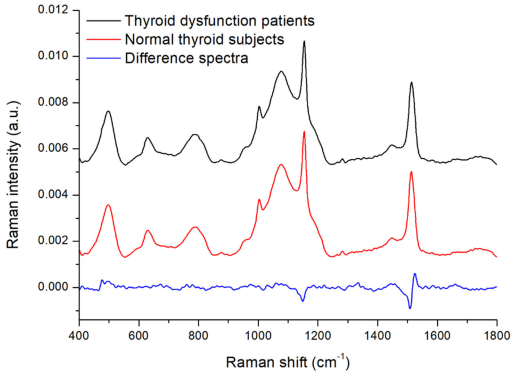
IEEE Photonics Journal
An IEEE Photonics Society Publication

Open Access

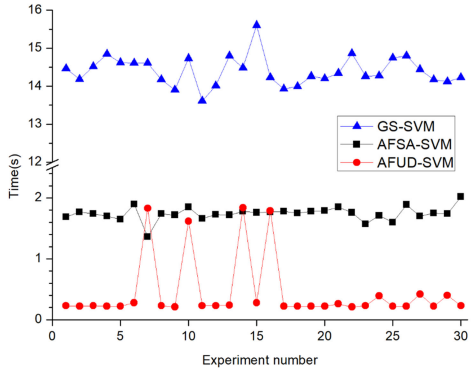
Rapid and Low-Cost Detection of Thyroid Dysfunction Using Raman Spectroscopy and an Improved Support Vector Machine

Volume 10, Number 6, December 2018

Xiangxiang Zheng
Guodong Lv
Guoli Du
Zhengang Zhai
Jiaqing Mo
Xiaoyi Lv



Comparison of the mean spectrum and difference spectrum.



Comparison of the optimization time among the three algorithms for 30 experimental iterations.

DOI: 10.1109/JPHOT.2018.2876686
1943-0655 © 2018 IEEE

Rapid and Low-Cost Detection of Thyroid Dysfunction Using Raman Spectroscopy and an Improved Support Vector Machine

Xiangxiang Zheng,[1] **Guodong Lv**,[2] **Guoli Du**,[2] **Zhengang Zhai**,[3] **Jiaqing Mo**,[1] **and Xiaoyi Lv** [1,4]

[1]College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China
[2]The First Affiliated Hospital of Xinjiang Medical University, Urumqi 830000, China
[3]Economy and Development Reform Commission, Urumqi High-Tech Industrial Development Zone, Urumqi 830046, China
[4]Institute of Health and Environmental Medicine of AMMS, Tianjin 300050, China

DOI:10.1109/JPHOT.2018.2876686

Manuscript received August 29, 2018; revised October 14, 2018; accepted October 15, 2018. Date of publication October 18, 2018; date of current version November 2, 2018. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61765014, in part by the Reserve Talents Project of National High-Level Personnel of Special Support Program under Grant QN2016YX0324, in part by the Urumqi Science and Technology Project under Grants P161310002 and Y161010025, and in part by the Reserve Talents Project of National High-Level Personnel of Special Support Program (Xinjiang [2014]22). *(Xiangxiang Zheng and Guodong Lv contributed equally to this work.)* Corresponding author: Xiaoyi Lv (e-mail: xiaoz813@163.com).
This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Abstract: This study presents a rapid and low-cost method to detect thyroid dysfunction using serum Raman spectroscopy combined with support vector machine (SVM). The serum samples taken from 34 thyroid dysfunction patients and 40 healthy volunteers were measured in this study. Tentative assignments of the Raman bands in the measured serum spectra suggested specific biomolecular changes between the groups. Principal component analysis (PCA) was used for feature extraction and reduced the dimension of high-dimension spectral data; then, SVM was employed to establish an effective discriminant model. To improve the efficiency and accuracy of the SVM discriminant model, we proposed artificial fish coupled with uniform design (AFUD) algorithm to optimize the SVM parameters. The average accuracy of 30 discriminant results reached 82.74%, and the average optimization time was 0.45 s. The results demonstrate that the serum Raman spectroscopy technique combined with the AFUD-SVM discriminant model has great potential for the detection of thyroid dysfunction. This technique could be used to develop a portable, rapid, and low-cost device for detecting thyroid function to meet the needs of individuals and communities.

Index Terms: Raman spectroscopy, optical diagnosis, thyroid dysfunction, support vector machine (SVM), parameter optimization.

1. Introduction

The prevalence of thyroid diseases is increasing each year worldwide, and the associated mortality rate is approximately 12%–17% [1]–[3]. Long-term uncontrolled thyroid dysfunction may affect cardiovascular, reproductive and other system functions [4], [5]. According to the epidemiological data of thyroid diseases in ten Chinese cities published in *Thyroid* in 2016, the prevalence of

hypothyroidism in China is 17.8%, and the prevalence of hyperthyroidism is 1.6%. Accordingly, the number of thyroid dysfunction patients in China may exceed 200 million [6]. It is particularly noteworthy that the symptoms of thyroid dysfunction are often hidden. There are many similarities between the symptoms of thyroid dysfunction and those of the suboptimal health states associated with today's fast-paced lifestyle. Most people attribute weight changes, irritability, anxiety, insomnia, and fatigue to their stress and lifestyle without considering that thyroid dysfunction could be a potential cause [4].

Thyroid disease patients need frequent examinations; hypothyroidism patients require life-long follow-up of thyroid function, and hyperthyroidism patients are generally followed-up once a month. Thus, factors of thyroid disease have become routine examination items in health examinations. The primary methods used to detect thyroid function currently include radioimmunoassay (RIA) and immunoluminescence (luminescence immunoassay). However, these two approaches require not only large biochemical detection equipment but also long detection times (5–6 h) and professional personnel for operation. As the prevalence rate continues to increase, only large hospitals can carry out thyroid function tests, which does not meet the needs of society. Therefore, for community, family and individual users, it is of great practical significance to develop a portable and rapid thyroid function detection device, such as a device similar to a blood glucose meter.

Raman spectroscopy is a non-invasive and sensitive optical analysis technology based on inelastic scattering. It also provides the conformation structure of molecules and specific fingerprint-type information, which have been widely used in the field of medical diagnostics and biomedical research [7]–[13]. Almost all types of diseases, e.g., cancers and infectious diseases, initially occur at the molecular level [14]. These changes in the Raman spectrum show that the position, intensity and line width of the characteristic peaks may change, which provides a basis for disease diagnosis.

