Journal of Information Hiding and Multimedia Signal Processing ©2021 ISSN 2073-4212
Ubiquitous International Volume **12**, Number **2**, June **2021**

Wavelet Kernel Twin Support Vector Machine

Qing Wu
School of Automation
Xi'an University of Posts and Telecommunications
Xi'an, Shaanxi, 710121, China
xiyouwuq@126.com

Bo-yan Zang
School of Automation
Xi'an University of Posts and Telecommunications
Xi'an, Shaanxi, 710121, China
boyanzang@126.com

Yu Zhang
School of Automation
Xi'an University of Posts and Telecommunications
Xi'an, Shaanxi, 710121, China
543454272@qq.com

Zong-xian Qi
School of Automation
Xi'an University of Posts and Telecommunications
Xi'an, Shaanxi, 710121, China
799432426@qq.com.

Received December 2018; revised February 2021

Abstract. *Twin support vector machine (TWSVM) is faster than standard support vector machine (SVM), and least square twin support vector machine (LSTSVM) further improve training speed. However, the previous works fail to achieve the robustness, especially for high dimensional heterogeneous data. To improve it, wavelet kernel TWSVM is introduced by combining wavelet kernel with TWSVM and LSTSVM. It keeps the advantages of TWSVM, LSTSVM and wavelet kernel, such as high training speed, approximating arbitrary nonlinear functions. Additionally, it achieves a good trade-off between robustness and high training speed. The theoretical analyses and experimental results show that wavelet kernel TWSVM has better performance than those existing works.*
Keywords: twin support vector machine, least square, nonlinear, kernel function, wavelet kernel.

1. **Introduction.** Support vector machine(SVM) proposed by Cortes and Vapnik [1] is one of the most popular machine learning algorithms based on structural risk minimization guidelines. SVM shows many unique advantages in solving small sample, nonlinear and high dimensional pattern recognition problems. Combining with other algorithms like deep learning, colony algorithm, hybrid kernel function, SVM has been applied in many fields, such as image recognition [2], image retrieval [3], network intrusion detection [4], interference classification [5] and so on. Twin support vector machine (TWSVM) was firstly proposed based on GEPSVM [6] by Jayadeva et al for binary classification [7]. TWSVM

generates two nonparallel planes such that each plane is close to one of two classes and as far as possible from the other. It is implemented by solving two smaller quadratic programming problems (QPPs) rather than a single large QPP, which makes the learning speed of TWSVM faster than the classical SVM. Now TWSVM and the improved TWSVM have been applied in many aspects, such as data recognition [8], function regression [9] and vehicle recognition [10] etc. In TWSVM, the inequality constraints are transformed into equality constraints, then least square twin support vector machine(LSTSVM) was put forward firstly by Xie [11], which has faster training speed than TWSVM.

For nonlinear situation, TWSVM uses kernel function mapping low dimensional data to higher dimensional space. There are some frequently-used kernel functions like polynomial kernel, Guassian kernel, RBF kernel and so on. Due to the fact that the wavelet technique shows promise for both non-stationary signal approximation and classification [12], it is valuable for us to study the problem of combination about wavelet technique and TWSVM.

Motivated by ideas and principles from multi-resolution and wavelet theory [13], we present a wavelet twin support vector machine(WTWSVM) and least square wavelet twin support vector machine(LSWTSVM) in this paper. Both of them have good classification performance since the wavelet kernel function can approximate arbitrarily a nonlinear function. The theoretical analyses and experimental results show the feasibility and validity of WTWSVM and LSWTSVM in classification.

2. **Wavelet analysis and wavelet kernel.** Now many scholars pay their attention to multi-kernel learning for its superior performance in multi-view learning since many kinds of information from multiple views can easily be combined. Wavelet decomposition emerges as a powerful tool for approximation [12, 13, 14], which means that the wavelet function is a set of bases that can almost approximate arbitrary function. Here, the wavelet kernel has the same expression as the multidimensional wavelet function.

