

Research Article

Micro Learning Support Vector Machine for Pattern Classification: A High-Speed Algorithm

Yu Yan,¹ Yiming Wang ,^{1,2} and Yiming Lei ¹

¹School of Economics, Peking University, Haidian District, Beijing 100871, China

²Key Laboratory of Mathematical Economics and Quantitative Finance, Peking University, Haidian District, Beijing 100871, China

Correspondence should be addressed to Yiming Lei; leiy0008@e.ntu.edu.sg

Received 6 June 2022; Revised 30 June 2022; Accepted 7 July 2022; Published 3 August 2022

Academic Editor: Zaher Mundher Yaseen

Copyright © 2022 Yu Yan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The support vector machine theory has been developed into a very mature system at present. The original support vector machine to solve the optimization problem is transformed into a direct calculation formula of line in this paper and the model is $o(n^2)$ time complexity. In the model of this article, weited theory, multiclassification problem and online learning have all become the direct inference, and we have applied the new model to the UCI data set. We hope that in the future, this model will be useful in real-world problems such as stock forecasting, which require nonlinear hi-speed algorithms.

1. Introduction

Since the establishment of the support vector machine ([1–3]) in 1995, Vapnik et al., the support vector machine(SVM) has been the focus of researchers in data mining. The classical SVM processing is a classic binary classification problem, where SVM labels unlabeled points by solving an optimal line. One of the basic principles of SVM is to use kernel technology so that the specific mapping method cannot be known. Throu the simple inner product of the nonlinear problem in the original space, we can solve this problem with a linear problem in another space that is mapped to. Moreover, the model can significantly improve the time complexity of the small sample problems throu dual theory, and one of the classic model thinking: the maximum interval is also widely used in various models¹.

In 2007, Jayadeva et al. ([4–7]) established the model as the twin support vector machine (TWSVM), which solves the classic two classification problem. Unlike SVM, TWSVM is mainly used to solve nonparallel problems. The support surface of the SVM is two parallel hyperplanes, and TWSVM is the solution to the two nonparallel hyperplanes. This model no longer uses the maximum interval principle in SVM. TWSVM solves for two straight lines that are as close

as possible to the two classes of points. Classification is performed by determining which line a new point is close to.

SVM has a time complexity of $o(n^3 + d)$ when the number of samples is n and the number of characteristics is d , making SVM very suitable for solving hi-dimensional small sample problems. Some other algorithms are also suitable for small samples, but they have mutual advantages and disadvantages with SVM². The kernel technique also makes SVM very suitable for solving nonlinear problems and hi-dimensional problems. There are also some algorithms that are suitable for hi-dimensional problem, but they have mutual advantages and disadvantages with SVM³. However, a common problem with a range of algorithms based on SVM is the inability to solve large-sample problems due to the limitations of the optimization algorithm. Therefore, we want to provide algorithm that maintains the good properties of SVM while reducing the time required to solve large sample sizes. We give a nonoptimal SVM model with a time complexity of $o(n^2 + d)$. This improvement will go some way towards circumventing the problem that SVM cannot be applied to large-scale data. Our model can be applied to many hi-dimensional or large sample problems⁴ and has substantial implications for solving real-

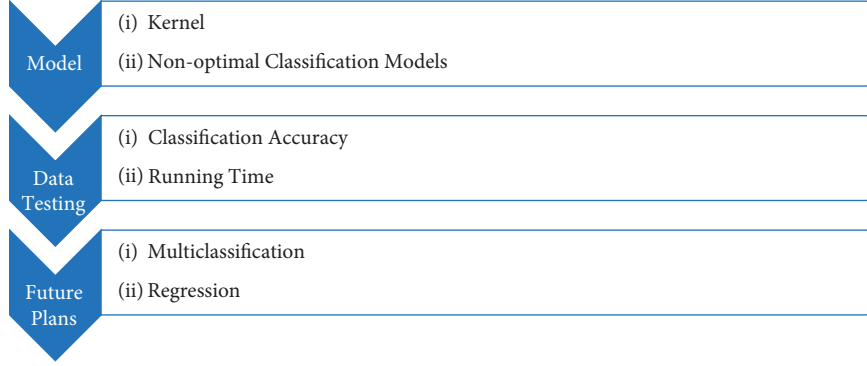


FIGURE 1: Details of this paper.

world problems with hi-dimensional, large samples, and hi-time demands.