In practical applications, because the Raman scattering signal is very weak and can be influenced by strong fluorescence, the signal-to-noise ratio (SNR) of the obtained spectral signal is not ideal, and it is difficult to directly observe differences among spectra. Therefore, robust and powerful spectral data mining algorithms are urgently needed to extract useful information and improve distinguishing accuracy [10], [14]. In recent years, support vector machines (SVMs) have been widely used in disease diagnosis based on Raman spectra and have yielded better discriminant results [14]–[17]. For example, Saranjam Khan *et al.* explored the use of Raman spectra together with SVM for the analysis of nasopharyngeal cancer and hepatitis B virus infection in human blood sera [14]. Shaoxin Li *et al.* used Raman spectra combined with SVM algorithm to diagnose prostate cancer [17]. In this article, SVM was selected as the classification algorithm for serum Raman spectroscopy.

However, the selection of SVM parameters has a direct impact on the accuracy of classification. Currently, in the field of disease diagnosis, the selection of SVM parameters is mainly based on the grid search (GS) method [17]–[19]. However, with the decrease in the search step length or the increase in the optimization parameters, the optimization time will greatly increase, resulting in reduced diagnostic efficiency. To solve this problem, we proposed the artificial fish coupled uniform design (AFUD) algorithm for the first time and applied it to SVM parameter optimization. The AFUD algorithm is based on the artificial fish swarm algorithm (AFSA) [20], [21], coupled with the uniform design (UD) method [22]. This algorithm not only has the advantages of the AFSA of fast optimization speed and stronger parallel optimization performance but also has the characteristics of the UD method, including better local optimization, allowing it to solve the problem of SVM parameter selection and optimization.

We present the use of Raman spectroscopy combined with SVM for the analysis and classification of thyroid dysfunction based on human blood sera. Initially, principal component analysis (PCA) was used for feature extraction and to reduce the high dimension of the spectral data. Then, SVM was introduced to establish an effective spectral data classification model, which was considered an effective data classification algorithm [14]–[16]. To solve the problem of SVM parameter optimization, we proposed the AFUD algorithm to optimize the parameters of the SVM and used the algorithm to classify Raman spectral data. For comparison, standard AFSA and GS methods were also used to optimize the SVM parameters and process the same spectral data.

2. Materials and Methods

2.1 Experimental Materials

In our experiment, 34 samples of fresh blood from thyroid dysfunction patients and 40 samples of fresh blood from normal thyroid function subjects were obtained. Three milliliters of fresh blood were collected from each patient without any anticoagulant, and the samples were placed in a clean environment at 37 °C 1 h for condensing. Then, a high-speed centrifuge (4000 r/min) was used to centrifuge the blood at 4 °C, and the top level of clear liquid was extracted to obtain the serum. The serum was sub-packed in a centrifuge tube and stored in a refrigerator at −20 °C for experiments. All samples were from the First Affiliated Hospital of Xinjiang Medical University. Approval of the ethics committee was obtained to study human blood serum samples.

2.2 Experimental Instrumentation

The serum Raman spectra were recorded using a confocal Raman micro-spectrometer (LabRAM HR Evolution RAMAN SPECTROMETER, HORIBA Scientific Ltd.) with a spectral resolution of approximately 0.35 cm^{-1}. Raman spectra of all the serum samples were measured from 400 to 1800 cm^{-1} with an excitation wavelength of 532 nm by an Ar^{+} laser, and the laser power was 100 mW. The software package LabSpec6 was applied for spectral acquisition and analysis. The spectra data were acquired in 5 s with a 50× objective, and a numerical aperture (NA = 0.25) was used. Two spectra were recorded from each sample in different positions, and the mean spectrum for each sample was taken for further analysis. A total of 74 serum Raman spectra were acquired, of which 34 were from thyroid dysfunction patients and 40 were from normal persons.

2.3 Data Analysis and Processing

The raw Raman spectra acquired from serum samples contained marked fluorescence backgrounds and noise (Fig. S1). To extract the pure Raman signals, a Vancouver Raman algorithm based on a fifth-order polynomial was employed to fit the thyroid serum fluorescence background (Fig. S2), and then, this polynomial was subtracted to correct the baseline [23]. Each background-subtracted Raman spectrum was also normalized by the integrated area under the curve. Thus, the influence of spectral intensity variability generated by possible laser power fluctuations could be reduced, and the spectral shapes and Raman peak intensities could be compared between the different groups of serum samples.

The spectral data after processing contained a set of 856 intensity variables from 400 cm^{-1} to 1800 cm^{-1}. To simplify and improve the diagnosis accuracy of SVM, it is necessary to reduce the dimension of high-dimensional spectral data. PCA has been widely used for dimension reduction and feature extraction of Raman spectrum data [12], [16], [18]. In this study, PCA was also used for dimension reduction and feature extraction of serum Raman spectra.

The PCA and SVM procedures were implemented with MATLAB language. The LIBSVM toolbox 3.22 created by Lin and Chang was used for SVM discriminant analysis [24].

2.4 AFSA Algorithm Optimization SVM Model

SVM is a robust and powerful machine learning algorithm and has become one of the most promising discriminant analysis algorithms [16]–[18]. Due to the existence of kernel functions, SVM can handle linearly inseparable problems by mapping the data into a high-dimensional space. This method has been proven to exceed traditional linear classifying algorithms [17], [18]. The most important parameters of SVM are kernel function and the parameter that control the priority of the slack variables' size constraint [18]. The four most frequently used kernel functions are linear kernels, polynomial kernels, Gaussian radial basis functions (RBFs) and sigmoid kernels. RBFs are suitable for both high-dimensional and low-dimensional data. In this study, a RBF was selected as the kernel function of the SVM. RBF is defined as $\exp(-g|x_i - x_j|^2)$ (where x_i and x_j are the two generic sample

data vectors). Thus, the parameters to be optimized for SVM were kernel parameter g and penalty parameter C, which was used for overfitting [18].

