5: The fitness curve of optimal parameters group searching by MKL-SVM-PSO algorithm.

speed, so that the average fitness values reach the best fitness value index quickly and smoothly.

$$\omega(k) = \omega_{\text{start}} - (\omega_{\text{start}} - \omega_{\text{end}}) \left( \frac{k}{T_{\text{max}}} \right)^2 \quad (18)$$

$$\omega(k) = \omega_{\text{start}} - (\omega_{\text{start}} - \omega_{\text{end}}) \left( \frac{2k}{T_{\text{max}}} - \left( \frac{k}{T_{\text{max}}} \right)^2 \right) \quad (19)$$

$$\omega(k) = \omega_{\text{end}} \left( \frac{\omega_{\text{start}}}{\omega_{\text{end}}} \right)^{1/(1+ck/T_{\text{max}})} \quad (20)$$

Using the nonlinear parameter of the above three methods, respectively, the fitness function curve represents total recognition accuracy of pulmonary nodules as shown in Figure 5.

In order to compare the influence of different kinds of inertia weight on the parameter optimization, Figure 6 describes the curve of five kinds of dynamic weight corresponding to (16)–(20) changing along with the number of iterations. In the early stage of iteration, the larger inertia weight can make the algorithm maintain a strong global search ability, and the small inertia weight can make the algorithm search precisely in the late stage of iteration. As we know from the variation curve of several dynamic weights, the dynamic weight of (18) changes slowly in the early stage, and the value is larger so as to maintain the global search ability of the proposed algorithm; moreover, the dynamic weights change rapidly in the late stage and improve greatly the local searching ability of the algorithm; furthermore, the parameter optimization has also got a good result with the

TABLE 1: Comparison of various indexes in parameter optimization stage of different algorithms.

	Different inertial weight algorithm	Average parameter optimization time (s)	Average optimal fitness value	Average ACC value obtained from test set	Average SEN value obtained from test set
1	The constant is 1	370.7950	94.1176%	90.45%	86.85%
2	(16)	462.4134	94.1176%	91%	88.89%
3	(17)	457.0022	94.1176%	91%	88.89%
4	(18)	416.0204	94.1176%	91%	88.89%
5	(19)	448.1536	94.1176%	91%	88.89%
6	(20)	450.4456	94.1176%	91%	88.89%
7	Grid search algorithm	3096.1427	94.1176%	91%	88.89%

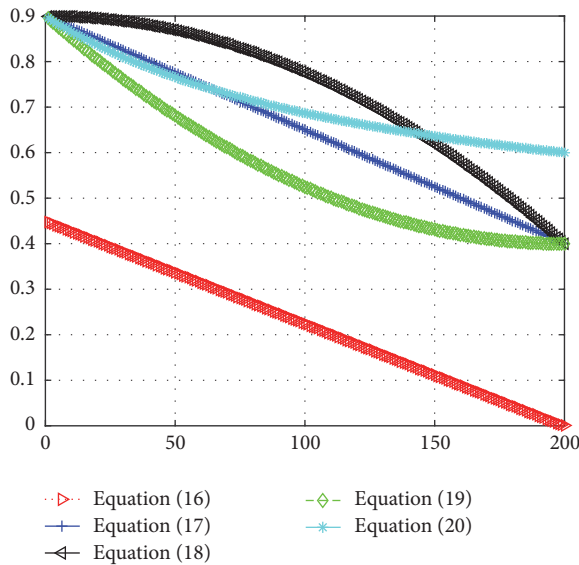


FIGURE 6: The variation curves of several dynamic inertia weights.

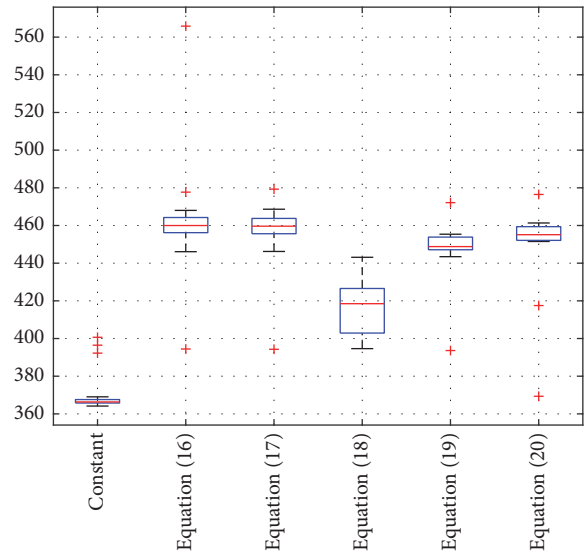


FIGURE 7: The statistical boxplot of parameter optimization time.

corresponding fitness curve, which is the optimal way of dynamic weights.

