

Research Article

PSO-Based Support Vector Machine with Cuckoo Search Technique for Clinical Disease Diagnoses

Xiaoyong Liu^{1,2} and Hui Fu¹

¹ Department of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, Guangdong 510665, China

² School of Business Administration, South China University of Technology, Guangzhou, Guangdong 510640, China

Correspondence should be addressed to Xiaoyong Liu; liugucas@gmail.com

Received 16 April 2014; Accepted 4 May 2014; Published 25 May 2014

Academic Editor: Xin-She Yang

Copyright © 2014 X. Liu and H. Fu. 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.

Disease diagnosis is conducted with a machine learning method. We have proposed a novel machine learning method that hybridizes support vector machine (SVM), particle swarm optimization (PSO), and cuckoo search (CS). The new method consists of two stages: firstly, a CS based approach for parameter optimization of SVM is developed to find the better initial parameters of kernel function, and then PSO is applied to continue SVM training and find the best parameters of SVM. Experimental results indicate that the proposed CS-PSO-SVM model achieves better classification accuracy and F-measure than PSO-SVM and GA-SVM. Therefore, we can conclude that our proposed method is very efficient compared to the previously reported algorithms.

1. Introduction

Accurate diagnosis and effective treatment of disease are important issues in life science research and have a positive meaning for human health. Recently, medical experts pay more attention to early diagnosis of disease and propose many new methods to deal with disease diagnosis problem. Using machine learning methods to diagnose disease is rapid development of a novel research branch of machine learning. Researchers have applied artificial intelligence and computer technology to develop some medical diagnostic systems, which improve the efficiency of diagnosis and become practical tools.

It is shown that support vector machine has good generalization ability and has been widely used in many research areas, such as signal classification [1], image processing [2], and disease diagnosis [3–6]. Illán et al. [3] showed a fully automatic computer-aided diagnosis (CAD) system for improving the accuracy in the early diagnosis of the AD. The proposed approach is based firstly on an automatic feature selection and secondly on a combination of component-based support vector machine (SVM) classification and a pasting votes technique of assembling SVM classifiers.

Sartakhti et al. [4] proposed a novel machine learning method that hybridized support vector machine (SVM) and simulated annealing (SA) for hepatitis disease diagnosis. The obtained classification accuracy of SVM-SA method was 96.25% and it was very promising with regard to the other classification methods in the literature for this problem. The approach proposed by Ramírez et al. [5] was based on image parameter selection and support vector machine (SVM) classification. The proposed system yielded a 90.38% accuracy in the early diagnosis of Alzheimer's disease and outperformed existing techniques including the voxel-as-features (VAF) approach. Abdi and Giveki [6] developed a diagnosis model based on particle swarm optimization (PSO), support vector machines (SVMs), and association rules (ARs) to diagnose erythematosquamous diseases. The proposed model consists of two stages: first, AR is used to select the optimal feature subset from the original feature set. Then, a PSO-based approach for parameter determination of SVM is developed to find the best parameters of kernel function (based on the fact that kernel parameter setting in the SVM training procedure significantly influences the classification accuracy and PSO is a promising tool for global searching).

Support vector machine is a machine learning algorithm based on statistical learning theory and has the strong predictive ability for nonlinear problems. However, SVM prediction performance is closely related to the quality of the selected parameters. Parameter optimization algorithms currently used are particle swarm optimization and genetic algorithms, but these algorithms have their shortcomings and affect the accuracy of disease prediction.

Cuckoo search (CS) is a new swarm intelligent optimization algorithm. Preliminary studies show that cuckoo search algorithm is simple and efficient, easy to implement and has less parameters [7]. Cuckoo search algorithm is able to provide a new method for the SVM parameter optimization. This paper proposes a disease diagnosis model based on cuckoo search, particle swarm optimization (PSO), and support vector machine.

The structure of this paper is as the following. Section 2 firstly introduces related algorithms, such as support vector machine and cuckoo search, and then presents the novel models, CS-PSO-SVM. Section 3 gives results of different models in two real disease diagnoses datasets from University of California Irvine Machine Learning Repository. Finally, conclusions are presented in Section 4.