In this paper, we consider a kind of nonoptimal machine learning model from the point of view that there is only one positive and one negative. The model takes a point from the positive and negative points to train a model, and then uses all the combinations of positive and negative points. Finally, the model is considered by using multiple models.

The basic logic of the model is that it can be used to train several classification models. In this case, we construct a classification model which can be used in kernel technology. It is interesting to note that in this model, the problems of machine learning, such as the classification problem, the weighted problem, and the fitting problem, will be straightforward to operate.

The details of our research are shown in Figure 1.

2. Classical Model

Consider the classic two classification problem: given training set $(x_i, y_i) \in \mathbb{R}^d \times \{-1, 1\}, i = 1, 2, 3, \dots, n$ where y_i is the label, and we have to look for the decision function $f(x)$ to infer any new input x corresponding output y . In order to facilitate the representation, we use the following notation: A represents a data set of positive class points, and B represents the data set of negative class points.

First, we review the classical linear SVM model. The model aims to establish a straight line between positive and negative, two types of samples. One of the principles of SVM is the principle of maximum distance, that is, to maximize the distance between two support planes. We assume that the dividing surface is $wx + b = 0$, and the two support surfaces are $wx + b = 1$ and $wx + b = -1$. The problem of solving the problem of SVM is changed into the following optimization problem:

$$\begin{aligned} \min \frac{1}{2} \|w\|^2, \\ \text{s.t. } y_i (wx_i + b) \geq 1. \end{aligned} \quad (1)$$

On the other hand, the linear optimization problem is as follows:

$$\begin{aligned} \min \frac{1}{2} \|w\|^2, \\ \text{s.t. } y_i (wx_i + b) \geq 1, \\ q_i \geq 0. \end{aligned} \quad (2)$$

Then, we consider the kernel technology in the dual problem and transform $x_i x_j$ into $K(x_i, x_j)$. The Euclidean space is mapped into another space, and the nonlinear problem is transformed into a linearly separable problem in a hi-dimensional space.

We look back on another machine learning algorithm: the TWSVM. The TWSVM focuses on solving the non-parallel problem. Two types of sample points are enriched near the two parallel lines. The model aims to find two nonparallel straight lines, which can be used to determine the type of line in the classification. The optimization problem of the model is as follows:

$$\begin{aligned} (\text{TWSVM1}) \min \frac{1}{2} (Aw_1 + b_1)^T (Aw_1 + b_1) + c_1 e_2^T b q_1 \\ \text{s.t. } -(Bw_1 + b_1)^T + q \geq e_2, q \geq 0, \\ (\text{TWSVM2}) \min \frac{1}{2} (Bw_2 + b_2)^T (Bw_2 + b_2) + c_2 e_1^T b q_2 \\ \text{s.t. } (Aw_2 + b_2)^T + q \geq e_1, q \geq 0. \end{aligned} \quad (3)$$

The dual problem of the model is as follows:

$$\begin{aligned} (\text{DTWSVM1}) \max e_2^T \alpha - \frac{1}{2\alpha^T} G (H^T H) G^T \alpha \\ \text{s.t. } 0 \leq \alpha \leq c_1, \\ (\text{DTWSVM2}) \max e_1^T \gamma - \frac{1}{2\gamma^T} H (G^T G) H^T \gamma \\ \text{s.t. } 0 \leq \gamma \leq c_2, \end{aligned} \quad (4)$$

where $H = [A, e1]$ and $G = [B, e2]$.

In order to introduce kernel technology, we consider replacing the two straight lines $xw^1 + b^1 = 0$ and $xw^2 + b^2 = 0$ (X is the sum of A and B).