Based on wavelet theory, any signal can be approximately expressed by a family of functions generated by dilations and translation of function called the mother wavelet,

$$h_{a,c}\left(x\right) = \left|a\right|^{-1/2} h\left(\frac{x-c}{a}\right) \tag{1}$$

where $a, c, x \in R$, a is dilation factor, and c is translation factor. Therefore the wavelet transform of the function $f\left(x\right) \in L_2\left(R\right)$ is written as:

$$W_{a,c}(f) = \langle f(x), h_{a,c}\left(x\right)\rangle \tag{2}$$

where $< \cdot, \cdot >$ denotes the dot product in $L_2\left(R\right)$. Equation (2) means that $W_{a,c}\left(f\right)$ is the decomposition of the function $f\left(x\right)$ on wavelet basis $h_{a,c}\left(x\right)$. Here the mother wavelet function is necessary to satisfy the following condition,

$$W_h = \int_0^{\infty} \frac{\left|H\left(\omega\right)\right|}{\left|\omega\right|} d\omega < \infty \tag{3}$$

where $H\left(\omega\right)$ is the Fourier transform of the mother function $h\left(\omega\right)$. So we can reconstruct the function $f\left(x\right)$ as follows:

$$f\left(x\right) = \frac{1}{W_h} \int_{-\infty}^{\infty} \int_0^{\infty} 1/a^2 W_{a,c}\left(f\right) h_{a,c}\left(x\right) da dc \tag{4}$$

If we take the final term to approximate (4), then:

$$\widehat{f}\left(x\right) = \sum_{i=1}^{l} W_i h_{a_i,c_i}\left(x\right) \tag{5}$$

where $\widehat{f}(x)$ is an approximation of $f(x)$.

For a common multidimensional wavelet function, we can write it as the product of one-dimensional (1-D for short) wavelet functions:

$$h(X) = \prod_{i=1}^{N} h(x_i) \tag{6}$$

As a kernel function, the wavelet kernel must obey the Mercer theorem [15].

Theorem 2.1. *Φ is a map from Euclidean space R^N to Hilbert space H:*

$$\Phi : R^N \to H \tag{7}$$

and the function $K(x_1, x_2)$ must obey the equation as follows which could be called kernel function

$$K(x_1, x_2) = \langle \Phi(x_1) \cdot \Phi(x_2) \rangle \tag{8}$$

where $\Phi(x)$ is a mapping function, and $< \cdot >$ is the inner product in Hilbert space H.

All wavelet kernels also obey the following theorems [15].

Theorem 2.2. *Let $h(x)$ be a mother wavelet, and let a and c denote the dilation and translation respectively. Here $x, a, c \in R$, and if $X, X' \notin R^N$, the dot-product kernels are set to*

$$K(X, X') = \prod_{I=1}^{N} h\left(\frac{x_i - c_i}{a}\right) h\left(\frac{x'_i - c'_i}{a}\right) \tag{9}$$

and translation-invariant wavelet kernels satisfying the translation invariant kernel theorem are set to

$$K(X, X') = \prod_{i=1}^{N} h\left(\frac{x_i - x'_i}{a}\right) \tag{10}$$

3. **Twin support vector machine.** To improve training speed of SVM, Javadeva et.al proposed twin support vector machine (TWSVM) inspired by GEPSVM [7]. For a binary classification problem, the goal of TWSVM is to find a pair of nonparallel hyperplanes. Suppose that data points belonging to positive class denoted by $A_1 \in R^{m_1 \times n}$, where each row $A_i \in R^n$ represents a data point. Similarly, $A_2 \in R^{m_2 \times n}$ represents all of negative points.

3.1. **Linear twin support vector machine.** For the linear case, the two nonparallel hyperplanes are generated by TWSVM as follows:

$$\begin{cases} f_+(x) = w_1^T x + b_1 = 0 \\ f_-(x) = w_2^T x + b_2 = 0 \end{cases} \tag{11}$$

where $w_1, w_2 \in R^n$, $b_1, b_2 \in R$. The TWSVM seeks two nonparallel hyperplanes (11) such that each hyperplane is closer to one of the two classes and as far as possible from the other [6]. The distance of a point from two hyperplanes can determine that a data point belongs to negative class or positive class. Formally, the TWSVM can be described as the following QPPs.

