Diagnosing Anorexia Based on Partial Least Squares, Back Propagation Neural Network, and Support Vector Machines

C. Y. Zhao,[†] R. S. Zhang,*[,†,‡] H. X. Liu,[†] C. X. Xue,[†] S. G. Zhao,[†] X. F. Zhou,[†] M. C. Liu,[†] and B. T. Fan[§]

Departments of Chemistry and Computer Science, Lanzhou University, Lanzhou 730000, China, and Université Paris 7-Denis Diderot, ITODYS 1, Rue Guy de la Brosse, 75005 Paris, France

Received April 7, 2004

Support vector machine (SVM), as a novel type of learning machine, for the first time, was used to develop a predictive model for early diagnosis of anorexia. It was based on the concentration of six elements (Zn, Fe, Mg, Cu, Ca, and Mn) and the age extracted from 90 cases. Compared with the results obtained from two other classifiers, partial least squares (PLS) and back-propagation neural network (BPNN), the SVM method exhibited the best whole performance. The accuracies for the test set by PLS, BPNN, and SVM methods were 52%, 65%, and 87%, respectively. Moreover, the models we proposed could also provide some insight into what factors were related to anorexia.

1. INTRODUCTION

Anorexia in infancy or early childhood is one of the most familiar diseases in pediatrics, especially in the children ranging in age from 9 months to 7 years old. Infant and young children's anorexia was first described by Chatoor and Egan,[1] as a separation disorder and later defined as a transactional feeding disorder. It is characterized by the infants' persistent food refusal for more than 1 month, malnutrition of the infant. However, anorexia is usually considered a psychogenic disorder, and there were no nationally defined criteria for the diagnosis of feeding disorders. Different authors used a variety of diagnostic methods and assigned different labels to address this problem. Furthermore, the clinic diagnosis is always determined by well-trained experts by observing patients for 1 or 2 months for the patients.[2,3] Such lengthy reduction in food intake always leads to weight loss, poor food efficiency, depletion of protein, growth retardation, and worsening of the child's condition.[4] Therefore, early diagnosis is an important factor to alleviate the harm to an infant or a young child.

In recent years, reports have proposed some possible interactions between some elements in the human body and anorexia in infancy or early childhood. According to reported studies,[5,6] human health may be affected by the amount of certain elements available from food, drinking water, and the atmosphere. Some, such as Zn, Fe, Ca, and Mg, play important roles in the metabolism of neuronal cells and their appendages. Metal ions perform catalytic roles in numerous enzymes and have antioxidant action and various other processes. There is considerable information that the respective element deficiency or excessive intake may disturb some elemental homeostasis, thus resulting in different diseases, such as anorexia.[7] It is apparent that concentrations of these elements have significant potential to be a valid monitor and preventive tools for early diagnosis of anorexia.[8−10] Previous studies of anorexia mainly focused on the anorexia nervosa,[11] disease-associated anorexia,[12] or the treatment and physical recovery of anorexia.[13] To the best of our knowledge, there are no quantitive methods for the diagnosis of anorexia described in the literature.

Many different methods can be used to perform quantitative studies, such as Least Squares (PLS), back-propagation neural network (BPNN), and support vector machine (SVM). PLS is a linear technique, commonly used for regression and classification.[14] BPNN (back-propagation neural networks) is a popular method for classification and multivariate calibration problems.[15] The flexibility of neural networks enables them to discover more complex nonlinear relationships in experimental data. However, these neural systems have some problems inherent to its architecture in relation to overtraining, overfitting, and network optimization. SVM is a new algorithm from the machine learning community. Overfitting of data can be avoided by using the SVM algorithm effectively. Due to its remarkable generalization performance, the SVM has attracted attention and gained extensive applications in many areas, i.e., isolated handwritten digit recognition,[16] object recognition,[17] drug design,[18] prediction of protein structure,[19] identifying genes,[20] and diagnosing breast cancer,[21] etc.

In the present work, the models based on the correlation between the six elements (Zn, Fe, Ca, Mg, Mn, and Cu) and anorexia were built by three methods: PLS, BPNN, and SVM. The work reported here aimed to (1) build predictive models for anorexia and compare the modeling efficacies of the three methods and (2) find the factors related to early diagnosing of anorexia.

2. METHODOLOGY

2.1. Partial Least Squares. PLS analysis is a method based on principal component-like latent variables. It allows

* Corresponding author phone: +86-931-891-2578; fax: +86-931-891-2582; e-mail address: ruison@public.lz.gs.cn.
[†] Department of Chemistry, Lanzhou University.
[‡] Department of Computer Science, Lanzhou University.
[§] Université Paris 7-Denis Diderot.

10.1021/ci049877y CCC: $27.50
Published on Web 09/14/2004

modeling of several dependent variables simultaneously and avoids the collinear problem among independent variables by regressing the dependent variables on orthogonal latent variables. PLS can be used for classification by rounding the output value obtained to an integer. More details on PLS can be found in ref 22.