2. Methods and Materials

2.1. Methods

2.1.1. SVM. Support vector machines [8–10] (SVMs) are a set of related supervised learning methods used for classification and regression. A support vector machine constructs a hyperplane or set of hyperplanes in a high-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data points of any class (so-called functional margin), since, in general, the larger the margin, the lower the generalization error of the classifier.

In order to extend the SVM methodology to handle data that is not fully linearly separable, we relax the constraints slightly to allow for misclassified points; the formulation is following (1). This is done by introducing a positive slack variable ξ_i , $i = 1, 2, \dots, L$:

$$\begin{aligned} x_i \cdot w + b &\geq +1 - \xi_i \quad (y_i = +1), \\ x_i \cdot w + b &\leq -1 + \xi_i \quad (y_i = -1) \end{aligned} \quad (1)$$

which can be combined into

$$y_i (x_i \cdot w + b) - 1 + \xi_i \geq 0, \quad (2)$$

where $\xi_i \geq 0$.

In this soft margin SVM, data points on the incorrect side of the margin boundary have a penalty that increases with the distance from it. As we are trying to reduce the number

of misclassifications, a sensible way to adapt our objective function from the previously mentioned is to find

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^L \xi_i \\ \text{s.t.} \quad & y_i (x_i \cdot w + b) - \varepsilon + \xi_i \geq 0, \end{aligned} \quad (3)$$

where the parameter C controls the trade-off between the slack variable penalty and the size of the margin. Reformulating as a Lagrange, which as before we need to minimize with respect to w , b , and ξ_i and maximize with respect to α (where $\alpha_i \geq 0$, $u_i \geq 0$),

$$\begin{aligned} L_P \equiv & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^L \xi_i \\ & - \sum_{i=1}^L \alpha_i [y_i (x_i \cdot w + b) - 1 + \xi_i] - \sum_{i=1}^L \mu_i \xi_i. \end{aligned} \quad (4)$$

Differentiating with respect to w , b , and ξ_i and setting the derivatives to zero,

$$\begin{aligned} \frac{\partial L_P}{\partial w} = 0 &\implies w = \sum_{i=1}^L \alpha_i y_i x_i, \\ \frac{\partial L_P}{\partial b} = 0 &\implies \sum_{i=1}^L \alpha_i y_i = 0, \\ \frac{\partial L_P}{\partial \xi_i} = 0 &\implies C = \alpha_i + \mu_i. \end{aligned} \quad (5)$$

So, we need to find

$$\begin{aligned} \max_{\alpha} \quad & \left[\sum_{i=1}^L \alpha_i - \frac{1}{2} \alpha^T H \alpha \right] \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^L \alpha_i y_i = 0. \end{aligned} \quad (6)$$

When applying SVM to nonlinear dataset, we need to define a feature mapping function $x \mapsto \phi(x)$. The feature mapping function is called kernel function. In the feature space, optimal hyperplane (Figure 1) can be gotten.

There are three common kernel functions:

polynomial kernel

$$k(x_i, x_j) = (x_i \cdot x_j + a)^b, \quad (7)$$

radial basis kernel

$$k(x_i, x_j) = e^{-\|x_i - x_j\|^2 / 2\sigma^2}, \quad (8)$$

sigmoidal kernel

$$k(x_i, x_j) = \tanh(ax_i \cdot x_j - b), \quad (9)$$

where a and b are parameters defining the kernel's behavior.

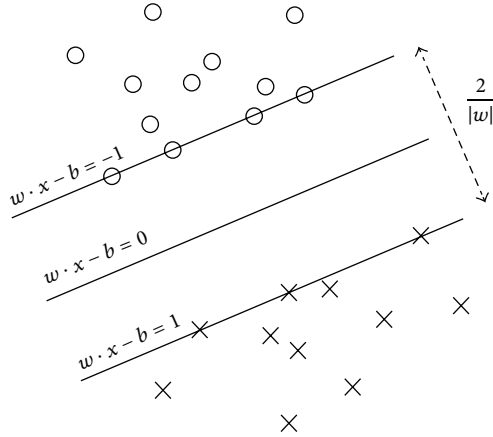


FIGURE 1: Optimal hyperplane.