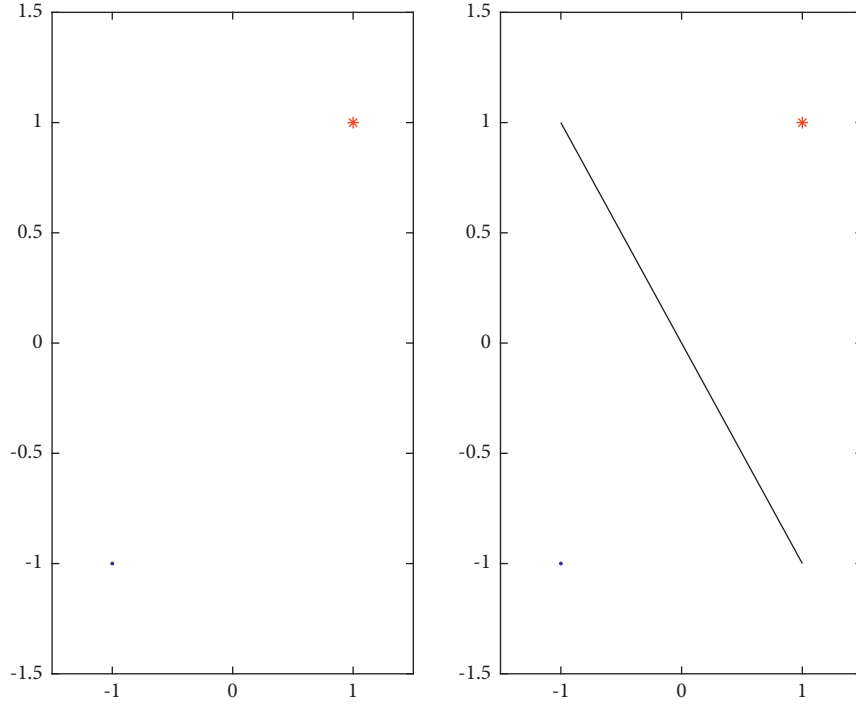


FIGURE 2: Two-point classification.

$$\begin{aligned} k(x, X)w_1 + b_1 &= 0, \\ k(x, X)w_2 + b_2 &= 0. \end{aligned} \quad (5)$$

The dual problem is as follows:

$$\begin{aligned} (\text{KDTWSVM1}) \text{Max}_{\alpha} & \frac{1}{2\alpha^T} R(S^T S) R^T \alpha \\ \text{s.t. } & 0 \leq \alpha \leq c_1, \\ (\text{KDTWSVM2}) \text{Max}_{\gamma} & \frac{1}{2\gamma^T} S(R^T R) S^T \gamma \end{aligned} \quad (6)$$

$$\text{s.t. } 0 \leq \gamma \leq c_2,$$

where $S = [k(A, X^T), e1]$ and $R = [k(B, X^T), e2]$;

3. New Model

Firstly, we consider the process of learning. If we only have a sample point, for example, if our problem is how to determine whether a person is male or female and the training set is just a lady in the picture. Then, we cannot judge another picture of the characters in the male and female. It is difficult to classification when we have only one class of points. By the same token, even thou our training focuses on ten thousand women's photographs, without a single photo of men, it is still unable to train a model that can distinguish between men and women. It is difficult for us to compare the difference between a man and ten thousand women in our normal human thinking, and we can only compare the difference between a male and a female.

Therefore, we consider a positive point and a negative point. The training set has only one positive and one negative point. Using the idea of the maximum interval of SVM, we can obtain that the optimal line is the two points of the vertical bisector of the line, and the functional distance between the two points and a strait line is 1 in Figure 2.

Obviously, we can get the dividing line as follows:

$$\frac{2(x^+ - x^-)x}{(x^+ - x^-)^2} - \frac{(x^+)^2 - (x^-)^2}{(x^+ - x^-)^2} = 0. \quad (7)$$

Then, we consider a very interesting classification problem, as shown in figure two in Figure 3.

From Figure 3, we can see that each point in the positive point and negative point in the class make the points of the line, and finally, the combination of points is a reasonable way. So we consider the following algorithm.

Then, we consider the general situation, that is, the number of positive and negative points (in order to facilitate the consideration, we assume that there are M positive points and N negative points). Take a positive point and a negative class point, and we can get its vertical bisector.

$$\frac{2(x^+ - x^-)x}{(x^+ - x^-)^2} - \frac{(x^+)^2 - (x^-)^2}{(x^+ - x^-)^2} = 0. \quad (8)$$

We consider the calculation of all the points, and then use each of the subline to consider the classification problem.