2.2. Back-Propagation Neural Network. Back-Propagation Neural Network (BPNN) is one of the most popular neural networks.[23] The BPNN model is composed of a large number of simple processing elements (PE) or neuron nodes organized into a sequence of layers. The first layer is the input layer with one node for each variable or feature of the data. The last layer is the output layer consisting of one node for each variable to be investigated. Between the two layers are a series of one or more hidden layer(s) consisting of a number of nodes, which are responsible for learning. Nodes in any layer are fully or randomly connected to nodes of a succeeding layer. Each connection is represented by a number called a weight (w). BPNN is most often used to analyze nonlinear multivariable data. In this network, signals are propagated from the input layer through the hidden layer(s) to the output layer. A node thus receives signals *via* connections from other nodes (or the outside world in the case of the input layer). For the criterion of classification, the value of output should be rounded to an integer.

2.3. SVM. What follows is a brief description of the SVM algorithm. A more detailed description can be found in Vapnik's[24] and Burges's articles.[25]

Support vector machine is a learning system that uses a hypothesis space of linear functions in a high dimensional feature space, trained with a learning algorithm from the optimization theory. Unlike artificial neural networks (ANN) that try to minimize the error on the training data, SVM attempts to minimize the upper bound on the generalization error based on the principle of structural risk minimization (SRM), which has been found to be superior to the empirical risk minimization (ERM) principle employed in ANN.

For a two-binary classification problem, assume that we have a set of samples, i.e. a series of input vectors $\bar{x_i} \in R^d$ ($i = 1, 2, ..., N$), with corresponding labels $y_i \in \{-1, +1\}$ ($i = 1, 2, ..., N$). Here, $+1$ and -1 indicate the two classes. The goal here is to construct one binary classifier or derive one decision function from the available samples, which has a small probability of misclassifying a future sample. Both the basic linear separable case and the most useful linear nonseparable case (for most real life problems) are considered here.

For a linear separable case, there exists a separating hyperplane whose function is $\bar{w}\bullet\bar{x} + b = 0$ which implies the following:

$$y_i(\bar{w}\bullet\bar{x_i} + b) \geq 1 \quad i = 1, 2, ..., N$$

By minimizing $(1/2)||\bar{w}||^2$ subject to this constraint, the SVM approach tries to find a unique separating hyperplane, which maximizes the distance between the hyperplane and the nearest data points of each class. The classifier is called the largest margin classifier.

By introducing Lagrange multipliers a_i, the SVM training procedure amounts to solving a convex quadratic programming (QP) problem. The solution is a unique globally optimized result and can be shown as the following formula:

$$\bar{w} = \sum_{i=1}^{N} y_i a_i \bar{x_i}$$

Only if the corresponding $a_i > 0$, these x_i are called Support Vectors. When a SVM is trained, the decision function can be written as

$$f(\bar{x}) = sign(\sum_{i=1}^{N} y_i a_i (\bar{x}\bullet\bar{x_i}) + b)$$

For a linear nonseparable case, allowing for training errors can be done by introducing positive slack variables ξ_i ($i = 1, 2, ..., N$) in the constraints, which then become

$$y_i(\bar{w}\bullet\bar{x_i} + b) \geq 1 - \xi_i \quad \xi_i \geq 0, i = 1, 2, ..., N$$

We want to simultaneously maximize the margin and minimize the number of misclassifications. This can be achieved by changing the objective function from $(1/2)\,||\bar{w}||^2$ to

$$\frac{1}{2}||\bar{w}||^2 + C\sum_{i=1}^{N}\xi_i^k \text{ Minimize } \frac{1}{2}||\bar{w}||^2 + C\sum_{i=1}^{N}\xi_i^k$$

Subject to

$$y_i(\bar{w}\bullet\bar{x_i} + b) - 1 + \xi_i \geq 0, \; i = 1, 2, ..., N$$

$$\xi_i \geq 0 \quad i = 1, 2, ..., N$$

The error weight C is a regularization parameter to be chosen by the user, it controls the size of penalties assigned to errors. The optimization problem is convex for any positive integer k. For $k = 1$ and $k = 2$ it is also a quadratic programming problem. This is called the Soft Margin Generalization of the OSH, while the original concept with no errors allowed is called Hard Margin.