In order to use SVM to solve a classification or regression problem on dataset that is nonlinearly separable, we need to first choose a kernel and relevant parameters which you expect they might map the nonlinearly separable data into a feature space where it is linearly separable. This is more of an art than an exact science and can be achieved empirically, for example, by trial and error. Sensible kernels to start with are polynomial, radial basis, and sigmoid kernels.

For classification, we need the following.

Create H , where $H_{ij} = y_i y_j \phi(x_i) \phi(x_j)$.

Choose how significantly misclassifications should be treated, by selecting a suitable value for the parameter C .

Find α so that

$$\begin{aligned} \max_{\alpha} \quad & \left[\sum_{i=1}^L \alpha_i - \frac{1}{2} \alpha^T H \alpha \right] \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^L \alpha_i y_i = 0. \end{aligned} \quad (10)$$

Calculate

$$w = \sum_{i=1}^L \alpha_i y_i \phi(x_i). \quad (11)$$

Determine the set of support vectors V , by finding the indices such that $0 \leq \alpha_i \leq C$.

Calculate

$$b = \frac{1}{N_v} \sum_{v \in V} \left(y_v - \sum_{i \in V} \alpha_i y_i \phi(x_i) \phi(x_v) \right). \quad (12)$$

Each new point x' is classified by evaluating $y' = \text{sgn}(w\phi(x') + b)$.

2.1.2. PSO-SVM. Particle swarm optimization is an evolutionary computation technique proposed by Kennedy and Eberhart. It is a population-based stochastic search process, modeled after the social behavior of a bird flock [11, 12]. It is similar in spirit to birds migrating in a flock toward some

destination, where the intelligence and efficiency lie in the cooperation of an entire flock [13]. PSO algorithms make use of particles moving in an n -dimensional space to search for solutions for n -variable function optimization problem. All particles have fitness values which are evaluated by the fitness function to be optimized and have velocities which direct the flying of the particles. The particles fly through the problem space by following the particles with the best solutions so far. PSO is initialized with a group of random particles (solutions) and then searches for optima by updating each generation [14].

SVM also has a drawback that limits the use of SVM on academic and industrial platforms: there are free parameters (SVM hyperparameters and SVM kernel parameters) that need to be defined by the user. Since the quality of SVM regression models depends on a proper setting of these parameters, the main issue for practitioners trying to apply SVM is how to set these parameter values (to ensure good generalization performance) for a given training dataset.

SVM based on PSO optimizes two important hyperparameters C and ϵ using PSO. The hyperparameter C determines the trade-off between the model complexity and the degree to which deviations larger than ϵ are tolerated. A poor choice of C will lead to an imbalance between model complexity minimization (MCM) and empirical risk minimization (ERM). The hyperparameter ϵ controls the width of the ϵ -insensitive zone, and its value affects the number of SVs used to construct the regression function. If ϵ is set too large, the insensitive zone will have ample margin to include data points; this would result in too few SVs selected and lead to unacceptable “flat” regression estimates [15].

2.1.3. Cuckoo Search [16]. Cuckoo search is an optimization algorithm developed by Xin-she Yang and Suash Deb [16–18]. It was inspired by the obligate brood parasitism of some cuckoo species by laying their eggs in the nests of other host birds (of other species). Some host birds can engage direct conflict with the intruding cuckoos. For example, if a host bird discovers that the eggs are not their own, it will either throw these alien eggs away or simply abandon its nest and build a new nest elsewhere. Some cuckoo species such as the New World brood-parasitic *Tapera* have evolved in such a way that female parasitic cuckoos are often very specialized in the mimicry in colors and pattern of the eggs of a few chosen host species [19].

Cuckoo search idealized such breeding behavior and thus can be applied for various optimization problems. It seems that it can outperform other metaheuristic algorithms in applications [20].

Cuckoo search (CS) uses the following representations.