Here, we proposed the AFUD algorithm, which couples the AFSA and UD methods to solve the parameter optimization of SVM. AFSA is a novel intelligent optimization algorithm, introduced by Dr. Li in 2002 [20], [21]. Its inspiration come from the natural social behavior of fish in searching, swarming and following. Each fish forages for food based in its own way. Information searching is transmitted to others, and the group achieves a global optimum [25]. UD method is a kind of space filling design that can be used in computer and industrial experiments. The UD method seeks the design points to be uniformly scattered on the domain [22]. It is an application of pseudo Monte Carlo method in number theory.

The AFUD algorithm not only has the advantages of the AFSA of fast optimization speed and stronger parallel optimization performance but also has the characteristics of the UD method, including better local optimization, which can effectively solve the problem of SVM parameter selection and optimization. The idea of AFUD algorithm is improved because, in the iterative optimization process of the AFSA, the threshold of the objective function and the margin of the difference between the two generations of target values are set as the algorithm jump condition. When the target value obtained after the AFSA optimization exceeds the given threshold and the given limit is satisfied, the AFSA is considered to have found a near-optimal solution to the problem. The algorithm then exits the AFSA iterative optimization process and uses the local search optimal method; otherwise, it continues to perform the AFSA iterative optimization. The local optimization described in this paper uses the UD method, which transforms the parameter setting problem of a SVM into the optimal design problem of two factors and some horizontal numbers, a penalty parameter C and a kernel parameter g [26]. For the near-optimal solution of the AFSA in the earlier stage, the optimal solution of the problem can be found quickly and accurately by setting the appropriate values of the horizontal numbers.

The AFUD algorithm was used to optimize the binary parameter (C, g) of the SVM, and the maximum classification accuracy was used as the optimization target of SVM. The specific steps are as follows (the flowchart is shown in Fig. S3):

Step 1: Data acquisition and preprocessing. Different pretreatment methods are required for different data. In this paper, the background of the fluorescence and the area normalization are deducted for the Raman spectrum data.

Step 2: Determination of the training set and prediction set. After analysis of the preprocessed data, the training set D_t and the prediction set D_p of the SVM are determined.

Step 3: Parameter setting. Parameter setting includes population size N_{Gen}, maximum evolution times N_{Iter}, perceived distance D_{Vis}, moving steps S_{Leg}, maximum probing times N_{Try} and crowding degree factor μ, the ranges of C and g in the SVM, threshold value θ, the difference between the accuracies of the two generations ϑ, the number of tests in the UN method N_{UD}, and the search area.

Step 4: Initialization of the AFSA. Each artificial fish in the AFSA represents the optimal combination of parameters (C, g) for the SVM to be optimized. The fish groups are randomly initialized according to the values of C and g in Step 3 so that N_{Gen} artificial fishes can optimize the parameter combinations in parallel.

Step 5: Calculation of the concentration of each artificial fish in the initial fishes. Based on the optimization principle of cross-validation of training set accuracy maximization, the concentration of each artificial fish is calculated and compared with the others. The maximum artificial fish concentration is taken as the initial fish optimal target value, and the optimal parameter combination (C, g) of artificial fish is kept.

Step 6: Conduction of fish behavior. The artificial fish in the fish group define the fish behavior (prey, follow, swarm or randomly move).

Step 7: Determination of the optimal target value of the fish group. If the optimal value of the current fish group is larger than the best value previously saved, the best value is replaced by the current value, and the corresponding parameter combination is recorded as (C, g);

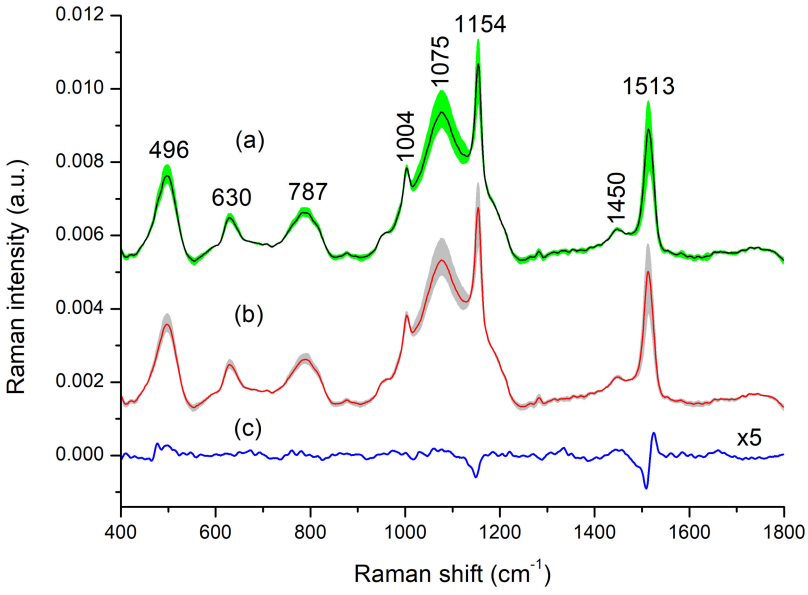


Fig. 1. Comparison of the normalized mean serum Raman spectra between thyroid dysfunction patients and normal thyroid subjects. (a) Thyroid dysfunction patients. (b) Normal thyroid subjects. (c) Difference spectra. For clarity, the difference spectra are enlarged by a factor of five. (The shaded areas represent the standard deviations of the means).

otherwise, the optimal value and the corresponding parameter combination (C, g) remain unchanged.

Step 8: Determination of whether the jump condition is satisfied. When the classification accuracies of two generations of fish groups before and after meet the error limit, if the maximum iteration number of the current fish group reaches the iteration upper limit or if the classification accuracy is greater than or equal to the given threshold, the maximum classification accuracy of the AFSA and the near-optimal combination of parameters (C, g) are recorded; otherwise, the next iteration is begun, and Step 6 is executed.

Step 9: Local optimization by the UD method. According to the parameters set in Step 3 and the neighborhood of the near-optimal solution σ obtained by Step 8, a new parameter combination of the N_{UD} group (C, g) is generated. The classification accuracy of each point is calculated, and the maximum value is selected as the optimal target value of the AFUD.

Step 10: Output the maximum classification accuracy and the optimal solution of the SVM optimized by AFUD (C_{best}, g_{best}).