$$\begin{array}{ll} \min & \frac{1}{2}(Aw_1 + b_1)^T (Aw_1 + b_1) + c_1 e_2^T \xi_1 \\ s.t. & -(Bw_1 + b_1) + \xi_1 \geq e_2, \xi_1 \geq 0 \end{array} \tag{12}$$

$$\begin{array}{ll} \min & \frac{1}{2}(Bw_2 + b_2)^T (Bw_2 + b_2) + c_2 e_1^T \xi_2 \\ s.t. & -(Aw_2 + b_2) + \xi_2 \geq e_1, \xi_2 \geq 0 \end{array} \tag{13}$$

where $c_1, c_2 > 0$ are the pre-specified penalty factors, and e_1, e_2 are vectors of ones of appropriate dimensions. By introducing Lagrangian multipliers α and β, we construct the Lagrange function as follows.

$$\begin{array}{c} L(\omega_1, b_1, \xi_1, \alpha, \beta) = \frac{1}{2}(A\omega_1 + e_1 b_1)^T (A\omega_1 + e_1 b_1) + c_1 e_2^T \xi_1 \\ +\alpha^T (B\omega_1 + e_2 b_1 - \xi_1 + e_2) - \beta^T \xi_1 \end{array} \tag{14}$$

Then we can get the following Karush-Kuhn-Tucher conditions:

$$A^T (A\omega_1 + e_1 b_1) + B^T \alpha = 0 \tag{15}$$

$$e_1^T (A\omega_1 + e_1 b_1) + e_2^T \alpha = 0 \tag{16}$$

$$c_1 e_2 - \alpha - \beta = 0 \tag{17}$$

$$-(B\omega_1 + e_2 b_1) + \xi_1 \geq e_2, \xi_1 \geq 0 \tag{18}$$

$$\alpha^T (B\omega_1 + e_2 b_1 - \xi_1 - e_2) = 0, \beta^T \xi_1 = 0 \tag{19}$$

$$\alpha \geq 0, \beta \geq 0 \tag{20}$$

Combing equation (5) with (6), we can obtain:

$$\begin{bmatrix} A^T & e_1^T \end{bmatrix} \begin{bmatrix} A & e_1 \end{bmatrix} \begin{bmatrix} \omega_1 & b_1 \end{bmatrix}^T + \begin{bmatrix} B^T & e_2^T \end{bmatrix} \alpha = 0 \tag{21}$$

Define $H = \begin{bmatrix} A & e_1 \end{bmatrix}$ and $G = \begin{bmatrix} B & e_2 \end{bmatrix}$. Now, the Wolfe dual of the QPPs can be described respectively as follows:

$$\begin{array}{ll} \max & e_2^T \alpha - \frac{1}{2}\alpha^T G \left(H^T H\right)^{-1} G^T \alpha \\ s.t. & 0 \leq \alpha \leq c_1 e_2 \end{array} \tag{22}$$

$$\begin{array}{ll} \max & e_1^T \beta - \frac{1}{2}\beta^T H \left(G^T G\right) H^T \beta \\ s.t. & 0 \leq \beta \leq c_2 e_1 \end{array} \tag{23}$$

where $G = \begin{bmatrix} B & e_2 \end{bmatrix}$, $H = \begin{bmatrix} A & e_1 \end{bmatrix}$, $\alpha = R^{m_2}$ and $\beta = R^{m_1}$, are Lagrangian multipliers.

By defining $v_1 = [w_1, b_1]$ and $v_2 = [w_2, b_2]$, we can obtain the following results.

$$v_1 = -\left(H^T H\right)^{-1} G^T \alpha \tag{24}$$

$$v_2 = \left(G^T G\right)^{-1} H^T \beta \tag{25}$$

The new sample point is assigned to positive class or negative class depending on the function (26).

$$i = \arg \min_{k=1,2} \frac{\left|w_k^T x + b_k\right|}{\|w_k\|} \tag{26}$$

3.2. **Nonlinear twin support vector machine.** For the nonlinear case, the TWSVM uses kernel function to map training data from low dimensional input space to high dimensional space like SVM. Two nonparallel hyperplanes are as follows:

$$\begin{array}{l} f_+ = K\left(x^T, C^T\right) u_1 + b_1 = 0 \\ f_- = K\left(x^T, C^T\right) u_2 + b_2 = 0 \end{array} \tag{27}$$

where $C = [A, B]^T$, and $K(\cdot)$ is a kernel function. One of two nonparallel hyperplanes can be obtained by solving the following QPP.