Each egg in a nest represents a solution, and a cuckoo egg represents a new solution. The aim is to use the new and potentially better solutions (cuckoos) to replace a not-so-good solution in the nests. In the simplest form, each nest has one egg. The algorithm can be extended to more complicated cases in which each nest has multiple eggs representing a set of solutions.

```

Objective function:
 $f(x), X = (x_1, x_2, x_3, \dots, x_d)$ 
Generate an initial population of  $n$  host nests;
While ( $t < \text{MaxGeneration}$ ) or (stop criterion)
  Get a cuckoo's nest  $i$  randomly and replace its solution by performing Lévy flights;
  Evaluate its quality/fitness  $F_i$ 
  Choose a nest  $j$  among  $n$  randomly;
  if ( $F_i > F_j$ ),
    Replace  $j$  by the new solution;
  end if
  A fraction ( $P_a$ ) of the worse nests are abandoned and new ones are built;
  Keep the best solutions/nests;
  Rank the solutions/nests and find the current best;
  Pass the current best solutions to the next generation;
end while

```

ALGORITHM 1

CS is based on three idealized rules:

- (1) each cuckoo lays one egg at a time and dumps its egg in a randomly chosen nest;
- (2) the best nests with high quality of eggs will be carried over to the next generation;
- (3) the number of available hosts nests is fixed, and the egg laid by a cuckoo is discovered by the host bird with a probability $P_a \in (0, 1)$. Discovering operate on some set of worst nests, and discovered solutions dumped from farther calculations.

In addition, Yang and Deb discovered that the random-walk style search is better performed by Lévy flights rather than simple random walk.

The pseudocode can be summarized as in Algorithm 1.

An important advantage of this algorithm is its simplicity. In fact, compared with other population- or agent-based metaheuristic algorithms such as particle swarm optimization and harmony search, there is essentially only a single parameter P_a in CS (apart from the population size n). Therefore, it is very easy to be implemented.

2.1.4. CS-PSO-SVM. CS-PSO-SVM consists of two stages: firstly, a CS based approach for parameter optimization of SVM is developed to find the better initial parameters of kernel function and then PSO is applied to continue SVM training and find the best parameters of SVM. CS-PSO-SVM algorithm can be shown as follows in detail.

Step 1. Initializing cuckoo and PSO with population size, inertia weight, generations, and the range of hyperparameters C and ϵ .

Step 2. Applying CS to find the better initial parameters C and ϵ .

Step 3. Evaluating the fitness of each particle.

Step 4. Comparing the fitness values and determining the local best and global best particle.

Step 5. Updating the velocity and position of each particle till the value of the fitness function converges.

Step 6. After converging, the global best particle in the swarm is fed to SVM classifier for training.

Step 7. Training and testing the SVM classifier.

The flowchart of CS-PSO-SVM algorithm is shown in Figure 2 in detail.

2.2. Materials. In this study, numerical experiments use two datasets, heart disease dataset, and breast cancer dataset from UCI Machine Learning Repository [21].

Statlog (heart) dataset has 270 instances. In this dataset, there are thirteen numerical attributes, including age, sex, chest, blood (Rbp), serum (mg/dL), sugar (Fbs), electrocardiographic (ECG), heart rate, angina, old peak, slope, vessels (0-3), and THAL. Number of the presence of heart disease in the patient is 150, and number of the absence of heart disease in the patient is 120. Breast cancer dataset has 699 instances. Number of attributes of each instance is nine. There are thirteen numerical attributes including radius (mean of distances from center to points on the perimeter), texture (standard deviation of gray-scale values), smoothness (local variation in radius lengths), compactness, concavity (severity of concave portions of the contour), concave points (number of concave portions of the contour), symmetry, and fractal dimension. There are 458 as "benign" and 241 as "malignant." In the above two datasets, "1" denotes one people as "benign" and "2" denotes one people as "malignant." Table 1 shows the detail of two datasets. Heart disease dataset and breast cancer dataset choose 190 and 549 instances as train datasets, respectively. The rest of two datasets are test datasets.