3. Results

3.1 Raman Spectral Analysis

Fig. 1(a) and 1(b) show the mean Raman spectra with ± 1 standard deviations of thyroid dysfunction patients and normal thyroid subjects in the range of 400 cm^{-1} to 1800 cm^{-1}, respectively. The solid lines indicate the mean Raman spectra, and the shaded areas represent one standard deviation. The figure clearly displays that the characteristic spectra of the normal people and patients are very similar in shape and intensity. Therefore, a powerful and robust data analysis algorithm is needed to distinguish the normal and thyroid dysfunction serum spectra [10], [14]. Raman peaks at 496 cm^{-1}, 630 cm^{-1}, 787 cm^{-1}, 1004 cm^{-1}, 1075 cm^{-1}, 1154 cm^{-1}, 1450 cm^{-1} and 1660 cm^{-1} can be observed in both groups, with the strongest intensities at 1075 cm^{-1}, 1154 cm^{-1} and 1513 cm^{-1}. The mean intensities of Raman peaks at 496 cm^{-1}, 630 cm^{-1}, 787 cm^{-1}, 1004 cm^{-1}, 1075 cm^{-1}

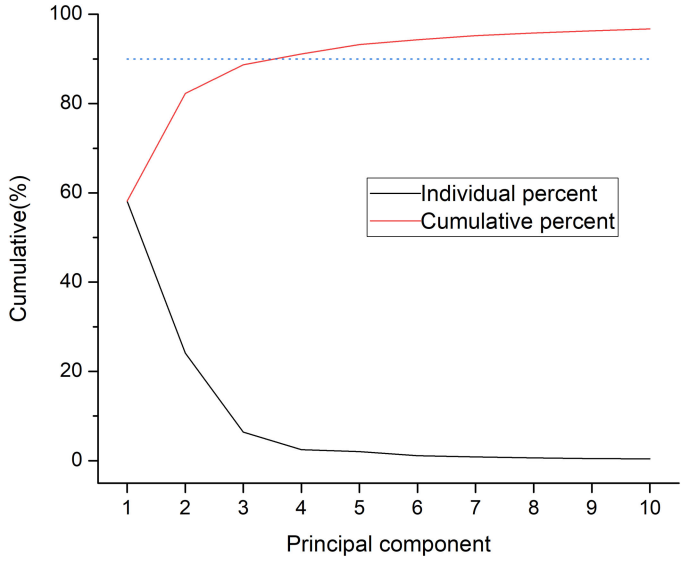


Fig. 2. Scree plot of cumulative and individual percentages of the first 10 PCA scores.

and 1450 cm^{-1} are more intense for thyroid dysfunction patients than for normal thyroid subjects, while the band intensities located at 1154 cm^{-1} and 1513 cm^{-1} are higher in normal subjects. These intensity differences in the Raman spectrum can be viewed more clearly at the bottom of Fig. 1(c) and S4, which implies that there is great potential to distinguish thyroid dysfunction patients from normal subjects using serum Raman spectroscopy.

3.2 PCA Feature Extraction

PCA was utilized on high-dimensional data to extract the key information from the background removed Raman spectra. Fig. 2 shows the first 10 principal components (PCs) and their cumulative percent of variance to the total variation. The figure shows that the total variation of the first 4 PCs accounted for 91% of the total variation (PC1: 58%; PC2: 24%; PC3: 6%; PC4: 3%). The number of PCs can be selected based on criteria that allow for retention of 90% of the total variance [12], [18], so the first 4 PCs were selected as the input vector of SVM.

3.3 Diagnosis Potential of SVM

The 74 spectral data of PCA feature extraction of Section 3.2 were loaded, and a random number column vector from 1–74 was created to disorganize the spectral data [27]. We chose the first 40 vectors as the training set; the remaining 34 were used as the prediction set to validate the diagnostic accuracy of the SVM model optimized by the AFSA and AFUD algorithms. As a comparison, the parameters of SVM were also optimized using the GS method under the same experimental conditions.

The values of the penalty parameter C and the kernel parameter g in the SVM are (0, 10] and (0, 5], respectively. In the AFSA, the population size is $N_{\text{Gen}} = 5$, the maximum evolution time is $N_{\text{Iter}} = 50$, the perceived distance is $D_{Vis} = 0.5$, the moving step is $S_{Leg} = 0.1$, the maximum probing time is $N_{Try} = 5$ and the crowding degree factor is $\mu = 0.5$. In the UD method, 100 iterations are performed, and the neighborhoods of C and g are 1 and 0.2, respectively. The condition for jumping out of the AFSA is that the diagnostic accuracy difference between the two generations is 0.1. The mean of the 30 accuracies of the predictive set is selected as the threshold of the AFSA, which is 80 in this paper. The GS method parameter settings are as follows: The ranges of C and g are (0, 10] and (0, 5], respectively, and the search steps are 0.1 and 0.01, respectively.

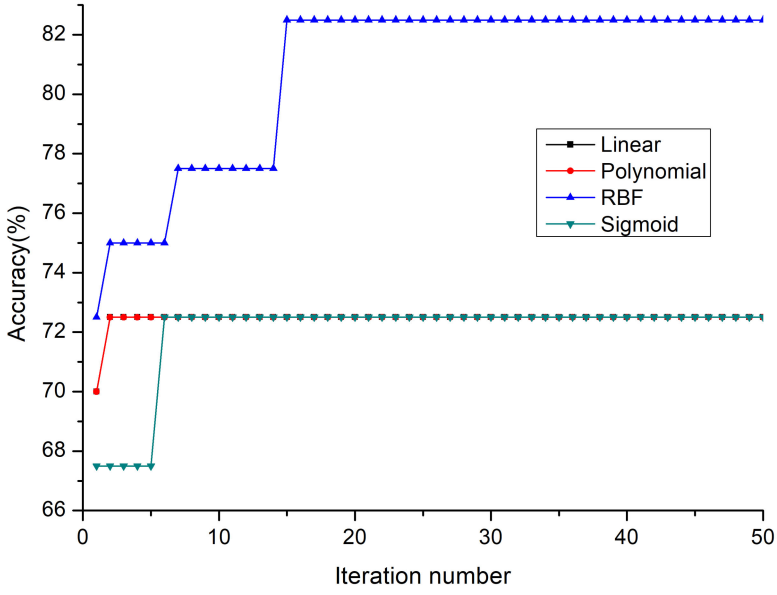


Fig. 3. Comparison of diagnostic accuracies among different kernel functions.