$$\begin{array}{ll} \min & \frac{1}{2}\big(K\left(A, C^T\right) u_1 + e_1 b_1\big)^T \big(K\left(A, C^T\right) u_1 + e_1 b_1\big) + c_1 e_2 \xi_1 \\ s.t. & -\big(K\left(B, C^T\right) u_1 + e_2 b_1\big) + \xi_1 \geq e_2, \xi_1 \geq 0 \end{array} \tag{28}$$

Similarly to linear TWSVM, the dual problem of (28) can be represented as follows:

$$\begin{array}{ll} \max & e_2^T \alpha - \frac{1}{2}\alpha^T R\big(S^T S\big)^{-1} R^T \alpha \\ s.t. & 0 \leq \alpha \leq c_1 e_2 \end{array} \tag{29}$$

where $S = [K(A, C^{\mathrm{T}}), e_1]$, $R = [K(B, C^{\mathrm{T}}), e_2]$.

Define $z_1 = [u_1, b_1]$ and $z_2 = [u_2, b_2]$. We can solve the dual problems (29) and get the following result.

$$z_1 = -(S^T S)^{-1} R\alpha \tag{30}$$

Similarly, we can obtain the other result.

$$z_2 = (R^T R)^{-1} S\beta \tag{31}$$

4. **Least Squares Twin support vector machine.** Suykens proposed least square support vector machine (LSSVM) in 1999 [18]. Now LSSVM has attracted much attention since it has faster training speed than that of SVM.

In this paper, just the nonlinear case was considered. LSTWSVM needs to find two nonparallel hyperplanes based on kernel functions like TWSVM as follows:

$$\left\{ \begin{array}{l} K\left(x, C^T\right) u_1 + b_1 = 0 \\ K\left(x, C^T\right) u_2 + b_2 = 0 \end{array} \right. \tag{32}$$

where $C^T = \left[A^T, B^T\right]$, $K(\cdot)$ and is a kernel function. LSTWSVM needs to determine two QPPs as follows.

$$\begin{array}{ll} \min & \frac{1}{2}\left\|K\left(A, C^T\right) u_1 + e_1 b_1\right\|^2 + \frac{1}{2}\xi_1^2 \\ s.t. & -\left(K\left(B, C^T\right) u_1 + e_2 b_1\right) + \xi_1 = e_2 \end{array} \tag{33}$$

$$\begin{array}{ll} \min & \frac{1}{2}\left\|K\left(B, C^T\right) u_2 + e_2 b_2\right\|^2 + \frac{1}{2}\xi_2^2 \\ s.t. & \left(K\left(A, C^T\right) u_2 + e_1 b_2\right) + \xi_2 = e_1 \end{array} \tag{34}$$

We can convert the QPPs (33) and (34) into the following unconstrained optimization problems.

$$\left\{ \begin{array}{l} \frac{1}{2}\left\|K\left(A, C^T\right) u_1 + e_1 b_1\right\|^2 + \frac{c_1}{2}\left\|e_2 + \left(K\left(B, C^T\right) u_1 + e_2 b_1\right)\right\|^2 \\ \frac{1}{2}\left\|K\left(B, C^T\right) u_2 + e_1 b_2\right\|^2 + \frac{c_2}{2}\left\|e_1 - \left(K\left(A, C^T\right) u_2 + e_1 b_2\right)\right\|^2 \end{array} \right. \tag{35}$$

By KKT conditions, the optimal solutions can be described as follows:

$$\left\{ \begin{array}{l} (u_1, b_1)^T = -\left(H^T H + \frac{1}{c_1} G^T G\right)^{-1} H^T e_2 \\ (u_2, b_2) = \left(G^T G + \frac{1}{c_2} H^T H\right)^{-1} G^T e_1 \end{array} \right. \tag{36}$$

where $H = \left[K\left(A, C^T\right), e_1\right]$, $G = \left[K\left(B, C^T\right), e_2\right]$.