In summary, the Particle Swarm Optimization algorithm with constant weight has a fast convergence speed, but in the later stages it is easy to fall into local optimal solution with little accuracy. The linear inertia weights of (16) and (17) are easy to fall into local optimum. If we adopt several dynamic nonlinear inertia weight methods from (18) to (20), the algorithm converges slowly in initial stage, but in the later period local search ability is enhanced, which makes the algorithm jump out of local optimum and get the global optimal solution, so as to improve the accuracy of the algorithm. The form in (18) is the optimal nonlinear inertia weight.

In order to compare several kinds of Particle Swarm Optimization algorithms with different inertial weights and the parameters optimization time and recognition results of grid search algorithm, each algorithm is operated 20 times, and the average results of 20 times are listed in Table 1.

From the experimental results in Table 1, it can be seen that the parameter optimization time of the MKL-SVM grid search algorithm is the longest, which is almost 7 times of that of the MKL-SVM-PSO algorithm, and the running speed is much slower than that of the MKL-SVM-PSO

grid search algorithm. In the MKL-SVM-PSO algorithm, the inertia weight is, respectively, set to a constant value, a linear dynamic weight, and a nonlinear dynamic weight, and the parameter optimization time has a little difference as well. When the inertia weight is set to a constant value, the average running time is the shortest, and the optimum fitness is 94.1176% in the training stage, but its generalization ability is not as good as the experimental result obtained by dynamically searching the inertia weight. The average ACC value obtained from the test set is 90.45% and the average SEN is only 86.85%. The optimal fitness value is always 94.1176% during each operation process in the training stage using dynamic inertia weight from (16) to (20). The same results of every test are obtained on the test set. The ACC value is 91% and the SEN is 88.89%; only the training times are different. Among them, the average running time of (18) is the shortest, that is, 416.0204 s, which is the best among all the algorithms. The nonlinear inertial weight algorithm is faster than the linear inertial weight algorithm, which should be due to the fast convergence of the dynamic inertial weight algorithm.

Figure 7 shows the 20 times statistical results corresponding to parameter optimization times. The statistical values of the boxplot are listed in Table 2.

TABLE 2: Statistical indexes corresponding to Figure 7.

	Different inertial weight algorithm	Upper adjacent(s)	Lower adjacent(s)	Median value(s)	Number of outliers
1	The constant is 1	369.049	364.151	366.343	3
2	(16)	468.018	446.098	459.983	3
3	(17)	468.646	446.225	459.6345	2
4	(18)	443.129	394.638	148.49	0
5	(19)	455.38	443.456	448.798	2
6	(20)	461.342	451.524	455.122	3

From the data in Figure 7 and Table 2, it can be seen that when the inertia weight is constant, the box is located at the bottom, and the three outliers are also very low, and also the training time is the shortest. Combined with Figure 2, the convergence rate of the algorithm is much slow, and no obvious generation number of convergence can be found. From the test results of test set in Table 1, the generalization ability is not as good as that of the dynamic inertial weight algorithm, so it is not the optimal one. When dynamic inertial weight algorithm is adopted, the boxplot of the model training time corresponding to the three nonlinear inertial weight algorithms is located under the boxplot corresponding to the linear inertial weight. It means that the training time (i.e., parameter optimization time) of nonlinear inertial weight is less than that of linear inertial weight, and the parameter optimization time obtained by (18) is the shortest and there are no outliers in it. The upper adjacent was 443.129 s, the lower adjacent was 394.638 s, and the median value was 418.49 s, respectively. The lower adjacent optimization time of the inertia weight obtained by (19) is 443.456 s, which

is close to the upper adjacent value of the box obtained by (18). Furthermore, the training time of the former method is second only to that of (18). The position of box obtained by (20) is superior to that of linear inertial weight, but it is not good at nonlinear weight. The positions of the two boxes obtained by the data of both (16) and (17) are very close. The upper adjacent and lower adjacent of the box corresponding to (17) are all higher than those of (16), but the median value of (17) box is lower than that of (16) box. Therefore, the parameter optimization time of (18) is optimal.