TABLE 1
Experimental Results of Parameter Optimization for the Three Algorithms (C Represents the Average Penalty Parameter, g Represents the Average Kernel Parameter, α Represents the Average Diagnostic Accuracy and t Represents the Average Diagnostic Time.)

Algorithm/category	C	g	α(%)	t(s)
AFSA-SVM	6.18±2.52	2.44±1.80	80.29±4.70	1.74
AFUD-SVM	6.74±2.15	2.80±1.50	82.74±4.49	0.45
GS-SVM	2.58±2.42	2.89±1.94	81.20±4.68	14.40

Under the same experimental conditions, the standard AFSA was used to optimize the parameters of the SVM, and the diagnostic accuracies of different kernel functions are shown in Fig. 3. The RBF kernel has the highest diagnostic accuracy, which verifies the reasonability of selecting the RBF kernel as the kernel function of the SVM in this paper.

Thirty times were repeated to optimize the parameters of the SVM using the AFSA and AFUD algorithms for the spectral data after dimension reduction, and the parameters of the SVM using the GS optimization method are also shown in Table 1. Fig. 4 shows the optimization efficiency diagram for the three algorithms.

From Table 1 and Fig. 4, we can see that the average diagnostic accuracies of the AFSA-SVM, AFUD-SVM and GS-SVM models for thyroid function are 80.29%, 82.74% and 81.20%, respectively, and the average optimization time is 1.74 s, 0.45 s and 14.40 s. As the accuracies of the three algorithms in optimizing the SVM parameters do not markedly differ, the AFSA and AFUD algorithms can shorten the diagnostic time by several-fold or even tens of folds compared with the GS method, thus greatly improving the diagnostic efficiency of the SVM model. The standard deviation of the AFUD-SVM diagnostic model is 4.49%, which is less than that of the AFSA-SVM and GS-SVM models. Therefore, the AFUD algorithm we proposed to optimize the diagnostic model of SVM parameters can greatly improve not only the diagnostic efficiency but also the diagnostic accuracy and the stability of the model.

To better understand and analyze the diagnostic models, Fig. 5 shows the representative optimization result of each diagnostic model. As shown in Fig. 5(a), the AFSA is not ideal for the

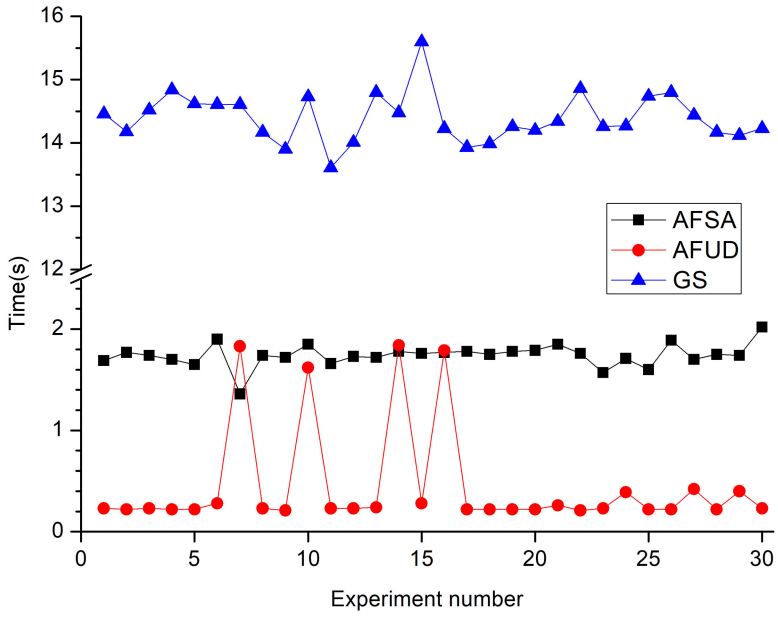


Fig. 4. Comparison of the optimization time among the three algorithms for 30 experimental iterations.

selection of the initial value (at the beginning of the iterations, the classification accuracy is not high, corresponding to only 75%). After 10 iterations, a near-optimal solution can still be achieved at a faster rate, which illustrates that the AFSA has advantages of insensitivity to initial values and strong parallel optimization ability. Fig. 5(b) clearly shows that the diagnostic accuracy of the model can reach 82.5% for the 100 test points selected by the UD method, which demonstrates that the AFSA can find the near-optimal solution of the target problem in the early stage of the iteration. In the small neighborhood of the near-optimal solution, with the UD method, the global optimal solution can be achieved with fewer trials. As shown in Fig. 5(c), after the standard AFSA iterates to 30 generations, the optimal diagnostic accuracy is achieved. Combined with Fig. 5(a), we can see that when the AFSA optimizes the parameters of the SVM, the number of iterations needed to achieve optimal results differs. It can be seen from Fig. 5(d) that when different penalty parameters C and kernel parameters g are chosen, the diagnostic accuracy is not fixed. In addition, higher diagnostic accuracy can only be achieved for C and g within a small interval, but in most of the ranges, the accuracy is less than 70%, and even drops to 42.5% in some areas. The selection of appropriate parameters of SVM is important to improve the diagnostic accuracy of the model.

To further verify the reliability of the three models for thyroid dysfunction diagnosis, we plotted the receiver-operating characteristic (ROC) curve for evaluation. The ROC curve combines the specificity and sensitivity of different models for the diagnosis of thyroid dysfunction, the integration area under the ROC curve (AUC) represents the diagnostic accuracy of the model, and larger AUC values suggest greater diagnostic accuracy [28]. It can be seen from Fig. 6 that the AUC values of the three diagnostic models are all above 0.8, which indicates that these three diagnostic models have high reliability in the discrimination of normal thyroid and abnormal subjects.