For a two-binary nonlinear classification problem, SVM performs a nonlinear mapping $\Phi(\bullet)$ of the input vector x_i from the input space R^d into a higher dimensional Hilbert space H and constructs an Optimal Separating Hyperplane. In the linear separable case, we know that the algorithm only depends on inner products between training examples and test examples. So we can generalize to nonlinear case. The inner products are substituted by the kernel function $k(\bar{x_i}, \bar{x_j}) = \Phi(\bar{x_i})\bullet\Phi(\bar{x_j})$, in the input space. Then, the decision function implemented by SVM can be written as

$$f(x) = sign(\sum_{i=1}^{N} y_i a_i k(\bar{x}, \bar{x_i}) + b)$$

Two typical kernel functions are listed below:

$$\text{Polynomial function } k(\bar{x_i}, \bar{x_j}) = (\bar{x_i}\bullet\bar{x_j} + 1)^d$$

$$\text{Radial basis function (RBF) } k(\bar{x_i}, \bar{x_j}) = \exp(-\gamma||\bar{x_i} - \bar{x_j}||^2)$$

3. DATA AND EXPERIMENT

3.1. Data Set. A data set containing 90 cases (62 nonanorexic cases and 28 anorexic cases) was collected from

Table 1. Cases, Features, and the Diagnosis[a]

no.	para 1	para 2	para 3	para 4	para 5	para 6	para 7	para 8	diagnosis	no.	para 1	para 2	para 3	para 4	para 5	para 6	para 7	para 8	diagnosis
1[b]	1	1	142	54	76	93	759	2.1	1	46	1	1	100	40	35	10	1000	2.33	1
2	0	6	156	31	70	52	820	3.65	1	47	1	7	105	42	36	10	1100	2.31	1
3	0	5	70	63	54	32	850	1.35	1	48	0	6	105	45	36	10	1050	1.26	1
4[b]	0	4	72	60	53	32	920	1.35	1	49[b]	0	7	115	32	50	10	750	2.36	1
5	0	3	220	38	142	25	2300	4.56	1	50	0	3	120	60	120	10	1300	1.28	1
6[b]	1	2	75	40	65	22	1000	3.21	1	51	0	4	120	31	125	10	1320	3.21	1
7	1	4	69	67	36	21	1200	1.26	1	52[b]	1	1	122	59	59	10	686	1.35	1
8	1	5	142	54	76	20	844	2.12	1	53	0	3	145	40	85	10	920	2.44	1
9	1	1	82	40	138	17	1600	1.83	1	54	0	2	150	60	200	10	3600	1.45	1
10[b]	1	5	65	77	16	16	780	1.35	1	55	0	3	160	21	160	10	2500	3.56	1
11	1	7	85	36	165	16	1650	0.96	1	56[b]	1	2	170	35	170	10	1700	1.35	1
12[b]	0	3	98	32	102	16	1920	3.21	1	57	0	5	105	37	66	9	980	1.63	1
13	1	5	110	80	78	16	1300	1.45	1	58	1	1	140	95	65	9	650	1.62	1
14	0	7	170	15	164	16	1630	3.58	1	59	0	4	146	36	150	9	2300	2.65	1
15	1	5	54	85	46	15	1300	0.69	1	60	1	1	85	60	75	8	250	6.4	1
16[b]	0	1	55	32	94	15	1200	2.62	1	61	1	3	125	40	50	8	900	2.35	1
17	1	3	65	29	100	15	1500	1.65	1	62	1	5	128	36	54	8	1300	1.28	1
18	0	3	100	35	100	15	1900	3.52	1	63	0	1	68	62	55	35	900	3.21	2
19[b]	1	1	103	105	80	15	900	1.82	1	64	0	1	75	30	75	10	550	2.31	2
20	1	1	105	105	75	15	1250	2.1	1	65	0	5	66	63	54	32	850	0.95	2
21	0	6	105	35	75	15	1250	3.22	1	66	0	2.5	76	40	60	22	1100	2.14	2
22	0	1	106	48	56	15	1034	1.84	1	67	0	1	60	75	45	15	800	5.65	2
23[b]	0	6	110	48	60	15	1100	2.12	1	68	0	1.5	71	16	155	14	1100	2.35	2
24	1	1	140	60	110	15	2000	2.12	1	69	0	3	52	56	52	14	620	0.47	2
25	1	1	150	35	150	15	2000	2.56	1	70	1	4	60	34	66	13	1100	1.08	2
26	0	6	180	21	45	15	3400	3.87	1	71[b]	0	4	85	86	36	13	900	1.89	2
27[b]	0	1.5	185	95	45	15	4000	1.2	1	72	0	1	23	48	24	12	336	1.44	2
28	1	6	60	62	35	14	660	1.24	1	73	1	2.5	70	34	75	12	840	1.23	2
29	1	5	98	36	110	14	1860	2.31	1	74[b]	1	5	138	35	28	11	520	1.44	2
30	0	4	100	32	110	14	1850	2.28	1	75	1	1	135	32	35	10.5	500	1.26	2
31	1	4	148	32	162	14	1960	2.65	1	76	1	0.8	55	31	35	10	860	1.38	2
32	1	2	42	32	120	13	1800	1.86	1	77	0	1	50	32	50	10	750	2.25	2
33	1	1	66	53	43	13	672	3.2	1	78[b]	1	1	87	17	20	6	650	2.11	2
34	1	3	110	70	130	13	1500	0.95	1	79	1	1.5	80	35	40	10	650	1.36	2
35	1	5	150	35	150	13	2010	2.26	1	80	1	2	40	28	20	10	350	1.23	2
36[b]	0	3	165	80	160	13	3000	1.28	1	81[b]	0	2	95	10	36	10	1200	1.45	2
37	0	7	90	36	90	12.5	1600	2.31	1	82	1	2.5	40	35	25	10	360	0.88	2
38	1	4	90	32	85	12.5	1540	2.1	1	83	1	4	80	32	35	10	600	1.28	2
39	0	1	75	32	150	12	1500	1.65	1	84	1	3	100	50	50	9	750	0.68	2
40[b]	0	6	100	25	65	12	1230	1.58	1	85[b]	1	3.5	50	32	45	9	1200	2.32	2
41[b]	0	1	100	26	65	11	1200	2.36	1	86	1	0.8	50	42	35	8	395	1.33	2
42	1	4	132	34	42	11	500	2.36	1	87	0	1	66	36	65	8	760	1.25	2
43	1	1	160	115	90	11	850	1.5	1	88[b]	1	3.5	65	20	36	8	760	1.15	2
44	1	1	136	36	55	8	1250	3.24	1	89[b]	0	4	54	32	94	8	1100	0.68	2
45[b]	0	3	78	40	100	10	600	2.04	1	90	1	0.8	45	11	8	6	550	1.27	2