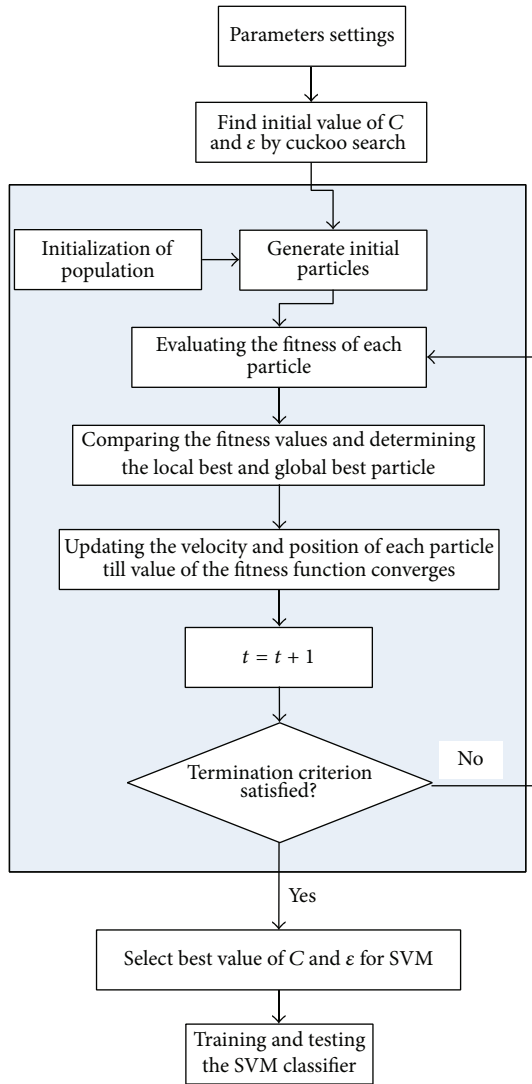


FIGURE 2: The flowchart of CS-PSO-SVM.

TABLE 1: Credit dataset.

Dataset	Number of instances	Number of attributes	Benign	Malignant
Heart disease dataset	270	13	150	120
Breast cancer dataset	699	9	458	241

Continuous attributes are normalized firstly and then used to train and test.

3. Results and Discussion

For the comparison of performance between traditional GA-SVM, PSO-SVM, and CS-PSO-SVM, these models are run several times. The program of the new algorithm is written by MATLAB 2012b and run on a computer with 2.0 GHz CPU, 1 GB DDR RAM. Figures 3 and 5 show the different curves

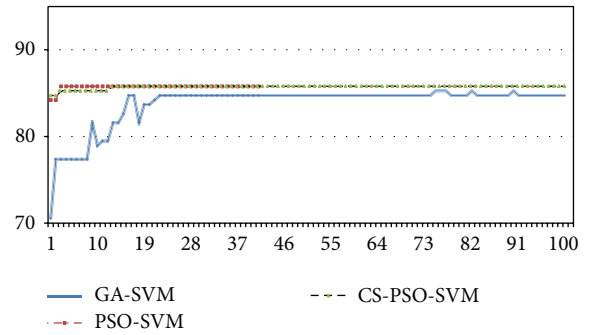


FIGURE 3: Population best fitness value in heart disease dataset.

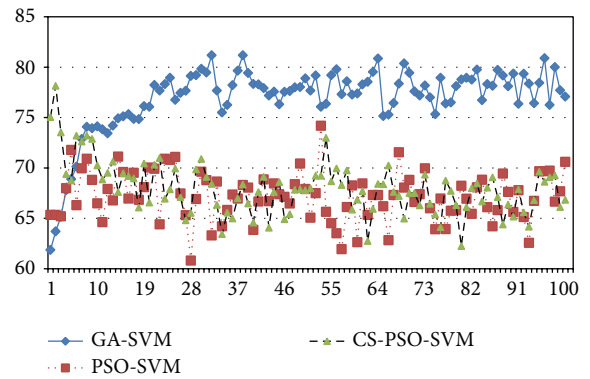


FIGURE 4: Population average fitness value in heart disease dataset.

TABLE 2: (a) Parameters setting of GA-SVM algorithm. (b) Parameters setting of PSO-SVM algorithm. (c) Parameters setting of CS-PSO-SVM algorithm.