4. Discussion

Here, we reported the use of the serum Raman spectrum combined with a PCA-SVM algorithm for the detection of thyroid dysfunction patients. Through the mean spectrum analysis of thyroid dysfunction patients and normal thyroid subjects, we found that there were specific differences in serum Raman spectra between the two groups. Because of these differences, it is possible to utilize

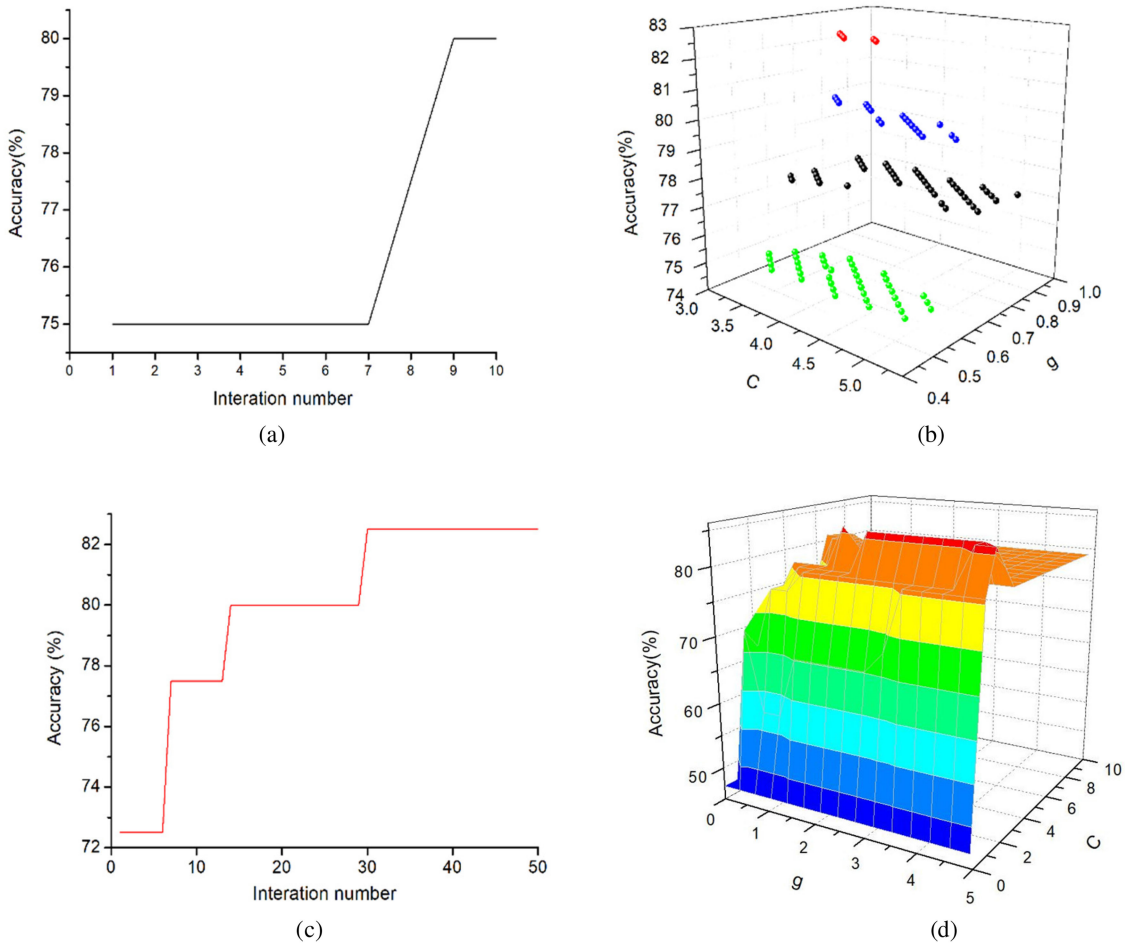


Fig. 5. (a) Accuracy curve of the AFUD-SVM diagnostic model in the early stage. (b) Schematic diagram of the local search with the AFUD-SVM diagnostic model. Different colors represent different accuracy rates. The red circles, blue circles, black circles and green circles correspond to accuracy rates of 82.5%, 80%, 77.5% and 75%, respectively. (c) Accuracy curve of the AFSA-SVM diagnostic model. (d) 3D accuracy map of the GS-SVM diagnostic model with different C and g values.

the serum Raman spectrum combined with the SVM algorithm for the detection and diagnosis of thyroid dysfunction patients.

In our exploratory study, we found that the serum Raman spectra reflect many vibrational modes of various biomolecules, such as cellulose, glycerol and phenylalanine molecules, which may change in conformation or quantity with changes in thyroid hormone content. Table 2 lists tentative assignments for the dominant spectral bands according to the previous literature [10], [29]–[32]. The observed serum spectral differences between normal subjects and patients could reflect the cellular and molecular alterations associated with malignant transformation [10], [29]. For example, the Raman peak at 630 cm^{-1} corresponded to the ring vibration of glycerol [30], which was higher in thyroid function patients. The increase in glycerol levels may be related to the decrease in lipid metabolism, e.g., hypothyroidism patients. The peak at 1004 cm^{-1} was attributed to the C-C symmetric stretch of phenylalanine [29] and was lower in normal thyroid function subjects than in patients with thyroid dysfunction. Phenylalanine is an essential amino acid in the human body that can be converted to tyrosine by a certain transformation, and some of tyrosine is used to synthesize thyroxine [33]. Thyroid hormones play an important role regulating metabolism, development, protein synthesis, and affecting other hormone functions. The two main hormones produced by the thyroid gland are

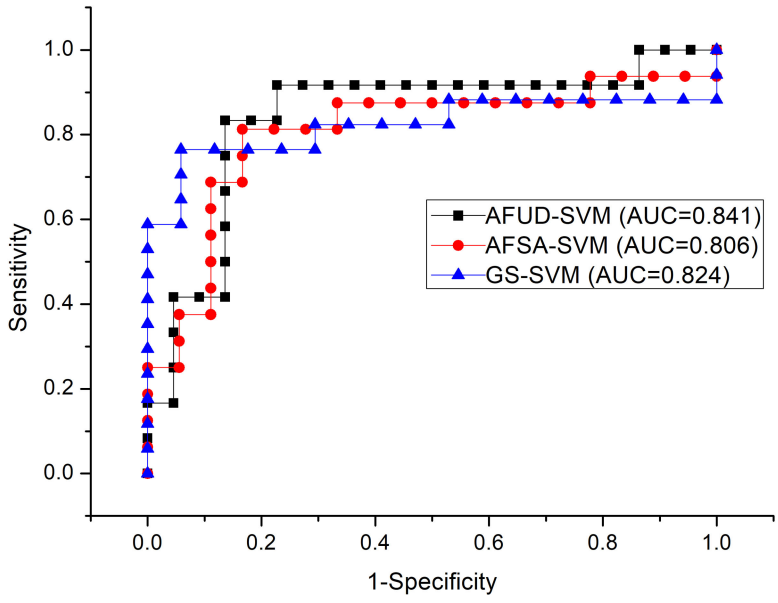


Fig. 6. ROC curves of classification results with the three algorithms.