Since the value of each inertial weight is different, the maximum number of convergence generations is different in various algorithms. Therefore, it is not reasonable to compare the Euclidean norm error between the optimal fitness value and the average fitness value after convergence, because it is difficult to express the merits and demerits of each algorithm. In order to reasonably express the Euclidean distance between the optimal fitness curve and the average fitness curve after convergence, we define the normalized Euclidean norm error as follows:

$$F_{gy} = \left\| V_{\text{best}}(t : \text{max gen}) - V_{\text{Average}}(t : \text{max gen}) \right\|_2$$

$$= \frac{\sqrt{(V_{\text{best}}(t) - V_{\text{Ave}}(t))^2 + (V_{\text{best}}(t+1) - V_{\text{Ave}}(t+1))^2 + \dots + (V_{\text{best}}(\text{max gen}) - V_{\text{Ave}}(\text{max gen}))^2}}{\text{max gen} - t + 1}, \quad (21)$$

where the average fitness curve converges from the  $t$ th generation.  $V_{\text{best}}$  and  $V_{\text{ave}}$  are both maxgen dimensional vectors and represent the optimal and average fitness value, respectively. Here the normalized Euclidean norm error is defined as  $F_{gy}$  and used to express the Euclidean distance between the average fitness value and the optimal fitness value of each generation after convergence.

The average values of the indexes obtained after 20 operation times are shown in Table 3. We compared and analyzed several key parameters such as the convergence generation number of the average fitness curve, the normalized Euclidean distance between the average fitness value and the optimal fitness value after reaching the convergence generation number, the median value, the maximum, the mean value, and total Euclidean norm of error vector with 200 generations in the average fitness curve. Each index is the statistical average of 20 operation results.

From above experimental results, it can be seen that when the inertia weight is constant, it is very difficult to find which generation of curves converges obviously in Table 3, so it is impossible to calculate the Euclidean norm of normalized error vector after reaching maximum number of convergence generations. The median and mean value of the average fitness curve are also the lowest among all the algorithms, which further indicates that the effect of the constant inertia weight is not as good as that of dynamic inertia weight. In the dynamic inertial weight algorithm, the maximum, mean, and median values of the average fitness curve are the highest in all the algorithms, and the Euclidean distance between the mean fitness curve and the optimal fitness curve is the smallest; that is, the least squares norm of the error vector is the smallest. The convergence is earlier, and the normalized Euclidean norm of the error vector is the smallest. However, there exists the particle premature phenomenon, so it is easy



TABLE 3: Comparison of various indexes under different inertia weights.

Different inertial weight algorithm	Maximum of $V_{best}$	Mean value of $V_{ave}$	Median value of $V_{ave}$	The Euclidean norm of the global error vector	Convergence generation number	The Euclidean norm of the normalized error vector after reaching the convergent generation number
The constant is 1	93.6044%	90.1465%	63.3698%	90.6721	—	—
(16)	94.0279%	93.6873%	93.7169%	8.4488	9	0.0294
(17)	93.9941%	93.2348%	93.6419%	28.2732	30	0.0348
(18)	94.0290%	93.2050%	93.6719%	31.5911	36	0.0320
(19)	94.0044%	93.2996%	93.6618%	27.2316	22	0.0324
(20)	93.9926%	93.2350%	93.6206%	28.1948	27	0.0341

to fall into the local optimal solution. It is further proven that the global search performance in the early stage is limited. The mean value of average fitness corresponding to nonlinear inertial weight of (18) is 93.2050%, and the maximum is 94.0290%; the median value is 93.6719%. The mean value is low, and also the convergence generation number is large, and the Euclidean norm of global error vector is large because of the sharp oscillation in the initial iteration. It also reflects that the global search ability of the algorithm in the initial iteration prevents fully the particle premature convergence. Thus, the three indexes, the maximum, the median value, and the Euclidean norm of the normalized error vector after reaching convergence, are able to reflect better convergence performance after convergence. The indexes obtained by (19) and (20) are basically superior to that of (17); that is, the nonlinear inertial weight algorithm is superior to the linear inertial weight algorithm generally.