(a)					
Parameter	PopSize	Iteration	P_c	P_m	
Value	20	100	0.5	0.005	
(b)					
Parameter	PopSize	Iteration	c_1	c_2	
Value	20	100	1.5	1.7	
(c)					
Parameter	PopSize	Iteration	P_a	PSO- c_1	PSO- c_2
Value	20	100	0.25	1.5	1.7

of population best fitness value in heart disease dataset and breast cancer dataset by GA-SVM, PSO-SVM, and CS-PSO-SVM. Figures 4 and 6 show the different curves of population average fitness value in heart disease dataset and breast cancer dataset by the above three algorithms. Table 2 lists the appropriate values of these parameters in three algorithms.

Table 3 shows the accuracy comparison of GA-SVM, PSO-SVM, and CS-PSO-SVM. For the heart disease dataset, in the test subset, the accuracy of CS-PSO-SVM is 85% and PSO-SVM and GA-SVM are 80%. In the training subset, the accuracy of CS-PSO-SVM is 100%.

TABLE 3: Comparison of models.

Dataset	Algorithm	Forecasting accuracy total in training set	Forecasting accuracy total in test set	Precision	Recall	F-measure
Heart disease dataset	GA-SVM	85.7895%	80%	76.09%	87.50%	81.40%
	PSO-SVM	86%	80%	76.09%	87.50%	81.40%
	CS-PSO-SVM	100%	85%	81.82%	90.00%	85.71%
Breast cancer dataset	GA-SVM	99.8179%	90.0%	92.08%	62.00%	74.10%
	PSO-SVM	97.4499%	90.0%	88.99%	64.67%	74.90%
	CS-PSO-SVM	98.3607%	91.3333%	91.43%	64.00%	75.29%

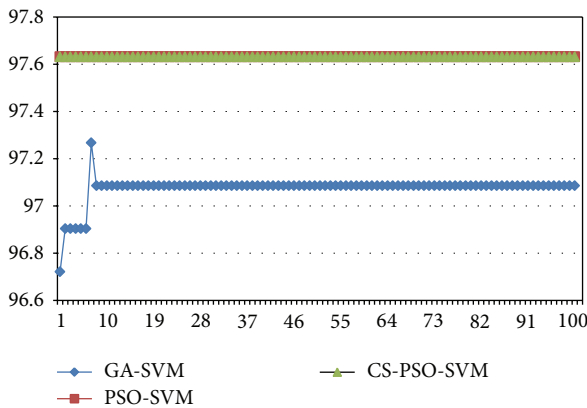


FIGURE 5: Population best fitness value in breast cancer dataset.

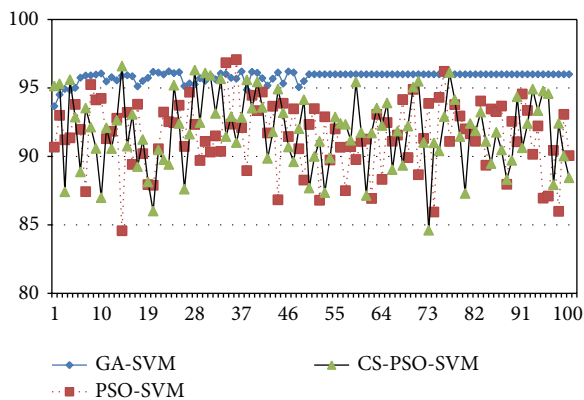


FIGURE 6: Population average fitness value in breast cancer dataset.

For the breast cancer dataset, in the test subset, the accuracy of CS-PSO-SVM is 91.333% and PSO-SVM and GA-SVM are 90%. The results of empirical analysis showed that the predictive ability of all the models is acceptable. However, the CS-PSO-SVM results outperformed the other methods. Therefore, CS-PSO-SVM is a more effective model to predict disease in two datasets.

4. Conclusion

In the last few decades, several disease diagnosis models have been developed for the disease prediction. The objective

of disease diagnosis models is to make a definite diagnosis from patients' laboratory sheet early as soon as possible and initiate timely treatment. Accurate diagnosis has an important meaning for human health. In this paper, we design a new disease diagnosis model, CS-PSO-SVM. The experimental research results show that the novel algorithm is better than GA-SVM and PSO-SVM.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to thank anonymous reviewers for their constructive and enlightening comments, which improved the paper. This work has been supported by Grants from Program for Excellent Youth Scholars in Universities of Guangdong Province (Yq2013108). The authors are partly supported by the Key Grant Project from Guangdong Provincial Party Committee Propaganda Department, China (LLYJ1311).