The core idea of this algorithm is to take each positive point and each negative class point out to build a subline, and then all the points out of an average. Take the positive and negative values as the classification results and we consider the classification results are as follows:

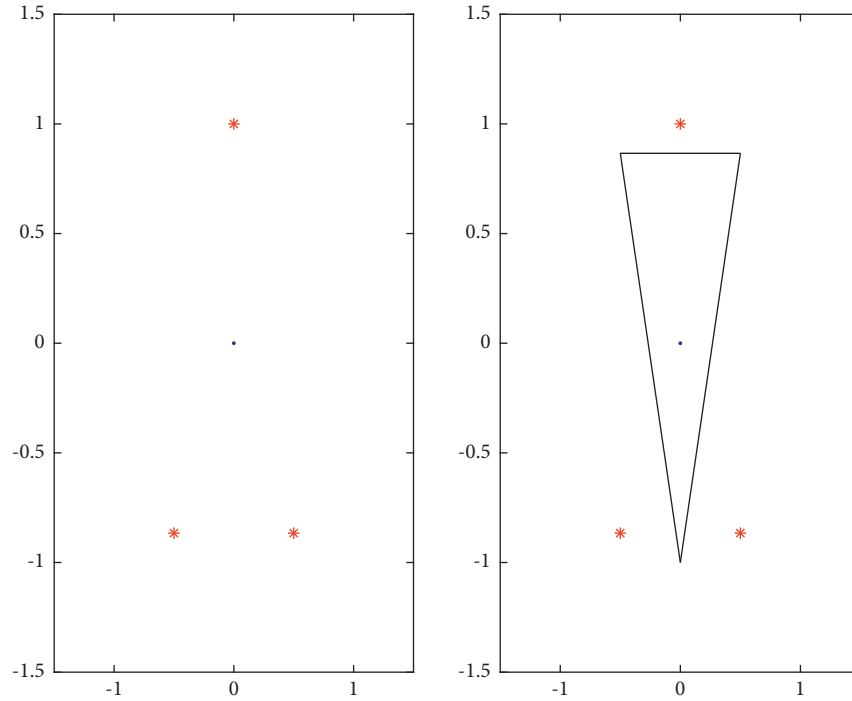


FIGURE 3: Some-point classification.

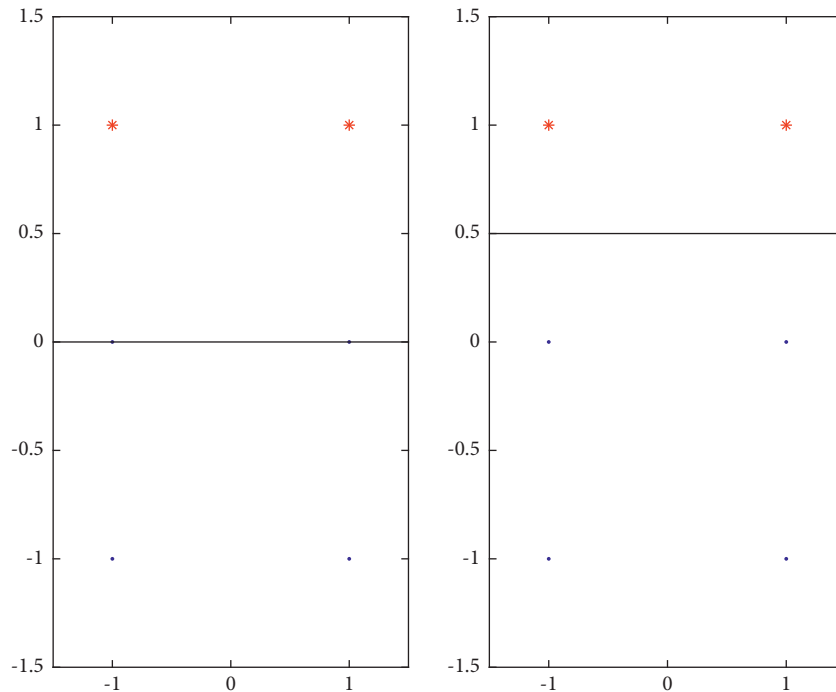


FIGURE 4: Change of Cmn.

$$\text{sign} \left(\sum_{i=1}^m \sum_{j=1}^n \frac{2(x^i - x^j)x}{(x^i - x^j)^2} - \frac{(x^i)^2 - (x^j)^2}{(x^i - x^j)^2} \right). \quad (9)$$

After we discuss an improvement of the model, we consider the following sample points in Figure 4.

In the training of the sample points, if we consider the model we calculate, it will lead to the left side of the figure, resulting in the training model is not reasonable. The foundation of our model is the two point training division. We consider the linear translation, namely, the introduction of a parameter C in $(-1, 1)$ two, making the line training as follows:

TABLE 1: Linear.