5. **Wavelet kernel Twin support vector machine.** The wavelet kernel (10) used in this article are given in part 2. Without loss of generality, we construct a translation-invariant mother wavelet kernel by a wavelet function adopted in [12]:

$$h(x) = \cos(1.75x)\exp\left(-\frac{x^2}{2}\right) \tag{37}$$

Lemma 5.1. *Given the mother wavelet (10) and the dilation $a, c, x \in R$, if $X, X' \in R^N$, then the wavelet kernel of the mother wavelet is*

$$\begin{aligned} K(X, X') &= \prod_{i=1}^{N} h\left(\tfrac{x_i - c_i}{a}\right) \\ &= \prod_{i=1}^{N} \left(\cos\left(1.75 \times \tfrac{(x_i - x'_i)}{a}\right)\exp\left(\tfrac{\|x_i - x'_i\|^2}{2a^2}\right)\right) \end{aligned} \tag{38}$$

That is a kind of multidimensional wavelet kernel. TWSVM with wavelet kernel determines two nonparallel hyperplanes as follows.

$$\begin{cases} K\left(X, C^T\right)u_1 + b_1 = \sum\limits_{i=1}^{l} u_1(i) \prod\limits_{j=1}^{N} h\left(\frac{x^j - x_i^j}{a_i}\right) + b_1 \\ K\left(X, C^T\right)u_2 + b_2 = \sum\limits_{i=1}^{l} u_2(i) \prod\limits_{j=1}^{N} h\left(\frac{x^j - x_i^j}{a_i}\right) + b_2 \end{cases} \tag{39}$$

Now, the decision function of WTWSVM for classification is given as

$$i = \arg\min_{k=1,2} \frac{\left|\sum\limits_{i=1}^{l} u_i \prod\limits_{j=1}^{n} h\left(\frac{x^i - x_j^i}{a}\right) + b\right|}{\|w_k\|} \tag{40}$$

The test data can be classified based on result of the equation (40).

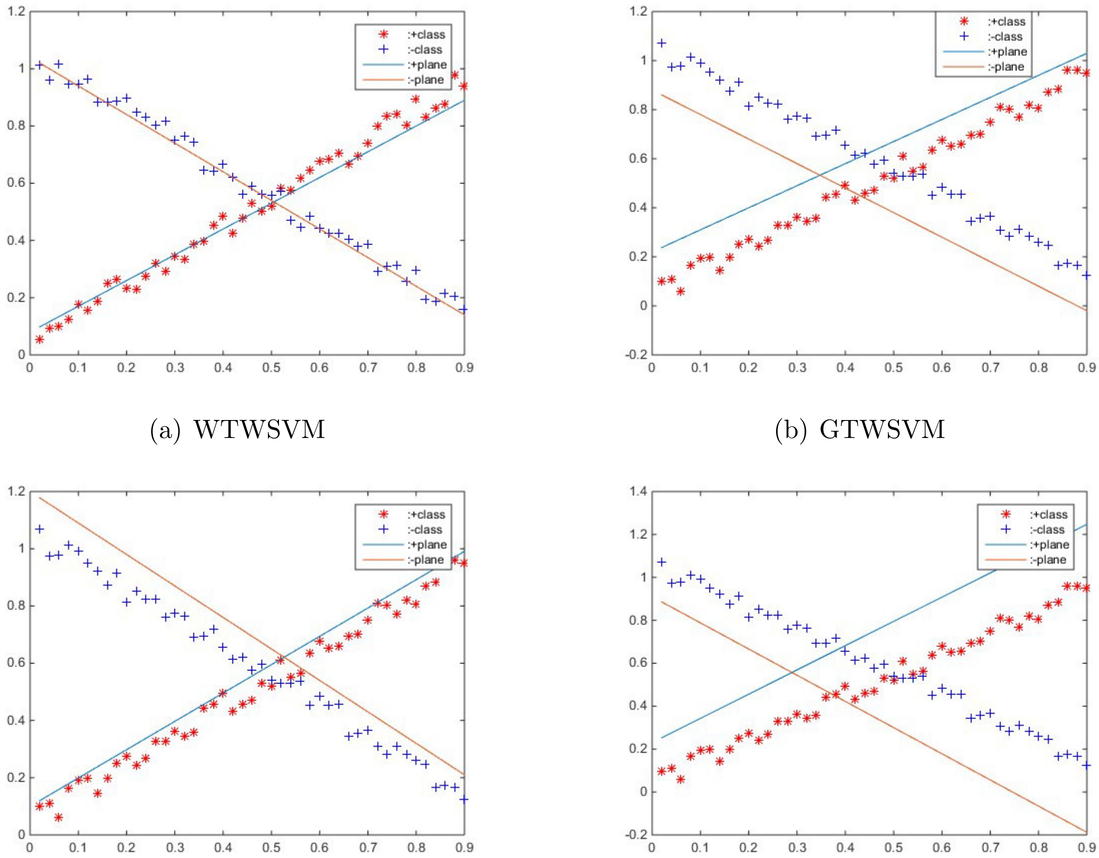
6. **Experiments.** To test the performance of our proposed approaches, we compared numerically WTWSVM and WLSTSVM with Gaussian kernel TWSVM (GTWSVM) and Gaussian kernel LSTSVM (GLSTSVM) respectively on a synthetic dataset and 12 datasets from UCI Repository [19]. All experiments were implemented by using MATLAB 8.4 on a personal computer with 1.6GHz and 4GB RAM.