[a] Parameter 1: gender of cases; parameter 2: age of cases; parameters 3−8: concentration of Zn, Fe, Mg, Cu, Ca, and Mn. [b] Test set.

Yaojie colliery, Lanzhou, China. Diagnoses were determined by experts of Hai Shiwan Hospital (Lanzhou, China), according to clinic symptoms for about 1 month. In this data set, according to the information given by the parents and doctors, the children had no mental disorders and organic diseases.

The data set contains two types of samples: nonanorexic and anorexic. Number 1 was used to represent nonanorexic cases and number 2 for anorexic ones. Each case was described by 8 features: parameter 1 (gender); parameter 2 (age); parameters 3−8 (concentrations of 6 elements: Zn, Fe, Mg, Cu, Ca, and Mn). To identify gender, number 0 expressed female and number 1 expressed male. The data set was randomly divided into two sets: the training set (67 cases) and the test set (23 cases), as shown in Table 1. The training set was used to adjust the parameters of the models, and the test set was used to evaluate its prediction ability. The leave-one-out (LOO) cross-validation procedure was used to estimate the modeling ability of the model. The predictive ability was evaluated in terms of the number of misclassified samples and the accuracy.

3.2. Experiment. Various methods can be used to assay the concentration of trace elements in the body, but hair analysis is a popular and accurate method. Hair analysis can verify extreme environmental conditions and significant changes in lifestyle, which induce detectable imbalances of some key trace elements.[5] The elements tested in this study were Zn, Fe, Mg, Ca, Cu, and Mn, which were chosen on the basis of their essential role in the human organism and the ensuing adverse effects caused by either their deficiency or excess.

The procedure adopted in this investigation was inspired by that already described in previous works.[6] Briefly, it consisted of the following steps:

1. Weighing the hair about 0.2 g from individuals that came from Yaojie colliery. All subjects were interviewed at the time of the sampling to obtain some general information on their health status, lifestyle, etc.

Table 2. Correlation Matrix

	para1	para2	para3	para 4	para 5	para 6	para 7	para 8
para1	1.000							
para2	−0.152	1.000						
para3	−0.082	0.158	1.000					
para4	0.122	−0.127	0.103	1.000				
para5	−0.184	0.042	0.452	−0.078	1.000			
para6	−0.081	0.050	0.127	0.147	0.022	1.000		
para7	−0.254	0.137	0.563	0.031	0.612	−0.021	1.000	
para8	−0.217	−0.052	0.284	−0.141	0.187	0.116	0.141	1.000

Table 3. Concentration of Six Elements in Nonanorexic Cases, Anorexic Cases, and All Cases

	Zn	Fe	Mg	Cu	Ca	Mn
av concn in all cases (μg/g)	99.90	45.01	76.92	14.63	1214.95	2.11
av concn in nonanorexic cases (μg/g)	113.79	48.34	89.82	15.66	1424.11	2.21
av concn in anorexic cases (μg/g)	69.14	37.64	48.36	12.34	751.82	1.89
standard	1	1	1	1	1	1
standardized factor	0.61	0.78	0.54	0.79	0.53	0.86

2. Sample pretreatment in order to remove exogenous material.

3. Analysis of hair samples according to the previously established procedures.

In detail, specimens were washed by 10% scour, followed by drying at 50 °C.

After dissolved in nitric acid, the hair samples were washed into a colorimetric tube of 10 mL by 5% hydrochloric acid. Finally, the concentration of elements was assayed by Inductively Coupling Plasma 400 (ICP) (Perkin Elmer, U.S.A.).

3.3. PLS, BPNN, and SVM Implementation and Computation Environment. All calculation programs implementing PLS and BPNN were written in M-files in the MATLAB software. All calculation programs implementing SVM were written in an R-file based on the R script for SVM.[26] The scripts were compiled using an R 1.7.1 compiler running on a Pentium IV PC with 256M RAM.