In summary, the Particle Swarm Optimization algorithm with dynamic inertia weight is better than the one with constant inertia weight, and the algorithm using nonlinear inertia weight is better than that one using linear inertia weight. The MKL-SVM-PSO algorithm has gained good results by use of dynamic nonlinear inertial weight of (18) in this paper. The algorithm has the ability of global searching at the beginning of iteration. After reaching the convergence generation number, the average fitness value can approach the optimal fitness value more smoothly and quickly, which makes it easier to find the global optimal solution.

## 5. Conclusion

In this paper, a MKL-SVM-PSO algorithm with nonlinear inertial weight is proposed to search the optimal parameter set of hybrid kernel Support Vector Machine quickly and accurately and achieved better effects in pulmonary nodule recognition. The main innovative work goes as follows:

- (1) The PSO algorithm is introduced into the mixture Kernels SVM algorithm and used for the discrimination of benign and malignant pulmonary nodules.
- (2) On the basis of changing dynamic weights, the similarities and differences between linear weights and nonlinear weights are discussed, and the optimal dynamic nonlinear weights are obtained. The average

fitness value of the algorithm is close to the optimal fitness value quickly and smoothly, so that the global optimal solution is easy to be obtained.

- (3) The Euclidean norm index of normalized error vector is proposed to measure the difference between the optimal fitness curve and the average fitness curve after convergence with different inertial weights. The index solves the problem that different convergence generations of different algorithms result in different dimensions of error vectors in various algorithms, and it is difficult to compare the discrepancy. The validity of dynamic inertial weight algorithm is verified from the point of view of statistics.

The experimental results show that the model training time of MKL-SVM-PSO algorithm is only 1/7 of the training time of MKL-SVM grid search algorithm with better recognition effect. It can be seen that the dynamic inertia weight is better than constant inertia weight in the MKL-SVM-PSO algorithm from Table 3. Compared with the linear inertial weight algorithm, the parameter optimization speed of nonlinear inertial weight algorithm is rapid, and the average fitness value after convergence is much closer to the optimal fitness value. The dynamic inertial weight corresponding to (18) is the optimal method in this paper.

Although ACC, as a fitness value, has obtained good experimental results in this method, medical attention is often paid to the SEN index to prevent missed detection. Our next job is to extend the proposed MKL-SVM-PSO algorithm to multitarget search in order to achieve accurate identification and nonmissed detection of pulmonary nodules.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

The authors gratefully acknowledge the support of the China Postdoctoral Science Foundation (no. 2016M591468), Jilin Province Postdoctoral Science Foundation (111900358),

the Jilin Provincial Science and Technology Development Plan funded project (20170520050JH), the Education Department of Jilin Province (no. JJKH20170575KJ and no. JJKH20181041KJ), Jilin Province Postdoctoral Science Foundation, and the Key Program of the National Natural Science Foundation of China (nos. NSFC11690012 and NSFC11631003).