	NEW		SVM		TWSVM	
Australian	0.816901	0.083721	0.517857	0.011664	0.624101	0.129945
BUPA	0.571429	0.156057	0.65942	0.082409	0.082143	0.102602
Diabetes	0.508	0.470447	0.510552	0.093228	0.67013	0.136054
Heart disease	0.362712	0.501019	0.762712	0.077671	0.450847	0.481994
Heartstatlog	0.486364	0.08046	0.771605	0.07011	0.75	0.067215
Herman	0.73871	0.082558	0.802419	0.046327	0.73871	0.082558
Sonar	0.661905	0.221057	0.492063	0.190972	0.528571	0.227502
Teaching	0.683871	0.097844	0.741935	0	0.825806	0.131825
Balance	0.889655	0.119824	0.479885	0.099444	0.463793	0.15534
Breast	0.345985	0.113855	0.466111	0.054133	0	0

Bold indicates best.

TABLE 2: RBF.

	NEW		SVM		TWSVM	
Australian	0.852113	0.062461	0.864286	0.018443	0.848921	0.044736
BUPA	0.642857	0.08165	0.572464	0.062616	0.610714	0.055174
Diabetes	0.748	0.230911	0.698052	0.070361	0.67013	0.070895
Heart disease	0.816949	0.067158	0.762712	0.070565	0.674576	0.437833
Heartstatlog	0.836364	0.039277	0.771605	0.032075	0.759091	0.044536
Herman	0.422581	0.218011	0.802419	0.046327	0.680645	0.082558
Sonar	0.590476	0.234134	0.555556	0.062994	0.528571	0.493265
Teaching	0.729032	0.11081	0.790323	0.06843	0.825806	0.195687
Balance	0.967241	0.030111	0.817529	0.022314	0.815517	0.087618
Breast	0.963504	0.031817	0.959333	0.032331	0.416058	0.183863

Bold indicates best.

TABLE 3: Time of linear.

Number	New	SVM	TWSVM
100	0.031	0.109	0.312
200	0.035	1.257	0.381
300	0.052	2.606	0.518
400	0.055	4.25	0.801
500	0.293	8.371	0.864
600	0.73	12.32	1.076
700	1.169	15.672	1.745
800	1.465	24.623	2.599
900	1.871	26.34	3.695
1000	2.274	33.828	4.225

TABLE 4: Time of RBF.

Number	New	SVM	TWSVM
100	0.099	0.473	0.501
200	0.352	1.511	0.973
300	0.743	3.114	1.927
400	1.2	5.611	2.943
500	1.996	9.308	4.484
600	2.968	13.462	6.642
700	4.224	27.574	8.26
800	8.507	32.041	11.58
900	20.755	38.874	24.928
1000	23.47	47.883	28.152

$$\text{sign} \left(\sum_{i=1}^m \sum_{j=1}^n \frac{2(x^i - x^j)x}{(x^i - x^j)^2} - \frac{(x^i)^2 - (x^j)^2}{(x^i - x^j)^2} + \text{Cmn} \right). \quad (10)$$

Then, we introduce the nucleus to each line. Since our model is only composed of the inner product of two vectors, we can use $K(x, y)$ to replace the xy , which can be obtained as follows:

$$\text{sign} \left(\sum_{i=1}^m \sum_{j=1}^n \frac{2(K(x^i, x) - K(x^j, x))}{K((x^i - x^j), (x^i - x^j))} - \frac{K(x^i, x^i) - K(x^j, x^j)}{K(x^i - x^j, x^i - x^j)} + \text{Cmn} \right). \quad (11)$$

It is evident that the time complexity of the model is $O(n^2)$. Because we have a large number of points to get the average score, the weighted sum of the sample points is the direct inference of the model. Similarly, if we want to get an online learning model or give up some of the sample points, the time complexity will be very low. Under this premise, the training complexity of the multiclassification problem is also very low.

4. Data Testing

First, we compute the linear kernel on the UCI dataset, and the accuracy and variance contrast is shown in Table 1.

We can see that the new model has a good advantage, and then we consider the computation of the nonlinear RBF kernel on the UCI data set in Table 2.

In the theory, we show that the new algorithm has strong superiority in time complexity. Here, we do some experiments to count. It can be seen that there are great advantages of some data sets. Then, we calculate the computation time of the linear kernel and the nonlinear kernel on the different number of data sets in Table 3.

In addition to the analysis of the time complexity of the linear problem, we then count the time of the nonlinear case. Then, we apply the nonlinear kernel in Table 4.