4. RESULTS AND DISCUSSION

4.1. Feature Selection. The quality of the model largely depends on the selection of variables which could represent the information of the interested object. The used variables should give the maximum of information both in statistics and in use. That was why the features selection was performed first in this work.

In the process of choosing features, one of two descriptors whose pairwise correlation coefficient exceeded 0.90 was removed to eliminate redundant information. Table 2 shows the correlation coefficients of these eight variables. The correlation coefficient value of any two variables is < 0.612, which means all the variables are independent.

However, some statistic processes were performed on the eight variables. Table 3 shows the average concentrations of these six elements in nonanorexic cases comparing with those in anorexic ones. It should be noted that, among the six elements, Zn, Fe, Cu, and Mn are generally viewed as trace elements, while Ca and Mg are not. Therefore, to show the differences in a more straightforward manner, the average concentrations in nonanorexic cases were regarded as the standard 1, while the corresponding ones in anorexic cases were transformed into a standardized factor, as shown in Table 3. It can be seen from the table that the concentration in nonanorexic cases was significantly higher than that in anorexic cases. It further indicated that the decrease in these elements had a great diagnostic value in governing the effect on anorexia. Other factors considered were the subjects' age and gender. In Table 4, the average ages of anorexic cases, of nonanorexic cases, and of the whole data set were compared. There were significant differences among them. Table 4 also showed the statistical results of the gender. The rates of male and female cases were almost the same in the whole data set, in the anorexic data set, and in the nonanorexic data set, and "gender " (parameter 1) was removed from the model.

Table 4. Statistical Results of the Age and the Gender

	whole data set	anorexic cases	nonanorexic cases
average age	3.08	2.28	3.44
rate of male cases and female cases	48/42 = 1.14	15/13 = 1.15	33/29 = 1.14

Besides the statistical results, another important factor that should be considered in the process of choosing proper features was their practical signification. Many studies have shown that the concentrations of the elements of Zn, Fe, Cu, Mn, Ca, and Mg play the important role for anorexia.[27−29] In general, these concentrations are hardly altered over a period of several months.[5] Consequently, the changes can represent the state of the health of an individual. Some experiments have shown that age might also affect the absorption of trace elements, leading to an eating disorder and anorexia.[30]

Therefore, parameters 2−8 (age of cases; concentration of Zn, Fe, Mg, Cu, Ca, and Mn) were selected as the input to the models for the early diagnosis of anorexia.

4.2. Results. 4.2.1. Data Processing. Data were preprocessed before building the predicting models. According to functions (1)−(3) listed below,[33] the selected 7 features were transformed into the standardized data *Zi*

$$hr = \frac{1}{2(X_{\max} - X_{\min})} \tag{1}$$

$$cv = \frac{1}{2(X_{\max} + X_{\min})} \tag{2}$$

$$Zi = \frac{Xi - cv}{hr} \tag{3}$$

where X_{max} and X_{min} are the maximum and minimum value for each feature X_i, and the standardized data Z_i will be used to build predicting models below.

4.2.2. Results of Partial Least Squares. For PLS, selecting an optimal number of latent variables (LV) plays a key role in obtaining a best result. We calculated the misclassified samples and training accuracy on LOO cross-validation on different LV, and the optimal number of LV was 6. Then, an additional test set was applied to this model to estimate the generalization performance. The misclassified numbers in the training set, the test set, and the whole data set were 31, 11, and 42, respectively, and the accuracy in three data sets was shown in Figure 3.

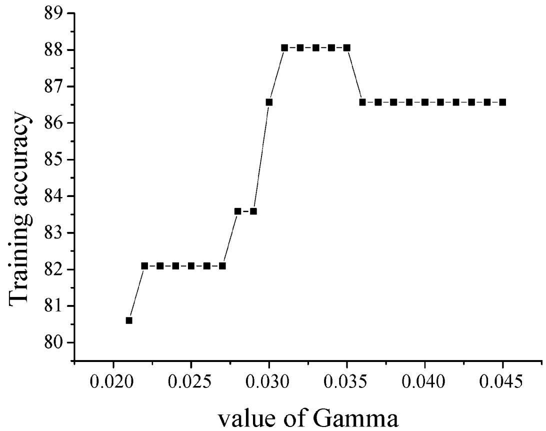

Figure 1. Training accuracy versus value of gamma from 0.02 to 0.045.

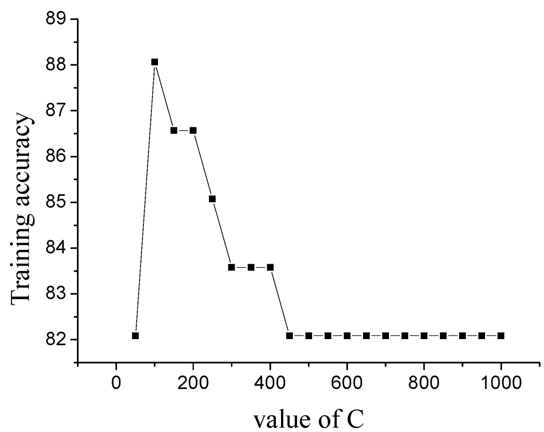

Figure 2. Training accuracy versus value of *C* from 50 to 1000.