TABLE 2
Peak Positions and Tentative Assignments of the Major Raman Bands of Human Serum [10], [29]–[32].

Raman shift (cm^{-1})	Vibrational mode	Tentative assignment
496	Ring vibration	Cellulose
630	Ring vibration	Glycerol
787	$\nu(C-C)$	Phosphatidylserine
1004	$\nu_s(C-C)$	Phenylalanine
1075	$\nu(C-N)$	Collagen
1154	$\nu(C-C)$	Carotenoid
1450	$\delta(CH_2)$	Phospholipid, collagen
1513	$\nu(C=C)$	Carotenoid

ν : stretching vibration; δ : bending vibration; ν_s : symmetric stretch

triiodothyronine (T3) and thyroxine (T4). Any dysfunction in the thyroid can affect the production of thyroid hormones (e.g., T3 and T4) which can be linked to various pathologies throughout the body [34]. The increase in phenylalanine content in the serum of patients may lead to an increase in thyroid hormones, e.g., T3 and T4, which could lead to hyperthyroidism [35]. Huang *et al.* also found an increase in phenylalanine content in malignant lung tissue using Raman spectroscopy [31]. The Raman bands at 1154 cm^{-1} and 1513 cm^{-1} were assigned to the vibrations of carotenoids and were lower in thyroid dysfunction patients than in normal subjects. Carotenoid is an antioxidant that plays a vital role in protecting cells and organisms from the effects of free radicals produced in the human body through reactive nitrogen species (RNS) and reactive oxygen species (ROS) [32]. M. Saleem *et al.* used Raman spectroscopy to study the serum of patients with dengue virus infection and found that a deficiency of carotenoids can eventually lead to the death of some cells, resulting in pathological changes [32]. Some cells in the serum die due to deficiency of carotenoids in the serum in patients with thyroid dysfunction, we can understand that some cells in the patient's serum die, leading to abnormal thyroid function. These changes indicated that there are relative

variations in some types of biomolecules in serum associated with thyroid dysfunction disease transformation.

PCA was used for feature extraction and reduced the high dimension of spectral data; a SVM was then chosen as the diagnostic algorithm. Table 1 and Figs. 4 and 6 show that AFUD algorithm achieved the best result in the parameter optimization of the SVM. The average accuracy and optimization time of 30 iterations were 82.72% and 0.45 s, respectively. Combined with Raman spectroscopy technology, the model can be used to diagnose thyroid dysfunction patients. The appeal of Raman spectroscopy lies in its promising potential for real-time clinical disease diagnosis [10]. Improving the diagnostic efficiency of the model from the perspective of the algorithm is one aspect of realizing real-time clinical diagnosis. In this paper, the AFUD algorithm was proposed to optimize the parameters of the SVM, and combined with Raman spectroscopy, a thyroid dysfunction diagnosis model with high diagnostic efficiency was established. The key to the improved diagnostic efficiency of this model is the development of a better parameter optimization algorithm, which also greatly increases the classification speed of the model.

5. Conclusion

In this study, 40 normal thyroid function subjects and 34 abnormal thyroid function patients were analyzed by their serum Raman spectra. The experimental results showed that the profile and peak intensities of the serum spectra were very similar between the two groups, while the subtle differences imply that it was possible to preliminarily screen thyroid function patients through a powerful data analysis algorithm.

To improve the diagnostic efficiency of serum Raman spectroscopy technology combined with SVM, we established two efficient diagnostic models, AFSA-SVM and AFUD-SVM, which were used for the diagnosis of thyroid dysfunction. Compared with the popular GS-SVM diagnostic model, the two diagnostic models we established in this study led to greatly improved diagnostic efficiency while ensuring diagnostic accuracy.

The two diagnostic models we established combined with serum Raman spectroscopy do not have ideal thyroid dysfunction diagnostic accuracy, which needs to be further studied in our future work. We plan to collect more patient samples to verify the reliability of this diagnostic model. However, its fast, noninvasive and low-cost detection characteristics make this technique very promising and interesting, and our model may have great significance for early screening and prevention of thyroid dysfunction diseases.

Acknowledgment

The authors wish to thank the anonymous reviewers for their valuable suggestions.

Competing Interests

The authors declare no competing financial interests.

References

[1] A. G. Unnikrishnan and U. V. Menon, "Thyroid disorders in India: an epidemiological perspective," *Indian J. Endocrinology Metabolism*, vol. 15, no. 6, pp. S78–S81, 2011.

[2] K. C. Rajendra, S. Khatiwada, K. D. Mehta, P. Pandey, M. Lamsal, and S. Majhi, "Cardiovascular risk factors in subclinical hypothyroidism: a case control study in Nepalese population," *J. Thyroid Res.*, vol. 2015, no. 1, 2015, Art. no. 305241.

[3] D. Vaishali, B. Anish, I. Vagesh, J. Harish, J. P. Dholye, and P. K. Varthakavi, "Prevalence, clinical and biochemical profile of subclinical hypothyroidism in normal population in Mumbai," *Indian J. Endocrinology Metabolism*, vol. 17, no. 3, pp. 454–459, 2013.

[4] S. W. Zhang, "More than 200 million patients with thyroid dysfunction in China," 2017. [Online]. Available: http://news.sciencenet.cn/htmlnews/2017/5/377432.shtm

[5] J. P. Walsh *et al.*, "Subclinical thyroid dysfunction as a risk factor for cardiovascular disease," *JAMA Internal Med.*, vol. 165, no. 21, pp. 2467–2472, 2005.