The time complexity of our model is $o(n^2 + d)$. The time complexity of the SVM and the TWSVM is $o(n^3 + d)$. Based on Tables 3 and 4, our model is still faster than the SVM and the TWSVM, even in small sample problems with less than 1000 samples. Also, since our algorithm does not need to solve the optimization problem, the time of our algorithm is stable with respect to the growth of the number of samples. It can be expected that in large-scale samples, our algorithm will be much faster than SVM and TWSVM in computation.

5. Conclusion

Based on the point-to-point model, this paper establishes a micro learning support vector machine. The model is different from the traditional SVM in the way of solving it, and it is not necessary to solve the optimization problem. This makes the model have some difference with the traditional machine learning algorithm. Both neural networks and SVM are unknown time to calculate. The microlearning support vector machine is in a fixed time when the length of the orientation and the number of sample points are fixed. This is of great benefit to the stability of our design and practical applications. From the view of time complexity, the algorithm is better than SVM. Extending the micro learning support vector machine to weighted problems, multiclassification problems, and fitting problems is very simple and straightforward.

Our algorithm outperforms both SVM and TSVM in terms of model accuracy and computation time, and it also has good nonlinear generalisation due to the fact that we also

use the kernel function. The computation time of this algorithm is explicit because it does not require solving an optimization problem. Overall, this algorithm is well suited to problems such as stock prediction and face recognition, which require nonlinear, hi-dimensional data, and hi computational speed.

Based on [8, 9], we can extend the model to a semi-supervised problem in the future. We just have not come up with a suitable modelling idea yet. We believe that the ideas used to build our model can also be extended to the field of feature extraction in the future. And it can be applied to many related problems (e. g. [10, 11]).

We likewise believe that our model can be used to solve problems related to regression after a SVM-to-SVR-like transformation (e. g. [12–17]). Of particular interest is the fact that our algorithms are well suited for applications in the field of financial forecasting (e. g. [18–20]). The field of financial forecasting requires algorithms with controllable computation times and good performance for nonlinear problems.

In the same way as SVM, our algorithm can be used to solve multiclassification problems. We hope that other researchers will apply our algorithm to multiclassification-related problems in the future (e. g. [21–24]). It is worth noting that this algorithm can be used for face recognition (e. g. [25, 26]). Similarly, our model can be used to solve the multilabel problem (e. g. [27, 28]).

Data Availability

Data are from the UCI dataset.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

The authors approved for the submission.

References

- [1] P. S. Bradley and O. L. Mangasarian, "Massive data discrimination via linear support vector machines," *Optimization Methods and Software*, vol. 13, no. 1, pp. 1–10, 2000.
- [2] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
- [3] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [4] R. Khemchandani and S. Chandra, "Twin support vector machines for pattern classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 5, pp. 905–910, 2007.
- [5] Z. Qi, Y. Tian, and Y. Shi, "Robust twin support vector machine for pattern classification," *Pattern Recognition*, vol. 46, no. 1, pp. 305–316, 2013.
- [6] Y. H. Shao, C. H. Zhang, X. B. Wang, and N. Y. Deng, "Improvements on twin support vector machines," *IEEE Transactions on Neural Networks*, vol. 22, no. 6, pp. 962–968, 2011.