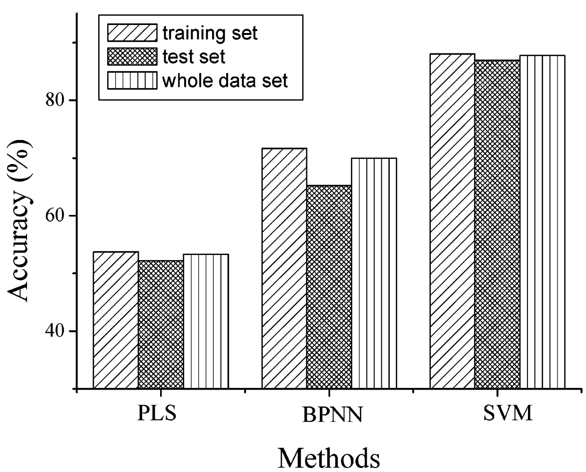

Figure 3. Comparison of the accuracy by PLS, BPNN, and SVM.

4.2.3. Results of Back Propagation Neural Network. In this work, a three-layered feed-forward network was used. The neurons were arranged in layers (the input layer, the hidden layer, and the output layer), in which seven neurons constitute the input layer, one neuron constitutes the output layer, and the hidden layer contains a variable number of neurons. The architectures of BPNN with a different number of neurons in hidden layer should be trained. Besides neurons, the performance of this neural network is also affected by other parameters, such as learning rate and momentum term.

Since it is difficult to choose the optimal value of these three parameters, it is necessary to train networks with a different value of them and choose the best configuration in practice. Networks with hidden nodes ranging from 1 to 10, a learning rate ranging from 0.1 to 0.9, and a momentum term ranging from 0.1 to 0.9 were tested. All weights were initialized to random values between −0.1 and 0.1. Through the above process, it was found that the network consisting of two hidden nodes usually gave the best classification performance on the LOO cross-validation in the training set, and the optimal learning rate and the momentum term were selected as 0.6 and 0.4, respectively.

From the best network, the inputs of the independent test set were presented with it, and the results of BPNN were obtained. The misclassified number in the training set, the test set, and the whole data set were 19, 8, and 27, respectively, and the corresponding accuracy was shown in Figure 3.

4.2.4. Results of Support Vector Machine. The quality of the SVM models depends on the kernel type, various parameters that control the kernel shape. Using a quadratic programming algorithm, SVM offers a unique maximal separation hyperplane. As other multivariate statistical models used in chemometrics, there are no clear guidelines for selecting the optimum set of theoretical parameters and decision function (kernel type and associated parameters). Therefore, the only practical way of finding an optimally predictive SVM model is through extensive experiments.

In this work, SVM training included the selection of capacity parameter *C*, the corresponding parameters of the kernel function.

Capacity parameter *C* is a regularization parameter that controls the tradeoff between maximizing the margin and minimizing the training error. If *C* is too small, then insufficient weight will be placed on fitting the training data. If *C* is too large, then the algorithm will overfit the training data. However, ref 31 indicats that the prediction error was scarcely influenced by C. To make the learning process stable, a large value should be set up for *C* first (e.g, C=100).

To select the type of kernel function, which determines the sample distribution in the mapping space, many studies indicated that the radial basis function is commonly used because of its good general performance and few parameters to be adjusted.[32] Therefore, in this work, the radial basis function (RBF) was used, the form of which in R is as follows

$$exp(-\gamma^{*}|\mathrm{u}-\mathrm{v}|^{2})$$

where γ is a parameter of the kernel and u and v are two independent variables.

The γ of kernel function greatly affects the number of support vectors, which has a close relation with the performance of the SVM and training time. Too many support vectors can produce overfitting and make the training time longer. The γ also controls the amplitude of the RBF function and, therefore, controls the generalization ability of SVM. Thus, to find the optimal parameter γ, experiments were performed using a different value of γ with the LOO procedure using the same training set used in PLS and BPNN. For the training data set, the first group of models, parameter γ was set in the range of 0.02 to 0.045 with 0.001

increment and $C = 100$. The plot of γ versus training accuracy is shown in Figure 1. As can been seen from the figure, when γ was 0.031, 0.032, 0.033, 0.034, and 0.035, the training accuracy reached the peak, and their corresponding SVs were 21, 22, 22, 23, and 23, respectively. The low number of support vectors prompted the selection of 0.031 as the optimal value of the gamma.

In addition, to test the effect of C, the second group of models using the same training data set were obtained with capacity parameter C from 50 to 1000, every 50 and a certain $\gamma = 0.031$. The curve of training accuracy and C value was shown in Figure 2. It can be seen from it that the selection of parameter C has some influence on the performance. The optimal C was found as 100 with a highest training accuracy of 88.05%.