References

- [1] A. Subasi and M. Gursay, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8659–8666, 2010.
- [2] A. Plaza, J. A. Benediktsson, J. W. Boardman et al., "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment*, vol. 113, no. 1, pp. S110–S122, 2009.
- [3] I. A. Illán, J. M. Górriz, M. M. López et al., "Computer aided diagnosis of Alzheimer's disease using component based SVM," *Applied Soft Computing*, vol. 11, no. 2, pp. 2376–2382, 2011.
- [4] J. S. Sartakhti, M. H. Zangoeei, and K. Mozafari, "Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (SVM-SA)," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 2, pp. 570–579, 2012.
- [5] J. Ramírez, J. M. Górriz, D. Salas-Gonzalez et al., "Computer-aided diagnosis of Alzheimer's type dementia combining support vector machines and discriminant set of features," *Information Sciences*, vol. 237, pp. 59–72, 2013.

- [6] M. J. Abdi and D. Giveki, "Automatic detection of erythematous diseases using PSO-SVM based on association rules," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 1, pp. 603–608, 2013.
- [7] X.-S. Yang and S. Deb, "Cuckoo search: recent advances and applications," *Neural Computing and Applications*, vol. 24, no. 1, pp. 169–174, 2014.
- [8] Tristan Fletcher, "Support Vector Machines Explained," 2014, <http://www.tristanfletcher.co.uk/SVM%20Explained.pdf>.
- [9] A. Reyaz-Ahmed, Y.-Q. Zhang, and R. W. Harrison, "Granular decision tree and evolutionary neural SVM for protein secondary structure prediction," *International Journal of Computational Intelligence Systems*, vol. 2, no. 4, pp. 343–352, 2009.
- [10] W. An, C. Angulo, and Y. Sun, "Support vector regression with interval-input interval-output," *International Journal of Computational Intelligence Systems*, vol. 1, no. 4, pp. 299–303, 2008.
- [11] V. J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, vol. 4, pp. 1942–1948, December 1995.
- [12] J. Kennedy, R. C. Eberhart, and Y. Shi, *Swarm Intelligence*, Morgan Kaufmann, 2002.
- [13] Y. Shi and R. C. Eberhart, "A modified particle swarm optimizer," in *Proceedings of the IEEE Congress on Evolutionary Computation*, pp. 69–73, May 1998.
- [14] F. Ardjani and K. Sadouni, "Optimization of SVM multiclass by particle swarm (PSO-SVM)," *International Journal of Modern Education and Computer Science*, vol. 2, no. 2, pp. 32–38, 2010.
- [15] Y. Ren and G. Bai, "Determination of optimal SVM parameters by using GA/PSO," *Journal of Computers*, vol. 5, no. 8, pp. 1160–1168, 2010.
- [16] http://en.wikipedia.org/wiki/Cuckoo_search.
- [17] I. Fister Jr., D. Fister, and I. Fister, "A comprehensive review of cuckoo search: variants and hybrids," *International Journal of Mathematical Modelling and Numerical Optimisation*, vol. 4, no. 4, pp. 387–409, 2013.
- [18] I. Fister Jr., X.-S. Yang, D. Fister, and I. Fister, "Cuckoo search: a brief literature review," in *Cuckoo Search and Firefly Algorithm*, X.-S. Yang, Ed., vol. 516 of *Studies in Computational Intelligence*, pp. 49–62, 2014.
- [19] R. B. Payne, M. D. Sorenson, and K. Klitz, *The Cuckoos*, Oxford University Press, 2005.
- [20] <http://www.scientificcomputing.com/news-DA-Novel-Cuckoo-Search-Algorithm-Beats-Particle-Swarm-Optimization-060110.aspx>.
- [21] UCI Machine Learning Repository, <http://archive.ics.uci.edu/ml>.