In first experiment, a synthetic dataset with cross noise is presented to demonstrate the effectiveness of the proposed approaches. Ten-fold cross-validation is carried out to determine the parameters. From a and b in Fig.1, we can see WTWSVM has better classification performance than standard TWSVM with Gaussian kernel. From c and d in Fig.1, it can be also easily seen that the classification performance of WLSTSVM is superior to that of LSTSVM with Gaussian kernel. So Fig.1 shows WTWSVM has best performance in four approaches.

We validate the effectiveness of WTWSVM and WLSTSVM by the second experiment. All the datasets are available from UCI Repository [19]. The selected datasets are listed in Table 1.

Table 2 compares the performance of the WTWSVM classifier with that of GTWSVM. It can be seen from Table 2 that WTWSVM has not higher classification precision than GTWSVM, but WTWSVM is faster obviously than GTWSVM on the vast majority of datasets. The results in Table 3 demonstrate that WLSTSVM has higher training speed than GLSTSVM.

In the two experiments, WTWSVM has better classification results than GTWSVM on most datasets, especially on high-dimensional datasets, which verifies that wavelet



(a) WTWSVM (b) GTWSVM

(c) WLSTSVM (d) GLSTSVM

Figure 1. Learning results of the four algorithms on the Cross-planes data set

Table 1. Attribute characteristics of the UCI datasets

Dataset	Dimension	Number	Dataset	Dimension	Number
australian	14	690	ionosphere	34	351
breast	9	277	pima	8	768
bupa	6	345	sonar	60	268
diabetes	8	768	vote	15	435
german	24	1000	wdbc	31	569
heart	13	270	wpbc	33	198

kernel can save more data distribution details and have better classification results than Gaussian kernel.

7. **Conclusion.** In this paper, a new wavelet kernel is proposed. Compared with Gaussian kernel, it is orthonormal or orthonormal approximately. Based on this construction, a WTWSVM and a WLSTSVM are introduced respectively. The theoretical analyses and experiment results show the feasibility and validity of the WLSTSVM and WTWSVM.

Acknowledgment. This work was supported in part by the National Natural Science Foundation of China under Grants (51875457), the International S&T Cooperation Program of Shaanxi Province (2019KW-056), the Natural Science Foundation of Shaanxi Province (2021JQ-701), and the Xi'an Science and Technology Plan Project (2020KJRC0109).

Table 2. Performance comparison of WTWSVM and GTWSVM

dataset	WTWSVM		GTWSVM	
	Accuracy(%)	Time(s)	Accuracy(%)	Time(s)
australian	80.75	67.2583	85.61	557.28
Breast	73.25	41.78	84.21	111.47
Bupa	70.15	53.8363	69.56	148.7
Diabetes	82.63	859	78.57	711.69
german	78.61	72.74	76	1248
heart	81.81	87.362	87.27	104.7
ionosphere	94.37	99.73	97.18	168.571
pima	77.27	195.44	85.32	713.83
sonar	88.10	50.23	95.34	78.6044
vote	97.73	107.66	97.72	236.63
wdbc	98.12	722.64	97.39	447.04
wpbc	90.02	58.81	85.36	70.91