The best choices for C and γ of the SVM were 100 and 0.031, with the support vector number of 21. The test set was presented using the SVM model. The SVM gave the misclassified number of 8 for the training set, 3 for the test set, and 11 for the whole data set, and the corresponding accuracy was also shown in Figure 3.

4.2.5. Comparison of the Three Classifiers. Figure 3 gives the comparison between the methods PLS, BPNN, and SVM. From the figure, it can be seen that the results of the SVM is the best one both on the LOO cross-validation of the training set and on the test set. It indicates that the SVM has the better generalization ability.

Another aspect that should be noted is the misclassified samples. According to the obtained results, it can be found that the samples "29" and "76" in the training set and the samples "1", "4", and "6" in the test set were misclassified by all the methods. It can be presumed that these samples could be mislabeled and need to be determined further.

5. CONCLUSION

In the present work, seven features, including the age and the concentration of elements Zn, Fe, Mg, Cu, Ca, and Mn, were used to build predictive models for the diagnosis of anorexia by the use of PLS, BPNN, and SVM. Compared with the results obtained by other two methods of PLS and BPNN, the model using SVM exhibited a better predictive ability with the minimal misclassified number of 8 for the training set, 3 for the test set, and 11 for the whole data set. It showed that the SVM method based on selected features can be used as a valid monitor and preventive tool for the diagnosis of anorexia before overt symptoms appear. More importantly, SVM was shown to be a very promising tool for classification due to the embodying of the structural risk minimization principle which minimizes an upper bound of the generalization error rather than minimizes the training error. This eventually leads to better generalization than neural networks which implement the empirical risk minimization principle. At the same time, the neural network may not converge to global solutions. In addition, there are fewer free parameters to be adjusted in the SVM, which made the model selecting process easy to be controlled. Therefore, the SVM is a very effective machine learning technique for many aspects and will gain more extensive application.

ACKNOWLEDGMENT

The authors thank the Association Franco-Chinoise pour la Recherche Scientifique & Technique (AFCRST) for supporting this study (Program PRA SI 02-02).

REFERENCES AND NOTES

(1) Chatoor, I. Diagnosis and treatment of infantile anorexia. *Infant Behav. Dev.* **1998**, *21*, 118.
(2) Chatoor, I.; Hirsch, R.; Ganiban, J.; Persinger, M.; Hambourger, E. Diagnosing infantile anorexia: the observation of mother-infant interactions. *J. Am. Acad. Child Adolesc. Psychiatry* **1998**, *37*, 959−967.
(3) Renouil, M.; Fourmaintraux, A.; Cartault, F.; Rodriguez, D.; Razafinarivo-Schoreitz, S.; Chaurand, G.; Wendling, C.; Bangui, A.; Ponsot, G. Severe anorexia in infants in Reunion: a new autosomal recessive disease? *Arch. Pediatr*. **1999**, *6*, 725−734.
(4) Essatara, M. B.; Levine, A. S.; Morley, J. E.; McClain, C. J. Zinc deficiency and anorexia in rats: normal feeding patterns and stress induced feeding. *Physiol. Behav.* **1984**, *32*, 469−474.
(5) Senofonte, O.; Violante, N.; D'Ilio, S.; Caimi, S.; Peri, A.; Carolia, S. Hair analysis and the early detection of imbalances in trace elements for members of expeditions in Antarctica. *Microchem. J*. **2001**, *69*, 231−238.
(6) Senofonte, O.; Violante, N.; Caroli, S. Assessment of reference values for elements in human hair of urban schoolboys. *J. Trace Elem. Med. Biol.* **2000**, *14*, 6−13.
(7) Gaetke, L. M.; Frederich, R. C.; Oz, H. S.; McClain, C. J. Decreased food intake rather than zinc deficiency is associated with changes in plasma leptin, metabolic rate, and activity levels in zinc deficient rats. *J. Nutr. Biochem*. **2002**, *13*, 237−244.
(8) Ryan, D. E.; Holzbecher, J.; Stuart, D. C. Trace elements in scalp-hair of persons with multiple sclerosis and of normal individuals. *Clin. Chem.* **1978**, *24*, 1996−2000.
(9) Hopps, H. C. The biologic bases for using hair and nail for analyses of trace elements. *Sci. Total Environ*. **1977**, *7*, 71−89.
(10) Sergeant, C.; Llabador, Y.; Deves, G..; Vesvres, M. H.; Simonoff, M.; Yefimova, M.; Courtois, Y.; Jeanny, J. C. Iron and other elements (Cu, Zn, Ca) contents in retina of rats during development and hereditary retinal degeneration. *Nucl. Instrum. Methods B* **2001**, *181*, 533−538.
(11) Kaye, W.; Gendall, K.; Strober, M. Serotonin Neuronal Function and Selective Serotonin Reuptake Inhibitor Treatment in Anorexia and Bulimia Nervosa. *Biol. Psychiatry* **1998**, *44*, 825−838.
(12) Konsman, J. P.; Dantzer, R. How the Immune and Nervous Systems Interact During Disease-Associated Anorexia. *Nutritio*. **2001**, *17*, 664−668.
(13) Fennig, S.; Fennig, S.; Roe, David. Physical recovery in anorexia nervosa: Is this the sole purpose of a child and adolescent medical-psychiatric unit? *Gen. Hosp. Psychiat.* **2002**, *24*, 87−92.
(14) Sorich, M. J.; Miners J. O.; McKinnon, R. A.; Winkler, D. A.; Burden, F. R.; Smith, P. A. Comparison of Linear and Nonlinear Classification Algorithms for the Prediction of Drug and Chemical Metabolism by Human UDP-Glucuronosyltransferase Isoforms. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2019−2024.
(15) Sboner, A.; Eccher, C.; Blanzieri, E.; Bauer, P.; Cristofolini, M.; Zumiani, G.; Forti, S.; A multiple classifier system for early melanoma diagnosis. *AI Med.* **2003**, *27*, 29−44.
(16) Cun, Y. L.; Jackel, L.; Bottou, L.; Cortes, C.; Denker, J.; Drucker, H.; Guyon, I.; Müller, U.; Säckinger, E.; Simard, P. Y.; Vapnik; V. Learning Algorithms for Classification: A Comparison on Handwritten Digit Recognition, Neural Networks. Neural Networks: The Statistical Mechanics Perspective. *World Sci*. **1995**, 261−276.
(17) Blanz, V.; Schölkopf, B.; Bülthoff, H.; Burges, C.; Vapnik, V.; Vetter, T. Comparison of view-based object recognition algorithms using realistic 3D models. In Artificial Neural Networks − ICANN'96. Malsburg, C. V. D., Seelen, W. V., Vorbrüggen, J. C., Sendhoff, B., Eds.; *Springer. Lect. Notes. Comput. Sci.* **1996**, *1112*, 251−256.
(18) Burbidge, R.; Trotter, M.; Buxton, B.; Holden, S. Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Comput. Chem.* **2001**, *26*, 5−14.
(19) Cai, Y. D.; Liu, X. J.; Xu, X. B.; Chou, K. C. Prediction of protein structural classes by support vector machines. *Comput. Chem.* **2002**, *26*, 293−296.
(20) Bao, L.; Sun, Z. R. Identifying genes related to drug anticancer mechanisms using support vector machine. *FEBS Lett*. **2002**, *521*, 109−114.
(21) Liu, H. X.; Zhang, R. S.; Luan, F.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T. Diagnosing Breast Cancer Based on Support Vector Machines. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 900−907.