[6] Z. Shan *et al.*, "The iodine status and prevalence of thyroid disorders after introduction of mandatory universal salt iodization for 16 years in China A cross-sectional study in 10 cities," *Thyroid Official J. Amer. Thyroid Assoc.*, vol. 26, no. 8, pp. 1125, 2016.
[7] E. B. Hanlon *et al.*, "Prospects for in vivo Raman spectroscopy," *Phys. Med. Biol.*, vol 45, no. 2, pp. R1–R59, 2000.
[8] Q. Tu and C. Chang, "Diagnostic applications of Raman spectroscopy," *Nanomed. Nanotechnol. Biol. Med.*, vol. 8, no. 5, pp. 545–558, 2012.
[9] C. Kallaway *et al.*, "Advances in the clinical application of Raman spectroscopy for cancer diagnostics," *Photodiagnosis Photodynamic Therapy*, vol. 10, no. 33, pp. 207–219, 2013.
[10] S. X. Li *et al.*, "Identification and characterization of colorectal cancer using Raman spectroscopy and feature selection techniques," *Opt. Exp.*, vol. 22, no. 21, pp. 25895–25908, 2014.
[11] S. X. Li *et al.*, "Characterization and noninvasive diagnosis of bladder cancer with serum surface enhanced Raman spectroscopy and genetic algorithms," *Sci. Rep.*, vol. 5, 2015, Art. no. 9582.
[12] X. Z. Li, T. Y. Yang, S. Q. Li, D. L. Wang, Y. T. Song, and S. Zhang, "Raman spectroscopy combined with principal component analysis and k nearest neighbour analysis for non-invasive detection of colon cancer," *Laser Phys.*, vol. 26, no. 3, 2016, Art. no. 035702.
[13] E. Ryzhikova *et al.*, "Raman spectroscopy of blood serum for Alzheimer's disease diagnostics: Specificity relative to other types of dementia," *J. Biophotonics*, vol. 8, no. 7, pp. 584–596, 2015.
[14] S. Khan *et al.*, "Analysis of hepatitis B virus infection in blood sera using Raman spectroscopy and machine learning," *Photodiagnosis Photodynamic Therapy*, vol. 23, pp. 89–93, 2018.
[15] S. Khan, R. Ullah, A. Khan, N. Wahab, M. Bilal, and M. Ahmed, "Analysis of dengue infection based on Raman spectroscopy and support vector machine (SVM)," *Biomed. Opt. Exp.*, vol. 7, no. 6, pp. 2249–2256, 2016.
[16] S. Khan, R. Ullah, S. Shahzad, S. Javaid, and A. Khan, "Optical screening of nasopharyngeal cancer using Raman spectroscopy and support vector machine," *Optik*, vol. 157, pp. 565–570, 2018.
[17] S. X. Li *et al.*, "Noninvasive prostate cancer screening based on serum surface-enhanced Raman spectroscopy and support vector machine," *Appl. Phys. Lett.*, vol. 105, no. 9, 2014, Art. no. 091104.
[18] X. Z. Li, T. Y. Yang, S. Q. Li, D. L. Wang, Y. T. Song, and K. D. Yu, "Different classification algorithms and serum surface enhanced Raman spectroscopy for noninvasive discrimination of gastric diseases," *J. Raman Spectrosc.*, vol. 47, no. 8, pp. 917–925, 2016.
[19] L. Wang *et al.*, "Raman spectroscopy, a potential tool in diagnosis and prognosis of castration-resistant prostate cancer," *J. Biomed. Opt.*, vol. 18, no. 8, 2013, Art. no. 087001.
[20] X. L. Li, Z. J. Shao, and J. X. Qian, "An optimizing method based on autonomous animats: fish-swarm algorithm," *Syst. Eng. Theory Pract.*, vol. 22, no. 11, pp. 32–38, 2002.
[21] X. L. Li, "A new intelligent optimization method: An artificial fish school algorithm," *Ph.D. dissertation*, Zhejiang Univ., Hangzhou, China, 2003.
[22] K. T. Fang, P. Winker, and Y. Zhang, "Uniform design: Theory and application," *Technometrics*, vol. 42, no. 3, pp. 237–248, 2000.
[23] J. H. Zhao, H. Lui, D. I. McLean, and H. S. Zeng, "Automated autofluorescence background subtraction algorithm for biomedical Raman spectroscopy," *Appl. Spectrosc.*, vol. 61, no. 11, pp. 1225–1232, 2007.
[24] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.
[25] W. Shen, X. Guo, C. Wu, and D. Wu, "Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm," *Knowl.- Based Syst.*, vol. 24, no. 3, pp. 378–385, 2011.
[26] C. M. Huang, Y. J. Lee, D. K. J. Lin, and S. Y. Huang, "Model selection for support vector machines via uniform design," *Comput. Statist. Data Anal.*, vol. 52, no. 1, pp. 335–346, 2007.
[27] T. Liu, C. S. Chen, X. Z. Shi, and C. Y. Liu, "Evaluation of Raman spectra of human brain tumor tissue using the learning vector quantization neural network," *Laser Phys.*, vol. 26, no. 5, 2016, Art. no. 055606.
[28] D. Faraggi and B. Reiser, "Estimation of the area under the ROC curve," *Statist. Med.*, vol. 21, no. 20, pp. 3093–3106, 2002.
[29] S. Y. Feng *et al.*, "Nasopharyngeal cancer detection based on blood plasma surface-enhanced Raman spectroscopy and multivariate analysis," *Biosens. Bioelectron.*, vol. 25, no. 11, pp. 2414–2419, 2010.
[30] Z. Movasaghi, S. Rehman, and I. U. Rehman, "Raman spectroscopy of biological tissues," *Appl. Spectrosc. Rev.*, vol. 42, no. 5, pp. 493–541, 2007.
[31] Z. W. Huang, A. McWilliams, H. Lui, D. I. McLean, S. Lam, and H. S. Zeng, "Near-infrared Raman spectroscopy for optical diagnosis of lung cancer," *Int. J. Cancer*, vol. 107, no. 6, pp. 1047–1052, 2003.
[32] M. Saleem, M. Bilal, S. Anwar, A. Rehman, and M. Ahmed, "Optical diagnosis of dengue virus infection in human blood serum using Raman spectroscopy," *Laser Phys. Lett.*, vol. 10, no. 3, pp. 035602, 2013.
[33] https://en.wikipedia.org/wiki/Thyroid_hormones
[34] M. Mohamedali, S. R. Maddika, A. Vyas, V. Iyer, and P. Cheriyath, "Thyroid disorders and chronic kidney disease," *J. Nephrology*, vol. 2014, 2014, Art. no. 520281.
[35] R. Moore and I. H. Mills, "Serum T3 and T4 levels in patients with anorexia nervosa showing transient hyperthyroidism during weight gain," *Clin. Endocrinology*, vol. 10, no. 5, pp. 443–449, 1979.