- [7] Y. H. Shao, N. Y. Deng, and Z. M. Yang, "Least squares recursive projection twin support vector machine for classification," *Pattern Recognition*, vol. 45, no. 6, pp. 2299–2307, 2012.
- [8] K. Chen, L. Yao, D. Zhang, X. Wang, X. Chang, and F. Nie, "A semisupervised recurrent convolutional attention model for human activity recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 5, pp. 1747–1756, 2020.
- [9] M. Luo, X. Chang, L. Nie, Y. Yang, A. G. Hauptmann, and Q. Zheng, "An adaptive semisupervised feature analysis for video semantic recognition," *IEEE Transactions on Cybernetics*, vol. 48, no. 2, pp. 648–660, 2018.
- [10] A. Glowacz, "Thermographic fault diagnosis of ventilation in bldc motors," *Sensors*, vol. 21, no. 21, p. 7245, 2021.
- [11] A. Glowacz, "Ventilation diagnosis of angle grinder using thermal imaging," *Sensors*, vol. 21, no. 8, p. 2853, 2021.
- [12] Q. Bo and W. Cheng, "Intelligent control of agricultural irrigation throu water demand prediction based on artificial neural network," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 7414949, 10 pages, 2021.
- [13] Q. Liang, "Application of convolution neural network (cnn) model combined with pyramid algorithm in aerobics action recognition," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 6170070, 2110 pages, 2021.
- [14] S. Mishra, T. Ahmed, V. Mishra et al., "Multivariate and online prediction of closing price using kernel adaptive filtering," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 6400045, 2114 pages, 2021.
- [15] Z. M. Yaseen, H. Faris, and N. A. Ansari, "Hybridized extreme learning machine model with salp swarm algorithm: a novel predictive model for hydrological application," *Complexity*, vol. 2020, Article ID 8206245, 14 pages, 2020.
- [16] T. Tiyasha, T. M. Tung, and Z. M. Yaseen, "A survey on river water quality modelling using artificial intelligence models: 2000–2020," *Journal of Hydrology*, vol. 585, Article ID 124670, 2020.
- [17] M. Zhu and Z. Meng, "Macroeconomic image analysis and gdp prediction based on the genetic algorithm radial basis function neural network (rbfn-ga)," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 2000159, 2110 pages, 2021.
- [18] B. Madhu, M. A. Rahman, A. Mukherjee, M. Z. Islam, R. Roy, and L. E. Ali, "A comparative study of support vector machine and artificial neural network for option price prediction," *Journal of Computer and Communications*, vol. 9, no. 5, pp. 78–91, 2021.
- [19] C. Xiao, W. Xia, and J. Jiang, "Stock price forecast based on combined model of ari-ma-ls-svm," *Neural Computing & Applications*, vol. 32, no. 10, pp. 5379–5388, 2020.
- [20] R. Yang, L. Yu, Y. Zhao et al., "Big data analytics for financial market volatility forecast based on support vector machine," *International Journal of Information Management*, vol. 50, pp. 452–462, 2020.
- [21] S. Kang, I. Kim, and P. J. Vikesland, "Discriminatory detection of ssdna by surface-enhanced Raman spectroscopy (sers) and tree-based support vector machine (tr-svm)," *Analytical Chemistry*, vol. 93, no. 27, pp. 9319–9328, 2021.
- [22] A. R. Mello, M. R. Stemmer, and A. L. Koerich, "Incremental and decremental fuzzy bounded twin support vector machine," *Information Sciences*, vol. 526, pp. 20–38, 2020.
- [23] D. Pradhan, B. Sahoo, B. B. Misra, and S. Padhy, "A multiclass svm classifier with teaching learning based feature subset selection for enzyme subclass classification," *Applied Soft Computing*, vol. 96, Article ID 106664, 2020.
- [24] Y. X. Xie, Y. J. Yan, G. F. Li, and X. Li, "Scintillation detector fault diagnosis based on wavelet packet analysis and multi-classification support vector machine," *Journal of Instrumentation*, vol. 15, no. 3, Article ID T03001, 2020.
- [25] J. A. A. Ahmed, X. Zhu, and M. Alaili, "Identifying Ethnicities of People throu Face Recognition: A Deep Cnn Approach," *Scientific Programming*, vol. 2020, Article ID 6385281, 2020.
- [26] J. Yang and H. Gao, "Cultural emperor penguin optimizer and its application for face recognition," *Mathematical Problems in Engineering*, vol. 2020, Article ID 9579538, 16 pages, 2020.
- [27] H. Wang and Y. Xu, "Sparse elastic net multi-label rank support vector machine with pinball loss and its applications," *Applied Soft Computing*, vol. 104, Article ID 107232, 2021.
- [28] Y. Zhang, Y. Xu, C. Xu, and P. Zhong, "Safe instance screening for primal multi-label prosvm," *Knowledge-Based Systems*, vol. 229, Article ID 107362, 2021.
- [29] D. Liu, Y. Shi, Y. Tian, and X. Huang, "Ramp loss least squares support vector machine," *Journal of computational science*, vol. 14, pp. 61–68, 2016.
- [30] L. Wang, H. Jia, and J. Li, "Training robust support vector machine with smooth ramp loss in the primal space," *Neurocomputing*, vol. 71, no. 13–15, pp. 3020–3025, 2008.
- [31] D. Zhang, L. Yao, K. Chen, S. Wang, X. Chang, and Y. Liu, "Making sense of spatio-temporal preserving representations for eeg-based human intention recognition," *IEEE Transactions on Cybernetics*, vol. 50, no. 7, pp. 3033–3044, 2020.