(22) Geladi, P.; Kowalski, R. B. Partial least-squares regression: a tutorial. *Anal. Chem. Acta*. **1986**, *185*, 1−17.

(23) Klawun, C.; Wilkins, C. L. A Novel Algorithm for Local Minimum Escape in Back-Propagation Neural Networks: Application to the Interpretation of Matrix Isolation Infrared Spectra. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 984−993.

(24) Vapnik, V. *Estimation of Dependencies Based on Empirical Data*; Springer: Berlin, 1982.

(25) Burges, C. J. C. A Tutorial on Support Vector Machines for Pattern Recognition. *Data. Min. Knowl. Discuss.* **1998**, *2*, 121−167.

(26) Venables, W. N. D.; Smith, M. *The R Development Core Team. R manuals;* 2003.

(27) Shay, N. F.; Manigan, H. F. Neurobiology of zinc-influenced eating behavior. *J. Nutr.* **2000**, *130*, 1493−1499.

(28) Iyengara, G. V.; Rappb, A. Human placenta as a ‘dual’ biomarker for monitoring fetal and maternal environment with special reference to potentially toxic trace elements. Part 2: Essential minor, trace and other nonessential elements in human placenta. *Sci. Total Environ*. **2001**, *280*, 207−219.

(29) Dalway, J. S.; Why trace elements are important. *Fuel. Process. Technol.* **2000**, *65*−*66*, 21−33.

(30) Keller, K. A.; Grider, A.; Coffield, J. A. Age-dependent influence of dietary zinc restriction on short-term memory in male rats. *Physiol Behav.* **2001**, *72*, 339−348.

(31) Wang, W.J.; Xu, Z. B.; Lu, W. Z.; Zhang, X. Y. Determination of the spread parameter in the Gaussian kernel for classification and regression. *Neurocomputing* **2003**, *55, 3*−*4*, 643−663.

(32) Bishop, C. *Neural Networks for Pattern Recognition*; Clarendon Press: Oxford, 1997.

(33) Moreda-Pineiro, A.; Marcos, A.; Fisher, A.; Hill, S. J. Evaluation of the effect of data pretreatment procedures on classical pattern recognition and principal components analysis: a case study for the geographical classification of tea. *J. Environ. Monitor.* **2001**, *3*, 352−360.

CI049877Y