Table 3. Performance comparison of WLSTWSVM and GLSTSVM

dataset	WLSTSVM		GLSTSVM	
	Accuracy(%)	Time(s)	Accuracy(%)	Time(s)
australian	62.59	8.6013	84.17	368.55
breast	78.57	1.8461	71.92	56.07
bupa	64.29	1.3827	75.36	79.98
diabetes	65.34	3.711	74.63	395.1
german	76.62	31.1324	74.5	730
heart	56.63	3.0566	85.45	51
ionosphere	80.28	21.9363	95.77	91.14
pima	69.88	3.8671	79.87	395.19
sonar	60.23	39.463	86.04	35.62
vote	69.32	6.1508	96.59	167
wdbc	61.22	30.7230	96.52	300.31
wpbc	70.73	11.7649	85.36	65.89

REFERENCES

[1] C. Cortes, V. Vapnik. *Support-vector networks.* Machine Learning, vol.20, no.3, pp. 273-297, 1995.

[2] Y. F. Li, J. B. Li, and J. S. Pan. *Hyperspectral image recognition using SVM combined deep learning*, Journal of Internet Technology, vol.20, no.3, pp. 851-859, 2019.

[3] D. P. Tian. *Support vector machine for content-based image retrieval: A comprehensive overview*, Journal of Information Hiding and Multimedia Signal Processing, vol. 9, no. 6, pp. 1464-1478, 2018.

[4] F. J. Kuang and S. Y. Zhang, *A Novel network intrusion detection based on support vector machine and tent chaos artificial bee colony algorithm*, Journal of Network Intelligence, vol. 2, no. 2, pp. 195-204, 2017.

[5] L. Q. Zhao, M. J. Gai and Y. F. Jia, *Classification of multiple power quality disturbances based on PSO-SVM of hybrid kernel function*, Journal of Information Hiding and Multimedia Signal Processing, vol. 10, no. 1, pp. 138-146, 2019.

[6] O. L. Mangasarian, E. W. Wild. *Multisurface proximal support vector machine classification via generalized eigenvalues*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.28, no.1, pp.69-74, 2006.

[7] K. Jayadeva, 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, 2007.

[8] X. Peng, D. Xu, and L. Kong. *L1-norm loss based twin support vector machine for data recognition*, Information Sciences, vol.340-341, pp.86-103, 2016.

[9] M. Tanveer, K. Shubham. *A regularization on Lagrangian twin support vector regression*, International Journal of Machine Learning and Cybernetics, vol.8, no.3, pp.807-821, 2017.

[10] G. Y. Zhu, C. G. Yang and P. Zhang. *Linear programming nonparallel support vector machine and its application in vehicle recognition*, Neurocomputing, vol.215, pp.212-216, 2016

[11] X. Xie. *Regularized multi-view least squares twin support vector machines*, Applied Intelligence, vol.17, no.1-8, 2018.

[12] I. Daubechies. *Ten lectures on wavelets*, Computers in physics. vol.6, no.3, pp1671-1671, 1995.

[13] I. Daubechies. *Biorthogonal bases of compactly supported wavelets*, Communications on Pure and Applied Mathematics, vol.45, no.5, pp.485-560, 2010.

[14] S. G. Malla. *A theory for multiresolution signal decomposition: The wavelet representation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.11, no.4, pp.674-693, 1989.

[15] J. Mercer. *Functions of positive and negative type and their connection with the theory of integral equations*, Philosophical Transactions of the Royal Society of London, vol.209, no.415-446, 1909

[16] I. Daubechies, A. Grossmann. *Meyer Y. Painless nonorthogonal expansions*, Journal of Mathematical Physics, vol.27, no.5, pp.1271-1283, 1986.

[17] H. H. Szu, B. Telfer. *Neural network adaptive wavelets for signal representation and classification*, Optical Engineering, vol.31, no.9, pp.1907-1916, 1992.

[18] J. A. K. Suykens, J. Vandewalle. *Least squares support vector machine classifiers*, Neural Processing Letters, vol.9, no.3, pp.293-300, 1999.

[19] UCI Machine learning repository, http://archive.ics.uci.edu/ml/index.php. Last accessed Nov. 10, 2018.