## References

- [1] Y. Li, D. Wen, K. Wang, and A. Hou, "Mixed kernel function SVM for pulmonary nodule recognition," *Lecture Notes in Computer Science*, vol. 8157, no. 2, pp. 449–458, 2013.
- [2] Q. Wang, W. Kang, C. Wu, and B. Wang, "Computer-aided detection of lung nodules by SVM based on 3D matrix patterns," *Clinical Imaging*, vol. 37, no. 1, pp. 62–69, 2013.
- [3] M. Javaid, M. Javid, M. Z. U. Rehman, and S. I. A. Shah, "A novel approach to CAD system for the detection of lung nodules in CT images," *Computer Methods and Programs in Biomedicine*, vol. 135, pp. 125–139, 2016.
- [4] S. Shen, A. A. T. Bui, J. Cong, and W. Hsu, "An automated lung segmentation approach using bidirectional chain codes to improve nodule detection accuracy," *Computers in Biology and Medicine*, vol. 57, pp. 139–149, 2015.
- [5] O. H. Madero, O. O. Vergara Villegas, V. G. Cruz Sanchez, D. H. J. Ochoa, and A. M. J. Nandayapa, "Automated system for lung nodules classification based on wavelet feature descriptor and support vector machine," *BioMedical Engineering OnLine*, vol. 14, pp. 1–20, 2015.
- [6] M. Firmino, G. Angelo, H. Morais, M. R. Dantas, and R. Valentim, "Computer-aided detection (CADE) and diagnosis (CADx) system for lung cancer with likelihood of malignancy," *Biomedical Engineering Online*, vol. 15, no. 1, article no. 2, pp. 1–17, 2016.
- [7] Y. Sui, Y. Wei, and D. Zhao, "Computer-aided lung nodule recognition by SVM classifier based on combination of random undersampling and SMOTE," *Computational and Mathematical Methods in Medicine*, vol. 12, Article ID 368674, pp. 1–13, 2015.
- [8] Z. Zhao, J. Chen, X. Yin, H. Song, X. Wang, and J. Wang, "Establishing assistant diagnosis models of solitary pulmonary nodules based on intelligent algorithms," *Cellular Physiology and Biochemistry*, vol. 35, no. 6, pp. 2463–2471, 2015.
- [9] A. Hussain, "Intensity-based statistical features for classification of lungs CT scan nodules using artificial intelligence techniques," *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 27, pp. 1–15, 2015.
- [10] T. Manikandan and N. Bharathi, "Lung cancer detection using fuzzy auto-seed cluster means morphological segmentation and SVM classifier," *Journal of Medical Systems*, vol. 40, no. 7, article no. 181, 2016.
- [11] B. R. Froz, A. O. D. C. Filho, A. C. Silva, A. C. D. Paiva, R. A. Nunes, and M. Gattass, "Lung nodule classification using artificial crawlers, directional texture and support vector machine," *Expert Systems with Applications*, vol. 69, pp. 176–188, 2017.
- [12] S. A. Khan, K. Kenza, M. Nazir, and M. Usman, "Proficient lungs nodule detection and classification using machine learning techniques," *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, vol. 28, no. 2, pp. 905–917, 2015.
- [13] T. Zhou, H. Lu, J. Zhang, and H. Shi, "Pulmonary nodule detection model based on SVM and CT image feature-level fusion with rough sets," *BioMed Research International*, vol. 2016, Article ID 8052436, pp. 1–13, 2016.
- [14] M. Bergtholdt, R. Wiemker, and T. Klinder, "Pulmonary nodule detection using a cascaded SVM classifier," *Spie Medical Imaging*, vol. 9785, 2016.
- [15] A. Sheeraz, J. M. Younus, A. M. Usman, Q. Usman, and H. Ali, "Pulmonary nodules detection and classification using hybrid features from computerized tomographic images," *Journal of Medical Imaging and Health Informatics*, vol. 6, no. 1, pp. 252–259, 2016.
- [16] Ö. Demir and A. Y. Çamurcu, "Computer-aided detection of lung nodules using outer surface features," *Bio-Medical Materials and Engineering*, vol. 26, pp. S1213–S1222, 2015.
- [17] B. M. Chang, H. H. Tsai, and C. Y. Yen, "SVM-PSO based rotation-invariant image texture classification in SVD and DWT domains," *Journal of medical imaging Health Informatics*, vol. 52, pp. 96–107, 2016.
- [18] L. Ma, X. Liu, L. Song, C. Zhou, X. Zhao, and Y. Zhao, "A new classifier fusion method based on historical and on-line classification reliability for recognizing common CT imaging signs of lung diseases," *Computerized Medical Imaging and Graphics*, vol. 40, pp. 39–48, 2015.
- [19] M. Li, X. Lu, X. Wang, S. Lu, and N. Zhong, "Biomedical classification application and parameters optimization of mixed kernel SVM based on the information entropy particle swarm optimization," *Computer Assisted Surgery*, vol. 21, pp. 132–141, 2016.
- [20] S. Fei-Yan, T. Yu-Bo, and R. Zuo-Lin, "Modeling the resonant frequency of compact microstrip antenna by the PSO-based SVM with the hybrid kernel function," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 29, no. 6, pp. 1129–1139, 2016.
- [21] S. S. Bucak, R. Jin, and A. K. Jain, "Multiple kernel learning for visual object recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1354–1369, 2014.
- [22] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1948, Perth, Australia, December 1995.
- [23] Y. Shi and R. C. Eberhart, "Empirical study of particle swarm optimization," *Congress on Evolutionary Computation*, vol. 3, pp. 1945–1950, 1999.